Engagement Dynamics in mHealth

Yikun Jiang* Kosuke Uetake† Nathan Yang‡

April 10, 2023

Abstract

We study the impact of premium adoption on mHealth user engagement dynamics with data from a popular mobile food and exercise tracking app. To estimate the causal effect of premium adoption we use a propensity score matching method. The analysis reveals that premium adoption is linked to elevated engagement levels for food and exercise calorie tracking, daily goal achievement, exercise calories, as well as weight loss. Moreover, these effects on engagement levels from premium adoption dampen quickly over time. We later confirm using a calibrated theoretical model that the observed usage dynamics are likely explained by mental accounting of sunk costs. A consequence of sunk costs, as we demonstrate, is that user engagement as well as their incentives to adopt the premium version, could increase if fees are raised.

Keywords: Behavioral Economics; Customer Dynamics; Freemium Design; Mental Accounting; Mobile Health and Wellness; Subscription Service; Sunk Cost Fallacy

---

*University of California, Berkeley, Department of Economics. Email: jiangyk@berkeley.edu.

†Yale University, School of Management. Email: kosuke.uetake@yale.edu.

‡Cornell University, SC Johnson College of Business. Email: ncy6@cornell.edu.

Previously titled "Does Premium Version Adoption in mHealth Improve User Engagement and Health-Related Outcomes?" We thank Unnati Narang, Daiva Nielsen, and conference/seminar participants at the Canadian Centre for Health Economics, McGill University, and Marketing Science 2022 Conference for helpful comments and suggestions.

1 Introduction

Companies and policymakers are now turning to new technologies that help support healthy lifestyles. In particular, there has been rapid development in mobile health (mHealth) applications. Empirical (Aswani et al. 2019, Kapoor et al. 2022, Zhou et al. 2018, Labonté et al. 2022, Uetake and Yang 2018) and theoretical (Mintz et al. 2020) research has demonstrated some merit of these mHealth technologies, as they have the potential to improve health and wellness via behavioral change. However, despite these merits, diligent usage of mHealth remains an elusive outcome (Murnane et al. 2015).

A feature of mHealth that distinguishes it from traditional health and wellness programs is its reliance on the freemium business model (Faulkner 2019). In a freemium business model, users can access many of the basic core features of the technology for free, but need to pay additional monthly or yearly fees in order to use enhanced features offered in a premium version. In general, the freemium subscription-based model has emerged as a popular monetization strategy for mobile apps (Narang and Shankar 2019). While past research has helped us better understand the drivers of mHealth adoption (Bojd et al. 2022, Ghose et al. 2021, Kato-Lin et al. 2016, Tang et al. 2015), we know little about the impact of mHealth premium adoption on user engagement and weight loss outcomes. Understanding these effects is important as promotions of premium versions (e.g., first-month free subscription) might have ramifications on diligent usage in the longer run. These longer-run usage patterns will ultimately have a financial impact on user turnover and customer lifetime value (McCarthy et al. 2017). Our research aims to empirically identify the causal effect of premium version adoption.

Using large-scale data from a popular fitness-tracking mHealth application, we investigate the causal impact of premium adoption[1] on user engagement with the app (e.g., food and exercise tracking, goal adherence, exercise calories) and weight loss outcomes. As premium adoption is an endogenous decision made by the users, we estimate the local average treatment effect (LATE) of premium adoption using a propensity score matching algorithm that matches adopters to similar non-adopters based on initial user characteristics, weight loss outcome, and past engagement. With our propensity score matching, we match the adopters and non-adopters quite well. We then apply a differences-in-differences (DiD) anal-

[1]We are using the terms, premium adoption, premium version adoption, premium upgrade, adoption, upgrade, and upgrade to the premium version, interchangeably in this work.

ysis on the matched sample, where we compare the change in outcomes before and after the premium version adoption between adopters and non-adopters. We then compare the results with DiD analysis with the results with the full sample (without matching). The comparison reveals the importance of addressing potential endogeneity concern.

Our empirical analysis of the matched sample reveals that premium adoption does indeed have short-term positive effects on premium user engagement in the first week, such as food tracking (about 12.7 percentage points more), exercise tracking (about 8.8 percentage points more), goal adherence (about 9.8 percentage points more), and exercise calories (about 60.2 calories more), compared with similar users using the free version. A comparison of our results across the full and matched samples shows the potential overstatement of the premium effects on user engagement, which underscores the importance of addressing endogeneity issues. Furthermore, there is a noticeable dampening of the premium effect on engagement over time. For example, during the fourth week since the premium upgrade, the impact of premium adoption in lifting engagement falls to 9.1 percentage points for food tracking, 5.6 percentage points for exercise tracking, 7.1 percentage points for goal adherence, and 45.4 calories for exercises. And during the sixth week since the premium upgrade, the impact of premium adoption in lifting engagement falls to statistically insignificant for food tracking, exercise tracking and goal adherence, and only marginally significant for exercise calories. After the sixth week since the premium upgrade, the effect diminishes to statistically insignificant for all of our four engagement measures.

We also explore the short-term and long-term weight loss effectiveness of the premium version on the matched sample. Our analysis confirms the short-term and long-term effectiveness. More specifically, within the first week after premium adoption, the premium version users have an increased probability of reducing weight of 20.8 percentage points, and an increased amount of weight lost by 0.97 pounds, compared with similar users who use the free version. The effects of the premium version gradually increase over a few weeks, as weight loss takes time. After the sixth week since the premium upgrade, the premium version users still have an increased probability of reducing weight of 20.8 percentage points, although they do not have a statistically significant increase in the amount of weight lost. This is possibly due to premium users become inactive in weight loss after the short-term boost of engagement by the premium version, and users who use the free version gradually catch up on the weight lost progress over the long term.

In addition to the average effects, our analysis of the matched sample also shows the effects of the premium upgrade are heterogeneous among users. The effects of premium adoption on engagement dampen more for younger users over time, whereas the effects of premium adoption last longer for users of higher age. As for weight loss outcomes, younger male users potentially benefit more from premium adoption, in terms of having a higher chance of staying below the initial weight and a higher amount of weight loss.

Our main results are robust to different specifications and different matching samples. We run the same DiD analysis on the subsample of users who upgraded during our observation period, utilizing the different timing of upgrade for identification. We also match similar non-adopters to adopters based on user characteristics, engagement and weight loss variables, separately for users who adopted the premium version early (within the first seven days), users who adopted the premium version in the first 21 days, and users who adopted the premium version late (between the 36th and the 49th day), and run the same DiD on these three matched sample separately. These robustness checks allow us to downplay potential caveats of the propensity matching approach in the presence of endogenous timing of premium adoption. In addition, we conduct an analysis base on Oster (2019), to confirm the extent to which the selection in upgrades is controlled for.

The inferred dampening patterns (especially among the younger demographic) appear to be more consistent with behavioral mechanisms that relate to sunk costs, rather than hedonic decline, moral licensing, self-control, and goal pursuit. Uncovering this dampening pattern is also relevant for companies in mHealth, as efforts to boost user engagement via premium adoption (e.g., subscription promotions) might be short-lived, and companies may want to combine premium subscription with other strategies to maintain engagement for longer-term. Our findings add to the emerging literature that has uncovered evidence of sunk cost effects. In particular, we demonstrate that patterns consistent to sunk costs may extend beyond online education (Goli et al. 2022) settings. At the same time, our analysis complements Goli et al. (2022) and shows that different empirical settings might not have the exact same exhibited heterogeneity across users; for example, we find after the fourth week since premium adoption, older premium users have higher engagement level compared with younger premium users.[2] It is possible that in our setting, younger users' engagement is more motivated via sunk cost which is transient, whereas the engagement of users of higher age is more motivated by other mechanisms, leading to the difference in engagement

[2]However, very interestingly, we both find the heterogenous effects on user engagement begin to appear after the fourth week of the payment, suggesting the payment effect via sunk cost is transient.

to emerge after the fourth week since upgrade.[3]

Having established the causal effects of premium adoption on user engagement, we investigate the potential role of sunk cost fallacy as a mechanism behind the observed patterns. To investigate this mechanism further, we develop a theoretical model of mHealth engagement. This model allows users to make decisions about whether or not to adopt a premium version of the mHealth app, and after this adoption decision, how much they use the app over time. By calibrating the model using the data, we uncover a few key insights from simulation analysis. First, we show that dampening patterns can indeed be consistent with scenarios where sunk costs are present, as well as scenarios where users are rational (i.e., no sunk cost mental accounting). The simulations also reveal that the higher engagement we observe among premium users as compared with basic users only applies to the scenario for which sunk costs are accounted for. Moreover, not only will engagement increase with fees, but so too will be users' underlying incentives to adopt the paid premium version provided that users mentally account for sunk costs. Therefore, this simulation analysis suggests that sunk costs likely play an important role in the increased engagement we observe among premium users.

As for our paper's contribution, first, we contribute towards the broader discussion about the role of marketing in encouraging people towards healthier lifestyles, as there is growing interest in identifying contexts, nudges, or public policies that are associated with healthier lifestyles. For example, past studies have focused on social comparisons (Aral and Nicolaides 2017, Uetake and Yang 2020), self-regulation (Huang et al. 2015), nutrition labeling (Bollinger et al. 2022, Puranam et al. 2017), medical diagnoses (Ma et al. 2013), AI vs human coaching (Kapoor et al. 2022), taxes (Gordon and Sun 2015, Khan et al. 2016, Seiler et al. 2021), financial vs non-financial incentives (Narang 2022), omnichannel grocery shopping (Chintala et al. 2021, Huyghe et al. 2017), variety (Haws et al. 2017), and healthiness claims (Rao and Wang 2017). We add to this discussion by offering some insights about the role of mHealth (and its design) in facilitating health-related behaviors. In particular, our work is most related to Kapoor et al. (2022). Their study highlights the impact of AI versus human coaching on weight loss among mHealth users. Our work complements their study in the following ways. First, our mHealth app's premium version (at the time of our data period) did not have coaching of any meaningful form,[4] so the actual functional benefits

[3] In addition to Goli et al. (2022) that analyze learners' engagement after payments, we also analyze the short-term and long-term effectiveness of the premium version using outcome variables related to weight loss progress.

[4] Based on our frequent conversations with company representatives, their team did not have anyone with strong-

associated with premium versus basic might be more psychologically driven (i.e., the main difference between premium and basic is the fact that users pay for a few additional features on premium), in our case. Second, while our analysis (like theirs) also provides some insights about weight loss, many of the behavioral mechanisms we posit (e.g., sunk costs) rely on our results about the premium version's impact on engagement over time.

Second, our findings add to the general understanding about freemium design. Thus far, much of the attention has been centered around advertising/pricing strategies in freemium design (Appel et al. 2019, Lambrecht and Misra 2017, Li et al. 2019) and the extent to which they have an impact on user engagement and relevant outcomes (e.g., health-related behaviors).[5] Our empirical context is unique in the sense that users need to put in effort and use self-control to stay engaged and get closer to their weight loss goal. Our findings suggest that the subscription-based nature of mHealth apps might in itself have an impact on how engaged users are with health-related behaviors (e.g., tracking, goal adherence, weight loss).

2 Data

We obtained user data from a popular mHealth fitness app company headquartered in the United States. The mHealth industry largely consists of apps and wearable technologies, serving the purpose of diagnosis, monitoring, and/or treatment, and this industry is valued at over $51 billion in 2021, and is expected to reach $243 billion by the year 2030.[6] Our empirical context will largely focuse on an mHealth app aimed at monitoring calorie consumption and weight. Using the conceptual framework of Liu et al. (2022) for characterizing where mHealth fits in a user's patient journey, the app we study likely applies to users in prec-linic stages, whereby the digital technology affordance pertains to information transparency.

When users start using the app after downloading it, they need to type in their basic demographic information - they will report their current weight, height, goal weight, age, and gender. The application will come up with a daily calorie budget for users to stay within. Users can then track the foods they consume for each meal (the app then supplements caloric values), any exercise they engage in, and their current weight as long as they wish to. At the

enough computer science abilities needed to develop AI-powered coaching. Food recommendation systems designed and tested using historical observational data by external academic teams, akin to FOODVAR (Nielsen et al. 2022), were never implemented by the company due to a lack of internal company capabilities.

[5]This finding contributes to the broader literature about the impact of subscription services on consumer behavior (Iyengar et al. 2022).

[6]See, for example, https://www.precedenceresearch.com/mhealth-market.

end of each day, users will know whether they stay within their daily calorie budget or not. The notification is purely a message and there is no dynamic change by consuming beyond the budget.

In particular, the sample they provided us consists of users who were active for at least 30 days and did at least five weigh-ins, as those are users whom are considered serious in their weight loss effort and are viewed as "focal" users from the company's perspective. In total, there are 11,997 users aiming to lose weight in our sample. We observe their daily activity since registration during late 2015 and over the course of the year in 2016. On average, the users were active for 166 days.[7] We focus on the first 100 days since registration for each user to analyze user behavior during a period when they were generally active in weight loss efforts. Among about 12,000 users, 1,131 users (about 9.4%) upgraded to the premium version of the app during the period.[8] Figure 1 shows the total number of users who upgrade on each day since registration. Among those who upgrade, the average number of days from registration to upgrade is 19. Table 1 provides summary statistics for users included in our sample. The users in our sample consist of 29.3% male and 70.7% female. An average user in our sample is around 41 years old, with a starting weight of around 203 pounds, and a target weight of 162 pounds, aiming to lose around 42 pounds, which is around 18.9% (average normalized distance to the goal weight) of the starting weight. On the day of registration, the suggested budget of calories is on average around 1650 calories, and users logged around 1281 calories from foods and logged around 163 calories for exercises.

Table 2 provides some descriptive patterns between premium upgrades and observable user characteristics by running a multiple linear regression of each user's decision of upgrade or not on each user's characteristic variables and behavioral variables using the full sample[9]. These patterns suggest that a typical premium user can be "profiled." In particular, we show that users who upgrade are predominantly those who have higher initial distance to the goal weight, initial goal BMI, older, and logged higher food calorie levels in the first seven days. On the other hand, users who have consistently tracked their food calories in the first seven days appear less likely to be premium adopters. Therefore, endogeneity and selection may operate through user characteristics, engagement level and weight loss progress.

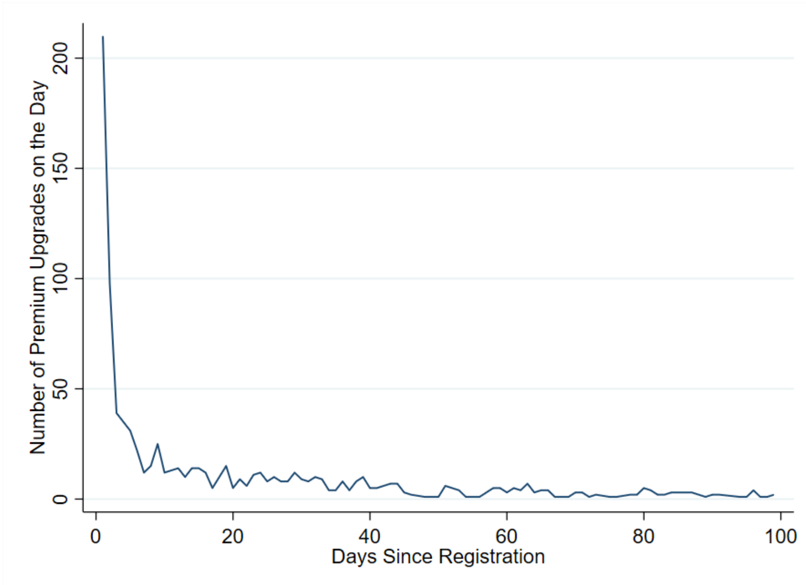
To upgrade to the premium version of the app, users need to pay $39.99 for access to the

[7]The number of active days is defined as the number of days between the user's registration date and the last day the user tracked anything on the app.

[8]The proportion of paid users is similar to FitBit, which has a premium adoption rate of about 4% based on its user/subscriber numbers in 2020.

[9]In all of our regression tables, * denotes $p < 0.10$, ** denotes $p < 0.05$, and *** denotes $p < 0.01$.

Figure 1: Number of Upgrades Each Day Since Registration



Notes: Upgrade defined as cases in which users have adopted the paid premium version of the mHealth app.

Table 1: Summary Statistics for Full Sample

Variable	Mean	Std. Dev.
Male	0.293	0.455
Age	41.604	14.329
Starting Weight (lb)	203.836	52.212
Target Weight (lb)	162.306	36.225
Initial Distance to the Goal Weight (lb)	41.53	32.378
Normalized Distance to the Goal Weight	0.189	0.11
Suggested Budget of Calories (calorie)	1649.775	434.024
Food Calories Logged (calorie)	1281.348	744.699
Exercise Calories (calorie)	163.278	324.616
Number of Active Days	166.248	82.50
N	11997	

Notes: Full sample of observations before matching.

Table 2: Linear Regression of Upgrade Decision on User Characteristics and Behavioral Variables (Full Sample)

	(1) Ever Upgraded
Starting Weight	-0.000500
	(0.000347)
Male	-0.0117
	(0.00835)
Height	0.00242
	(0.00201)
Age	0.00213***
	(0.000190)
Start Date	-0.0000209
	(0.0000403)
Initial Distance to the Goal Weight	0.00105***
	(0.000356)
Initial Goal BMI	0.00440*
	(0.00227)
Number of Days Tracked Food in the First Seven Days	-0.0117***
	(0.00348)
Number of Days Within Daily Calorie Budget in the First Seven Days	-0.00130
	(0.00217)
Number of Days Tracked Exercise in the First Seven Days	0.0000367
	(0.00145)
Total Exercise Calories Tracked in the First Seven Days	0.00000262
	(0.00000182)
Total Food Calories Tracked in the First Seven Days	0.00000260*
	(0.00000135)
Constant	-0.0389
	(0.261)
R-Squared	0.0191
Observations	11997

Notes: Linear probability regression where outcome is whether or not user has upgraded to the premium version of the mHealth app. Control variables include demographic characteristics (e.g., starting weight, gender, age), as well as behavioral usage patterns in first seven days.

premium version for a year.[10] It is worth noting that the core features are shared across the free and premium version, including food calorie tracking, exercise calorie tracking, weight tracking, goal setting, and calorie budgeting. The premium version of the app provides more features,[11] but those additional features can only help users through more usage of the app, and it may take a period of time before users benefit from those features. Therefore, if users use the app less after the premium upgrade, those features cannot function well.

[10] Users can cancel the yearly subscription at any time but they will not receive a refund for unused periods after the cancellation.

[11] Some of these include intuitive food logging, smart camera for tracking food, syncing capabilities with other health apps and wearable devices, community support, advanced tracking, meal and exercise planning, and more gamified rewards/celebrations.

3 Research Design

3.1 Threats to Identification

We are interested in estimating the causal effect of mHealth premium adoption on users' subsequent engagement and weight loss outcomes. An important threat to identification relates to endogeneity of the premium adoption decision. For example, users who are inherently motivated might be those who are more likely to adopt the premium version as well as be more engaged on the app. Therefore, this inherent motivation would confound the actual impact that the premium version itself has on user engagement levels.

To address the concerns about endogeneity, we use propensity score matching to match premium adopters who upgrade during the 8th to 21st days (in the second and third weeks) since registration with similar non-adopters (users that never upgrade) based on propensity score estimated by user characteristic variables and variables that summarizes user engagement and weight loss outcomes in the first seven days since registration. Because different users registered at different time, we also match adopters and non-adopters that registered around similar dates, to control for time-varying intensity to upgrade. We then use a differences-in-differences (DiD) approach on the matched sample based on propensity score. This allows us to estimate the local average treatment effect (LATE) of mHealth premium adoption. This identification strategy is similar to what others have used to study premium adoption effects on user engagement in other contexts, such as online education (Goli et al. 2022) and music streaming (Datta et al. 2018).

Our propensity score matching method uses a flexible method, generalized additive model [12], to predict the probability of premium upgrade for each individual based on initial individual characteristics (e.g., age, gender, height, initial weight, initial goal weight, initial goal BMI, start date) and behavioral variables before premium upgrade (e.g., total exercise calories tracked, total food calories tracked, distance to the goal weight, number of days with tracked food calories, number of days that the user tracked calories and stayed within the daily calorie budget, and number of days with tracked exercise calories). A generalized additive model (GAM) is a generalized linear model (GLM) in which the linear predictor

[12] We also tried several other approaches to predict the propensity score, including LASSO, Ridge, Elastic Net, Random Forest, and Generalized Gradient Boosting. We select Generalized Addictive Model for propensity score prediction because the matched sample based on its propensity score have the most balanced covariates across treatment and control groups.

is given by a user specified sum of smooth functions of the covariates plus a conventional parametric component of the linear predictor. Iyengar et al. (2022) and Goli et al. (2022) also use flexible methods to fit the probability of subscription. One of the advantages of our empirical setting is that we observe the current weight and distance to goal weight on each day for each user, which are relevant measures of user motivation, and thus potential drivers of the upgrade decision. By including these variables in our propensity score matching algorithm along with other user characteristics and engagement variables, we hope to be closer to satisfying the conditional independence assumption of propensity score matching.

The identification based on the propensity score matching fits our scenario, as it is not feasible to randomly upgrade users to the premium version, and randomly assigned financial incentives for upgrading will be prone to selection-into-treatment compliance biases (Daljord et al. 2022). There are of course important qualifiers for the validity of propensity score matching. First, the conditional independence assumption requires that the individual characteristics are good predictors of the treatment assignment (i.e., premium adoption). Second, the common support assumption requires that the relevant weights can indeed be calculated. We explore the empirical validity of these assumptions in subsection 3.2, along with potential caveats for our specific empirical context. Moreover, we provide additional sensitivity analysis that assesses the validity of the commonly-used assumption about selection on observable heterogeneity (i.e., conditional independence) by bounding the treatment effect along the lines of Oster (2019). Our main concern pertains to the extent that selection operates via unobservable heterogeneity. We argue in our analysis of robustness that the results from this bounding exercise suggest that the selection on observable heterogeneity is not too strong of an assumption.

Similar to the existing papers we mentioned above, we can not perfectly control for the endogeneity of the timing of the premium upgrade. For this reason, there might be systematic differences between early and late adopters of the premium version. We offer some additional analysis that confirms the robustness of our main findings for both early and late adopters (see subsubsection 4.3.2).

3.2 Empirical Specification

Using propensity score matching, we match each premium adopter with 10 comparable users that are likely to adopt but do not do so in our entire observation period. To implement the propensity matching process, we model the premium upgrade decision with a semiparametric generalized additive model and match premium users with similar free users via nearest neighbor matching without replacement.[13] Here is the function[14] we use to predict the propensity score:

$$\text{Upgrade}_i = \sum_{d=1}^{D} \sum_{k=1}^{3} \omega_k^d s_k(X_i^d) + \mathbf{Z_i}'\delta + \varepsilon_i, \tag{1}$$

where D is the number of variables included in X_i, $s_k(\cdot)$ denotes a smooth function using regression splines and ω_k^d is the weight associated to each function. We consider spline functions up to the third degree. Additional variables $\mathbf{Z_i}$ include dummy variables and variables with limited variations.[15] Below, we list the variables included in X_i and Z_i.

- X_i: starting weight, height, age, start date, initial goal weight, initial goal BMI, total exercise calories tracked in the first seven days, total food calories tracked in the first seven days, distance to goal weight on the seventh day. We denote each variable in X_i by X_i^d.

- $\mathbf{Z_i}$: gender, number of days tracked food in the first seven days, number of days within daily calorie budget in the first seven days, number of days tracked exercises in the first seven days.

In our main specification in this section, we match users who upgrade to the premium version between the 8th day and the 21st day since registration with similar users who never upgrade based on initial user characteristics and behavioral variables that summarize users' engagement and weight loss outcome before the upgrade. Because there is a substantial proportion of users who upgrade within the first week since their registration and a noticeable proportion of users who upgrade quite late, we separately analyze those early adopters and those late adopters in our robustness check (subsection 4.3).

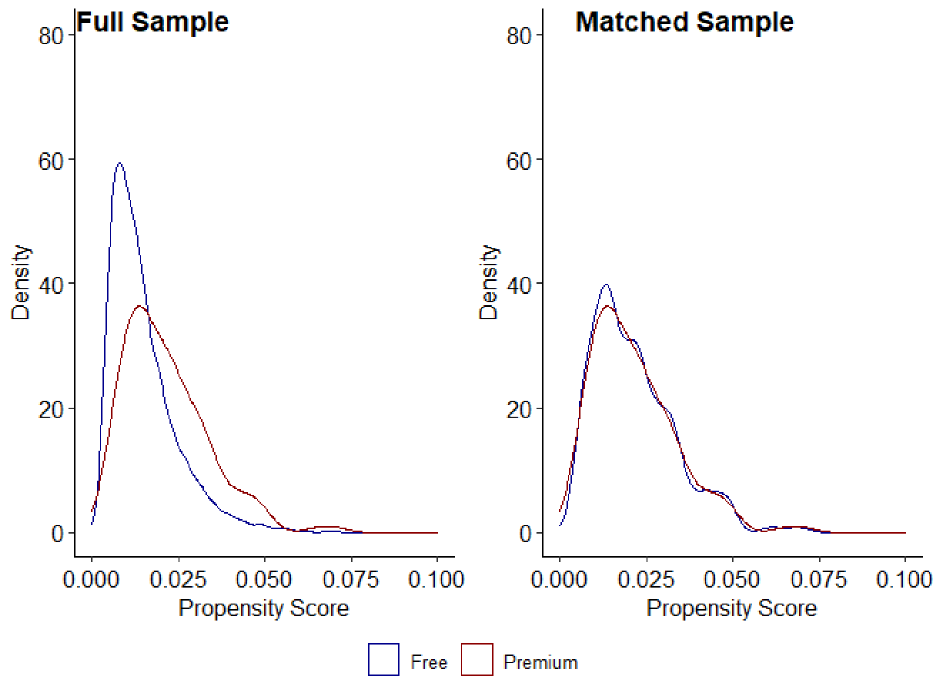
We plot the distributions of propensity score before and after matching for both specifications in Figure 2. These plotted distributions suggest that our algorithm can indeed distinguish between free users and premium users, and after matching, free users and premium users have similar propensity score distributions.

[13] For more information, we use R packages "mgcv"(Wood 2011) and "MatchIt"(Ho et al. 2011).

[14] We choose this function specification because it allows us to match each premium adopter with the most similar non-adopters, and still have good covariate balance for all of our matching variables across the treatment and control groups for all of our specifications, including the main specification and the alternative matching specifications in our robustness check.

[15] We tried alternative specifications, where the variables in Z_i are also included in X_i. The results do not change much with such an alternative specification.

Figure 2: Distribution of Propensity Scores Before and After Matching



Notes: Distributions are obtained non-parametrically.

In Table 3, we compute the covariate balance of our matching variables for those who upgrade between the 8th day and the 21st day (treatment group) and those who never upgrade (control group) in the full sample, via Welch Two Sample t-test. Similar to what we find in section 2, the treatment group have a higher starting weight, a higher age, a later start date, a larger initial goal weight and initial goal BMI, more days with tracked food in the first seven days, and a larger distance to the goal weight at the seventh day, all with statistical significance. The existence of these patterns underscores the importance of addressing endogeneity and selection issues when assessing the effectiveness of premium adoption on health and weight loss objectives. By using propensity score matching, we assume that it is possible to control for the possible unobserved factors that affect both engagement and upgrade decisions through matching on the rich set of variables that summarize user characteristics, engagement and weight loss outcome before upgrading.

To assess the quality of our matching, we compute the covariate balance of our matched sample (Table 4) for our matching variables via Welch Two Sample t-test, and compare it with the covariate balance of the full sample (Table 3). We notice that before matching, our full sample is not balanced in terms of almost all of our matching variables, but after matching, those who upgrade between the 8th day and the 21st day (treatment group) and those who never upgrade (control group) have very similar average in the matched sample.

We have a few observations about the validity of our propensity score matching approach from an empirical standpoint. First, Table 2 confirms that the observed individual characteristics have some ability to explain the treatment assignment (i.e., premium adoption), even with a very simple regression that different from the generalized addictive model we use to predict the propensity score. However, there might still be selection on unobservables that might further dampen the effects we document. Second, Table 4 and Figure 2 provide some evidence that the common support assumption might be satisfied, as the propensity score weights seem to work for all of the matching variables.

Table 3: Summary Statistics of the Matching Variables for the Treatment and Control Groups (Before Matching)

Variable	Treatment Mean	Control Mean	t-Statistic	p-Value
Starting Weight	218.2659	203.0753	3.9119	0.0001326
Male	0.2744	0.2932	-0.5343	0.5939
Height	66.5488	66.3633	0.6563	0.5125
Age	45.7683	41.1334	4.1699	0.00004872
Start Date	5575.6037	5562.7923	2.4126	0.01692
Initial Goal Weight	167.7576	161.8745	2.1086	0.03646
Initial Goal BMI	26.4511	25.6941	2.2583	0.02521
Number of Days Tracked Food in the First Seven Days	6.2561	6.0883	1.3124	0.1912
Number of Days Within Daily Calorie Budget in the First Seven Days	5.0427	4.8797	1.0528	0.2939
Number of Days Tracked Exercise in the First Seven Days	3.0244	3.0407	-0.08349	0.9336
Total Exercise Calories Tracked in the First Seven Days	1040.8872	1132.3933	-0.9895	0.3238
Total Food Calories Tracked in the First Seven Days	9406.4077	9017.2104	1.3074	0.1929
Distance to Goal Weight at the Seventh Day	48.9028	38.7652	3.8481	0.000169
Number of Users	1131	10866		

Notes: The t-stats reported follow the Welch's two-sample t-test.

Given these matches, we use a DiD empirical specification to estimate the causal effect of premium adoption on user engagement and weight loss outcomes. We compare the outcome measures of premium adopters and matched non-adopters, and we investigate whether the effects diminish or persist over time. More specifically, we estimate the following:

$$Y_{it} = \sum_{k=1}^{6} \beta_k D_{it}^{k} + \beta_7 I(\text{After the Sixth Week}) + \alpha_i + \gamma_t + \varepsilon_{it}. \tag{2}$$

Table 4: Summary Statistics of the Matching Variables for the Treatment and Control Groups (After Matching)

Variable	Treatment Mean	Control Mean	t-Statistic	p-Value
Starting Weight	218.2659	218.3592	-0.02307	0.9816
Male	0.2744	0.2780	-0.09980	0.9206
Height	66.5488	66.5817	-0.1120	0.9110
Age	45.7683	46.3402	-0.4950	0.6212
Start Date	5575.6037	5573.7793	0.3298	0.7419
Initial Goal Weight	167.7576	168.4738	-0.2465	0.8055
Initial Goal BMI	26.4511	26.5333	-0.2359	0.8137
Number of Days Tracked Food in the First Seven Days	6.2561	6.2793	-0.1750	0.8612
Number of Days Within Daily Calorie Budget in the First Seven Days	5.0427	4.9902	0.3259	0.7449
Number of Days Tracked Exercise in the First Seven Days	3.0244	3.0909	-0.3271	0.7440
Total Exercise Calories Tracked in the First Seven Days	1040.8872	1033.4728	0.07688	0.9388
Total Food Calories Tracked in the First Seven Days	9406.4077	9470.6382	-0.2079	0.8355
Distance to Goal Weight at the Seventh Day	48.9028	47.6248	0.4677	0.6405
Number of Users	164	1640		

Notes: The t-stats reported follow the Welch's two-sample t-test.

Y_{it} denotes the outcome of interest. For engagement, we specify Y_{it} as whether user i tracks food calories on day t, tracks exercise calories on day t, stays in the daily calorie budget on day t, and user i's amount of exercise calories on day t, respectively. For weight loss outcomes, we specify Y_{it} as whether user i has lower weight on day t compared with his/her initial weight, and how much accumulated weight loss user i has achieved up till day t, compared with the initial weight, respectively. The main short-term treatment of interest is each D_{it}^{k}, which is an indicator for whether user i is within the k-th weeks since he/she adopted the premium version at day t. For users who never adopted, D_{it}^{k} always equals to 0. To capture the long-term effect, we specify an indicator variable, I(After the Sixth Week), which is an indicator for whether user i is after the sixth week since he/she adopted the premium version at day t. This variable always equals to 0 for those who never adopted. In addition, α_i captures individual fixed effect, γ_t captures date fixed effect, and ϵ_{it} captures unobserved random shocks on user i on day t. We cluster the standard error at individual level in all of our regression specifications.

4 Empirical Results

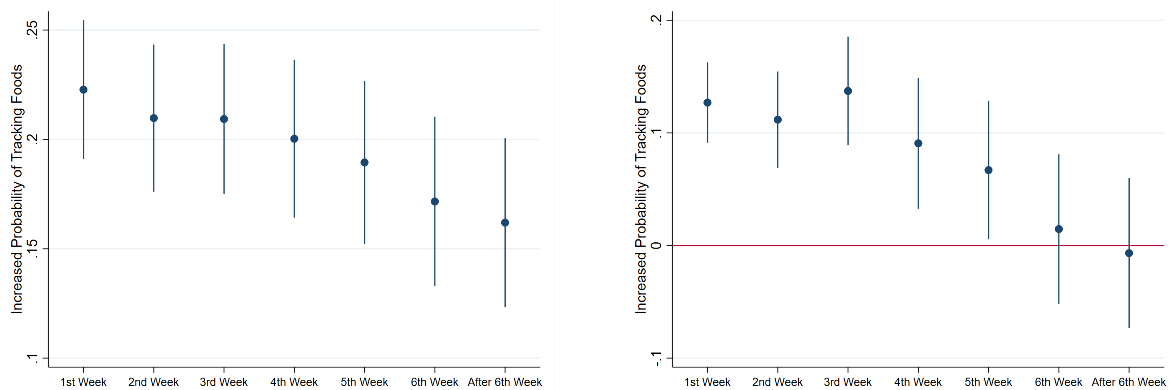
4.1 Baseline Specification

This section summarizes our main findings. The detailed estimates are provided in Table 5, Table 6 (analysis using full sample) and Table 7, Table 8 (analysis using matched sample). We report the results based on both the full sample and the matched sample to demonstrate how ignoring the potential endogeneity of upgrading affects estimation results.

To facilitate in the comparison across different specifications (e.g., full vs matched sample), we also illustrate, for both the full sample and the matched sample results, the impact of premium adoption on food tracking (Figure 3), exercise tracking (Figure 4), staying within calorie budget (Figure 5), exercise calories (Figure 6), lowering weight (Figure 7), and amount of weight lost (Figure 8).

Figure 3: Increased Probability of Tracking Food After Premium Upgrade

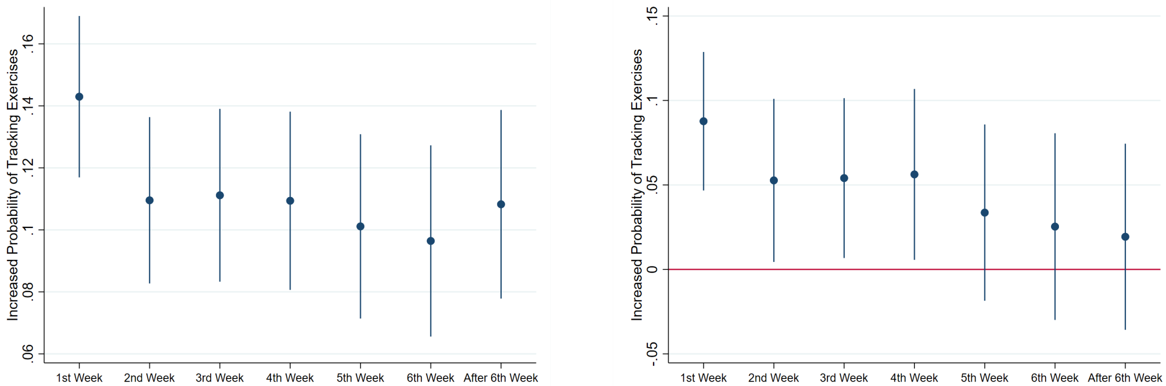
(a) Full Sample (b) Matched Sample



Notes: These plots show the effect of premium adoption on the probability of tracking food calories. The effects are plotted over time. Panel (a) provides the estimates from the full sample, while Panel (b) provides the estimates from the matched sample that corrects for endogeneity and selection biases.

Figure 4: Increased Probability of Tracking Exercises After Premium Adoption

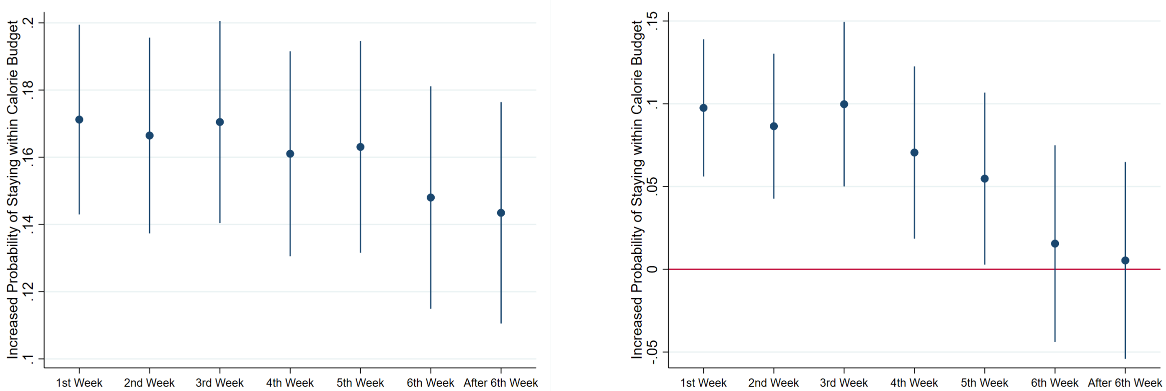
(a) Full Sample (b) Matched Sample



Notes: These plots show the effect of premium adoption on the probability of tracking exercise calories. The effects are plotted over time. Panel (a) provides the estimates from the full sample, while Panel (b) provides the estimates from the matched sample that corrects for endogeneity and selection biases.

Figure 5: Increased Probability of Staying Within Calorie Budget After Premium Upgrade

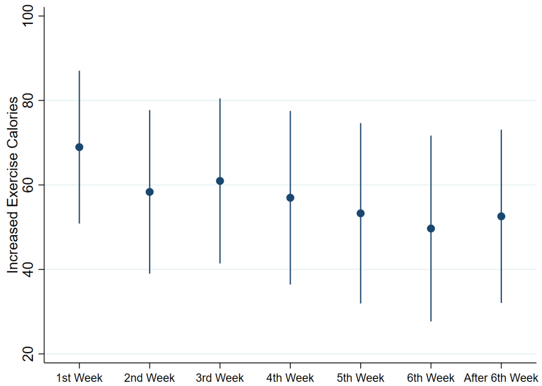
(a) Full Sample (b) Matched Sample



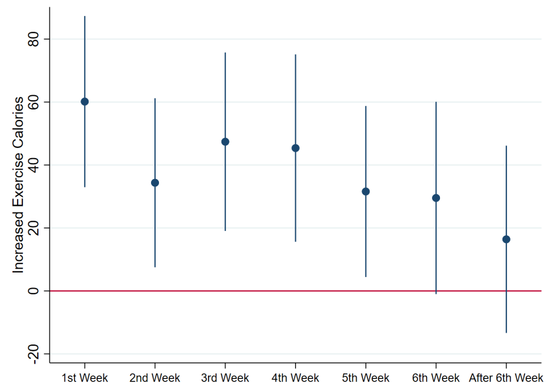
Notes: These plots show the effect of premium adoption on the probability of staying within the budget of calories. The effects are plotted over time. Panel (a) provides the estimates from the full sample, while Panel (b) provides the estimates from the matched sample that corrects for endogeneity and selection biases.

Figure 6: Increased Exercise Calories After Premium Upgrade

(a) Full Sample



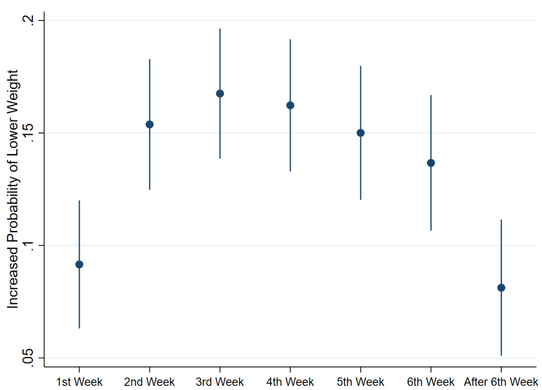
(b) Matched Sample



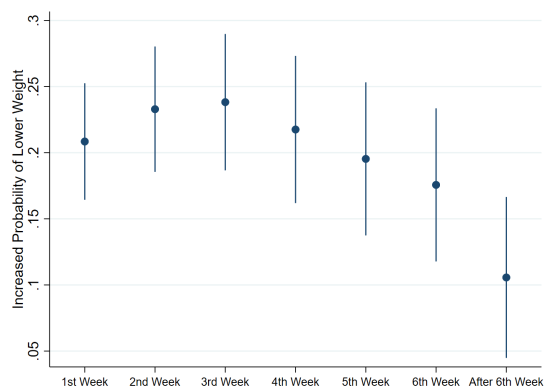
Notes: These plots show the effect of premium adoption on exercise calories. The effects are plotted over time. Panel (a) provides the estimates from the full sample, while Panel (b) provides the estimates from the matched sample that corrects for endogeneity and selection biases.

Figure 7: Increased Probability of Losing Weight After Premium Upgrade

(a) Full Sample



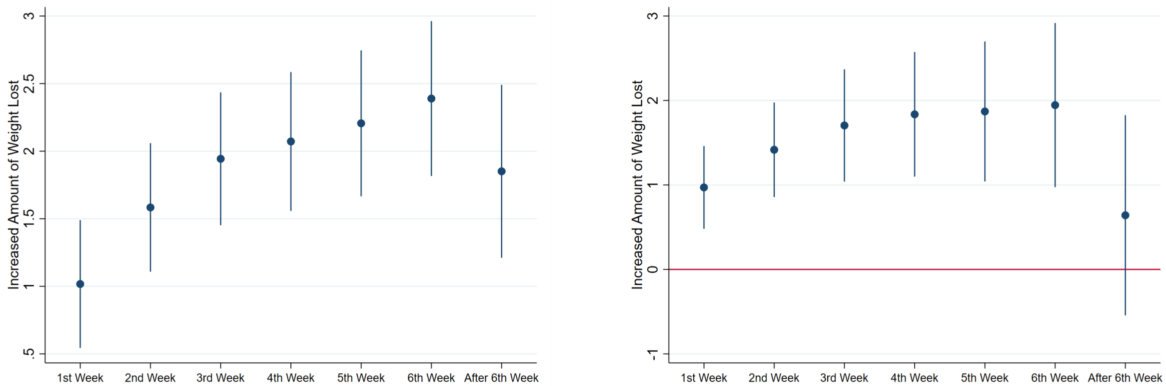
(b) Matched Sample



Notes: These plots show the effect of premium adoption on the probability of losing weight (relative to the initial weight). The effects are plotted over time. Panel (a) provides the estimates from the full sample, while Panel (b) provides the estimates from the matched sample that corrects for endogeneity and selection biases.

Figure 8: Increased Amount of Weight Lost

(a) Full Sample (b) Matched Sample



Notes: These plots show the effect of premium adoption on the amount of weight lost. The effects are plotted over time. Panel (a) provides the estimates from the full sample, while Panel (b) provides the estimates from the matched sample that corrects for endogeneity and selection biases.

The presented estimates in our figures and tables suggest that after upgrading to the premium version, users appear to be more engaged with the app and make more efforts in weight loss (through a higher probability of tracking food and exercises, tracking and staying within a daily calorie budget, and more exercise calories), compared with users who do not upgrade. In the full sample, those who upgrade to the premium version are 22.3 percentage points more likely to track foods within the first week, the increase in the probability of tracking food decreases to 20.0 percentage points within the fourth week, and 16.2 percentage points after the sixth week. Similarly, those who upgrade to the premium version are 14.3 percentage points more likely to track exercises within the first week and the increase in the probability of tracking exercises decreases to 10.8 percentage points after the sixth week. As for tracking and staying within the daily calorie budget, those who upgrade to the premium version are 17.1 percentage points more likely to track and stay within the daily calorie budget within the first week after upgrading, the increase in the probability of tracking and staying within the daily calorie budget decreases to 16.1 percentage points within the fourth week, and 14.3 percentage points after the sixth week. Lastly, those who upgrade to the premium version track 68.96 exercise calories more, and the increase in tracked exercise calories decreases to 56.97 and 49.69 calories within the fourth and the sixth week, respectively.

Next, we find that premium users are more successful in reducing weight than those who use the free version. In Table 6, using the full sample, we find that those who upgrade to the premium version are 9.16 percentage points more likely to lose weight from the initial weight within the first week, the increase in the probability of staying below the initial weight increases to 16.2 percentage points in the fourth week, but decrease to 8.12 percentage points after the sixth week. Similarly, those who upgraded to the premium version tend to have 0.95 pounds more weight loss within the first week, and the increase in the amount of weight loss increases to 2.08 pounds within the fourth week, 2.38 pounds within the sixth week, but decreases to 1.89 pounds after the sixth week.

Table 5: Regression Results for the Full Sample

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.223***	0.143***	0.171***	68.96***
	(0.0162)	(0.0133)	(0.0144)	(9.233)
Within the Second Week	0.210***	0.110***	0.166***	58.35***
	(0.0172)	(0.0137)	(0.0149)	(9.882)
Within the Third Week	0.209***	0.111***	0.170***	60.96***
	(0.0175)	(0.0142)	(0.0153)	(9.969)
Within the Fourth Week	0.200***	0.109***	0.161***	56.97***
	(0.0184)	(0.0147)	(0.0156)	(10.48)
Within the Fifth Week	0.189***	0.101***	0.163***	53.29***
	(0.0190)	(0.0152)	(0.0161)	(10.89)
Within the Sixth Week	0.172***	0.0964***	0.148***	49.69***
	(0.0198)	(0.0157)	(0.0169)	(11.22)
After the Sixth Week	0.162***	0.108***	0.143***	52.57***
	(0.0197)	(0.0155)	(0.0168)	(10.46)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
R-Squared Within	0.212	0.0642	0.121	0.0233
Observations	1199700	1199700	1199700	1199463
N. Individuals	11997	11997	11997	11997

Notes: Full sample consists of all observations before the matching procedure. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

To estimate the causal effect of the premium upgrade, we estimate the same regressions on the matched sample in order to estimate the LATE of premium upgrade on engagement and weight loss outcomes. As described in the previous section, we match users who upgraded during their 8th day to 21st day since registration with 10 similar users who never upgraded based on initial user characteristics and variables related to engagement and weight loss progress in the first seven days. Table 7 and Table 8 show that the inferred patterns from

Table 6: Effects of Premium Adoption on Weight Loss for Full Sample

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.0916***	0.950***
	(0.0145)	(0.233)
Within the Second Week	0.154***	1.581***
	(0.0148)	(0.242)
Within the Third Week	0.168***	1.950***
	(0.0147)	(0.250)
Within the Fourth Week	0.162***	2.079***
	(0.0150)	(0.261)
Within the Fifth Week	0.150***	2.201***
	(0.0152)	(0.275)
Within the Sixth Week	0.137***	2.379***
	(0.0154)	(0.292)
After the Sixth Week	0.0812***	1.892***
	(0.0154)	(0.325)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.124	0.350
Observations	1199700	1194252
N. Individuals	11997	11997

Notes: Full sample consists of all observations before the matching procedure. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

premium upgrades are similar to the results based on the full sample.

In the matched sample, compared with those who do not upgrade, after upgrading, those who upgrade to the premium version are 12.7 percentage points more likely to track foods within the first week, and the increase in the probability of tracking food decreases to 9.08 percentage points within the fourth week, and statistically insignificant after the fifth week. Similarly, those who upgraded to the premium version is 8.77 percentage points more likely to track exercises within the first week, and the increase in the probability of tracking exercises decreases to 5.62 percentage points within the fourth week. As for tracking and staying within the daily calorie budget, those who upgraded to the premium version are 9.75 percentage points more likely to track and stay within the daily calorie budget within the first week and the increase in the probability of tracking and staying within the daily calorie budget decreases to 7.05 percentage points within the fourth week. As for exercise calories, those who upgraded to the premium version tracked 60.15 exercise calories more and the

Table 7: Regression Results for the Matched Sample

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.127***	0.0877***	0.0975***	60.15***
	(0.0182)	(0.0209)	(0.0211)	(13.86)
Within the Second Week	0.112***	0.0527**	0.0864***	34.36**
	(0.0218)	(0.0246)	(0.0223)	(13.69)
Within the Third Week	0.137***	0.0541**	0.0997***	47.39***
	(0.0246)	(0.0241)	(0.0253)	(14.45)
Within the Fourth Week	0.0908***	0.0562**	0.0705***	45.38***
	(0.0296)	(0.0258)	(0.0265)	(15.17)
Within the Fifth Week	0.0670**	0.0336	0.0548**	31.58**
	(0.0313)	(0.0266)	(0.0265)	(13.85)
Within the Sixth Week	0.0146	0.0253	0.0155	29.54*
	(0.0339)	(0.0282)	(0.0303)	(15.57)
After the Sixth Week	-0.00677	0.0193	0.00534	16.40
	(0.0340)	(0.0281)	(0.0303)	(15.17)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
R-Squared Within	0.217	0.0664	0.121	0.0259
Observations	180400	180400	180400	180355
N. Individuals	1804	1804	1804	1804

Notes: Observations are only for users that are matched. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

increase in tracked exercise calories decreases to 45.38 calories within the fourth week. These effects are qualitatively similar to those based on the full sample as we report above.

Moreover, we find that the effects on weight loss in the matched sample are estimated to be generally similar to what we find based on the full sample. The current weight of premium users is lower than those who use the free version, suggesting the effectiveness of the premium version in helping users lose weight. For example, those who upgraded to the premium version are 20.8 percentage points more likely to stay below the initial weight within the first week, the increase in the probability of staying below the initial weight increases to 21.8 percentage points in the fourth week, decrease to additional 17.6 percentage points within the sixth week, and to 10.6 percentage points after the sixth week. However, we notice that in terms of the amount of weight lost, after the sixth week, the premium users do not have a significantly more amount of weight lost, compared with free users. This can be due

Table 8: Effects of Premium Adoption on Weight Loss Effectiveness for Matched Sample

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.208***	0.971***
	(0.0225)	(0.250)
Within the Second Week	0.233***	1.416***
	(0.0242)	(0.285)
Within the Third Week	0.238***	1.704***
	(0.0263)	(0.339)
Within the Fourth Week	0.218***	1.836***
	(0.0284)	(0.376)
Within the Fifth Week	0.195***	1.870***
	(0.0295)	(0.423)
Within the Sixth Week	0.176***	1.946***
	(0.0295)	(0.495)
After the Sixth Week	0.106***	0.641
	(0.0310)	(0.604)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.143	0.375
Observations	180400	180299
N. Individuals	1804	1804

Notes: Observations are only for users that are matched. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

to premium users becoming disengaged in weight loss efforts, after the short-term increase in engagement, while the free version users making progress and catching up in weight loss over time.

When comparing the estimates across specifications involving full versus matched samples, the dampened estimates in the matched sample specifications further confirm our aforementioned concerns about the important threats to identification revolving around endogeneity and selection. As Figure 4 shows, for instance, the increased probability of tracking exercise seems to stabilize after week 2 in the full sample, while it continues to decrease even after week 2 in the matched sample. The difference between the two specifications implies that users who upgrade are indeed different and they select the upgrading option. Thus, failure to account for these biases may lead researchers and the company to overstate the benefits of the premium adoption.

We also note that the inferred increase in engagement is declining in magnitude over time after the premium upgrade, which is qualitatively consistent with the findings that the sunk cost associated with payments can encourage engagement but its effects diminish quickly over time (Gourville and Soman 1998, Goli et al. 2022).

4.2 Potential Heterogeneity in Premium Adoption Effect

This section explores potential heterogeneity in the premium adoption effect across observable demographic groups, based on age[16] and gender. Table 9 provides the results from analysis about tracking, goal adherence and exercise outcomes, while Table 10 provides the results from analysis about weight loss outcomes. For all of this analysis, we use the matched sample via our propensity score approach.[17]

Table 9 shows that user engagement is very similar across age groups during the first three weeks since premium adoption, as the interaction between age and adoption week indicators is statistically insignificant for the most part. However, during and after the fourth week, we do see a pattern indicating that the longer-term effects of premium adoption on engagement seem present primarily among the older demographic. That is, the inferred dampening pattern we find in our main analysis (section 4) appears to be more applicable to the younger demographic for most proxies of user engagement (except exercise tracking). Furthermore, younger users who upgrade are markedly less engaged than similar users who never upgrade (except exercise tracking), during and after their fourth week since upgrading. These findings are consistent with a conjecture that older users have more incentive to maintain their health.

Table 10 shows that the effectiveness of premium adoption on health-related outcomes (e.g., weight loss and especially the amount of weight loss), might be muted for the older demographic. In contrast, weight loss amounts upon adopting the premium version appear more pronounced for the male demographic. These patterns appear consistent with the literature in nutrition sciences that has demonstrated slower basal metabolism rates with age (Kirkwood and Austad 2000), which can make older users slower in losing weight, even if they have higher long-term engagement compared with younger users. Thus, even though older users might have stronger incentives to lose weight, they potentially face greater physiological challenges to achieve weight loss.

[16] We use standardized age, which is defined as each user's age minus the sample mean, and then divided by the sample standard deviation.

[17] To gain enough sample size for power, in the DiD analysis with heterogeneity, we use users who upgrade within the first 21 days since registration, matched with similar users who never upgrade, based on initial users characteristics and engagement variables on the first day. We present this matching specification in our robustness check too.

Taken together, these patterns suggest that the premium version of the app will likely have similar effects on engagement across genders, and that much of the observed heterogeneity manifests through age. The results suggest that mHealth companies may want to provide more information on the tips for weight loss targeting to older users.

4.3 Robustness and Sensitivity Analysis

This section explores the robustness of our main findings to different specifications and matching samples. We also investigate the extent to which selection on unobservables might impact the main findings.

4.3.1 Sub-Sample of Users Who Upgrade

Here, we consider using only the sub-sample of users who have upgraded to premium, in case those who chose to upgrade are different from those who did not in unobservable factors. Since the users who upgrade (even at different timing) may be similar to each other, this analysis allows us to test the robustness of our main results. The results of the analysis can be found in Table 11 and Table 12. The results are qualitatively similar to the main results and hence our results are robust to the selection of matched samples.

4.3.2 Matching Sample

Next, we explore alternative matching samples for the DiD analysis. More specifically, we run the same DiD analysis on three different matched samples, matching early adopters (those who upgraded within the first seven days) with similar non-adopters (Table 13, Table 14), matching adopters who upgraded within the first 21 days with similar non-adopters (Table 15, Table 16) and matching late adopters (those who upgraded between the 36th and the 49th day) (Table 17, Table 18) with similar non-adopters.[18] The series of regressions in this section reveal how much the treatment effects vary by the timing of adoption.

We first match early adopters with similar non-adopters because there is a substantial proportion of premium users in our setting that upgraded to the premium version early on (within the first seven days since registration). Second, we consider adopters within the first 21 days because this specification can include a majority of the adopters in our observation period. Lastly, we examine late adopters (those who upgraded between the 36th and the

[18] The corresponding propensity score distribution before and after matching can be found in online appendices.

49th day) to see how a longer observation period of 35 days before the premium upgrading affects the treatment effects. Note that this allows us to construct engagement variables based on behavior within the first 35 days since registration (beyond engagement within the first seven days since registration, as in our main specification), along with the initial user characteristics, to match adopters and non-adopters. Exploring the robustness of results to these different matching processes also informs us about the potential caveats of the propensity matching approach due to endogenous timing. Moreover, in the Appendix, we explore the extent to which the common support assumption might hold for these alternative specifications as well.

Overall, these alternative specifications demonstrate the robustness of our main findings about engagement and effectiveness to various matching assumptions. There are however some subtle differences. We noticed that the premium users still have a significantly more amount of weight lost compared with free users, after the sixth week of the premium adoption. The magnitude of the increase after the sixth week is mostly smaller than the magnitude of the increase in the sixth week, suggesting the potential of free users catching up in weight loss in the long term.

4.3.3 Selection on Unobservables

An important qualifier for our matching approach is that the selection operates primarily via observed heterogeneity, as opposed to unobserved heterogeneity. This assumption is potentially strong, as selection on unobservables is plausible. To explore the validity of this assumption, we employ a bounds approach proposed by Oster (2019). This approach essentially confirms the extent to which the selection in upgrades is controlled for via the observed heterogeneity that our matching algorithm relies on.

To implement this bounds approach, we consider a simplified version of the main specification we estimate. That is, we label the treatment as simply the adoption of the premium version. Furthermore, this sensitivity analysis is conducted using the full (rather than matched) sample with the panel structure. For this reason, the magnitudes of the estimates from this analysis are not directly comparable to our main estimates; that said, we use this analysis to confirm whether or not the direction of the treatment effect is sensitive to selection on unobservables.

Table 19 provides a summary of the results from this sensitivity analysis. Along with the estimated "controlled" treatment effects, we provide standard errors obtained via bootstrapping. These estimates show that the direction for the upgrade effect seems unlikely to be purely driven by selection via unobservables, as the coefficients are all inferred to be positive (as in our main findings).

In summary, while selection via observables (only) is a strong assumption that remains commonly used in the causal inference literature with matching, we provide some analysis that may potentially assuage some of these concerns for our specific empirical context.

Table 9: Effects of Premium Adoption on Tracking, Goal Adherence, and Exercise Across Demographic Groups

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.0730***	0.0389*	0.0628***	28.35***
	(0.0185)	(0.0200)	(0.0187)	(9.150)
Within the Second Week	0.0637***	-0.00345	0.0613***	16.98*
	(0.0206)	(0.0212)	(0.0199)	(10.27)
Within the Third Week	0.0659***	0.00252	0.0677***	21.35**
	(0.0217)	(0.0216)	(0.0213)	(10.35)
Within the Fourth Week	0.0547**	0.00274	0.0685***	18.82*
	(0.0233)	(0.0220)	(0.0213)	(10.97)
Within the Fifth Week	0.0519**	0.000141	0.0712***	20.40*
	(0.0241)	(0.0225)	(0.0220)	(10.82)
Within the Sixth Week	0.0274	-0.000420	0.0496**	16.26
	(0.0255)	(0.0232)	(0.0235)	(10.61)
After the Sixth Week	0.0186	0.00469	0.0492**	20.25*
	(0.0247)	(0.0227)	(0.0233)	(10.46)
Within the First Week * Standardized Age	0.00272	-0.00799	-0.00490	4.292
	(0.0138)	(0.0150)	(0.0147)	(9.524)
Within the Second Week * Standardized Age	0.00982	0.00263	-0.00302	6.358
	(0.0158)	(0.0167)	(0.0153)	(9.447)
Within the Third Week * Standardized Age	0.0263	0.00690	0.0172	14.66
	(0.0167)	(0.0169)	(0.0161)	(9.893)
Within the Fourth Week * Standardized Age	0.0300	0.0179	0.0169	15.95*
	(0.0182)	(0.0171)	(0.0163)	(9.314)
Within the Fifth Week * Standardized Age	0.0442**	0.0224	0.0192	29.18***
	(0.0189)	(0.0176)	(0.0171)	(8.730)
Within the Sixth Week * Standardized Age	0.0400**	0.0154	0.0225	24.97**
	(0.0200)	(0.0180)	(0.0178)	(10.50)
After the Sixth Week * Standardized Age	0.0547***	0.0248	0.0330*	27.36***
	(0.0192)	(0.0175)	(0.0176)	(10.31)
Within the First Week * Male	-0.0210	0.0316	-0.0488	30.93
	(0.0311)	(0.0315)	(0.0312)	(23.64)
Within the Second Week * Male	0.0000951	0.0589*	-0.0300	36.59
	(0.0327)	(0.0348)	(0.0324)	(24.38)
Within the Third Week * Male	0.00927	0.0546	-0.0137	40.17
	(0.0365)	(0.0349)	(0.0350)	(24.52)
Within the Fourth Week * Male	0.0204	0.0521	-0.0376	36.77
	(0.0399)	(0.0375)	(0.0366)	(24.95)
Within the Fourth Week * Male	0.0179	0.0211	-0.0182	15.74
	(0.0407)	(0.0383)	(0.0372)	(24.71)
Within the Fourth Week * Male	0.00476	0.0184	-0.0305	4.756
	(0.0433)	(0.0392)	(0.0390)	(25.54)
After the Sixth Week * Male	0.0304	0.0567	-0.0115	21.97
	(0.0424)	(0.0384)	(0.0387)	(25.28)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
R-Squared Within	0.216	0.0633	0.124	0.0221
Observations	688600	688600	688600	688508
N. Individuals	6886	6886	6886	6886

Notes: Observations are only for users that are matched. Standardized age is defined as each user's age minus the sample mean, and then divided by the sample standard deviation. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

Table 10: Effects of Premium Adoption on Weight Loss Across Demographic Groups

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.200***	1.017***
	(0.0232)	(0.316)
Within the Second Week	0.310***	1.406***
	(0.0243)	(0.298)
Within the Third Week	0.321***	1.462***
	(0.0253)	(0.318)
Within the Fourth Week	0.308***	1.388***
	(0.0258)	(0.335)
Within the Fifth Week	0.296***	1.451***
	(0.0261)	(0.354)
Within the Sixth Week	0.278***	1.393***
	(0.0260)	(0.381)
After the Sixth Week	0.211***	0.241
	(0.0260)	(0.426)
Within the First Week * Standardized Age	-0.0163	-0.467**
	(0.0177)	(0.226)
Within the Second Week * Standardized Age	-0.0268	-0.472**
	(0.0184)	(0.228)
Within the Third Week * Standardized Age	-0.0180	-0.507**
	(0.0193)	(0.240)
Within the Fourth Week * Standardized Age	-0.0242	-0.611**
	(0.0199)	(0.250)
Within the Fifth Week * Standardized Age	-0.0316	-0.658**
	(0.0202)	(0.264)
Within the Sixth Week * Standardized Age	-0.0432**	-0.746***
	(0.0200)	(0.283)
After the Sixth Week * Standardized Age	-0.0336*	-0.544*
	(0.0196)	(0.329)
Within the First Week * Male	0.0487	0.00637
	(0.0404)	(0.562)
Within the Second Week * Male	0.0262	0.552
	(0.0422)	(0.549)
Within the Third Week * Male	0.0463	1.380**
	(0.0411)	(0.554)
Within the Fourth Week * Male	0.0612	1.788***
	(0.0417)	(0.576)
Within the Fourth Week * Male	0.0660	2.008***
	(0.0419)	(0.605)
Within the Fourth Week * Male	0.0753*	2.413***
	(0.0416)	(0.649)
After the Sixth Week * Male	0.0639	3.056***
	(0.0414)	(0.796)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.129	0.335
Observations	688600	687957
N. Individuals	6886	6886

Notes: Observations are only for users that are matched. Standardized age is defined as each user's age minus the sample mean, and then divided by the sample standard deviation. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

Table 11: Regression Results for the Upgraded Sample

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.146***	0.0922***	0.104***	46.33***
	(0.0155)	(0.0132)	(0.0138)	(8.873)
Within the Second Week	0.108***	0.0405***	0.0765***	28.57***
	(0.0172)	(0.0139)	(0.0148)	(9.868)
Within the Third Week	0.0797***	0.0243	0.0573***	23.23**
	(0.0184)	(0.0148)	(0.0160)	(10.15)
Within the Fourth Week	0.0449**	0.00555	0.0257	13.16
	(0.0202)	(0.0158)	(0.0170)	(10.88)
Within the Fifth Week	0.00947	-0.0198	0.00600	2.102
	(0.0218)	(0.0169)	(0.0183)	(11.50)
Within the Sixth Week	-0.0354	-0.0420**	-0.0330*	-9.030
	(0.0238)	(0.0182)	(0.0200)	(12.08)
After the Sixth Week	-0.151***	-0.0985***	-0.131***	-35.91***
	(0.0267)	(0.0197)	(0.0225)	(12.28)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
R-Squared Within	0.157	0.0366	0.0772	0.0131
Observations	113100	113100	113100	113094
N. Individuals	1131	1131	1131	1131

Notes: Observations are only for users that upgraded to the premium version. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

Table 12: Effects of Premium Adoption on Weight Loss for the Upgraded Sample

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.101***	1.341***
	(0.0149)	(0.239)
Within the Second Week	0.169***	2.131***
	(0.0159)	(0.255)
Within the Third Week	0.188***	2.642***
	(0.0165)	(0.274)
Within the Fourth Week	0.187***	2.908***
	(0.0176)	(0.299)
Within the Fifth Week	0.181***	3.191***
	(0.0187)	(0.327)
Within the Sixth Week	0.170***	3.497***
	(0.0200)	(0.355)
After the Sixth Week	0.125***	3.503***
	(0.0234)	(0.434)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.145	0.375
Observations	113100	112409
N. Individuals	1131	1131

Notes: Observations are only for users that upgraded to the premium version. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

Table 13: Regression Results for the Matched Early Adopters

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.0400**	0.0137	0.0410**	24.08*
	(0.0164)	(0.0209)	(0.0186)	(12.60)
Within the Second Week	0.0410**	-0.0199	0.0498**	20.24
	(0.0183)	(0.0223)	(0.0195)	(14.10)
Within the Third Week	0.0398*	-0.0153	0.0611***	23.22
	(0.0204)	(0.0234)	(0.0211)	(14.31)
Within the Fourth Week	0.0454**	-0.0178	0.0608***	19.08
	(0.0213)	(0.0241)	(0.0214)	(15.15)
Within the Fifth Week	0.0479**	-0.0272	0.0782***	16.77
	(0.0217)	(0.0246)	(0.0228)	(15.54)
Within the Sixth Week	0.0256	-0.0271	0.0559**	6.221
	(0.0234)	(0.0252)	(0.0232)	(14.54)
After the Sixth Week	0.0247	-0.00372	0.0615***	22.46
	(0.0234)	(0.0249)	(0.0237)	(14.97)
Individual FE	Yes	Yes	Yes	Yes
R-Squared Within	0.210	0.0602	0.121	0.0212
Observations	508200	508200	508200	508157
N. Individuals	5082	5082	5082	5082

Notes: Observations are only for users that are matched based on the early adopters. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

Table 14: Regression Results for the Matched Early Adopters

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.351***	1.045***
	(0.0213)	(0.257)
Within the Second Week	0.480***	1.618***
	(0.0234)	(0.241)
Within the Third Week	0.503***	1.930***
	(0.0234)	(0.263)
Within the Fourth Week	0.500***	1.931***
	(0.0230)	(0.285)
Within the Fifth Week	0.492***	2.075***
	(0.0227)	(0.309)
Within the Sixth Week	0.476***	2.110***
	(0.0227)	(0.337)
After the Sixth Week	0.400***	1.154***
	(0.0225)	(0.425)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.134	0.332
Observations	508200	507718
N. Individuals	5082	5082

Notes: Observations are only for users that are matched based on the early adopters. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

Table 15: Regression Results for the Matched Adopters
(Within the First 21 Days)

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.0663***	0.0478***	0.0479***	38.05***
	(0.0151)	(0.0159)	(0.0152)	(9.310)
Within the Second Week	0.0640***	0.0145	0.0526***	28.55***
	(0.0163)	(0.0170)	(0.0160)	(9.914)
Within the Third Week	0.0693***	0.0190	0.0645***	34.16***
	(0.0176)	(0.0173)	(0.0172)	(9.994)
Within the Fourth Week	0.0617***	0.0186	0.0575***	30.52***
	(0.0191)	(0.0180)	(0.0175)	(10.30)
Within the Fifth Week	0.0582***	0.00583	0.0666***	25.34**
	(0.0198)	(0.0184)	(0.0181)	(10.24)
Within the Sixth Week	0.0295	0.00433	0.0410**	17.56*
	(0.0210)	(0.0190)	(0.0192)	(10.33)
After the Sixth Week	0.0292	0.0221	0.0469**	27.20***
	(0.0208)	(0.0187)	(0.0192)	(10.36)
Individual FE	Yes	Yes	Yes	Yes
R-Squared Within	0.216	0.0632	0.123	0.0219
Observations	688600	688600	688600	688508
N. Individuals	6886	6886	6886	6886

Notes: Observations are only for users that are matched based on the those who adopted within the first 21 days. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

Table 16: Regression Results for the Matched Adopters (Within the First 21 Days)

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.214***	0.927***
	(0.0194)	(0.268)
Within the Second Week	0.315***	1.493***
	(0.0203)	(0.258)
Within the Third Week	0.334***	1.817***
	(0.0205)	(0.268)
Within the Fourth Week	0.325***	1.875***
	(0.0210)	(0.281)
Within the Fifth Week	0.314***	2.010***
	(0.0212)	(0.297)
Within the Sixth Week	0.299***	2.084***
	(0.0212)	(0.319)
After the Sixth Week	0.229***	1.148***
	(0.0212)	(0.375)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.128	0.334
Observations	688600	687957
N. Individuals	6886	6886

Notes: Observations are only for users that are matched based on the those who adopted within the first 21 days. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

Table 17: Regression Results for the Matched Adopters
(During the 36th to 49th Day)

	(1) Tracking Foods	(2) Tracking Exercises	(3) Staying Within Budget	(4) Exercise Calories
Within the First Week	0.269***	0.180***	0.206***	50.12**
	(0.0450)	(0.0375)	(0.0368)	(19.51)
Within the Second Week	0.250***	0.120***	0.170***	25.97
	(0.0521)	(0.0386)	(0.0467)	(21.10)
Within the Third Week	0.214***	0.0857**	0.161***	12.09
	(0.0546)	(0.0380)	(0.0487)	(19.61)
Within the Fourth Week	0.217***	0.110**	0.170***	34.07
	(0.0580)	(0.0437)	(0.0484)	(22.38)
Within the Fifth Week	0.160***	0.151***	0.146***	39.31*
	(0.0608)	(0.0417)	(0.0507)	(20.85)
Within the Sixth Week	0.166***	0.128***	0.152***	46.66*
	(0.0613)	(0.0445)	(0.0513)	(25.53)
After the Sixth Week	0.125*	0.0980*	0.129**	16.48
	(0.0655)	(0.0507)	(0.0553)	(22.41)
Individual FE	Yes	Yes	Yes	Yes
R-Squared Within	0.142	0.0424	0.0787	0.0189
Observations	66000	66000	66000	65998
N. Individuals	660	660	660	660

Notes: Observations are only for users that are matched based on the late adopters. Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in.

Table 18: Regression Results for the Matched Adopters (During the 36th to 49th Day)

	(1) Lower Weight	(2) Amount of Weight Lost
Within the First Week	0.0875**	1.723***
	(0.0415)	(0.570)
Within the Second Week	0.0528	2.289***
	(0.0449)	(0.707)
Within the Third Week	0.0379	2.554***
	(0.0438)	(0.783)
Within the Fourth Week	0.0244	2.715***
	(0.0404)	(0.891)
Within the Fifth Week	-0.0157	2.873***
	(0.0399)	(1.009)
Within the Sixth Week	-0.0396	3.067***
	(0.0410)	(1.096)
After the Sixth Week	-0.0723*	2.896**
	(0.0429)	(1.202)
Individual FE	Yes	Yes
Date FE	Yes	Yes
R-Squared Within	0.160	0.285
Observations	66000	65920
N. Individuals	660	660

Notes: Observations are only for users that are matched based on the late adopters. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far.

Table 19: Sensitivity Analysis via Oster (2019)

	Tracking Foods	Tracking Exercises	Staying Within Budget	Exercise Calories	Lower Weight	Weight Lost
Upgrade	0.488	0.762	0.775	642.777	0.410	0.649
	(0.00655)	(0.0147)	(0.0110)	(20.480)	(0.0103)	(0.0242)

Notes: The full sample is used for the Oster (2019) sensitivity analysis. The effects reported are for the controlled treatment effects, where treatment is defined as the adoption of premium (not necessarily adoption at a specific time). Tracking foods outcome is an indicator for whether user has tracked food calories in a given week, tracking exercises outcome is an indicator for whether user has tracked exercise calories in a given week, staying within budget is an indicator for whether user stayed within the suggested calorie limit (provided by the app upon registration), and exercise calories outcome is the amount of exercise user has engaged in. Lower weight outcome is an indicator for whether their weight is lower than the starting weight, while weight loss is the amount weight they have lost so far. Standard errors are obtained via bootstrapping.

5 Theoretical Analysis of Sunk Costs in mHealth

While we have established causal empirical evidence of dampening premium adoption effects on mHealth usage, the behavioral mechanism still remains unclear. In particular, might sunk costs be a candidate explanation for the patterns we see in the causal inference analysis?

To get more clarity on this matter, we develop a behavioral model of premium adoption and app usage that is used to better understand the predicted behavior under the presence of sunk costs. While this analysis will not perform a comprehensive decomposition of all plausible behavioral mechanisms, we aim to focus on understanding the extent to which behavior among premium adopters can be explained by the sunk cost fallacy relative to otherwise rational behavior.

5.1 Model

To investigate the possibility of sunk cost fallacy as a behavioral mechanism, we generalize a model originally proposed by Ho et al. (2018) to analyze the theoretical properties of mHealth usage behavior in the presence of mental accounting for sunk cost. By doing so, we compare this model with one where users are rational and not subject to sunk cost. Our model contains two main stages. In the first stage, users decide whether to adopt the basic or premium version of the app. Once this decision has been made, in stage two, users then proceed to use the app for tracking exercise and food calories.

5.1.1 Premium Adoption

Users decide upon registration whether or not to adopt the premium version of the app, as indicated by their decision $a = 1$ for adopting premium and $a = 0$ for not adopting premium (i.e., basic). Users will anticipate how much they engage with the app upon adoption of basic or premium, and this anticipated value for each app version option is given by $V^a(\mathbf{q}^*) - a\kappa$, where $V^a(\mathbf{q}^*)$ represents the user's expected long-run discounted utility from using the app, while $\mathbf{q}^* = \{q_t^*\}_{\forall t}$ represents the sequence of optimal usage frequencies across time t. Here, κ represents the sunk cost disutility associated with the fees from adopting the premium version. Moreover, the optimal usage frequencies are determined in the second stage of this model, as we will describe in the subsequent section.

$$V^1(\mathbf{q}^*) - V^0(\mathbf{q}^*) - \kappa > 0. \tag{3}$$

It is this premium adoption step where our theoretical framework departs from Ho et al. (2018), as users in our model have the *choice* about whether or not to pay the sunk cost (i.e., fee). This added flexibility we include in the theoretical analysis might help us better understand the extent to which sunk costs might act as an *ex ante* commitment device.

5.1.2 App Usage

Once the user has decided upon which version of the app will be adopted (i.e., basic vs premium), they then begin using the app by tracking their exercise and food calories. We will now describe two main scenarios our model aims to disentangle. The first is a situation where users are rational (i.e., no sunk costs), while the second is a situation where users exhibit mental accounting of sunk costs. By specifying these two scenarios, we will establish testable hypotheses based on the model primitives.

Rational behavior. Here, we will focus on users who have adopted the premium version, and investigate patterns of the usage trajectories over time. The decision for the user then is regarding *tracking frequency* decisions using the mHealth app (i.e., q_t) in each time period t over a planning horizon, 1,..., T. In each time period t (i.e., week since registering for app), the user's utility is written as:

$$U^a(q, t) = B^a(q, t), \tag{4}$$

where $B^a(q, t)$ captures the benefit, specific to app version, from usage. We let the benefit of usage from the basic and premium versions to be represented as a flexible functions of q_t and t as follows,

$$B^0(q_t, t) = \theta_1^0 q_t - \theta_2^0 q_t^2 + (\psi_1^0 t + \psi_2^0 t^2) q_t, \tag{5}$$

$$B^1(q_t, t) = \theta_1^1 q_t - \theta_2^1 q_t^2 + (\psi_1^1 t + \psi_2^1 t^2) q_t, \tag{6}$$

where $\theta_1^a, \theta_2^a > 0$, such that the marginal benefit with respect to usage is positive, and this marginal benefit diminishes with usage. The term $(\psi_1^a t + \psi_2^a t^2)$ represents the effect of time on marginal benefit, where the marginal benefit might decline with time. For example, as users are becoming more successful at weight loss with the help of the mHealth app, there may be less benefit from calorie tracking (i.e., less weight to lose). Another example for diminishing marginal benefit might be satiation/boredom with the app over time.

$$U^a(q,t) = \theta_1^a q_t - \theta_2^a q_t^2 + (\psi_1^a t + \psi_2^a t^2) q_t. \tag{7}$$

If the users are forward-looking, they will choose usage frequency q_t so as to maximize the cumulative utility from using the app, $\sum_{\tau=1}^{T} U(q_\tau, \tau)$, which can be written as

$$\sum_{\tau=1}^{T} U^a(q_\tau, \tau) = \sum_{\tau=1}^{T} \{\theta_1^a q_\tau - \theta_2^a q_\tau^2 + (\psi_1^a \tau + \psi_2^a \tau^2) q_\tau\}, \tag{8}$$

where maximizing with respect to q_t yields the optimal usage q_t^*:

$$q_t^* = \frac{1}{2\theta_2^a}(\theta_1^a + \psi_1^a t + \psi_2^a t^2). \tag{9}$$

Mental accounting for sunk costs. This section will provide a model that explicitly allows for the possibility of mental accounting for sunk costs associated with paying a fee. We note that the utility associated with not adopting, $U^0(q_t, t)$, is the same in the case with rational behavior (i.e., no mental accounting of fixed costs).

In this scenario, the utility for users who have adopted the premium version will be a function of usage and mental accounting, as given by

$$U^1(Q_t, q_t, t) = \begin{cases} B^1(q_t, t) - M(\kappa, Q_t), & \text{if } t \leq T_\kappa \\ B^1(q_t, t), & \text{if } t > T_\kappa. \end{cases}$$

Here, $Q_t = \sum_{\tau}^{t} q_\tau$ represents the cumulative mHealth usage, while κ is the psychological disutility of carrying the mental account of sunk costs (κ) associated with the fee. Behavioral economic theory posits that mental accounting for sunk costs will occur for a finite period, T_κ. Past this time horizon, the user's utility reverts back to the rational scenario. Similar to Ho et al. (2018), we define $M(\kappa, Q_t)$ to be,

$$M(\kappa, Q_t) = \lambda_1 Q_t + \lambda_2 \kappa Q_t, \tag{10}$$

where $M(\kappa, Q_t)$ decreases with Q_t and approaches 0 as $t \to T_\kappa$. With increased usage frequency of the mHealth app, the sunk cost becomes less salient. Therefore, within the mental accounting horizon, we can write utility as

$$U^1(Q_t, q_t, t) = \theta_1^1 q_t - \theta_2^1 q_t^2 + (\psi_1^1 t + \psi_2^1 t^2)q_t - \lambda_1 Q_t - \lambda_2 \kappa Q_t. \tag{11}$$

As before, we will assume that users are forward-looking, so their objective is to maximize $\sum_{\tau=1}^{T} U^1(Q_\tau, q_\tau, \tau)$. The user accounts for the effect of q_t on future utility through cumulative usage up to time t, $Q_t = \sum_{\tau}^{t} q_\tau$. Within the horizon T_κ, the optimal mHealth usage can be obtained via backward induction. That is, obtain $q^*_{T_\kappa}$, and then work backward to solve $q^*_{T_\kappa - 1}$, and so on. Using the recursive structure, we can solve for optimal usage within the mental accounting horizon to be:

$$q_t^* = \frac{1}{2\theta_2^1}(\theta_1^1 + \psi_1^1 t + \psi_2^1 t^2) - \lambda_1(T_\kappa - t + 1) - \lambda_2\kappa(T_\kappa - t + 1). \tag{12}$$

Thus, with mental accounting of sunk costs, the optimal usage of the mHealth app is

$$q_t^* = \begin{cases} \frac{1}{2\theta_2}[\theta_1^1 + \psi_1^1 t + \psi_2^1 t^2 - \lambda_1(T_\kappa - t + 1) - \lambda_2\kappa(T_\kappa - t + 1)], & \text{if } t \leq T_\kappa \\ \frac{1}{2\theta_2^1}(\theta_1^1 + \psi_1^1 t + \psi_2^1 t^2), & \text{if } t > T_\kappa. \end{cases}$$

Based on q_t^*, we see that if $\lambda_1 = \lambda_2 = 0$, the optimal tracking frequency is the same for the rational and mental accounting scenarios.

Given the sequence of optimal usage and cumulative usage patterns patterns, we define $V^1(\mathbf{q}^*) = \sum_{\tau=1}^{T} U^1(Q_\tau^*, q_\tau^*, \tau)$ and $V^0(\mathbf{q}^*) = \sum_{\tau=1}^{T} U^0(q_\tau^*, \tau)$. We use these definitions as utility is the same for the model without and with sunk costs when the premium version is not adopted, and in the case where premium is adopted, the model without sunk costs is nested within the model with sunk costs. When $\lambda_1 = \lambda_2 = 0$, the model with sunk costs collapses to the model without sunk costs.

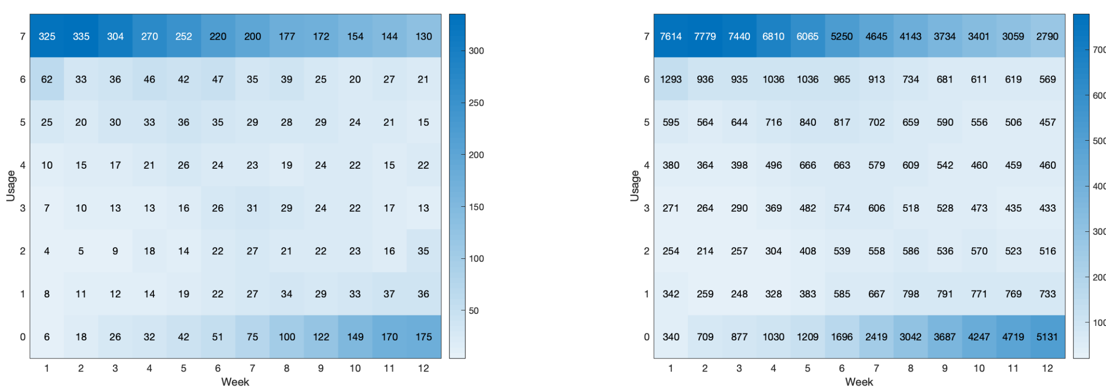
5.2 Insights from the Calibrated Model

We now explore some theoretical properties of usage behavior and the extent to which they are impacted by sunk costs. With the calibrated model, user engagement dynamics can be simulated. To obtain the parameters, we calibrate the model using our data about mHealth usage and premium adoption. For this calibration, we will leverage the entire sample of data (i.e., not just the matched users). For the time horizon, we set $T_\kappa = 12$ weeks, and t represents the week since a user signed up to use the app (basic or premium).

The model is calibrated by searching for model (i.e., target) parameters such that the differences between observed usage behavior q_t in week t and theoretically-predicted usage q_t^* in week t is minimized. To measure usage behavior, we use the number of days that food calories are tracked in week t. Furthermore, the calibrated model parameters need to satisfy incentive compatibility with respect to premium adoption, where these parameter search criteria are optimized using a minimum-distance estimator. Key informative moments in the data are summarized in Figure 9 and Figure 10 for observed q_t and Q_t, respectively. The variation in q_t and Q_t, their correlations with each other, and the extent to which this variation is different for premium versus basic users ultimately help us calibrate the target parameters. A general pattern that emerges from these heat maps is a dampening of usage (both weekly and cumulatively). The number of weeks for which all days of the week are tracked drops by about 50% for both premium and basic users. If a subset of the premium users are potentially subject to sunk cost biases, this pattern suggests that the dampening pattern itself is unlikely to be distinct from rational behavior, which all basic users are presumably following as they do not pay fees (i.e., sunk costs).

Figure 9: Heatmap Distribution for q_t Across Time

(a) Premium users | (b) Basic users



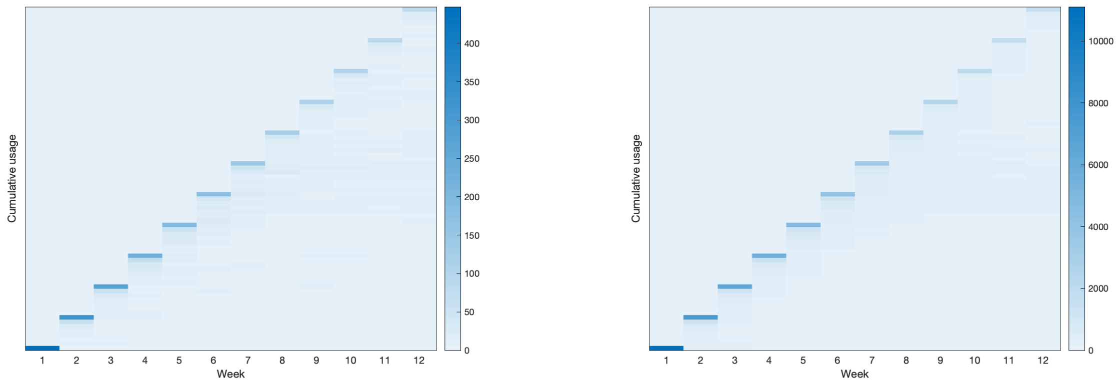
Notes: These plots present the distribution of observed q_t for premium (Panel a) and basic (Panel b) users.

The calibrated parameters are summarized in Table 20. These calibrated parameters confirm the presence of sunk costs, as $\kappa > 0$. Furthermore, users appear to mentally account for these sunk costs in their engagement decisions, as reflected by the non-zero parameters $\lambda_1 < 0$ and $\lambda_2 < 0$.

An assessment of model fit confirms that our calibrated model closely tracks the informative moments in our data. In particular, the usage patterns implied by our calibrated

Figure 10: Heatmap Distribution for Q_t Across Time

(a) Premium (b) Basic

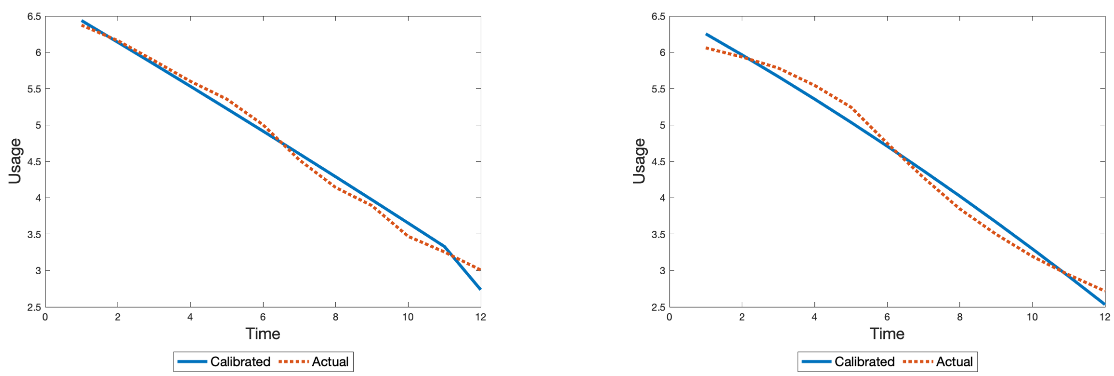


Notes: These plots present the distribution of observed Q_t for premium (Panel a) and basic (Panel b) users.

model mimic well the actual usage and cumulative usage patterns we observe in the data (Figure 11 and Figure 12), respectively. Here, we calculate the expected actual usage and cumulative usage patterns with cross-sectional averages over time.

Figure 11: Calibrated Model Fit Assessment for Usage

(a) Premium users (b) Basic users



Notes: These plots provide model fit assessments for usage among premium (Panel a) and basic (Panel b) users. Solid blue line represents the usage patterns over time as predicted by calibrated model, while the dotted red line represents the user-averaged actual usage patterns over time.

With the calibrated model, we simulate the usage trajectory q_t^* in a scenario with sunk costs, as well as for a scenario with no sunk costs (i.e., $\lambda_1 = \lambda_2 = 0$). The results from this simulation analysis are provided in Figure 13.

A few insights emerge from this simulation. First, the dampening patterns of usage emerge for scenarios with and without sunk costs. However, we see that usage is initially higher in the scenario with sunk costs, and this usage dampens to the point that it is reduced

Table 20: Calibrated Model

	Targets
θ_1^0	0.277
θ_2^0	0.021
ψ_1^0	-0.012
ψ_2^0	-0.000
θ_1^1	2.121
θ_2^1	0.328
ψ_1^1	-0.017
ψ_2^1	-0.001
λ_1	-0.107
λ_2	-0.038
κ	1.848

Notes: Calibrated parameters obtained by searching for parameters such that the differences between observed usage behavior q_t in week t and theoretically-predicted usage q_t^* in week t is minimized. To measure usage behavior, we use the number of days that food calories are tracked in week t. The calibrated model parameters also satisfy incentive compatibility with respect to premium adoption, where these parameter search criteria are optimized using a minimum-distance estimator.

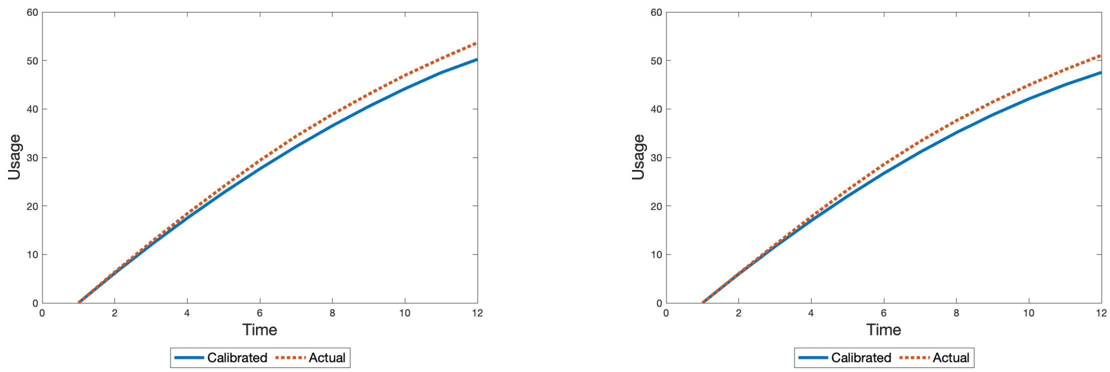
by about half the initial level. Absent the presence of sunk costs, premium users appear to not be particularly engaged with the app and start out close to the converged level. A similar plot, but for basic version users shows that the level of engagement seems high in general. This pattern suggests that users who have adopted the premium version might be purposely doing so as a commitment device via mental accounting of sunk costs. Thus, selection into the premium version with the intent to leverage mental accounting of sunk costs might offer one potential explanation for why our empirical findings in subsection 4.1 appear quite similar for full and matched sample analysis.

A plot of the difference (Figure 14) in usage for premium versus basic users reveals that this difference depends on whether sunk costs are present (solid blue line) or not (dashed red line). A comparison of these trajectories show that the increase in usage from premium adoption is noticeably more pronounced when sunk costs are present. In fact, the difference between premium and basic version usage is positive for the entire 12 weeks when sunk costs are present. When sunk costs are not present, we do not see an increase in engagement between premium and basic users. This finding suggests that sunk costs may be a key driver behind the increased engagement we observe when users adopt the premium version.

Therefore, we demonstrate that the inferred differences in mHealth engagement between premium and basic users can indeed be explained by mental accounting of sunk costs, espe-

Figure 12: Calibrated Model Fit Assessment for Cumulative Usage

(a) Premium users (b) Basic users



Notes: These plots provide model fit assessments for usage among premium (Panel a) and basic (Panel b) users. Solid blue line represents the cumulative usage patterns over time as predicted by calibrated model, while the dotted red line represents the user-averaged actual usage patterns over time.

cially so during the user's initial interactions with the app. While diminishing usage can be reflective of rational behavior, our model suggests that sunk costs is the key driver behind why premium users are more engaged with the app as compared with basic users. Moreover, our simulation results suggest that users might be *selecting* into premium adoption as a credible commitment to frequent engagement via anticipated mental accounting of sunk costs. This finding further corroborates the notion that the premium version of the app does not convey obvious functional benefits for the purpose of nutrition tracking, and that most of the benefits via improved engagement are likely psychological.

5.3 Implications on Premium Fee Design

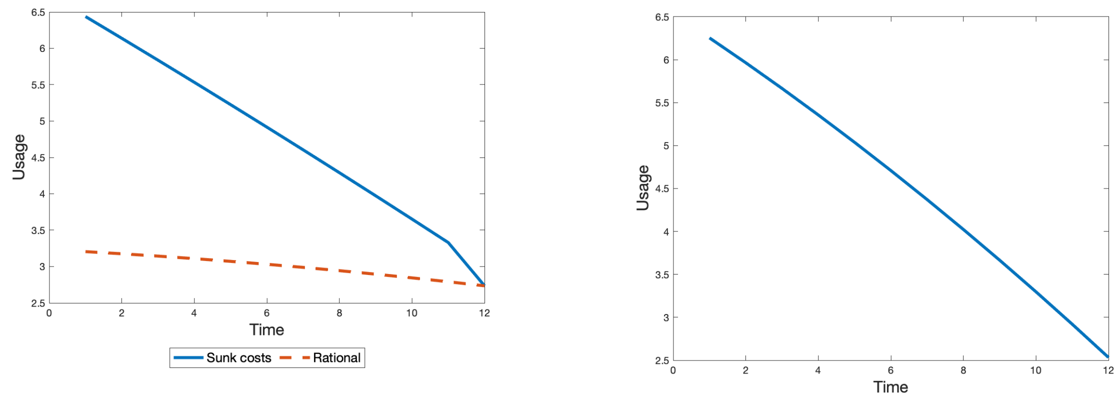
As we have demonstrated the role of sunk costs in explaining the observed increases in engagement patterns upon adopting premium, we now explore an immediate managerial implication. That is, how will changes in the premium fees impact user engagement as well as their incentives to adopt premium in the first place.

To investigate this managerial implication, we run additional simulations in which the sunk costs are increased or decreased by $r\%$. For each of these hypothetical scenarios, we then simulate the trajectory of usage patterns $\mathbf{q}^*$. With these simulated usage patterns, we can also quantify the incentives to adopt premium, as reflected by the incentive compatibility condition $V^1(\mathbf{q}^*) - V^0(\mathbf{q}^*) - \kappa$.

Simulated engagement across these hypothetical scenarios are summarized in Figure 15.

Figure 13: The Calibrated Impact of Sunk Costs Mental Accounting on Usage Dynamics

(a) Premium users (b) Basic users



Notes: These plots compare and assess the differences in usage with and without premium. Panel (a) provides the usage trajectories for premium users, while Panel (b) provides the usage trajectories for basic users. The scenario with rational behavior (i.e., no sunk costs) is obtained by setting $\lambda_1 = \lambda_2 = 0$.

Table 21: Incentives to Adopt Premium and Sunk Costs

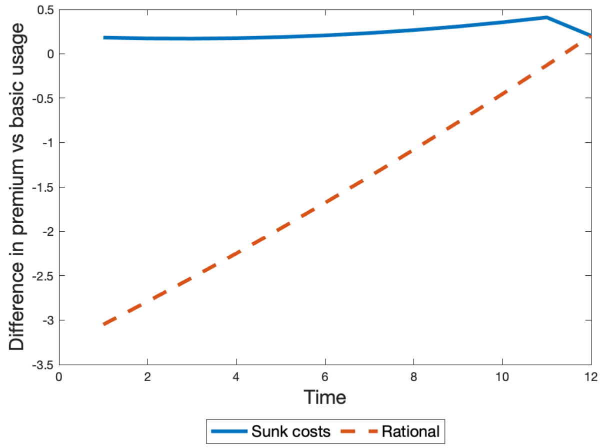
	Incentives
Baseline	75.888
Increase sunk cost	81.408
Decrease sunk cost	70.596

Notes: The difference in values from adopting premium versus not adopting (i.e., choosing basic) are presented for the baseline (row 1), fee increase (row 2), and fee decrease (row 3) scenarios. This incentive to adopt is calculated using $V^1(\mathbf{q}^*) - V^0(\mathbf{q}^*) - \kappa$, where $\mathbf{q}^*$ pertains to the engagement levels at κ (row 1), $\kappa \times (1+r)$ (row 2) or $\kappa \times (1-r)$ (row 3).

The plotted $\mathbf{q}^*$ pertains to the engagement levels at κ (solid blue line), $\kappa \times (1+r)$ (up-facing triangle marker) or $\kappa \times (1-r)$ (down-facing triangle marker). Here, $r = 0.25$, so fee changes are 25% of the calibrated κ. These plots show that an increase in fees (i.e., sunk costs) will lead to an increase in user engagement levels, while a decrease in fees will have an opposite effect. This finding confirms that the mHealth app can potentially use fees as a way to nudge users towards consistent usage of the app.

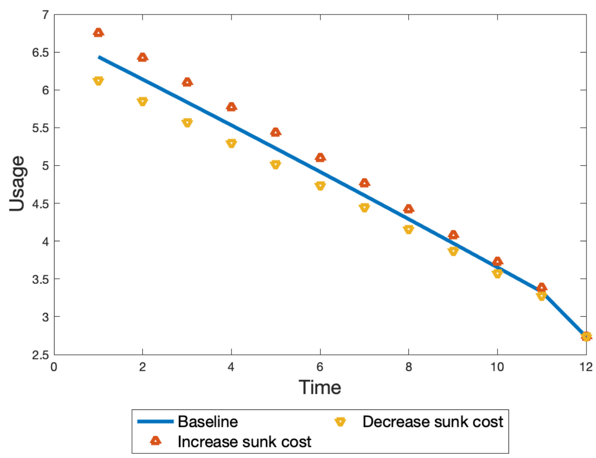
In fact, the anticipated elevated usage due to higher sunk costs might actually make actually make the premium version more appealing to users, as greater usage of the app will also raise the lifetime utility they receive from using the app. When we quantify the incentives to adopt, $V^1(\mathbf{q}^*) - V^0(\mathbf{q}^*) - \kappa$, it is apparent that raising the fees can actually increase these incentives Table 21. Therefore, the mHealth app might be able to attract *more* users to the premium option with *higher* fees. So in the presence of sunk cost mental

Figure 14: Differences in Premium and Basic Usage



Notes: This plot provides the differences in premium versus basic usage over time when sunk costs are present (solid blue line) versus when users are rational (dashed red line). The scenario with rational behavior (i.e., no sunk costs) is obtained by setting $\lambda_1 = \lambda_2 = 0$.

Figure 15: Usage Patterns at Different Fee Levels

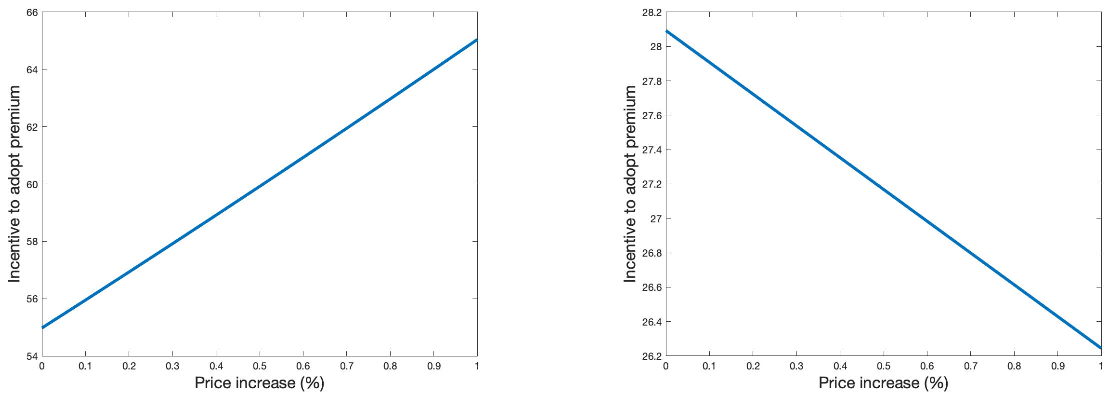


Notes: The plotted $\mathbf{q}^*$ pertains to the engagement levels at κ (solid blue line), $\kappa \times (1 + r)$ (up-facing triangle marker) or $\kappa \times (1 - r)$ (down-facing triangle marker). Here, $r = 0.25$, so fee changes are 25% of the calibrated κ.

accounting, the demand curve for the premium version might have the "counter-intuitive" upward price elasticity slope.

Figure 16: Incentives to Adopt Premium as Fees Increase

(a) Sunk costs　　　　(b) Rational



Notes: The difference in values from adopting premium versus not adopting (i.e., choosing basic) are presented in the plots. This incentive to adopt is calculated using $V^1(\mathbf{q}^*) - V^0(\mathbf{q}^*) - \kappa$, where $\mathbf{q}^*$ pertains to the engagement levels at $\kappa \times (1+r)$.

The counter-intuitive positive price elasticity for the premium version disappears completely if we eliminate the mental accounting of sunk costs. We demonstrate this property by considering fee increases, $\kappa \times (1+r)$ for $r \in [0, 1]$. For each hypothetical increase in the sunk costs, we consider users who are subject to mental accounting of sunk costs (Panel a), and users who are rational (Panel b). These plots are provided in Figure 16. These simulations show that rational users are less likely to adopt the premium version of the app as the fee increases, unlike users who are impacted by sunk costs. Therefore, an important boundary condition for a strategy of boosting user engagement via fee increases would be limited by the number of users who are rational.

5.4 Interpretation of Selection into Premium

Our empirical analysis in section 4 largely relied on matched samples via propensity score matching. We now explore how the use of the matched sample might impact the main theoretical insights from our calibrated simulations of the model. Many of the key patterns about the engagement dynamics and calibrated model fit remain unchanged, so we will focus our discussion on results that appear to be distinct from the calibration based on full sample.[19] Note that we focused our main discussions around the full sample, as it offers a

[19]The plots for the engagement dynamics using the matched sample calibration are virtually indistinguishable from the plots generated via full sample calibration. Details of these plots can be found in the Appendix.

Table 22: Calibrated Model using Matched Sample

	Targets
θ_1^0	1.337
θ_2^0	0.102
ψ_1^0	-0.050
ψ_2^0	-0.001
θ_1^1	1.678
θ_2^1	0.272
ψ_1^1	-0.007
ψ_2^1	-0.001
λ_1	-0.080
λ_2	-0.005
κ	13.386

Notes: Data used is the matched sample from the propensity score analysis (subsection 4.1). Calibrated parameters obtained by searching for parameters such that the differences between observed usage behavior q_t in week t and theoretically-predicted usage q_t^* in week t is minimized. To measure usage behavior, we use the number of days that food calories are tracked in week t. The calibrated model parameters also satisfy incentive compatibility with respect to premium adoption, where these parameter search criteria are optimized using a minimum-distance estimator.

more general perspective about the usage dynamics.

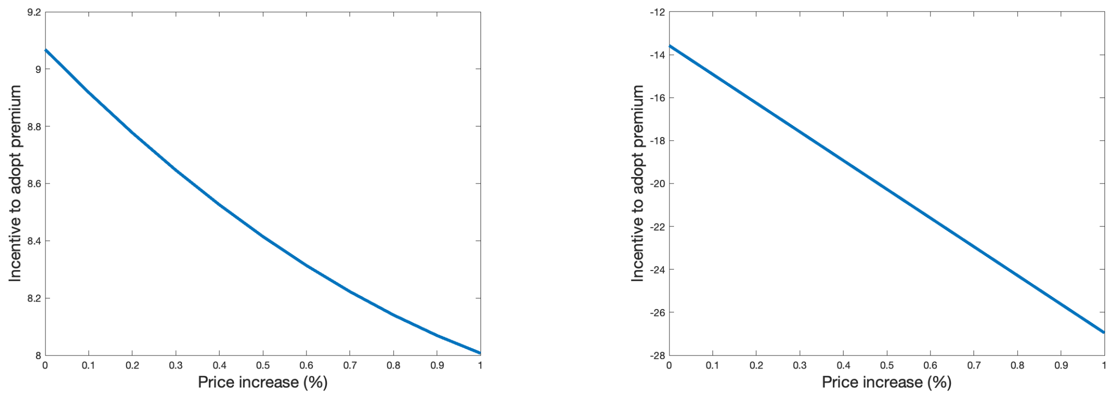
The first contrasting insight we point out pertains to the estimates relating to the sunk cost. For the matched sample (Table 22), the sunk cost is calibrated to be larger than in the calibration involving the full sample (Table 20). Therefore, much of the selection into premium adoption that we see might operate through sunk costs.

A consequence of the larger sunk costs is that the accrued utility via mental accounting of sunk costs is unlikely to compensate for increases in the fees, as reflected by the downward-sloping incentives for adopting premium (Figure 17). Furthermore, we see that in the scenario without mental accounting (i.e., rational behavior), the incentive to adopt premium virtually disappears as the plotted line is everywhere negative. In contrast, this plotted line remains above zero when there is mental accounting of sunk costs.

We interpret these findings as illustrating the important role of sunk costs in both engagement dynamics (as usage patterns remain unchanged) as well as premium adoption incentives. If selection is primarily operating through sunk costs, the mHealth app may still want to take note of our main findings using the full sample, as ignoring individuals with lower sunk costs may preclude the app from extracting benefits from users via fee increases.

Figure 17: Incentives to Adopt Premium as Fees Increase for Matched Sample

(a) Sunk costs (b) Rational



Notes: Calibrated model based on the matched sample from the propensity score analysis (subsection 4.1). The difference in values from adopting premium versus not adopting (i.e., choosing basic) are presented in the plots. This incentive to adopt is calculated using $V^1(\mathbf{q}^*) - V^0(\mathbf{q}^*) - \kappa$, where $\mathbf{q}^*$ pertains to the engagement levels at $\kappa \times (1 + r)$.

6 Conclusion

Using data from a popular freemium mHealth fitness tracker, we study the impact of premium adoption on user engagement and weight loss outcomes in the app. Our DiD analysis confirms that adoption will indeed have a positive impact on various dimensions of user engagement, such as food tracking, exercise tracking, budget adherence, exercise calories and weight loss. Although the premium version of the mHealth app is effective in helping users achieve sustainable weight loss, our results point to potential dampening effects over time of premium adoption on user engagement; especially so among the younger demographics. At the same time, the actual effectiveness of premium adoption on weight loss outcomes is heterogeneous among users. Moreover, the premium version can positively impact users' short-term engagement, thereby facilitating weight loss progress sooner for premium users as compared with free users in the short term. In the long-term, the effects of the premium version on weight loss decline slightly, as free version users gradually catch up in weight loss. We later demonstrate via simulations of a calibrated theoretical model of mHealth usage that the higher usage we observe for premium users (as compared with basic users) can indeed be explained by mental accounting of sunk costs. Absent mental accounting of sunk costs, our calibrated model shows that premium users would not exhibit higher engagement patterns with the mHealth app; this finding reinforces the notion that the premium version of the mHealth app likely offers little functional value to users, and that any value that users extract out of the app are likely psychological. Moreover, the presence of sunk cost mental accounting allows the mHealth app raise user engagement by increasing the premium version fees; in fact, we show that not only will engagement increase with fees, but so too will be users' incentives to adopt the paid premium version provided that they are subject to the mental accounting of sunk costs. In summary, our findings illustrate the potentially important role of sunk costs as a mechanism behind user engagement on mHealth apps.

The inferred dampening patterns of the premium adoption effect suggest that any impact on user engagement might be short-lived. The lack of long-run persistence in these effects suggest that the managerial implications of using premium adoption as a way to retain active users might be limited. To confirm this conjecture, a natural follow up question would be to assess the extent to which active engagement (i.e., diligent tracking) is actually associated with retention (i.e., not completely quitting on the app), and thus, CLV. We believe a study that contains data from a wider time horizon might be better suited to answer such a question. In addition, the heterogeneous effects of premium upgrade on user engagement and weight loss outcomes suggests that mHealth companies can use different strategies towards different user segments to enhance outcomes.

This research has largely focused on the extensive margins of user engagement (i.e., tracking, goal adherence, weight loss). However, there might be more qualitative intensive margins worth studying, such as the quality of tracking and content of tracked food/exercise. For example, are premium users more accurate in entering the foods that they actually ate. Furthermore, are the foods logged by premium users of healthier quality? Answering these questions will allow researchers to form a closer link between freemium app design and health outcomes.

We also acknowledge the potential limitations of relying on the assumptions behind propensity score matching, such as conditional independence. In particular, future work using better data could explore plausibly exogenous variation in the availability of the premium version, for example, due to app outages. Existing research about mobile marketing suggests that app outages can potentially impact user behavior (Narang et al. 2020). Furthermore, having access to each user's geographic coordinates would allow researchers to use weather shocks to instrument for the total exercise calories tracked (Aral and Nicolaides 2017, Uetake and Yang 2020).

As data from this mHealth app is potentially useful for food recommendation system design (Ling et al. 2022, Nielsen et al. 2022), it might be meaningful to take into consideration the observation that premium users are in general more engaged with the app relative to basic users. While basic users would still constitute a majority of the data used for training these recommendation systems, many of the "repeat" observations of the same user might be disproportionately represented by the premium users. Without good controls for intrinsic motivation, training these recommendation systems might be fraught by analogous issues as those pertaining to endogeneity that motivated our empirical approach.

References

Appel G, Libai B, Muller E, Shachar R (2019) Retention and the monetization of apps. *International Journal of Research in Marketing* 20:79.

Aral S, Nicolaides C (2017) Exercise contagion in a global social network. *Nature communications* 8(1):1–8.

Aswani A, Kaminsky P, Mintz Y, Flowers E, Fukuoka Y (2019) Behavioral modeling in weight loss interventions. *European journal of operational research* 272(3):1058–1072.

Bojd B, Song X, Tan Y, Yan X (2022) Gamified challenges in online weight-loss communities. *Information Systems Research* .

Bollinger B, Liebman E, Hammond D, Hobin E, Sacco J (2022) Educational campaigns for product labels: Evidence from on-shelf nutritional labeling. *Journal of Marketing Research* 59(1):153–172.

Chintala SC, Liaukonyte J, Yang N (2021) Browsing the aisles or browsing the app? how online grocery shopping is changing what we buy. *How Online Grocery Shopping is Changing What We Buy (December 23, 2021)* .

Daljord Ø, Mela CF, Roos JM, Sprigg J, Yao S (2022) The design and targeting of compliance promotions. *Marketing Science* forthcoming.

Datta H, Knox G, Bronnenberg BJ (2018) Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science* 37(1):5–21.

Faulkner C (2019) Fitbit's new premium subscription service hopes to sway you with personalized data, challenges, and more. *The Verge* .

Ghose A, Guo X, Li B, Dang Y (2021) Empowering patients using smart mobile health platforms: Evidence from a randomized field experiment. *arXiv preprint arXiv:2102.05506* .

Goli A, Chintagunta PK, Sriram S (2022) Effects of payment on user engagement in online courses. *Journal of Marketing Research* 59(1):11–34.

Gordon BR, Sun B (2015) A dynamic model of rational addiction: Evaluating cigarette taxes. *Marketing Science* 34(3):452–470.

Gourville JT, Soman D (1998) Payment depreciation: The behavioral effects of temporally separating payments from consumption. *Journal of consumer research* 25(2):160–174.

Haws KL, Liu PJ, Redden JP, Silver HJ (2017) Exploring the relationship between varieties of variety and weight loss: When more variety can help people lose weight. *Journal of Marketing Research* 54(4):619–635.

Ho DE, Imai K, King G, Stuart EA (2011) MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software* 42(8):1–28, URL http://dx.doi.org/10.18637/jss.v042.i08.

Ho TH, Png IP, Reza S (2018) Sunk cost fallacy in driving the world's costliest cars. *Management Science* 64(4):1761–1778.

Huang G, Khwaja A, Sudhir K (2015) Short-run needs and long-term goals: A dynamic model of thirst management. *Marketing Science* 34(5):702–721.

Huyghe E, Verstraeten J, Geuens M, Van Kerckhove A (2017) Clicks as a healthy alternative to bricks: how online grocery shopping reduces vice purchases. *Journal of Marketing Research* 54(1):61–74.

Iyengar R, Park YH, Yu Q (2022) The impact of subscription programs on customer purchases. *Journal of Marketing Research* .

Kapoor A, Manchanda P, Narayanan S (2022) Can a human coach help you lose more weight than an ai coach: Empirical evidence from a mobile fitness tracking app. *Working paper* .

Kato-Lin YC, Abhishek V, Downs JS, Padman R (2016) Food for thought: The impact of m-health enabled interventions on eating behavior. *Available at SSRN 2736792* .

Khan R, Misra K, Singh V (2016) Will a fat tax work? *Marketing science* 35(1):10–26.

Kirkwood T, Austad S (2000) Why do we age? *Nature* 408:233–238.

Labonté K, Knäuper B, Dubé L, Yang N, Nielsen DE (2022) Adherence to a caloric budget and body weight change vary by season, gender, and bmi: an observational study of daily users of a mobile health app. *Obesity Science & Practice* .

Lambrecht A, Misra K (2017) Fee or free: When should firms charge for online content? *Management Science* 63(4):1150–1165.

Li H, Jain S, Kannan P (2019) Optimal design of free samples for digital products and services. *Journal of Marketing Research* 56(3):419–438.

Ling Y, Nie JY, Nielsen D, Knäuper B, Yang N, Dubé L (2022) Following good examples-health goal-oriented food recommendation based on behavior data. *Proceedings of the ACM Web Conference 2022*, 3745–3754.

Liu PJ, Inman JJ, Li B, Wong CA, Yang N (2022) Consumer health in the digital age. *Journal of the Association for Consumer Research* 7(2):198–209.

Ma Y, Ailawadi KL, Grewal D (2013) Soda versus cereal and sugar versus fat: drivers of healthful food intake and the impact of diabetes diagnosis. *Journal of Marketing* 77(3):101–120.

McCarthy DM, Fader PS, Hardie BG (2017) Valuing subscription-based businesses using publicly disclosed customer data. *Journal of Marketing* 81(1):17–35.

Mintz Y, Aswani A, Kaminsky P, Flowers E, Fukuoka Y (2020) Nonstationary bandits with habituation and recovery dynamics. *Operations Research* 68(5):1493–1516.

Murnane EL, Huffaker D, Kossinets G (2015) Mobile health apps: adoption, adherence, and abandonment. *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 261–264.

Narang U (2022) Incentives to work out: Evidence from field experiments. *Working paper* .

Narang U, Shankar V (2019) Mobile marketing 2.0: State of the art and research agenda. *Marketing in a Digital World* .

Narang U, Shankar V, Narayanan S (2020) How does mobile app failure affect purchases in online and offline channels? *Working paper* .

Nielsen D, Yang N, Dube L, Knäuper B, Ling Y, Nie JY (2022) Consumption variety in food recommendation. *Journal of the Association for Consumer Research* 7(4).

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2):187–204.

Puranam D, Narayan V, Kadiyali V (2017) The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science* 36(5):726–746.

Rao A, Wang E (2017) Demand for "healthy" products: False claims and ftc regulation. *Journal of Marketing Research* 54(6):968–989.

Seiler S, Tuchman A, Yao S (2021) The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. *Journal of Marketing Research* 58(1):22–49.

Tang J, Abraham C, Stamp E, Greaves C (2015) How can weight-loss app designers' best engage and support users? a qualitative investigation. *British journal of health psychology* 20(1):151–171.

Uetake K, Yang N (2018) Harnessing the small victories: Goal design strategies for a mobile calorie and weight loss tracking application. *Available at SSRN 2928441* .

Uetake K, Yang N (2020) Inspiration from the "biggest loser": Social interactions in a weight loss program. *Marketing Science* 39(3):487–499.

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1):3–36.

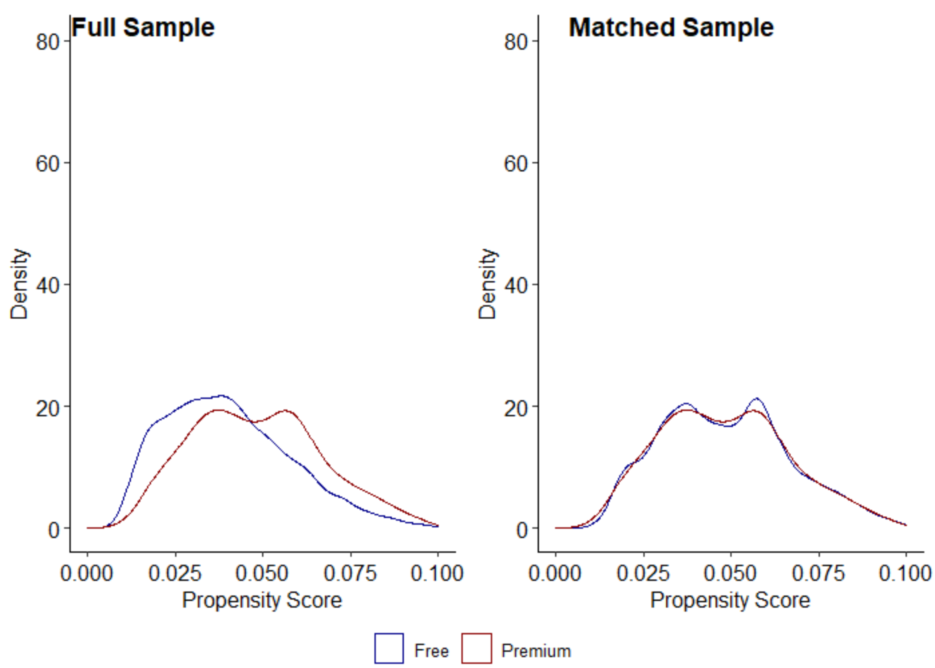
Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, Aswani A, et al. (2018) Evaluating machine learning–based automated personalized daily step goals delivered through a mobile phone app: Randomized controlled trial. *JMIR mHealth and uHealth* 6(1):e9117.

Online Appendix

Additional Figures

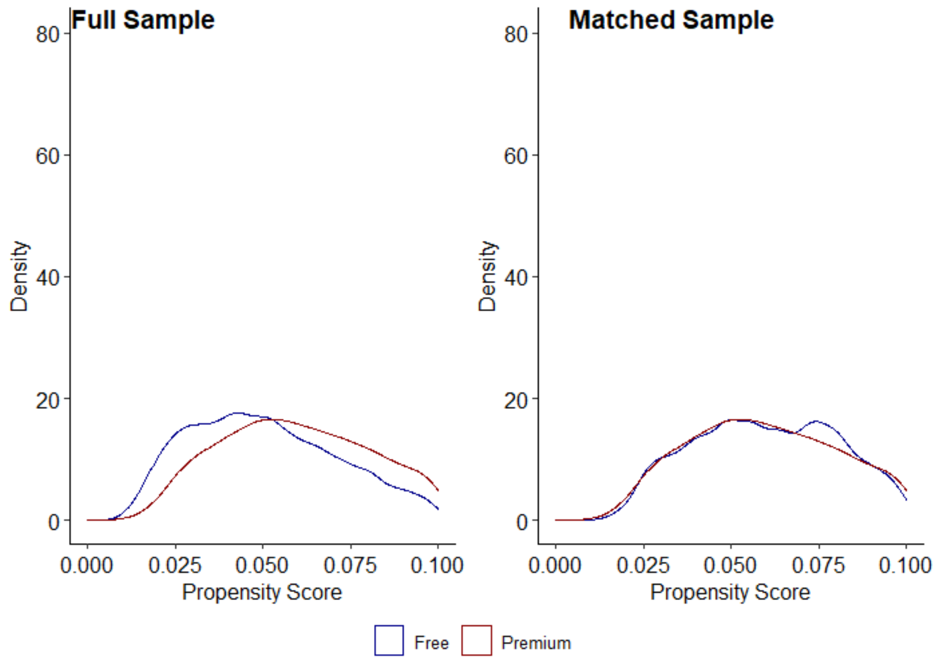
This section provides additional details about the alternative matching specifications we consider in our robustness check. Below are distribution of propensity score before and after matching, for the alternative approaches we use for matching include matching early adopters (those who upgraded within the first seven days) with similar non-adopters (Figure 18), matching adopters who upgraded within the first 21 days with similar non-adopters (Figure 19), and matching late adopters (those who upgraded between the 36th and the the 49th day) with similar non-adopters (Figure 20). In all of these alternative matching options, our density plots of the propensity score weights suggest that the common support assumption might hold for these alternative specifications as well.

Figure 18: Distribution of Propensity Scores Before and After Matching - Early Adopters



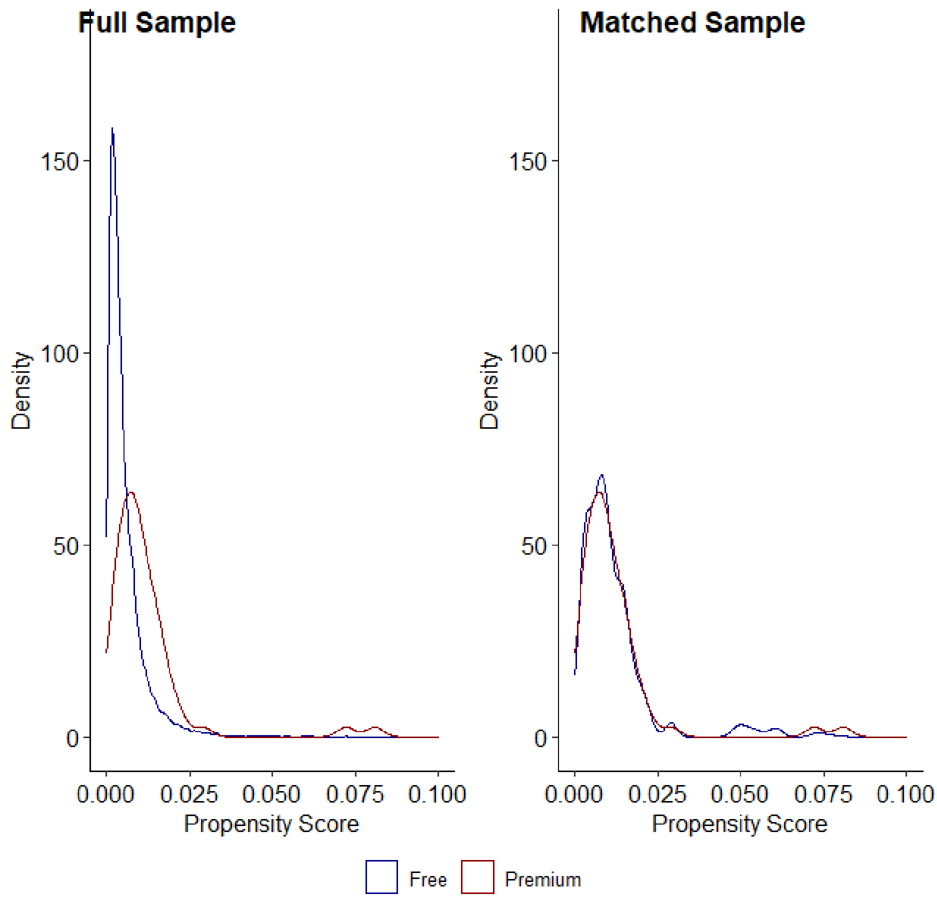
Notes: Distributions are obtained non-parametrically.

Figure 19: Distribution of Propensity Scores Before and After Matching - Adopters Within the First 21 days



Notes: Distributions are obtained non-parametrically.

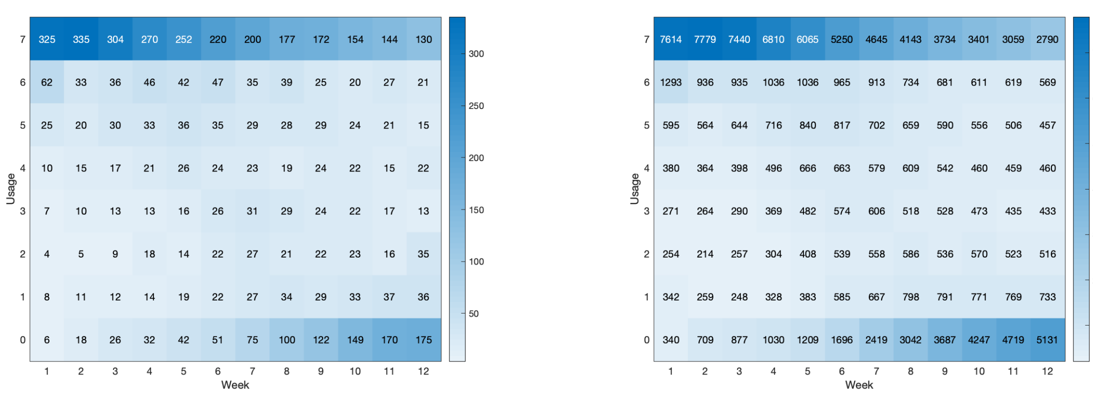
Figure 20: Distribution of Propensity Scores Before and After Matching - Late Adopters



Notes: Distributions are obtained non-parametrically.

Figure 21: Heatmap Distribution for q_t Across Time using Matched Sample

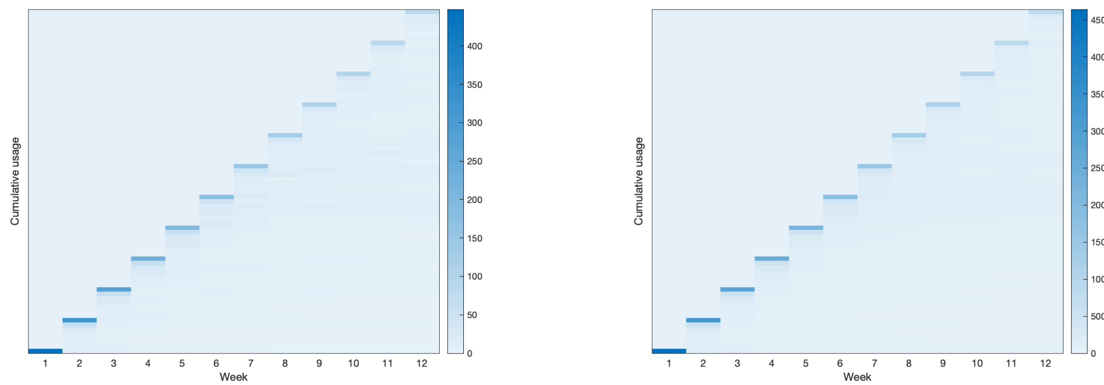
(a) Premium users (b) Basic users



Notes: These plots present the distribution of observed q_t for premium (Panel a) and basic (Panel b) users for the matched sample.

Figure 22: Heatmap Distribution for Q_t Across Time using Matched Sample

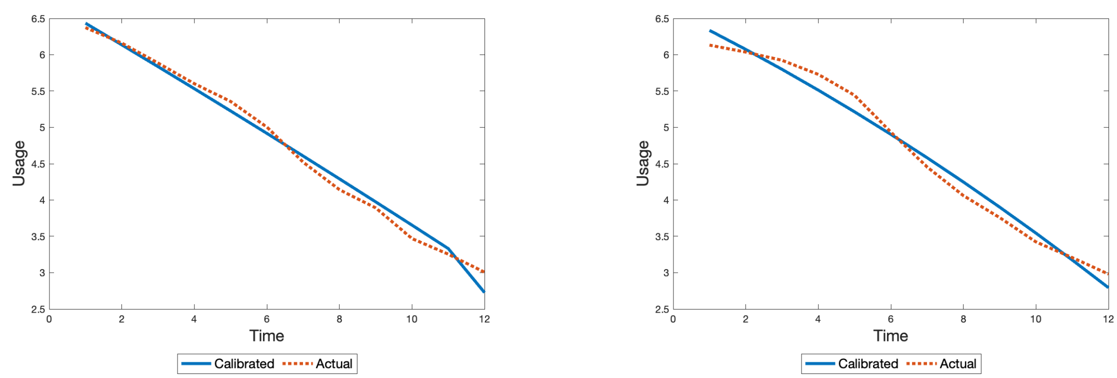
(a) Premium (b) Basic



Notes: These plots present the distribution of observed Q_t for premium (Panel a) and basic (Panel b) users for the matched sample.

Figure 23: Calibrated Model Fit Assessment for Usage using Matched Sample

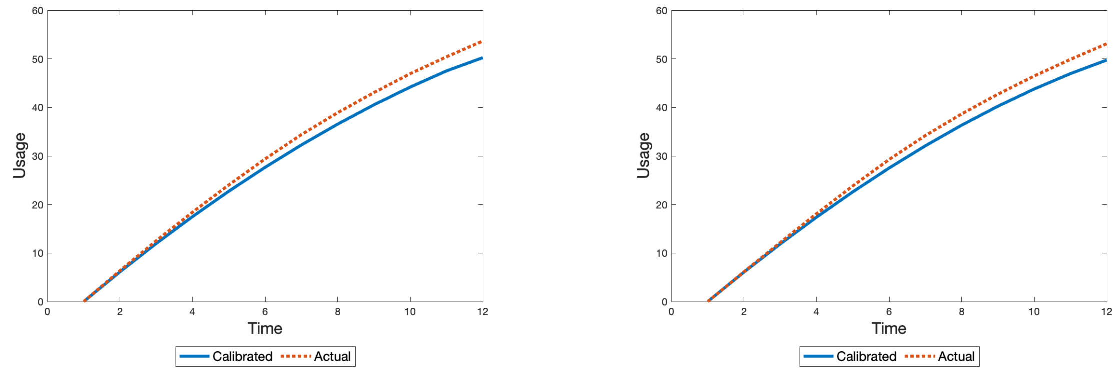
(a) Premium users (b) Basic users



Notes: Calibrated model based on the matched sample from the propensity score analysis (subsection 4.1). These plots provide model fit assessments for usage among premium (Panel a) and basic (Panel b) users. Solid blue line represents the usage patterns over time as predicted by calibrated model, while the dotted red line represents the user-averaged actual usage patterns over time.

Figure 24: Calibrated Model Fit Assessment for Cumulative Usage using Matched Sample

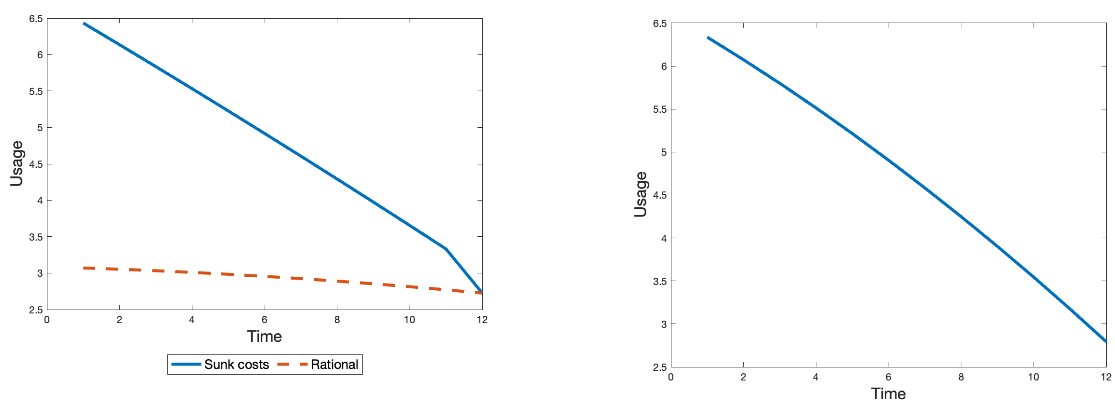
(a) Premium users (b) Basic users



Notes: Calibrated model based on the matched sample from the propensity score analysis (subsection 4.1). These plots provide model fit assessments for usage among premium (Panel a) and basic (Panel b) users. Solid blue line represents the cumulative usage patterns over time as predicted by calibrated model, while the dotted red line represents the user-averaged actual usage patterns over time.

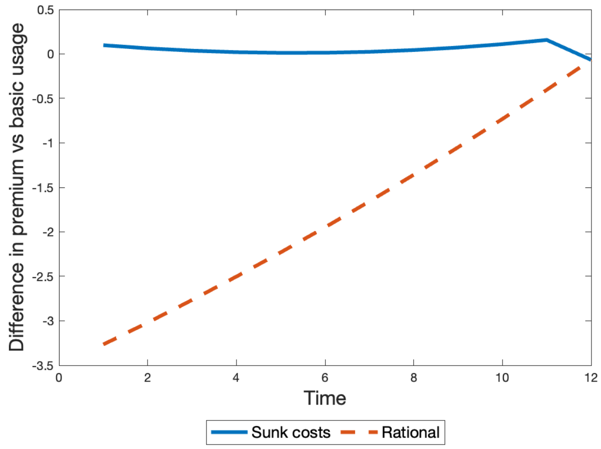
Figure 25: The Calibrated Impact of Sunk Costs Mental Accounting on Usage Dynamics using Matched Sample

(a) Premium users (b) Basic users



Notes: Calibrated model based on the matched sample from the propensity score analysis (subsection 4.1). These plots compare and assess the differences in usage with and without premium. Panel (a) provides the usage trajectories for premium users, while Panel (b) provides the usage trajectories for basic users. The scenario with rational behavior (i.e., no sunk costs) is obtained by setting $\lambda_1 = \lambda_2 = 0$.

Figure 26: Differences in Premium and Basic Usage using Matched Sample



Notes: Calibrated model based on the matched sample from the propensity score analysis (subsection 4.1). This plot provides the differences in premium versus basic usage over time when sunk costs are present (solid blue line) versus when users are rational (dashed red line). The scenario with rational behavior (i.e., no sunk costs) is obtained by setting $\lambda_1 = \lambda_2 = 0$.