Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song[*]

May 2, 2020

Abstract

Digital technologies such as smartphones and social media consume a large and growing share of leisure time. While these technologies provide obvious benefits, it is often argued that they can be addictive and harmful. We lay out a model of digital addiction and estimate its parameters using a randomized experiment involving about 2000 smartphone users. We also measure treatment effects of reduced smartphone use on survey measures of smartphone addiction and subjective well-being. This document is a pre-analysis plan for the experiment.

JEL Codes: D12, D61, D90, D91, I31, L86, O33.

Keywords: Behavioral welfare analysis, consumer surplus from digital technologies, demand for commitment, habit formation, self-control, naivete, temptation, randomized field experiments.

---

1 Introduction

People around the world use digital technologies in a large and growing share of their leisure time. In an annual 57-country survey, time spent browsing the internet on desktop and mobile devices grew to 170 minutes in 2019, exceeding television for the first time (Zenith Media 2019). Americans check their phones 50 to 80 times per day (Deloitte 2019; Vox 2019; New York Post 2019). The average unemployed young man in America spent almost an hour a day playing video games in 2015 (Aguiar et al. 2017).

A natural interpretation of these facts is that new digital technologies provide tremendous consumer surplus. Applying a shadow price of $5 per leisure hour to an average smartphone screen time of three hours per day (eMarketer 2019) would imply that smartphones generate $1.4 trillion per year in surplus for American adults. Consumers need to be paid substantial amounts of money to give up Facebook and other digital goods (Corrigan et al. 2018; Sunstein 2018; Brynjolfsson, Eggers, and Gannamaneni 2018; Allcott, Braghieri, Eichmeyer, and Gentzkow 2020), and calculations suggest that including the wefare gains from Facebook alone would add around 0.1 percentage points to U.S. annual GDP growth (Brynjolfsson et al. 2019).

However, such standard metrics may overstate the true consumer surplus gains from digital technologies if behavioral biases cause people to consume more than the amount that would maximize their utility. Many argue that smartphones, video games, and social media apps have addictive properties similar to classic temptation goods like cigarettes and gambling (Alter 2017; Eyal 2019; Newport 2019). The World Health Organization (2018) has listed digital gaming disorder as an official medical condition. Recent experimental studies find that deactivating social media can increase subjective well-being and reduce interest in using social media in the future, suggesting some form of learning or habit formation (e.g. Mosquera et al. 2019; Allcott, Braghieri, Eichmeyer, and Gentzkow 2020).

As additional motivating evidence, we ran a survey asking people what everyday activities they think they do "too little," "too much," or "the right amount." Figure 1 orders the activities from left to right by the absolute difference between the shares of "too little" and "too much." "Use my smartphone" and "browse social media" are two of the top five activities where people feel tension between what they do and what they would like themselves to do. Compared to the other three top activities (exercise, retirement savings, and sleep), behavioral issues around digital goods have received much less attention from economists.[1]

To what extent does digital addiction affect time use and the welfare gains from digital technologies? To answer this question, we formulate a model of digital addiction, use a randomized experiment to transparently identify the model parameters in the context of U.S. smartphone use, and estimate effects of digital addiction on equilibrium time use, consumer surplus, and subjective

---

[1]For example, see Charness and Gneezy (2009) and Carrera et al. (2019) on exercise, Madrian and Shea (2001) and Carroll et al. (2009) on retirement savings, and Bessone et al. (2020) on sleep.

well-being.

Our model of digital addiction builds on Gruber and Köszegi (2001), combining three standard concepts in behavioral economics: temptation, habit formation, and naiveté. Matching the suggestive survey results in Figure 1, we model smartphone use as a "temptation good" (Banerjee and Mullainathan 2010): the "present self" derives more utility from it than the "long-run self" considers in advance. Analogous to a model of present focus (Laibson 1997; Ericson and Laibson 2018), this generates time-inconsistency, as the long-run self would like to restrain the present self's consumption. We add a standard habit formation model in the spirit of Becker and Murphy (1988), and we allow consumers to be arbitrarily naive about their temptation and habit formation, nesting a model of projection bias (Loewenstein, O'Donoghue, and Rabin 2003).

To identify our model parameters and provide descriptive evidence on smartphone addiction, we carried out a field experiment in spring and summer 2020. We recruited about 2000 participants aged 18–64 using Facebook and Instagram ads and asked them to install Phone Dashboard, a behaviorally informed smartphone self-control app developed specifically for our study. Participants completed four surveys at three-week intervals—a baseline (Survey 1) and three follow-ups (Surveys 2, 3, and 4)—that included survey measures of smartphone addiction and subjective well-being as well as predictions of future app usage. Participants also responded to short text message survey questions three times each week and kept Phone Dashboard on their phone until six weeks after Survey 4.

Phone Dashboard can allow users to set daily time use limits on each app on their phone (effective tomorrow), and then to "snooze" the limit, allowing more time on an app that day—but only after a delay. This design provides a commitment device that restrains future behavior, while retaining flexibility through the snooze option. Beginning on Survey 2, we randomly assigned a Limit Treatment group that had access to the limit functionality. We further randomized the Limit Treatment group into five subgroups that had snooze delays of 0, 2, 5, or 20 minutes, or no ability to snooze. On Survey 3, we used a multiple price list (MPL) to elicit people's preferences between certain payments of different amounts and access to the limit functionality for the next three weeks; this allows us to measure people's valuations of the Limit functionality.

We also independently randomized a Screen Time Bonus, in which we paid people to reduce their use of six apps that people in our pilots had found to be especially tempting: Facebook, Instagram, Twitter, Snapchat, web browsers, and YouTube—hereafter, "FITSBY." We used a multiple price list on Survey 2 to elicit people's preferences between the Bonus and certain payments of different amounts; this allows us to determine people's valuations of the Bonus. Following Allcott, Kim, Taubinsky, and Zinman (2020), we use these bonuses to identify the magnitude of perceived temptation problems, because people who perceive that they will be tempted to overuse their smartphone apps will value the usage reduction induced by the Bonus. A randomly assigned Bonus Group was informed on Survey 2 that they would receive the Bonus for usage reductions beginning

three weeks later, for the three weeks after Survey 3.

This design allows us to transparently identify our model's structural parameters using experimental results instead of using a traditional dynamic estimation routine. We identify the demand slope primarily from the contemporaneous effect of the Bonuses. We identify perceived habit formation primarily from the anticipatory effect of the Bonus in the three weeks between Survey 2 and Survey 3 (before the incentive began), and we identify actual habit formation primarily from the effect the Bonus after the incentive ended. We identify naivete about temptation by comparing predicted versus actual usage in the Control conditions. We identify temptation in three ways: (i) the effect of the time limit functionality on time use, (ii) people's valuations of the Bonus, and (iii) people's valuations of access to the limit functionality. We develop simple analytical formulas that translate these parameter estimates into the effects of temptation and naivete on time use and consumer surplus.

Our work builds on several existing literatures. We extend a large literature documenting present focus in diverse settings including exercise, healthy eating, consumption-savings decisions, and laboratory tasks (Ericson and Laibson 2019).[2] Ours is one of a small handful of papers that estimates the parameters of a partially naive present focus model using field (instead of laboratory) behavior.[3] The digital self-control problems we study are particularly interesting because this is one of the few domains where market forces have created commitment devices, such as smartphone app time use limits and email and web blockers. As Laibson (2018) points out, such commercially available commitment devices are rare in other domains: "If one counts only pure commitment mechanisms that have not been created by behavioral economists, the list is small."

We also extend the empirical literature on habit formation. One set of papers documents persistent impacts of temporary interventions in settings such as exercise, energy conservation, voting, and hand-washing.[4] A second set of papers studies forward-looking habit formation—for example, the anticipatory effects of future cigarette tax increases.[5] Hussam et al. (2019) estimate the anticipatory and long-run effects of temporary experimental incentives for hand washing. To our knowledge, the only paper other than ours to estimate a model with present focus, habit formation, and naivete is Chaloupka, Levy, and White (2019).

---

[2]This includes Read and van Leeuwen (1998), Fang and Silverman (2004), Shapiro (2005), Shui and Ausubel (2005), Ashraf, Karlan, and Yin (2006), DellaVigna and Malmendier (2006), Paserman (2008), Gine, Karlan, and Zinman (2010), Duflo, Kremer, and Robinson (2011), Acland and Levy (2012), Andreoni and Sprenger (2012a; 2012b), Augenblick, Niederle, and Sprenger (2015), Beshears et al. (2015), Goda et al. (2015), Kaur, Kremer, and Mullainathan (2015), Laibson et al. (2015), Royer, Stehr, and Sydnor (2015), Augenblick (2018), Kuchler and Pagel (2018), Toussaert (2018), Augenblick and Rabin (2019), Schilbach (2019), John (forthcoming), and Sadoff, Samek, and Sprenger (forthcoming).

[3]To our knowledge, these are Allcott, Kim, Taubinsky, and Zinman (2020), Bai et al. (2018), Carrera et al. (2019), Chaloupka, Levy, and White (2019), and Skiba and Tobacman (2018).

[4]This includes Charness and Gneezy (2009), Allcott and Rogers (2015), Fujiwara, Meng, and Vogl (2016), and Allcott, Braghieri, Eichmeyer, and Gentzkow (2020).

[5]This includes Chaloupka (1991), Becker, Grossman, and Murphy (1994), and Gruber and Köszegi (2001). See Chaloupka and Warner (2001) for a review and Auld and Grootendorst (2004) for evidence on empirical challenges.

Finally, we extend two literatures that speak directly to digital addiction. The first is a set of experimental papers studying the effects of social media use on outcomes like subjective well-being and grades.[6] A second body of work studies the effects of online self-control devices (Marotta and Acquisti 2017; Acland and Chow 2018). Ours is the first paper to formally model digital addiction and estimate the effects on time use and consumer surplus.

Section 2 presents the experimental design. Sections 3, 4, and 5 detail the data, empirical strategy, and table and figure shells for the model-free results. The results quoted in the text and table and figure shells are either simulated data or pilot data.

2 Experimental Design

2.1 Experiment Overview

Figure 2 presents the experimental design and timeline, and Table 1 presents sample sizes at each step of our experiment. We refer to the baseline and subsequent three surveys as Surveys 1, 2, 3, and 4. We refer to period t as the three-week period following Survey t.

Between March 22 and April 8, we recruited participants using Facebook and Instagram ads. Our ads said, "Contribute to academic research and get paid!" Appendix Figure A1 presents the ads. To minimize sample selection bias, the ad did not hint at our research questions or suggest that the study was related to smartphone use or social media. XXX unique users were shown the ad, of whom 1,620 clicked on it.

Recruitment and intake survey. Clicking on the ad took the participant to a brief screening survey, which included several background questions, the consent form, and instructions on how to install Phone Dashboard. 1,253 people satisfied the inclusion criteria, by being a US resident between 18 and 64 years old whose primary phone is an Android and only uses one phone regularly. Of those people, 550 consented to participate in the study, and 804 installed Phone Dashboard and finished the intake survey.

Survey 1 (baseline). On April 12, we sent participants a link to the baseline survey. The survey recorded additional demographics and the outcome variables (described below) measuring smartphone addiction and subjective well-being. The survey also asked, *How interested are you to set limits on your phone use?* with categorical answers from "Not at all interested" to "Very interested"; we call this variable *interest in limits*. As practice for the multiple price lists on later surveys, the survey then introduced the idea of an MPL and walked people through an MPL in which they traded off receiving different survey completion payments at different times. 340 people finished the baseline survey.

---

[6]This includes Sagioglu and Greitemeyer (2014), Tromholt (2016), Hunt et al. (2018), Vanman, Baker, and Tobin (2018), Mosquera et al. (2018), Allcott, Braghieri, Eichmeyer, and Gentzkow (2020), and Collis and Eggers (2020).

Between Survey 1 and Survey 3, we sent three text message questions per week (described below).

141 people were dropped from the experiment because they reported already using another app to limit their phone use (five percent of the sample), did not promise to "provide my best answers" on our surveys, reported having idiosyncratic bugs with Phone Dashboard, failed to answer more than two text message questions between Survey 1 and Survey 2, have a device known to kill background apps such as Phone Dashboard, or were missing screen time data. The remaining 199 people were included in the randomization and invited to take Survey 2.

Survey 2. On May 3, three weeks after Survey 1, we sent participants a link to Survey 2. This survey implemented several approaches to identify temptation and naivete.

The first part of Survey 2 implemented the Screen Time Bonus, following the approach of Allcott, Kim, Taubinsky, and Zinman (2020). We began by eliciting predicted average daily FITSBY screen time over the next three weeks. We then introduced the Screen Time Bonus. The survey read,

You might have the chance to earn money by reducing your screen time on the FITSBY apps (Facebook, Instagram, Twitter, Snapchat, Browsers, and YouTube) over a 3-week period, starting either tomorrow or in 3 weeks.

If you're selected for the Screen Time Bonus, you would receive $50 for every hour you reduce your average daily FITSBY screen time below a Bonus Benchmark of [X] hours per day over the 3-week period, up to $150.

The Bonus Benchmark [X] is the participant's average daily FITSBY time over the baseline period, rounded up to the nearest integer. We gave several examples of possible earnings, including:

- *If you reduce your FITSBY screen time to $[X-1] hours and 30 minutes per day over the next 3 weeks, you would receive $25.*

- *If your FITSBY screen time is above $X hours per day, you would receive $0.*

We then asked people to predict by what percent they would reduce their FITSBY screen time, if they were selected for the Screen Time Bonus.

Next, the survey asked people to choose, hypothetically, between the Screen Time Bonus or a fixed payment equal to their expected earnings if offered the Bonus. The survey coached people as to how they might decide:

- *You might prefer $[expected earnings] instead of the Screen Time Bonus if you don't want any pressure to reduce your screen time.*

- *You might prefer the Screen Time Bonus instead of $[expected earnings] if you want to give yourself extra incentive to use your phone less.*

The survey then offered people a series of choices between the Screen Time Bonus and fixed payments of different amounts ranging from $150 to $0. The MPL was incentive compatible: we

randomly assigned 0.2 percent of participants (the Multiple Price List group) to receive the choice they made on a randomly selected row from the MPL. We informed people of this before beginning the MPL, saying,

Last week, the computer randomly selected some participants to receive what they choose on the multiple price list below, and also randomly selected one of the rows to be "the question that counts." If you were randomly selected to participate, you will be paid based on what you choose in that row. So you should answer carefully!

The survey then informed people of their Bonus assignment. This is the first point in the study where treatment groups are treated differently. 274 participants reached this point; we call this group our "analysis sample." The Bonus group was informed that they would receive the bonus *starting in three weeks*, after Survey 3. The Bonus Control group was told that they had not been selected for a bonus. The Multiple Price List group received what they chose on "the question that counts."

To identify forward-looking habit formation, it was crucial to ensure that the Bonus group was aware of the future price change. Before advancing, each participant had to answer a question correctly describing their bonus treatment condition. We also sent three text messages between Survey 2 and Survey 3 reminding the Bonus group of their future bonus:

Don't forget, we'll pay you $50 for every hour you reduce your average daily screen time between May 24 and June 14. There is no bonus for changing your screen time before then.

People were asked to respond to the text message to confirm that they had read it.

The second part of Survey 2 implemented the Phone Dashboard screen time limits for the Limit group. First, the Limit group was given instructions on how to set daily app usage limits for themselves. We asked people to report the time limit they would like to set on each of the FITSBY apps each day over the next three weeks. We then asked people to update their Phone Dashboard app and encouraged people to set the limits they had reported a moment earlier.

Finally, we reminded people of their treatment group assignments and re-elicited FITSBY screen time predictions for periods 2, 3, and 4.

Surveys 3 and 4. On May 24 and June 14, we sent participants links to Surveys 3 and 4, respectively. The surveys measured the same impact evaluation outcome variables as in Survey 1, as well as usage predictions for periods 3, 4, and 5. On Survey 3, we also used an MPL to measure the Limit group's valuation of access to the limit functionality for the next three weeks. The Multiple Price List group received what they chose on a randomly selected MPL row.

Text message surveys. For the nine weeks between Survey 1 and Survey 4, we sent three text messages per week to measure aspects of perceived phone use that were better measured in real time rather than retrospectively. We rotated through nine text message questions described below.

Payment. All participants received $5 for completing the baseline survey and a $25 completion

payment if they completed the remaining surveys, kept Phone Dashboard installed through the end of period 5 (six weeks after Survey 4), and responded to 75 percent of text messages. People who completed the study were also entered in a drawing for a $500 gift card, in which two winners were drawn.

Attrition management. In addition to back-loading the survey payments, we took extensive efforts to minimize post-randomization attrition. There are two full surveys (the intake and Survey 1) before the interventions start. At the beginning of Survey 2, just before the interventions start, we informed people that "anyone who drops out after this page can really damage the entire study," and offered them a choice to drop out now or commit to finishing the whole study. We sent daily email reminders for six days to people who had not completed any of Surveys 2–4, and four days after those surveys we began offering an additional payment for completing all remaining surveys. We also sent reminder emails to people who had failed to respond to two consecutive text messages.

2.2 Phone Dashboard App

Phone Dashboard is an Android app that was developed by a company called Audacious Software for our experiment. Figure 3 presents screen shots. It is available from the Android app store and has two key functionalities: screen time measurement and app blocking.

Phone Dashboard records the app that is in the foreground of a smartphone every five seconds. It does not record the content that the user is viewing within the app. The usage data are uploaded every night and transferred to our secure servers. Users can see their cumulative screen time by day and by week on the Phone Dashboard home screen, as in the top left panel of Figure 3. This usage information was designed to be useful for participants in the Limit or Bonus groups who might want to track their usage. In practice, it is unlikely that the usage information affected the Control group, as they opened the usage information screen an average of XXX times per week over periods 1–4.

Phone Dashboard's second functionality is app-specific time limits. The functionality was designed with our temptation model in mind, allowing the "long-run self" to exert control over the consumption of the "present self." As shown in the top middle and top right panels of Figure 3, the user can input a daily usage limit for each app, and then view all limits in one screen. Usage limits take effect the next day. Thus, limits do not affect the present self's consumption—they reveal the present self's preferences over future consumption. The Limit group was given a version of Phone Dashboard with this functionality; the Limit Control group was not.

Phone Dashboard serves five-minute and one-minute push notifications to users as an app's daily time limit approaches. When the limit arrives, users can "snooze" their limit and get an additional amount of time that they specify—but starting only after a delay of m minutes, where $m \in \{0, 2, 5, 20, \infty\}$ is randomly assigned across participants. We use $m = \infty$ to refer to a group that did not have the snooze functionality. For $0 < m < \infty$, the ability to snooze the limit again

does not affect the present self's consumption, but instead reveals the present self's preferences over future consumption that day. The variation in m helps identify preferences for commitment vs. flexibility as well as the length of time that present temptation might last. The bottom three panels of Figure 3 illustrate these functionalities. To reduce attrition and uninstallation, Phone Dashboard also allows people to permanently opt out of the limits.

Our experiment includes only Android smartphone users because a similar functionality cannot be implemented by third-party apps on iOS.

2.3 Welfare Survey Outcome Variables

For the impact evaluation, we consider the five outcomes described below. Appendix A.1 presents survey question text and descriptive statistics for each outcome variable and moderator.

Ideal use change. The survey said,

Some people say they use their smartphone too much and ideally would use it less. Other people are happy with their usage or would ideally use it more. How do you feel about your smartphone use over the past 3 weeks?

- *I use my smartphone too much.*

- *I use my smartphone the right amount.*

- *I use my smartphone too little.*

For people who said they used their smartphone "too much" or "too little," we then asked, *Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your smartphone use?* The *ideal use change* variable is the answer to this question, in percent. This question is designed to capture the perceived extent of overuse, and it plays some role in identifying our model.

Addiction scale. Our addiction scale is a battery of 16 questions modified from two well-established survey scales, the Mobile Phone Problem Use Scale (Bianchi and Phillips 2005) and the Bergen Facebook Addiction Scale (Andreassen et al. 2012). The questions attempt to measure the six core components of addition identified in the addiction literature: salience, tolerance, mood modification, relapse, withdrawal, and conflict (Griffiths 2005).

The survey asked, *In the past three weeks, how often have you ...*, with a matrix of 16 questions, such as

- *used your phone longer than intended?*

- *felt anxious when you don't have your phone?*

- *lost sleep due to using your phone late at night?*

Possible answers were Never, Rarely, Sometimes, Often, and Always, which we coded as 0, 0.25, 0.5, 0.75, and 1, respectively. *Addiction scale* is the sum of these numerical scores for the 16 questions.

SMS addiction scale. The SMS addiction questions were shortened versions of some addiction scale questions. Examples include:

- *In the past day, did you feel like you had an easy time controlling your screen time?*

- *In the past day, did you use your phone mindlessly?*

- *When you woke up today, did you immediately check social media, text messages, or email?*

People were instructed to text back their answers on a scale from 1 (not at all) to 10 (definitely). *SMS addiction scale* is the sum of these scores for the nine questions.

Phone makes life better. The survey asked, *To what extent do you think your smartphone use makes your life better or worse?* Responses were on a scale from -5 ("Makes my life worse") through 0 ("Neutral") to +5 ("Makes my life better"). This question is designed to qualitatively reflect the consumer surplus that people feel like they are deriving from their smartphones.

Subjective well-being. We use standard measures from the subjective well-being literature, mostly following the measures from our own earlier work (Allcott, Braghieri, Eichmeyer, and Gentzkow 2020). The survey asked,

Please tell us the extent to which you agree or disagree with each of the following statements. **Over the last three weeks,** with a matrix of seven questions:

- *... I was a happy person*

- *... I was satisfied with my life*

- *... I felt anxious*

- *... I felt depressed*

- *... I could concentrate on what I was doing*

- *... I was easily distracted*

- *... I slept well*

Possible answers were on a seven-point scale from "strongly disagree" through "neutral" to "strongly agree," which were coded as -1, -2/3, -1/3, 0, 1/3, 2/3, and 1, respectively. The variable *subjective well-being* is the sum of these numerical scores for the seven questions, after reversing *anxious*, *depressed*, and *easily distracted* so that more positive reflects better subjective well-being.

Indices. Define the *welfare survey index* as the sum of the five outcome variables described above, weighted by the baseline inverse covariance matrix as described by Anderson (2008). When

presenting results and constructing this index, we orient the variables so that more positive values imply normatively better outcomes. Thus, we multiply *addiction scale* and *SMS addiction scale* by (-1).

Define the *restriction index* as the sum of *interest in limits* (with the four categorical answers coded as 0, 1, 2, and 3) and *ideal use change*, after normalizing each into standard deviation units. Define the *addiction index* as the sum of *addiction scale* and *phone makes life better* after normalizing each into standard deviation units. We use these latter two indices for stratification and as moderators when testing for heterogeneous treatment effects.

2.4 Randomization

We use a factorial experimental design, as illustrated in Figure 2. The randomization assigned people to the Screen Time Bonus (the "Bonus" group), "Bonus Control," or the "Multiple Price List" group with 25, 75, and 0.2 percent probability, respectively. The randomization also independently assigned people to receive the screen time limit functionality (the "Limit" group) or "Limit Control" with 60 and 40 percent probability, respectively. Within the Limit group, people were randomly assigned to snooze delays of $m \in \{0, 2, 5, 20\}$ minutes or a "No Snooze" condition ($m = \infty$), each with 12 percent probability.

We balanced the Bonus, Limit, and snooze delay treatment groups within eight strata defined by above- versus below-median baseline FITSBY use, restriction index, and addiction index.

3 Data

All figures and tables below present results for the analysis sample, although sample sizes sometimes differ across outcomes due to missing data.

Table 2 quantifies the representativeness of our analysis sample on observables, by comparing their demographics to the U.S. adult population. Comparing column 1 to column 2, we see that our sample is relatively young and female. Appendix Table A2 shows that the treatment groups are balanced on observables. Appendix Table A1 presents descriptive statistics for the welfare survey outcomes.

Table 3 documents attrition.

We focus on FITSBY (Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube) instead of overall screen time because this allows us to narrow the experimental design and analysis to the most tempting apps. As we will see in Figure 5, FITSBY are six of the seven apps that people feel that they overuse the most.[7] Table 2 shows that FITSBY comprises XX percent of screen time during the baseline (period $t = 1$). Appendix Figure A2 presents most popular apps.

---

[7]The seventh is "games," but this category comprises many individual apps, which would have complicated our design and analysis.

During the baseline period, the average use of Facebook, browsers, YouTube, Instagram, Twitter, and Snapchat was 26, 12, 9, 5, 2, and 2 minutes per day, respectively. Appendix Figure A3 presents the distribution of baseline FITSBY usage.

3.1 Variable Construction

We define a variable called *tightness* to capture the stringency of screen time limits set by the Limit group. Define $f_{iadt}^{\cap}$ as the screen time of user i on app a on day d in period t. Define l_{iat} as the average screen time limit (in minutes per day) in place in period t, and define $N_{d \in t=1}$ as the number of days in the baseline period. *Tightness* L_{iat} represents the usage reduction that the limit l_{iat} would have caused during the baseline period in the absence of snoozes or other behavioral response:

$$L_{iat} = \frac{1}{N_{d \in t=1}} \sum_{d \in t=1} \max\left\{0, f_{iad1} - l_{iat}\right\}. \tag{1}$$

If the daily limit l_{iat} would not have been binding in baseline day d, the max function returns 0. If l_{iat} would have been binding in day d, then the max function returns the excess screen time on that day. We set $L_{iat} = 0$ if no limit is set or if user i does not have app a. This definition speaks to how much each app contributes to overall temptation, not how tempting each app is for the subset of people who use it. We aggregate over apps to construct user-level limit tightness: $L_{it} = \sum_a L_{iat}$.

We winsorize predicted usage at no more than 60 minutes per day more or less than actual usage in the corresponding period.

4 Empirical Strategy

4.1 Empirical Strategy

Define Y_{it} as an outcome for person i for period t. Y_{it} could represent a welfare survey outcome measured on survey $t \in \{3, 4\}$, or period t FITSBY usage $f_{it}^{\cap}$. Define L_i and B_i as indicators for the Limit and Bonus groups. Define $\boldsymbol{X}_i$ as a vector of exogenous covariates in Table 2: income, indicators for male, White, and college completion, as well as age, period 1 phone use, and period 1 FITSBY use. Define ν_{it} as a vector of the 8 stratum indicators for period t. We estimate the effects of the Limit and Bonus using the following regression:

$$Y_{it} = \tau_t^B B_i + \tau_t^L L_i + \rho_t \boldsymbol{X}_i + \beta_t Y_{i1} + \nu_{it} + \varepsilon_{it}. \tag{2}$$

We expected that the Limit would have similar ITT effects on welfare survey outcomes on Surveys 3 and 4, and we also expected that the Bonus would have no effect on welfare survey outcomes on Survey 3, as the bonus period starts after Survey 3. To parsimoniously present the

overall effects on the welfare survey outcomes in the below Figure 10, we thus constrain τ_t^L to be the same for $t = 3$ and $t = 4$ and report two coefficients: τ^L and τ_4^B.

To increase precision and simplify analyses of effects on welfare survey outcomes, we also estimate instrumental variables (IV) regressions with previous period usage $f_{i-1,i}^\cap$ as the endogenous variable:

$$Y_{it} = \tau f_{i,t-1}^\cap + \rho_t \boldsymbol{X}_i + \beta_t Y_{i1} + \nu_{it} + \varepsilon_i, \tag{3}$$

instrumenting for $f_{i,t-1}^\cap$ with B_i and L_i interacted with $t = 3$ and $t = 4$ indicators. We combine data from Surveys 3 and 4 and let all coefficients other than τ vary across the two periods. We think of this regression as a transparent way of combining the ITT of the Bonus and Limit interventions, weighting the ITTs by their effects on usage. Because our Limit intervention could affect survey outcomes through channels other than reduced FITSBY usage—for example, by giving people an increased feeling of control over their usage—we do not claim that the IV exclusion restriction necessarily holds.

4.2 Pre-Analysis Plan

We submitted our pre-analysis plan on May 2nd, as this was the final day before participants were informed of their treatment conditions on Survey 2. The pre-analysis plan specified the following things:

1. **Estimating equations.** Equation (2) would be our primary estimating equation for impact evaluation. We would also use Equation (3) to present combined effects of the Bonus and Limit treatments on welfare survey outcomes and as our primary specification to estimate heterogeneous treatment effects on the welfare survey outcomes.

2. **Moderators.** We planned to test for heterogeneous treatment effects by splitting the sample on above- versus below-median values of education, age, gender, baseline FITSBY use, *restriction index*, and *addiction index*.

3. **Variable construction.** We specified the construction of the welfare survey outcomes and indices, the *tightness* variable, and the winsorization of predicted usage.

The pre-analysis plan included table and figure shells for Tables 1–3, Appendix Tables A1 and A2, Figures 1–3, 4–11, and Appendix Figures A1-A11. The remaining figures were not included in the pre-analysis plan because we anticipated that they could change as we developed the model.

5   Empirical Results

5.1   Qualitative Evidence on Digital Addiction

Figure 4 presents qualitative evidence on digital addiction from the baseline survey. The top left panel shows that 36.2 percent of people reported being "moderately" or "very" interested in setting time use limits on their smartphone apps, while only 20.1 percent reported being "not at all" interested. The top right panel presents the distribution of responses to the *ideal use change* question. 45.2 percent of people say that they used their smartphone the right amount over the past three weeks, and only 0 percent said that they used it too little. Among people who said they used their smartphone too much, the average ideal reduction was 37.2 percent.

At baseline, we asked people to report their ideal use change by app or category. Figure 5 presents results. We code "I don't use this app at all" as 0, so this figure reflects how much each app contributes to overall temptation, not how tempting each app is for the subset of people who use it. Facebook is by far the most tempting app in our sample: our average participant would ideally reduce Facebook use by 23.2 percent.

The bottom left panel shows that while most people think that their smartphone use makes their life better, 18.1 percent think that their smartphone use makes their life worse. The bottom right panel shows the distribution of results on the 16-point addiction scale.

The top two panels are the variables in our restriction index, and the bottom two panels are the variables in our addiction index.

5.2   Blocker Use

Figure 6 presents tightness and snooze use by Snooze Delay group. We will use this to illustrate the tradeoff between commitment and flexibility.

5.3   Effects on FITSBY use

Figure 7 presents treatment effects on usage, using Equation (2).

Figure 8 presents effects on usage by Snooze Delay group. This allows us to further explore the commitment-flexiblity tradeoff.

Figure 9 presents effects on usage by app. Facebook has the largest effects, consistent with its heavy baseline usage and high self-reported overuse.

Appendix Figure A4 presents heterogeneous effects by time of day.

5.4   Effects on Welfare Survey Outcomes

Figure 10 presents the intent-to-treat effects of the Limit and Bonus.

Appendix Figure A5 presents local average treatment effects estimated using Equation (3), combining effects from both Treatments.

Appendix Figures A6–A11 study heterogeneity along the six pre-specified moderators.

5.5   Predicted versus Actual Use

Figure 11 presents predicted vs. actual use.

References

Acland, Dan and Matthew R. Levy. 2012. "Naivete, Projection Bias, and Habit Formation in Gym Attendance." Working Paper: GSPP13-002.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484):1481–1495.

Andreoni, James and Charles Sprenger. 2012a. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7):3333–3356.

———. 2012b. "Risk Preferences Are Not Time Preferences." *American Economic Review* 102 (7):3357–3376.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics* 121 (2):673–697.

Augenblick, Ned. 2018. "Short-Term Discounting of Unpleasant Tasks." Working Paper.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working Over Time: Dynamic Inconsistency In Real Effort Tasks." *The Quarterly Journal of Economics* 130 (3):1067–1115.

Augenblick, Ned and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *The Review of Economic Studies* 86 (3):941–975.

Bai, Liag, Benjamin Handel, Edward Miguel, and Gautam Rao. 2018. "Self-Control and Demand for Preventive Health: Evidence from Hypertension in India." NBER Working Paper No. 23727.

Becker, Gary S. and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4):675–700.

Beshears, John, James J. Choi, Christopher Harris, David Laibson, Brigitte C. Madrian, and Jung Sakong. 2015. "Self-Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" Working Paper: NBER WP No. 21474.

Brynjolfsson, Erik, Felix Eggers, and Avinash Gannamaneni. 2018. "Using Massive Online Choice Experiments to Measure Changes in Well-being." NBER Working Paper.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky. 2019. "How are Preferences For Commitment Revealed?" Working Paper.

Chaloupka, Frank J., Matthew R. Levy, and Justin S. White. 2019. "Estimating Biases in Smoking Cessation: Evidence from a Field Experiment." Working Paper: NBER WP No. 26522.

Charness, Gary and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3):909–931.

Corrigan, Jay R., Saleem Alhabash, Matthew Rousu, and Sean B. Cash. 2018. "How much is social media worth? Estimating the value of Facebook by paying users to stop using it." *PloS one* 13 (12):e0207101.

DellaVigna, Stefano and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3):694–719.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6):2350–2390.

Ericson, Keith Marzilli and David Laibson. 2019. *Intertemporal Choice*, vol. 2, chap. 1. Elsevier, 1 ed.

Fang, Hanming and Dan Silverman. 2004. "Time Inconsistency and Welfare Program Participation: Evidence from the NLSY." Cowles Foundation Discussion Paper No. 1465.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4):160–188.

Gine, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2:213–235.

Goda, Gopi Shah, Matthew R. Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2015. "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings." Working Paper: NBER WP No. 21482.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *Quarterly Journal of Economics* 116 (4):1261–1303.

Hunt, Melissa G., Rachel Marx, Courtney Lipson, and Jordyn Young. 2018. "No More FOMO: Limiting Social Media Decreases Loneliness and Depression." *Journal of Social and Clinical Psychology* 37 (10):751–768.

John, Anett. Forthcoming. "When Commitment Fails - Evidence from a Field Experiment." *Management Science* .

17

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy* 123 (6):1227–1277.

Kuchler, Theresa and Michaela Pagel. 2018. "Sticking to Your Plan: The Role of Present Bias for Credit Card Paydown." Working Paper: NBER WPNo. 24881.

Laibson, David. 2018. "Private Paternalism, the Commitment Puzzle, and Model-Free Equilibrum." *AEA Papers and Proceedings* 108:1–21.

Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2015. "Estimating Discount Functions with Consumption Choices over the Lifecycle." Working Paper.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4):1209–1248.

Marotta, Veronica and Alessandro Acquisti. 2017. "Online Distractions, Website Blockers, and Economic Productivity: A Randomized Field Experiment." Preliminary Draft.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2018. "The Economic Effects of Facebook." Available at SSRN 3312462.

Paserman, M. Daniele. 2008. "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation." *The Economic Journal* 118:1418–1452.

Read, Danieal and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2):189–205.

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics* 7 (3):51–84.

Sadoff, Sally, Anya Samek, and Charles Sprenger. Forthcoming. "Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts." *The Review of Economic Studies* .

Sagioglu, Christina and Tobias Greitemeyer. 2014. "Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it." *Computers in Human Behavior* 35:359–363.

Schilbach, Frank. 2019. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review* 109 (4):1290–1322.

Shapiro, Jesse M. 2005. "Is there a daily discount rate? Evidence from the food stamp nutrition cycle." *Journal of Public Economics* 89:303–325.

Shui, Haiyan and Lawrence M. Ausubel. 2005. "Time Inconsistency in the Credit Card Market." Working Paper.

Skiba, Paige Marta and Jeremy Tobacman. 2018. "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default." Working Paper.

Toussaert, Severine. 2018. "Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment." *Econometrica* 86 (3):859–889.

Tromholt, Morten. 2016. "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-Being." *Cyberpsychology, Behavior, and Social Networking* 19 (11):661–666.

Vanman, Eric J., Rosemary Baker, and Stephanie J. Tobin. 2018. "The Burden of Online Friends: The Effects of Giving Up Facebook on Stress and Well-Being." *The Journal of Social Psychology* 158 (4):496–507.

Wurmser, Yoram. 2019. "Time Spent with Media 2019." Tech. rep., eMarketer. https://www.emarketer.com/content/us-time-spent-with-mobile-2019.

Table 1: **Experiment Timeline and Sample Sizes**

Phase	Date	Sample size
Recruitment and intake	March 22 - April 8	[] shown ads
		1,620 clicked on ads
		1,253 passed screen
		550 consented
		804 finished intake survey
Survey 1 (baseline)	April 12	345 began Survey 1
		340 finished Survey 1
		199 were randomized
Survey 2	May 3	278 began Survey 2
		274 informed of treatment ("analysis sample")
		278 finished Survey 2
Survey 3	May 24	265 began Survey 3
		255 finished Survey 3
Survey 4	June 14	247 began Survey 4
		247 finished Survey 4
Completion	July 26	[] kept Phone Dashboard through July 26

Table 2: **Sample Demographics**

	(1) Analysis sample	(2) U.S. adults
Income ($000s)	38.9	33.7
College	0.59	0.29
Male	0.37	0.49
White	0.67	0.74
Age	33.7	47.0
Period 1 phone use (minutes/day)	112.2	175.0
Period 1 FITSBY use (minutes/day)	57.2	.

Notes: Column 1 presents average demographics for our analysis sample, and column 2 presents average demographics of American adults. In column 2, smartphone use is from market research company eMarketer (Wurmser, 2019); all other variables are from the 2018 American Community Survey.

Table 3: **Response Rates**

(a) **Limit Treatment and Control**

	(1) Control	(2) All limits	(3) Snooze 0	(4) Snooze 2	(5) Snooze 5	(6) Snooze 20	(7) No snooze	(8) F-test p-value
Completed Survey 3	0.98	0.92	1.00	0.89	0.92	0.82	0.96	0.04
Completed Survey 4	0.97	0.90	1.00	0.85	0.88	0.82	0.92	0.06
Have period 2 Usage	0.99	0.98	1.00	0.96	0.96	0.95	1.00	0.67
Have period 3 Usage	0.98	0.91	1.00	0.89	0.88	0.82	0.96	0.03
Have period 4 Usage	0.95	0.91	1.00	0.89	0.88	0.82	0.96	0.11
Have period 5 Usage	0.95	0.91	1.00	0.89	0.88	0.82	0.96	0.11

(b) **Bonus Treatment and Control**

	(1) Control	(2) Treatment	(3) t-test p-value
Completed Survey 3	0.95	0.98	0.36
Completed Survey 4	0.92	0.98	0.16
Have period 2 Usage	0.99	1.00	0.57
Have period 3 Usage	0.94	0.98	0.29
Have period 4 Usage	0.93	0.98	0.19
Have period 5 Usage	0.93	0.98	0.19

Notes: Columns 1 and 2 of Panel (a) present present response rates for Limit Control and Treatment groups. Columns 3–7 present response rates for each of the snooze delay conditions within the Limit Treatment group. Column 8 presents the p-value of an F-test of differences between the Limit Control and the separate snooze delay conditions. Columns 1 and 2 of Panel (b) present response rates for Bonus Control and Treatment groups. Column 3 presents the p-value of a t-test of differences between the Bonus Control and Treatment groups.

Figure 1: **Online and Offline Temptation**
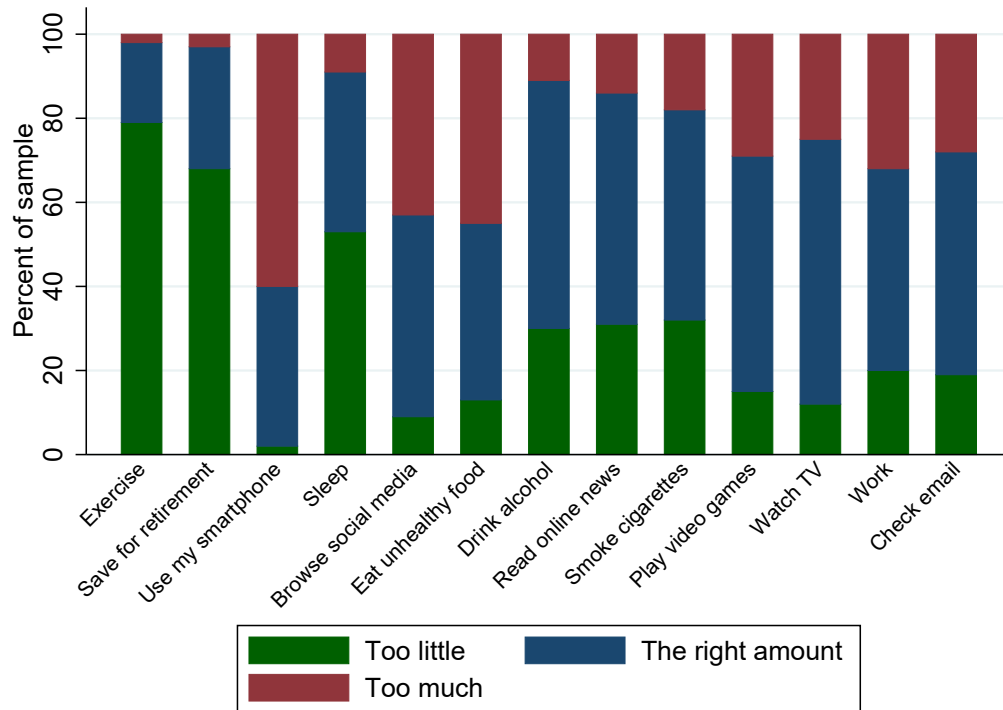
Notes: This figure presents responses to an online survey in which we asked, "For each of the activities below, please tell us whether you think you do it too little, too much, or the right amount." The bars are ordered from left to right in order of largest to smallest absolute value of (share "too little" – share "too much").
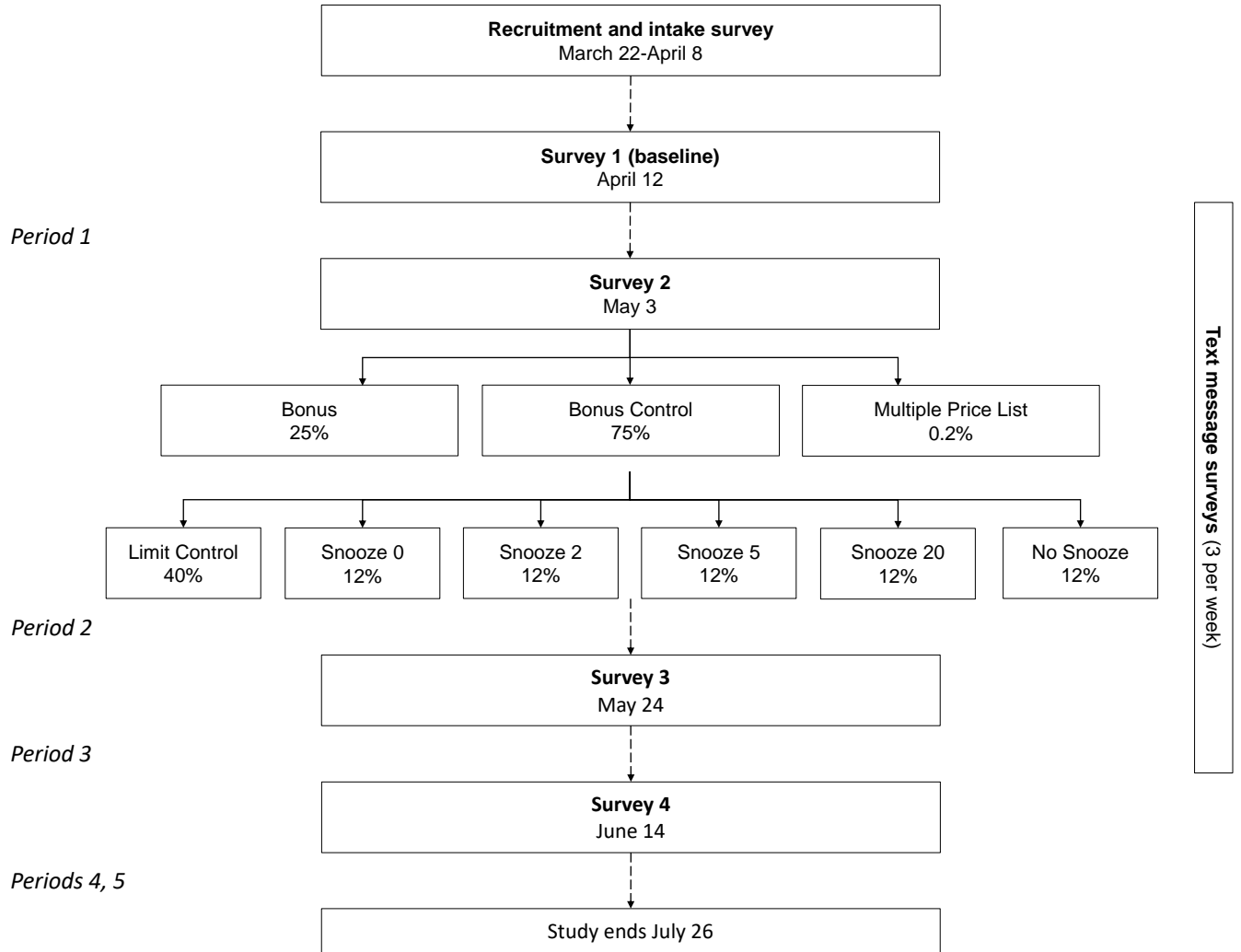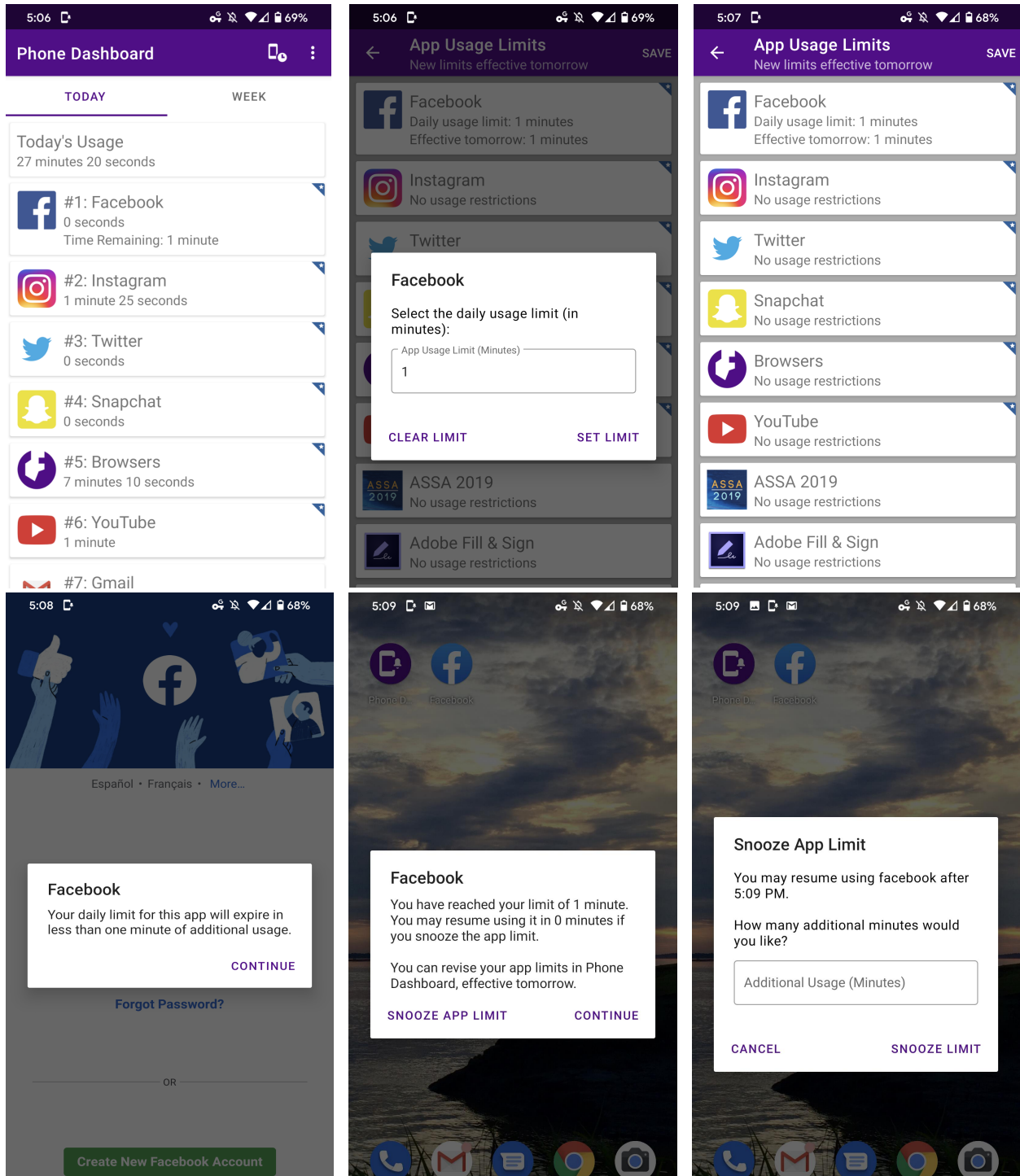
Figure 2: **Experimental Design**

Figure 3: **Phone Dashboard Screen Shots**

Notes: This figure presents screen shots of the Phone Dashboard app. The top left presents day's total usage by app. The top middle shows how a user can set daily a daily usage limit for each app, effective tomorrow. The top right shows the usage limits set for each app. The bottom left shows the warning users receive when they are within five minutes or one minute of their limit. The bottom middle shows the message users receive when they reach the limit. Users with the snooze functionality can resume using an app after a delay of $m \in \{0, 2, 5, 20\}$ minutes; this user has $m = 0$. The bottom right shows the option for a user to choose how many additional minutes to add to the daily limit after the snooze delay.

24

Figure 4: **Baseline Qualitative Evidence of Self-Control Problems**
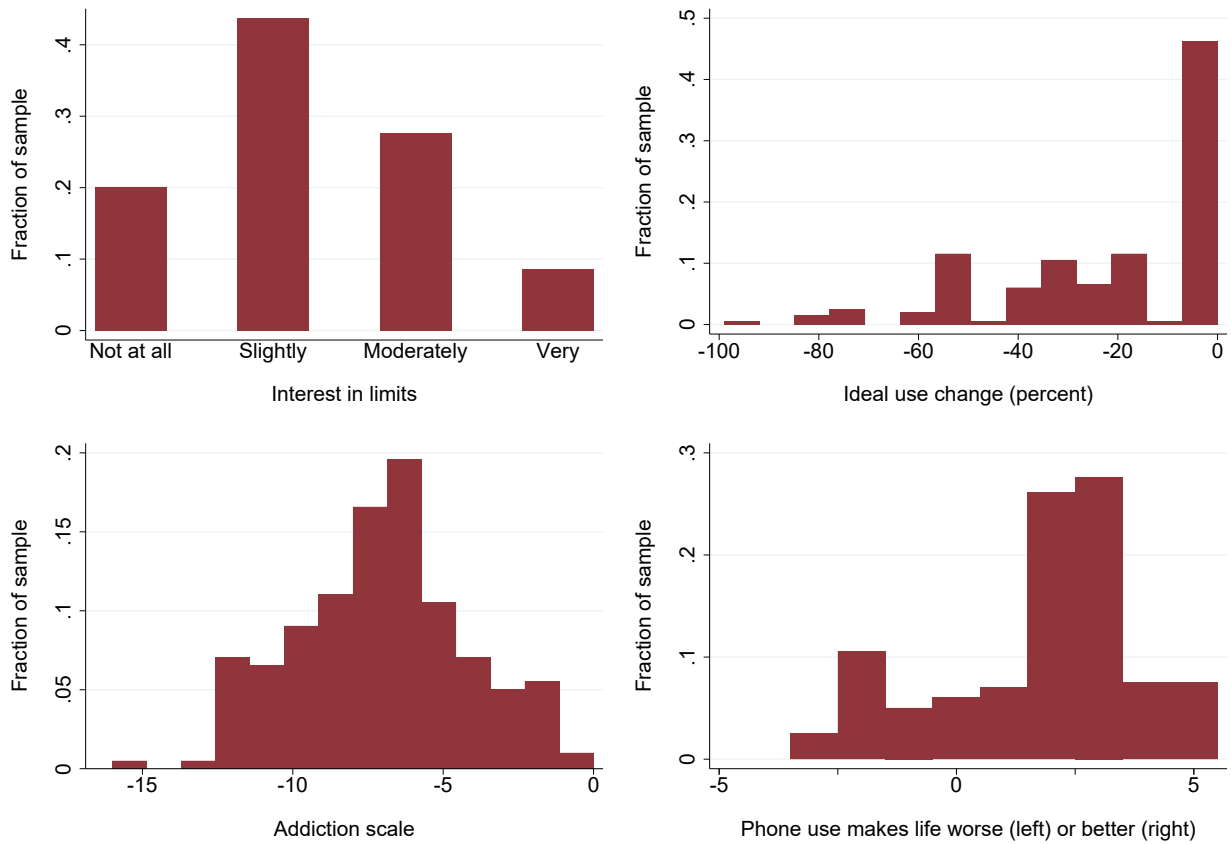
Notes: This figure presents the distributions of four measures of smartphone addiction from the baseline survey. *Interest in limits* is the answer to, "How interested are you to set limits on your phone use?" *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *Phone use makes life worse or better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?"

Figure 5: **Ideal Use Change by App**
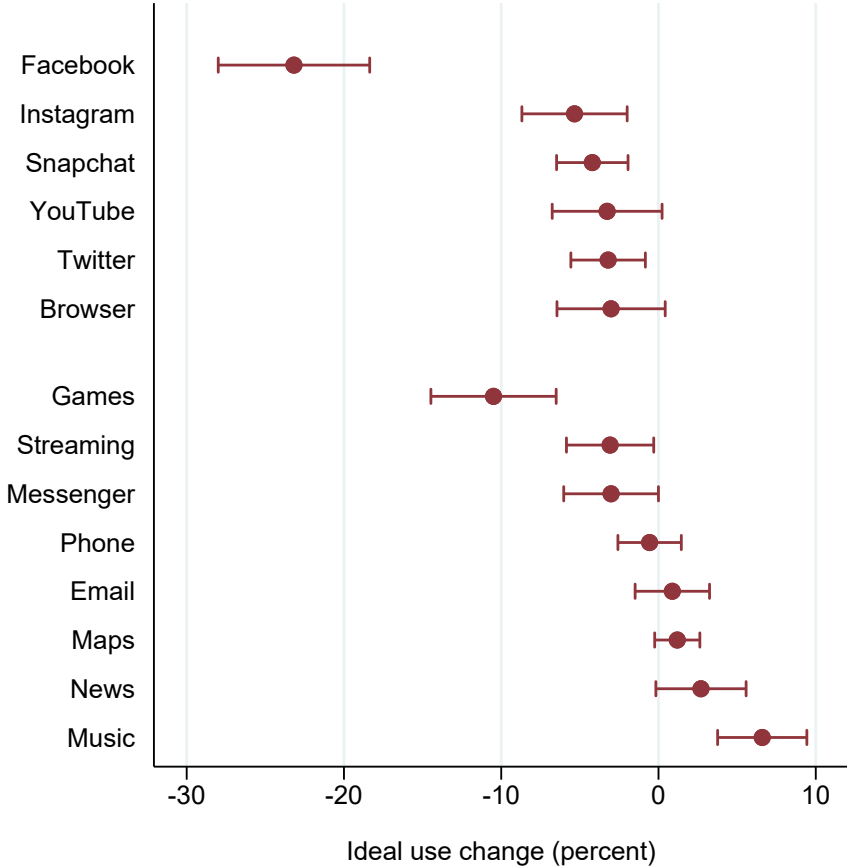
Notes: This figure presents mean *ideal use change* by app or app category at baseline. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?"

Figure 6: **Limit Tightness and Snooze Use by Snooze Delay Group**

(a) **Limit Tightness**

(b) **Snooze Use**
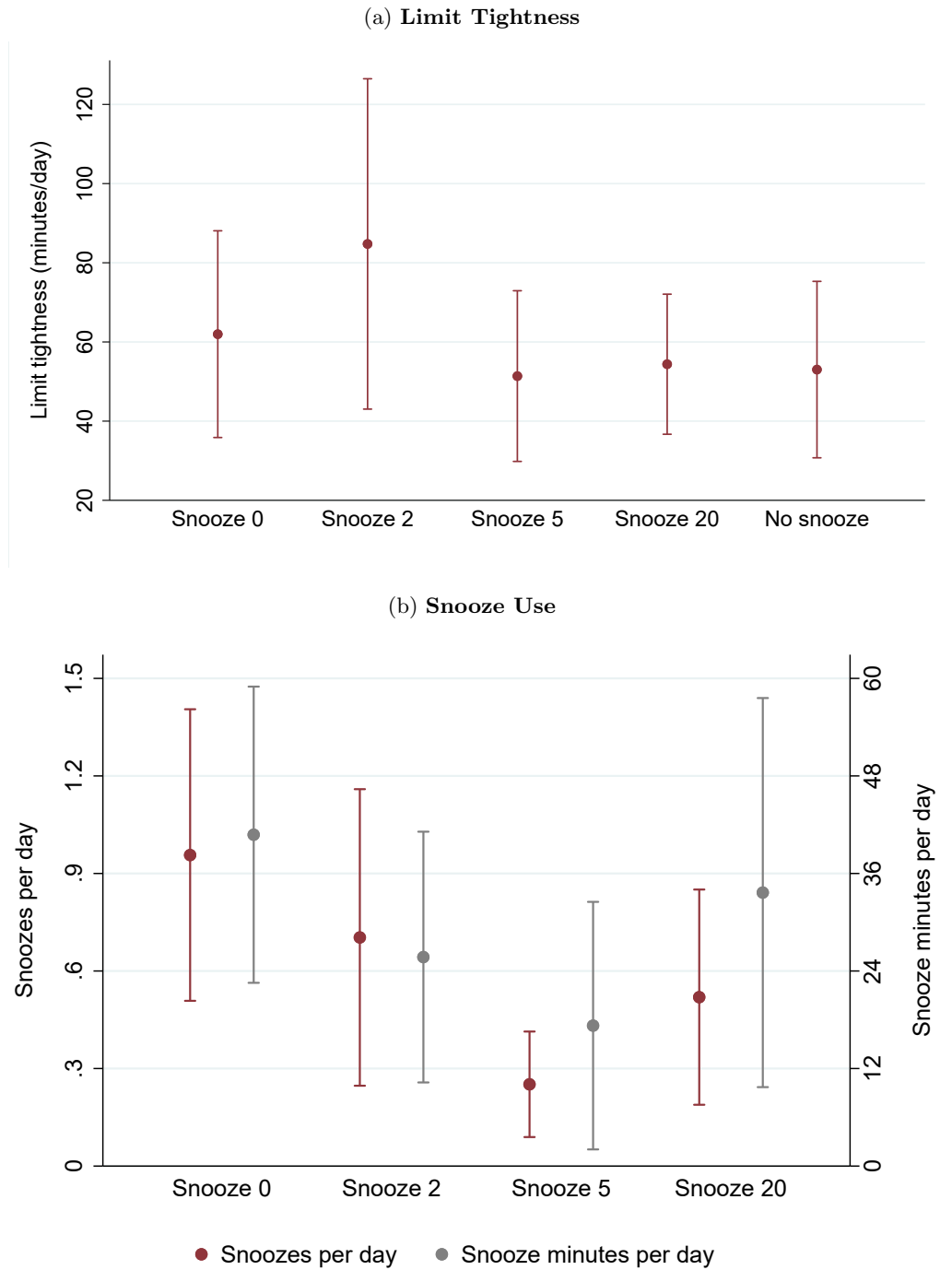
Notes: Panel (a) presents the average *user-level limit tightness* by snooze delay group. This is the sum over apps of the screen time reduction that would have resulted from the user's limits given their baseline usage patterns, in the absence of snoozes or other behavioral response. Panel (b) presents the average snooze use by snooze delay group.
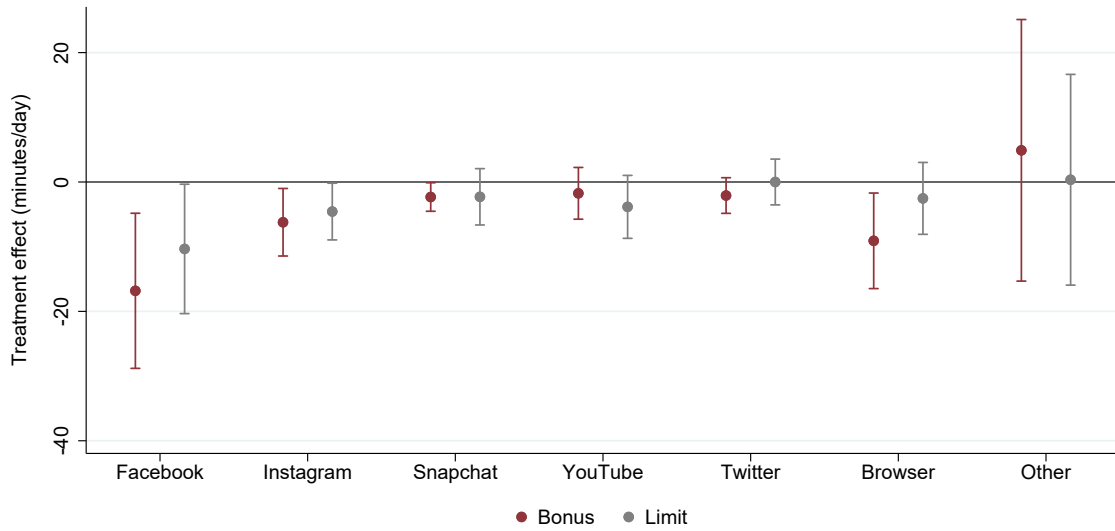
Figure 7: **Treatment Effects on FITSBY Use**
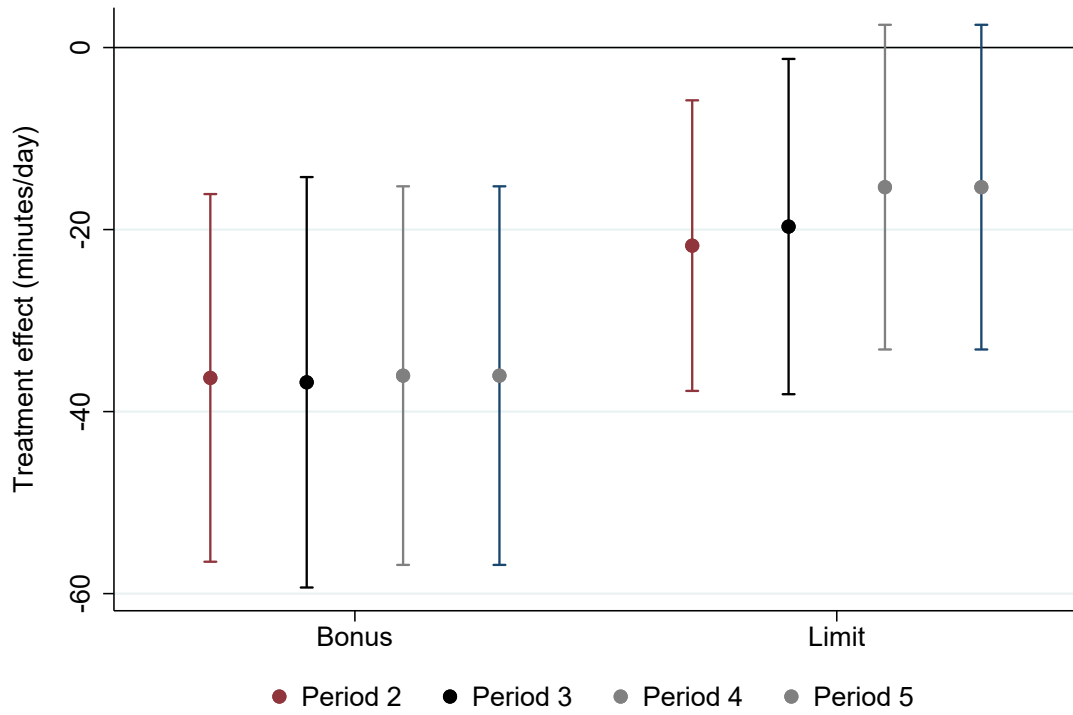
Notes: This figure presents effects of the Bonus and Limit treatments on FITSBY use using Equation (2). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.

Figure 8: **Treatment Effects of Limit on FITSBY Use by Snooze Delay Group**
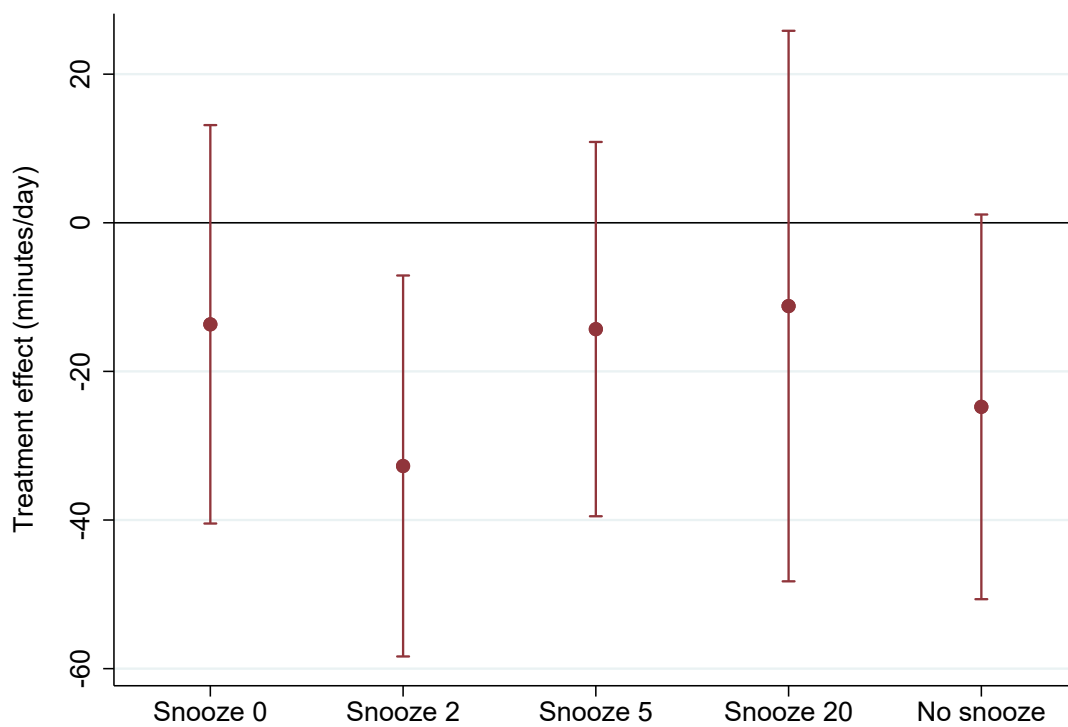
Notes: This figure presents effects of the Limit treatment on FITSBY use by snooze delay group using Equation (2). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.

Figure 9: **Effects on FITSBY Use by App**

Notes: This figure presents effects of the Bonus and Limit treatments on FITSBY use by app using Equation (2). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.

Figure 10: **Intent-to-Treat Effects of Limits and Bonus**
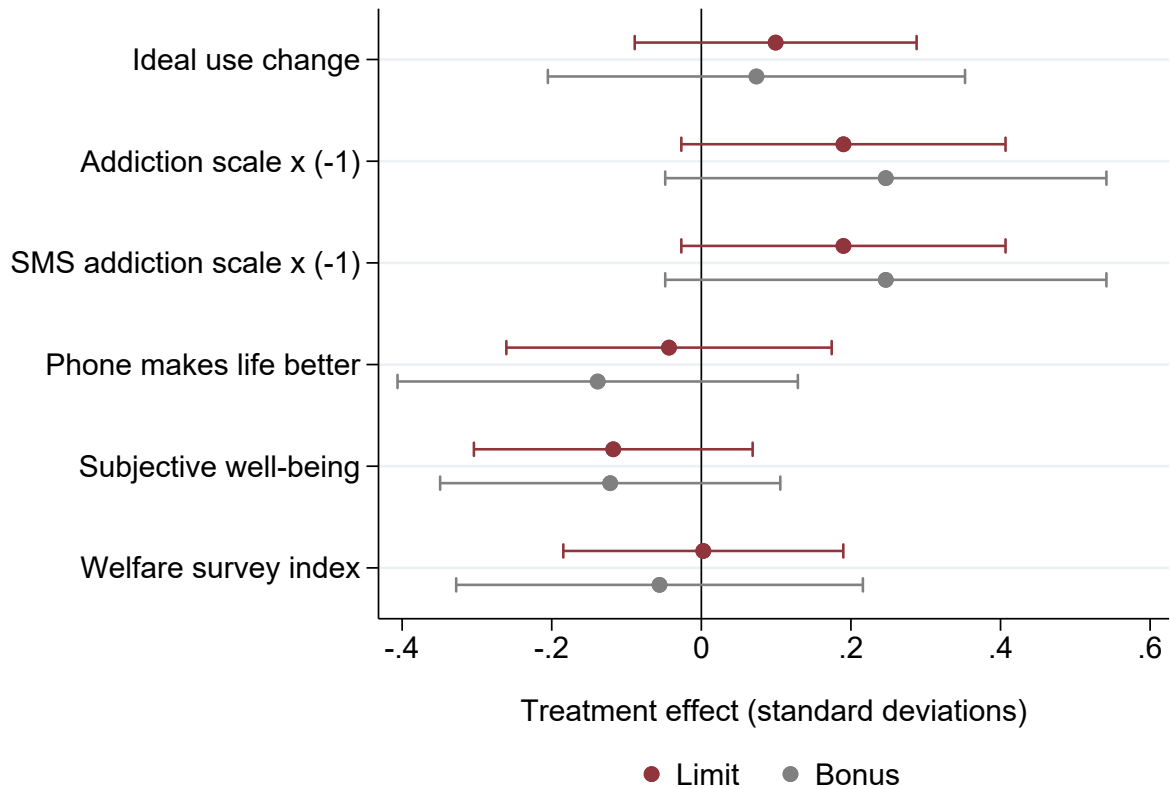
This figure presents effects of the Bonus and Limit treatments on survey welfare outcomes using Equation (2). *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure 11: **Predicted vs. Actual FITSBY Use in Control Conditions**
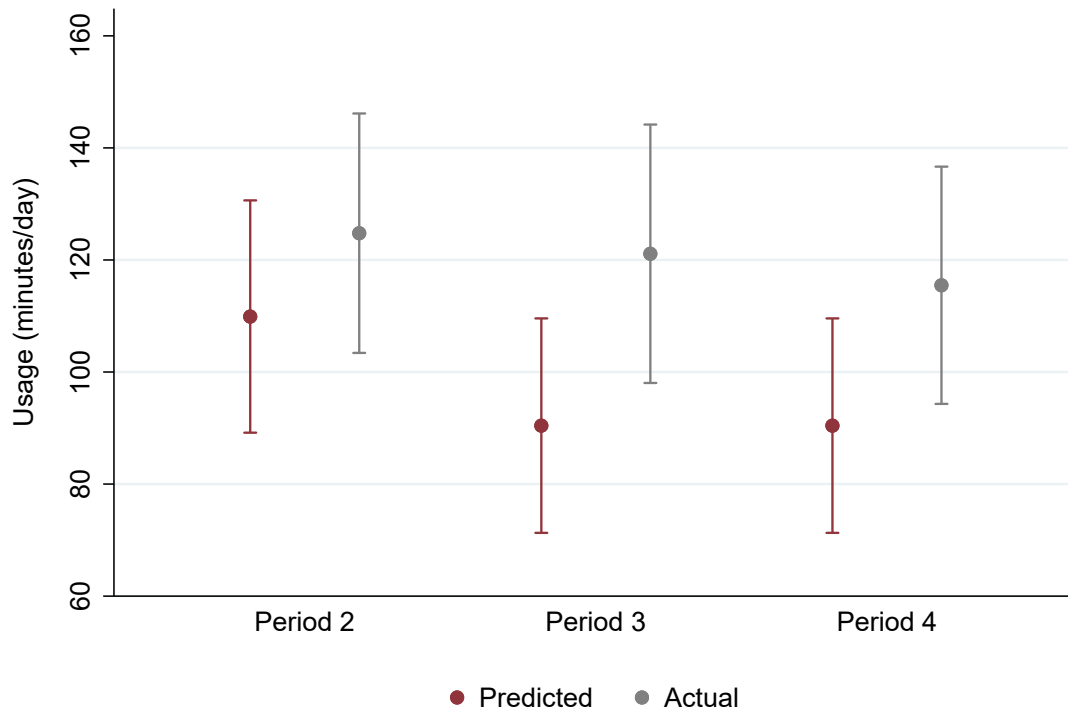
Notes: This figure presents average actual FITSBY use by period and average predicted FITSBY use on the survey at the beginning of the period for both Bonus Control and Limit Control. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.
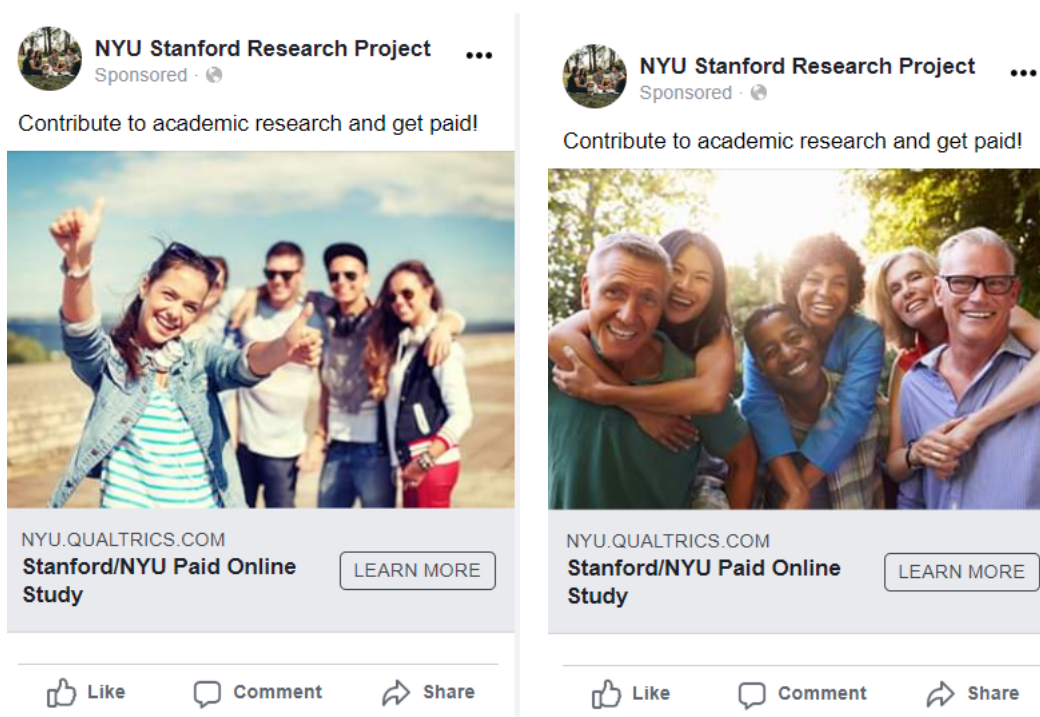
Online Appendix: Not for Publication

Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song

A   Appendix to Experimental Design Section

Figure A1: **Facebook Recruitment Ads**

Notes: The ads at left and right were shown to users aged 18–34 and 35–64, respectively.

A.1   Variable Definitions

Ideal use change. Some people say they use their smartphone too much and ideally would use it less. Other people are happy with their usage or would ideally use it more. How do you feel about your overall smartphone use over the past 3 weeks?

- I used my smartphone too much.

- I used my smartphone the right amount.

- I used my smartphone too little.

Relative to your actual use over the past 3 weeks, by how much would you ideally have [if "too much": reduced. If "too little": increased] your smartphone use? Please give a number in percent. ____ %

   Addiction scale. Over the past 3 weeks, how often have you...

- Been worried about missing out on things online when not checking your phone?

- Checked social media, text messages, or email immediately after waking up?

- Used your phone longer than intended?

- Found yourself saying "just a few more minutes" when using your phone?

- Used your phone to distract yourself from personal problems?

- Used your phone to distract yourself from feelings of guilt, anxiety, helplessness, or depression?

- Used your phone to relax in order to go to sleep?

- Tried to reduce your phone use without success?

- Experienced that people close to you are concerned about the amount of time you use your phone?

- Felt anxious when you don't have your phone?

- Found it difficult to switch off or put down your phone?

- Been annoyed or bothered when people interrupt you while you use your phone?

- Felt your performance in school or at work suffers because of the amount of time you use your phone?

- Lost sleep due to using your phone late at night?

- Preferred to use your phone rather than interacting with your partner, friends, or family?

- Put off things you have to do by using your phone?

Never, Rarely, Sometimes, Often, Always

SMS addiction scale.

- In the past 24 hours, did you use your phone longer than intended?

- In the past 24 hours, did your performance at school or work suffer because of the amount of time you used your phone?

- In the past 24 hours, did you feel like you had an easy time controlling your screen time?

- In the past 24 hours, did you use your phone mindlessly?

- In the past 24 hours, did you use your phone because you were feeling down?

- In the past 24 hours, did using your phone keep you from working on something you needed to do?

- In the past 24 hours, would you ideally have used your phone less?

- Last night, did you lose sleep because of using your phone late at night?

- When you woke up today, did you immediately check social media, text messages, or email?

Please text back your answer on a scale from 0 (not at all) to 10 (definitely).

Phone makes life better. To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?

11-point scale from -5 (Makes my life worse) to 0 (Neutral) to 5 (Makes my life better)

Subjective well-being. Please tell us the extent to which you agree or disagree with each of the following statements. Over the past 3 weeks, ...

- ... I was a happy person

- ... I was satisfied with my life

- ... I felt anxious

- ... I felt depressed

- ... I could concentrate on what I was doing

- ... I was easily distracted

- ... I slept well

7-point scale from strongly disagree to neutral to strongly agree

B   Appendix to Data Section

Table A1: **Descriptive Statistics**

	Mean	Standard deviation	Minimum value	Maximum value
Ideal use change	-20.3	22.7	-99	0
Addiction scale x (-1)	-7.1	2.9	-16	0
SMS addiction scale x (-1)	-7.1	2.9	-16	0
Phone makes life better	1.7	2.1	-3	5
Subjective well-being	-0.7	3.5	-16	13

Notes: This table present descriptive statistics for the survey welfare outcomes at baseline.

Table A2: **Covariate Balance**

(a) **Limit Treatment and Control**

Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	37.22 (34.43)	41.21 (33.24)	0.40
College	0.59 (0.49)	0.60 (0.49)	0.84
Male	0.38 (0.49)	0.35 (0.48)	0.69
White	0.67 (0.47)	0.68 (0.47)	0.85
Age	33.54 (9.34)	33.85 (9.40)	0.81
Period 1 FITSBY use (minutes/day)	59.00 (42.76)	54.72 (40.96)	0.46
N	124	88	
F-test of joint significance (p-value)			0.94
F-test, number of observations			207

(b) **Bonus Treatment and Control**

| Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	42.18 (36.55)	37.96 (33.23)	0.47
College	0.64 (0.48)	0.58 (0.50)	0.44
Male	0.32 (0.47)	0.37 (0.48)	0.51
White	0.68 (0.47)	0.67 (0.47)	0.93
Age	33.42 (8.96)	33.74 (9.57)	0.83
Period 1 FITSBY use (minutes/day)	54.90 (40.01)	56.73 (40.47)	0.78
N	50	159	
F-test of joint significance (p-value)			0.94
F-test, number of observations			204

Notes: Panels (a) and (b) present tests of covariate balance for the Limit and Bonus treatment and control groups.

Figure A2: **Most Popular Apps**
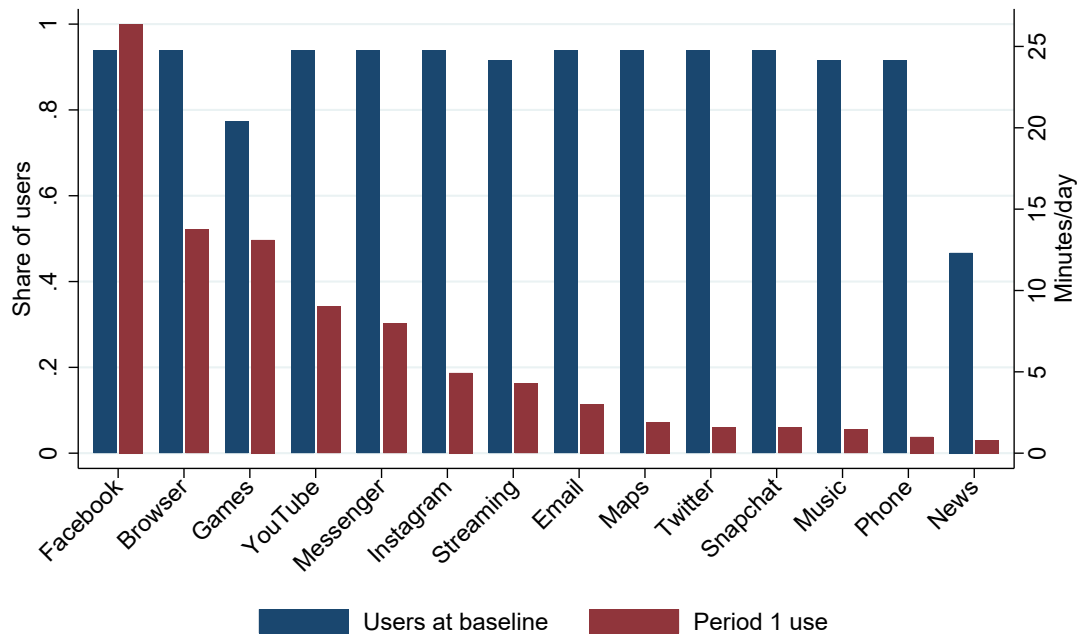
Notes: This figure presents the share of users that have each app and the average daily screen time in Period 1 (baseline). Period 1 use is across all users, not conditioning on whether or not they have the app.
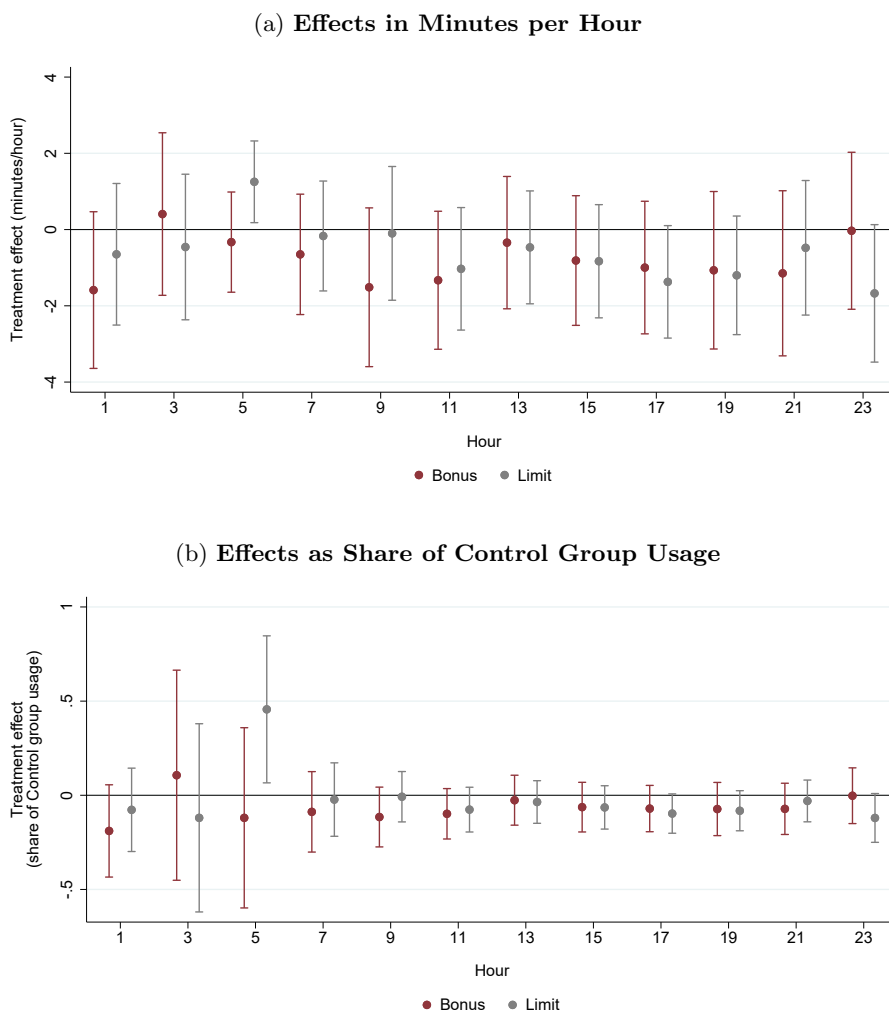
Figure A3: **Distribution of Baseline FITSBY Use**
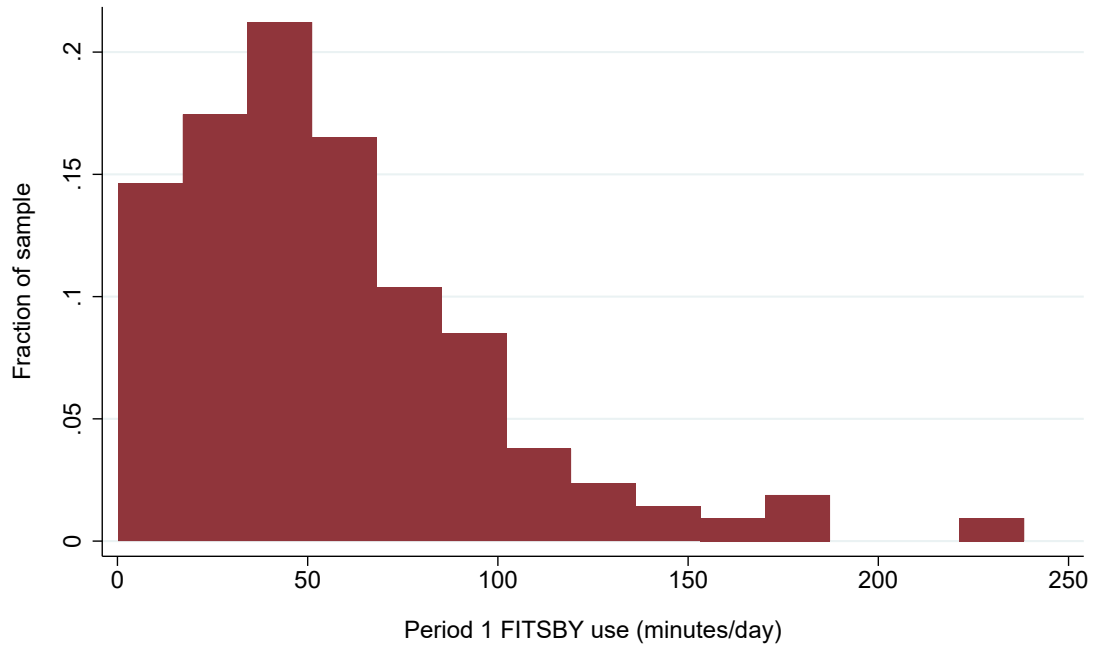
Notes: This figure presents a distribution of FITSBY use in Period 1 (baseline). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.

C   Appendix to Empirical Results Section

Figure A4: **Effects on FITSBY Use by Time of Day**

(a) **Effects in Minutes per Hour**

(b) **Effects as Share of Control Group Usage**

Notes: This figure presents effects of the Bonus and Limit treatments on FITBSY use by time of day using Equation (2). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.
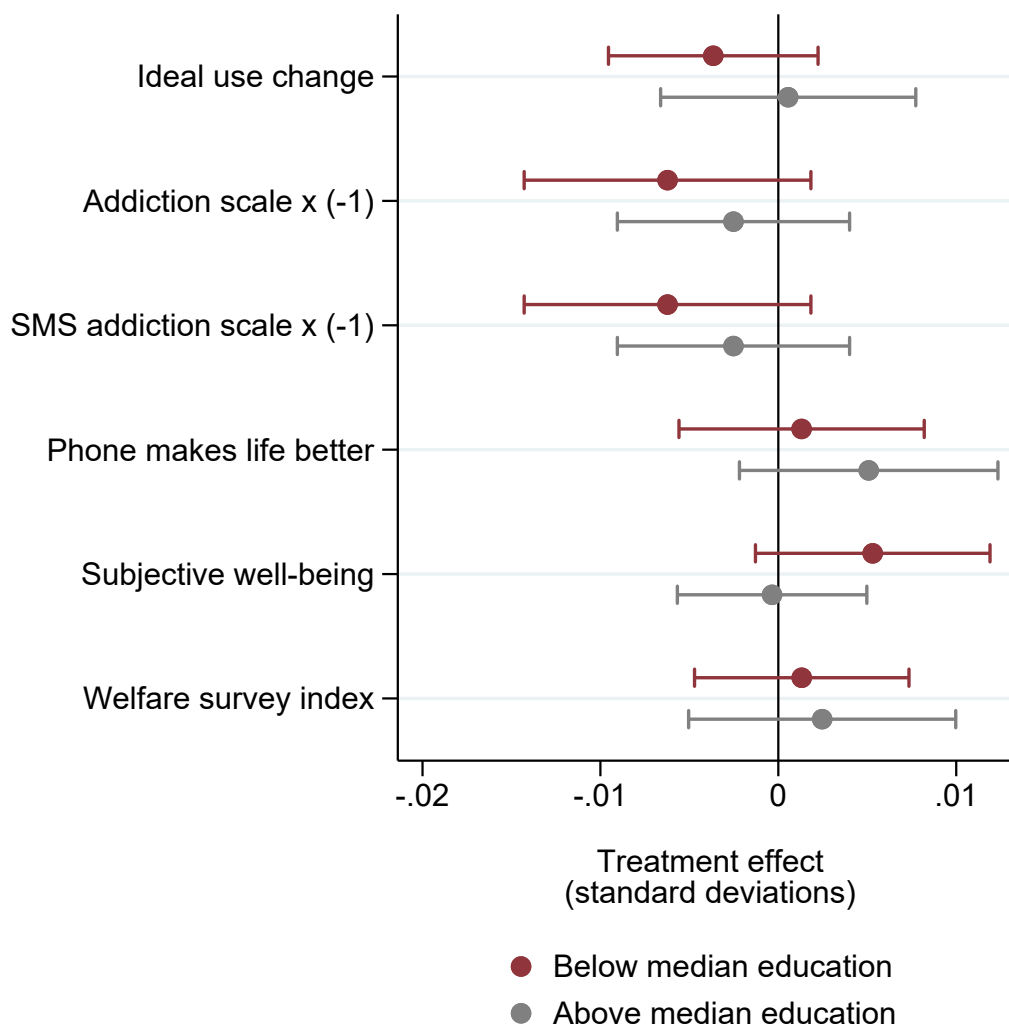
Figure A5: **Local Average Treatment Effects of FITSBY Use on Welfare Survey Outcomes**
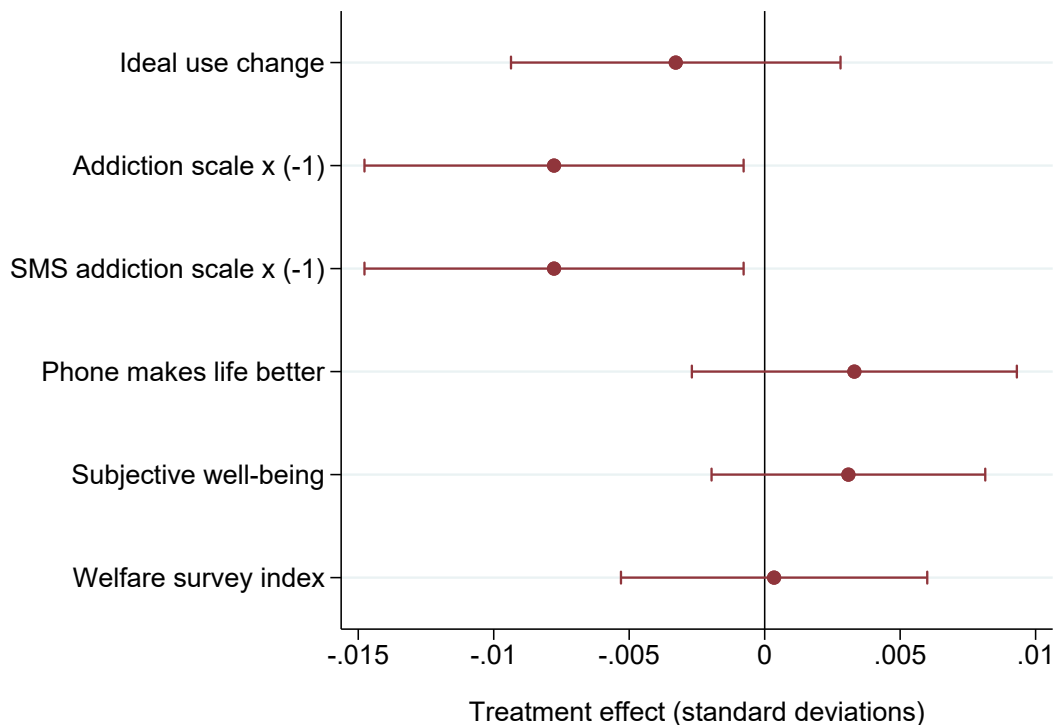
Treatment effect (standard deviations)

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3). We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

C.1 Heterogeneous Treatment Effects

Figure A6: **Heterogeneous Effects on Welfare Survey Outcomes by Education**

Treatment effect
(standard deviations)

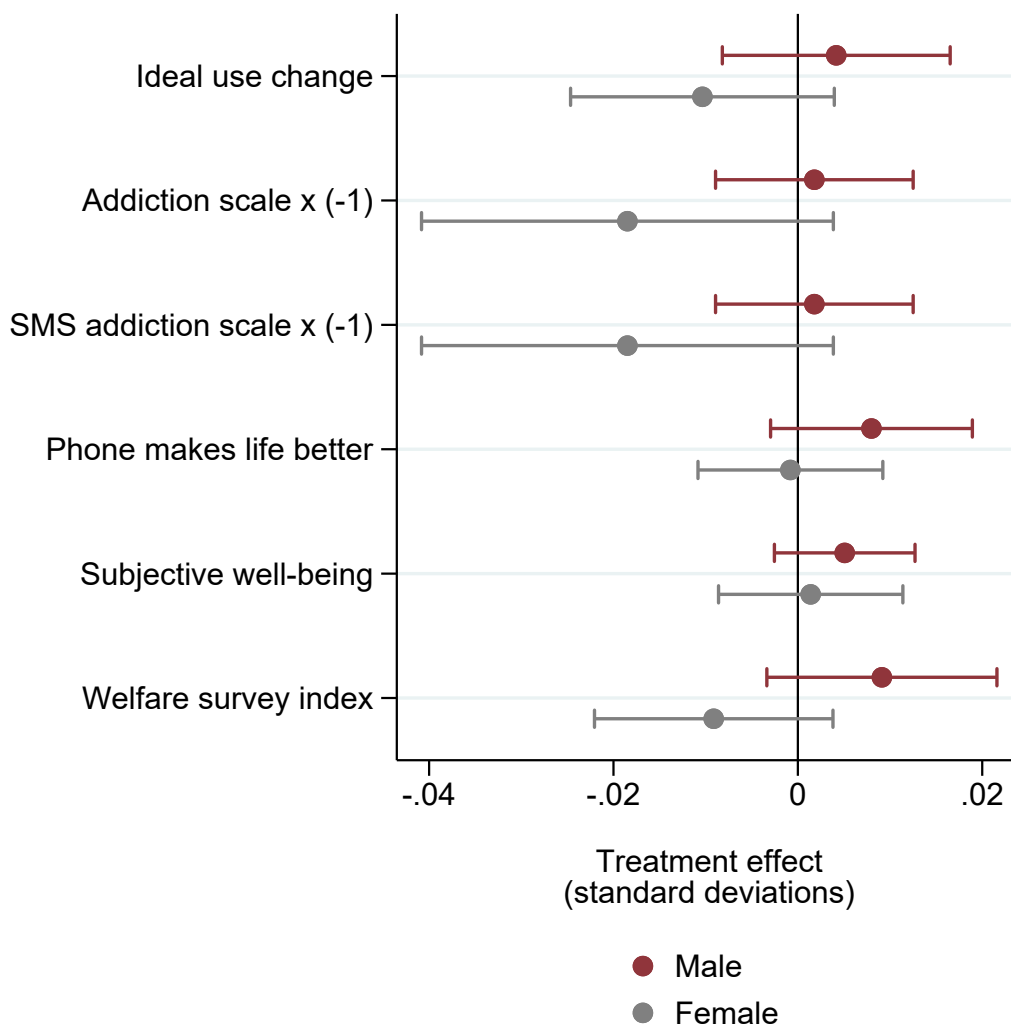● Below median education
● Above median education

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3), for above- and below-median education. We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A7: **Heterogeneous Effects on Welfare Survey Outcomes by Age**
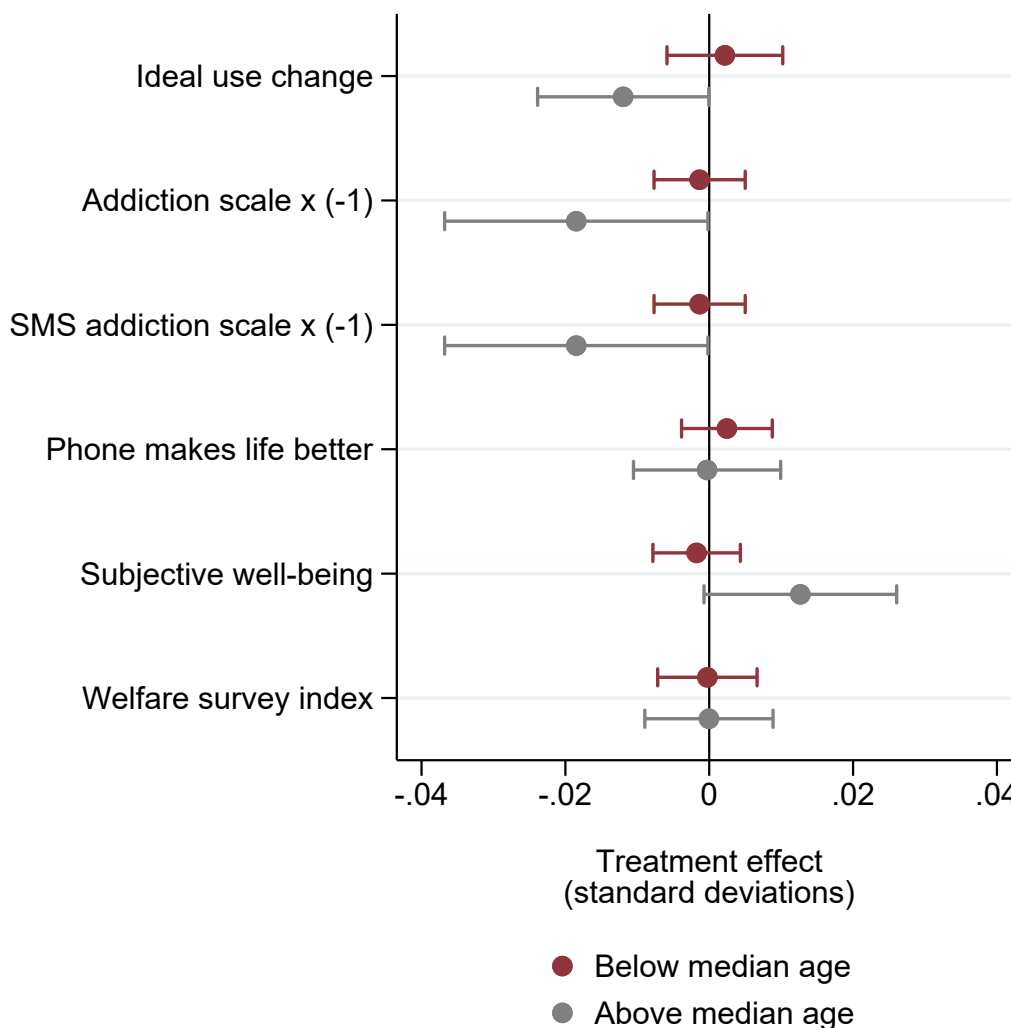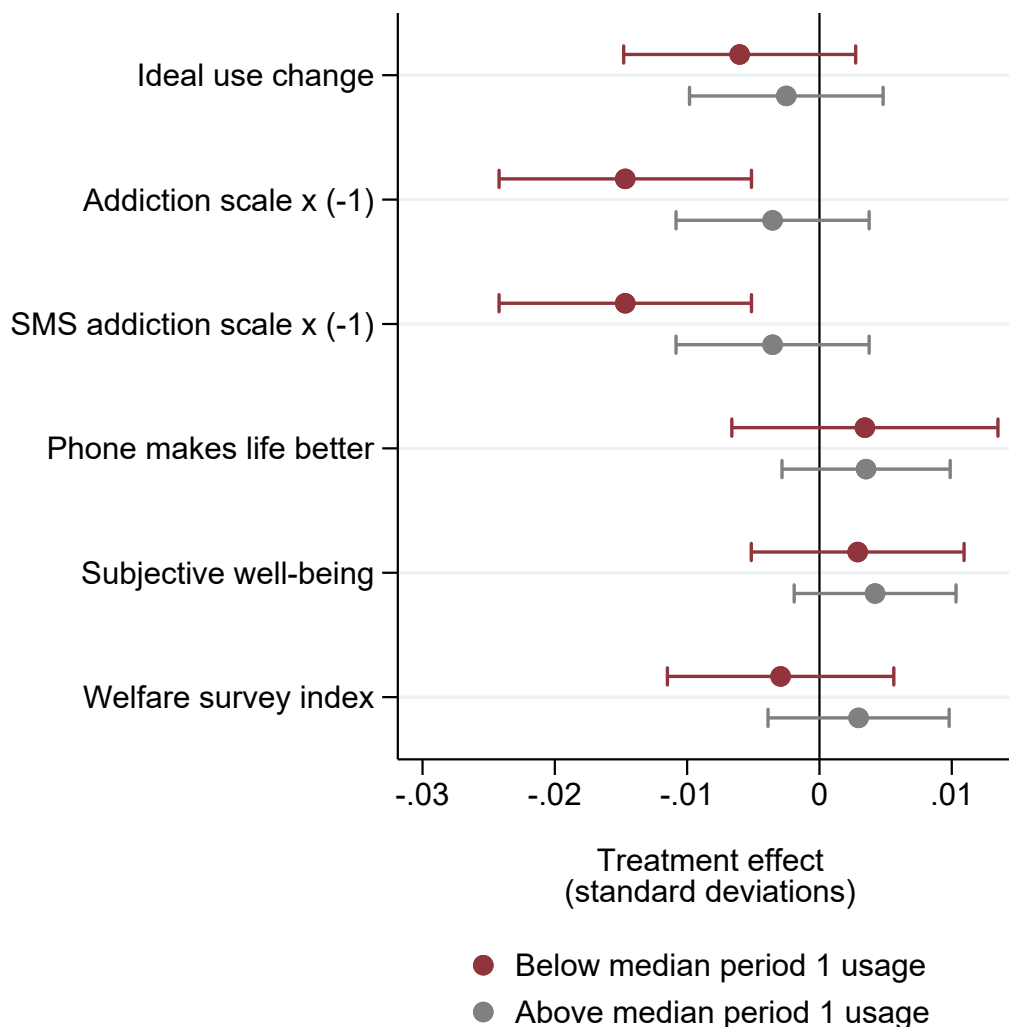

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3), for above- and below-median age. We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

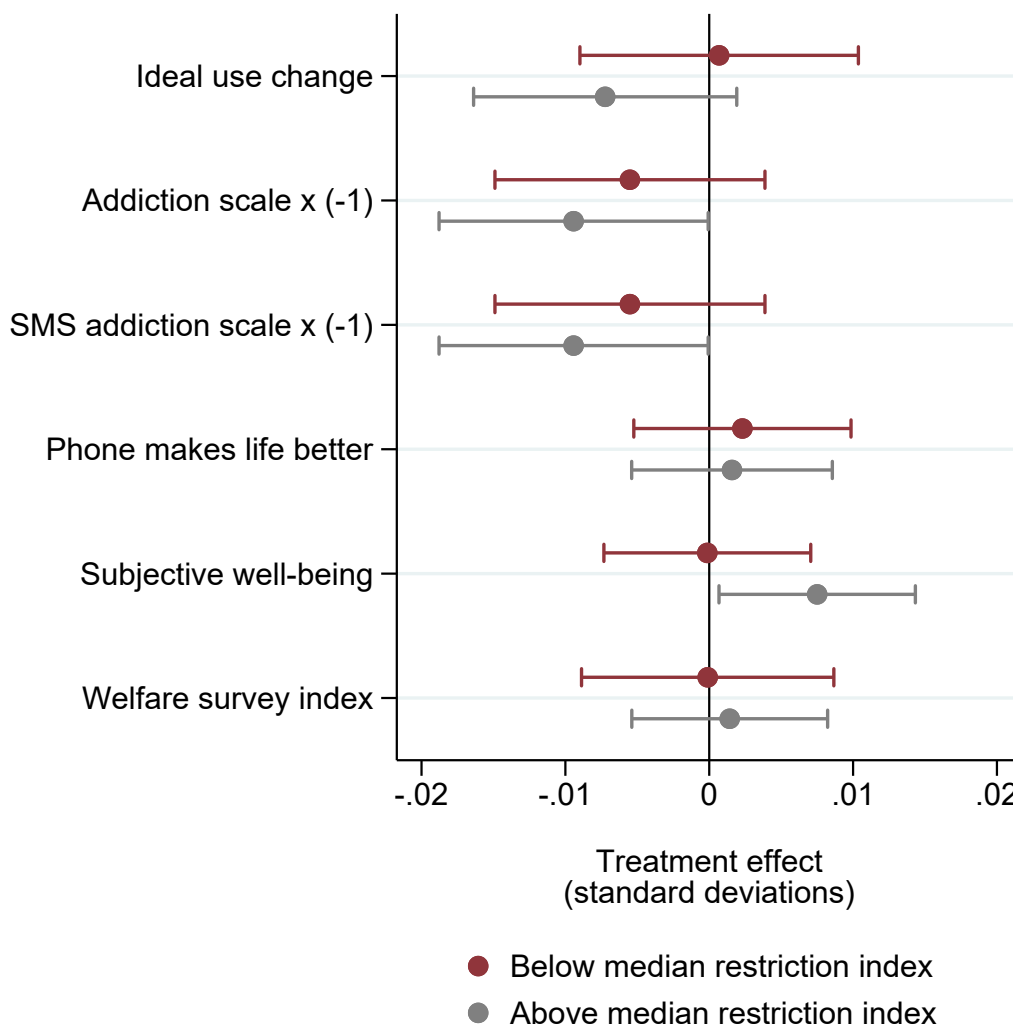Figure A8: **Heterogeneous Effects on Welfare Survey Outcomes by Gender**

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3), for men versus women. We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A9: **Heterogeneous Effects on Welfare Survey Outcomes by Baseline FITSBY Use**
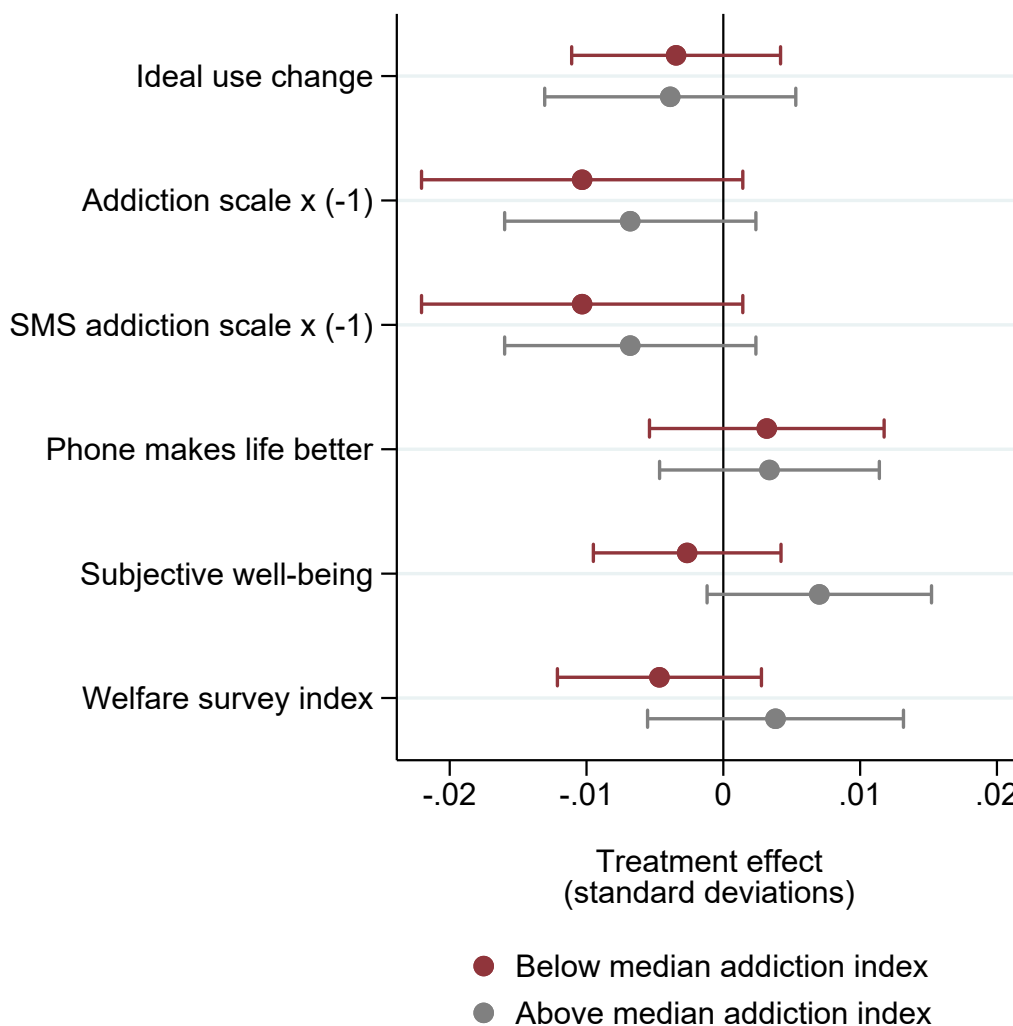

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3), for above- and below-median baseline FITSBY use. We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A10: **Heterogeneous Effects on Welfare Survey Outcomes by Restriction Index**

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3), for above- and below-median values of *restriction index*, a combination of *interest in limits* and *ideal use change*. We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A11: **Heterogeneous Effects on Welfare Survey Outcomes by Addiction Index**

● Below median addiction index
● Above median addiction index

Notes: This figure presents local average treatment effects of FITSBY use on welfare survey outcomes using Equation (3), for above- and below-median values of *addiction index*, a combination of *addiction scale* and *phone makes life better*. We instrument for FITSBY use with Bonus and Limit indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Welfare survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.