NBER WORKING PAPER SERIES

# WHY DON'T WE SLEEP ENOUGH?
# A FIELD EXPERIMENT AMONG COLLEGE STUDENTS

Mallory L. Avery
Osea Giuntella
Peiran Jiao

Working Paper 30375
http://www.nber.org/papers/w30375

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2022

We thank participants to seminars at the University of Pittsburgh, University of Rome 2 Tor Vergata, University of Milano Statale, Maastricht University, the Nuffield Centre for Experimental Social Sciences, the Pittsburgh Experimental Economics Lab, the Circadian Rhythms and Sleep Conference Grands, the Advances in Field Experiments Conference, the Workshop on Risky Health Behaviors, the American-European Health Study Group, the IV Workshop on Behavioral and Experimental Health Economics, the Center for Health Incentives and Behavioral Economics (CHIBE) Roybal Retreat, and the Maastricht Behavioral and Experimental Economics Symposium. We benefited from discussion with Domenico De Palo, David Dickinson, Daniel Hamermesh, Ben Handel, David Huffman, Gautam Rao, Silvia Saccardo, Frank Schilbach, Sally Sadoff, Heather Schofield, Jeffrey Shrader, Justin Sydnor, Severine Toussaert, and Lise Vesterlund. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Why Don't We Sleep Enough? A Field Experiment Among College Students
Mallory L. Avery, Osea Giuntella, and Peiran Jiao
NBER Working Paper No. 30375
August 2022
JEL No. B49,C93,I10

## ABSTRACT

This study investigates the mechanisms affecting sleep choice and explores whether commitment devices and monetary incentives can be used to promote healthier sleep habits. To this end, we conducted a field experiment with college students, providing them incentives to sleep and collecting data from wearable activity trackers, surveys, and time-use diaries. Monetary incentives were effective in increasing sleep duration with some evidence of persistence after the incentive was removed. We uncover evidence of demand for commitment. Our results are consistent with partially sophisticated time-inconsistent preferences and overconfidence, and have implications for the effectiveness of information interventions on sleep choice.

Mallory L. Avery
Monash University
Australia
mallory.avery@monash.edu

Peiran Jiao
Maastricht University
Netherlands
p.jiao@maastrichtuniversity.nl

Osea Giuntella
Department of Economics
University of Pittsburgh
Posvar Hall
Pittsburgh, PA 15260
and NBER
osea.giuntella@gmail.com

# 1 Introduction

Sleep deprivation is an emerging public health challenge. According to the Center for Disease Control and Prevention, more than a third of American adults sleep less than the recommended minimum of seven hours (Liu, 2016). Some scholars consider it the most prevalent risky behavior in modern societies and evidence suggests that in many countries people may be sleeping between one and two hours less than what their ancestors used to sleep one hundred years ago (Roenneberg, 2013). Growing evidence documents the causal effects of sleep deprivation on chronic diseases, health, cognitive skills, decision making, human capital, and productivity (e.g. Gibson and Shrader, 2018; Heissel and Norris, 2018; Giuntella and Mazzonna, 2019), and firms, athletes, and military training programs increasingly recognize how sleep deprivation can impair performance.[1]

Despite sleep being increasingly recognized as a fundamental contributor to health and human capital and despite economists' interest in time-use (Becker, 1965; Hamermesh, 2019), sleep behavior has received little attention in the economic literature. Given that we spend approximately a third of our time –one of our scarcest resources– sleeping, and given the substantial economic and health impacts of sleep deprivation, sleep behavior should be an object of natural interest to economists (Mullainathan, 2014). However, most economic models analyzing time allocation regard sleep as a predetermined and homogeneous constraint on time allocation. While for some individuals sleep duration and quality are influenced by medical conditions (insomnia, sleep apnea, etc.), for most individuals bedtime and sleep duration are choices. Individuals may optimally allocate less time to sleep and delay their bedtime (or wake up earlier) to work longer or enjoy more leisure. And indeed, the few pioneering studies analyzing sleep choice have assumed individuals choose hours of sleep optimally (Biddle and Hamermesh, 1990). Yet, according to the Royal Philips global sleep survey, 8 in 10 adults worldwide want to improve their sleep and a poll from YouGov suggests that, while 89% of Americans would like to sleep

[1] Recently Aetna, an American managed health care company, introduced incentives to increase workers' sleep (see https://www.cnbc.com/2016/04/05/why-aetnas-ceo-pays-workers-up-to-500-to-sleep.html). Concern has been raised regarding sleep deprivation among NBA players (see https://www.espn.com/nba/story/_/id/27767289/dirty-little-secret-everybody-knows-about). Finally, sleep is one of the three pillars of the army's performance triad, along with physical activity and nutrition (see https://armymedicine.health.mil/Performance-Triad).

for 7 hours or more each night, more than 40% report to sleep less than that.[2]

This study investigates sleep choice and the role of commitment devices and monetary incentives to promote healthier sleep habits. We conducted a field experiment among college students and collected data from wearable activity trackers, surveys, and time-use diaries. Eliciting preferences and randomizing incentives to go to bed earlier and sleep longer, we shed light on the role of present bias, biased beliefs, demand for commitment, and habit formation in sleep.

We recruited a total of 508 participants at the University of Oxford (163 subjects) and the University of Pittsburgh (345 subjects) across 6 waves of the experiment. They were given wearable devices (Fitbit) to collect data on their sleep, physical activity, and heart rate for 8 (waves 1-5) or 15 (wave 6) consecutive weeks.[3] In the treated group, subjects set bedtime and sleep duration targets for themselves each Monday of the 3 treatment weeks and were rewarded for each night (Monday through Thursday) that both targets were achieved based on Fitbit data. We elicited subjects' time and risk preferences in the lab, and integrated the data collected from wearable devices with weekly surveys and time-use diaries.

Our monetary incentives were effective in improving sleep behavior. Treated participants responded to monetary incentives by sleeping longer. They were 13% more likely to sleep the recommended number of hours (between 7 and 9) and 16% less likely to sleep less than 6 hours. This finding is robust to the inclusion of individual fixed effects, accounting for time-invariant individual heterogeneity. Furthermore, we document a persistent improvement in sleep. Even after the intervention was removed, treated participants were 9% more likely to sleep between 7 and 9 hours. Our intervention also had effects on sleep regularity, reducing variance in sleep duration, bedtimes, and (more weakly) wake-up times.

Given that participants positively responded to our incentive to sleep, a natural question is how they reallocated their time to achieve their targets. To address this question, we collected time-use diaries before, during, and after the intervention, and examined how individuals in the treatment group allocated their time when receiving incentives to go to bed earlier and sleep longer. The only activity that systematically and significantly changed during the intervention

[2]See `https://www.usa.philips.com/c-e/smartsleep/campaign/world-sleep-day.html` and `https://today.yougov.com/topics/health/articles-reports/2019/03/13/sleep-habits-americans-survey-poll`.

[3]Subjects were monitored for 2 pre-treatment weeks, 3 treatment weeks, and either 3 (waves 1-5) or 10 (wave 6) post-treatment weeks.

was screen time (i.e., watching TV, videos, or using smart devices etc.), which declined by approximately 40 minutes.

Despite only 43% of participants sleeping for 7 to 9 hours at baseline, more than 90% reported that their ideal sleep duration was longer than 7 hours and, among treated subjects, 92% reported that the sleep and bedtime target ranges provided in the experiment were in line with what was best for them. Combined with the aforementioned results on the decrease in screen time, this suggests that sleep decisions may be characterized by dynamic inconsistency, as delaying bedtime may have immediate benefits (i.e., the utility from watching a further episode of a TV series, or working an extra hour), but delayed costs (i.e., the lack of energy or alertness following a night of poor sleep). This suggests that there may be scope for incentives and commitment devices to promote optimal behavior (O'Donoghue et al., 2006). The effectiveness of incentives and commitment devices to promote optimal choices in the presence of self-control problems has been analyzed in the context of other health behaviors such as alcohol consumption, unhealthy eating, and exercising (Charness and Gneezy, 2009; Volpp et al., 2009; Just and Price, 2013; Acland and Levy, 2015; Royer et al., 2015). However, sleep is a particularly interesting domain in which to investigate the prevalence and persistence of behavioral biases. Firstly, it is an activity that people engage in every day, and about which they have received repeated feedback throughout their lives. Thus, sleep is a domain wherein demand for commitment might be highly relevant. Daily experience and feedback with time-inconsistent behavior may raise some individuals' awareness of their problem (Laibson, 1997; O'Donoghue and Rabin, 1999), and increase demand for commitment (Rabin et al., 1999; DellaVigna and Malmendier, 2006; Laibson, 2015; Schilbach, 2019). Secondly, sleep is also an interesting domain to study biased beliefs. If people are persistently overconfident about their own sleep even in the face of extensive experience and feedback (Huffman et al., 2018), this can help us disentangle different mechanisms of biased beliefs. For instance, if individuals have biased recall of own sleep and fail to debias in the face of repeated feedback, this may be consistent with motivated reasoning (Bénabou and Tirole, 2016) rather than misinformation, suggesting the potential of using incentives to mitigate the bias (Zimmermann, 2019).

We uncover evidence that participants voluntarily opted for commitment devices in the form of setting restrictive targets. Our findings are consistent with partially sophisticated time inconsistency and biased beliefs as key behavioral mechanisms underlying poor sleep choices. In total, 60% of our subjects took up some form of commitment. More present-biased subjects reported less sleep at baseline and were more likely to take up commitment devices. Among them, commitment devices reduced insufficient sleep. Meanwhile, many subjects were overconfident about their own achievement rates, over-remembered their own bedtime and sleep duration, over-placed their own sleep duration and quality among peers, and understated personal risk associated with sleep deprivation relative to the risk they predicted for peers. Overconfident subjects–defined as those reporting longer sleep duration than that measured using Fitbits–were more likely to be sleep deprived at baseline and selected overly optimistic targets. Present-biased individuals were more likely to achieve their targets if they were less overconfident. We also discuss the role of alternative explanations, such as planning fallacy and experimenter demand.

The rest of the paper is organized as follows. Section 2 discusses previous related work, our contribution, and the relevance of students for the study of sleep choice. Section 3 describes the experimental procedure, the design of our intervention, the empirical specification and the data. In Section 4, we present the results of our randomized experiment, discuss the effectiveness and persistence of incentives to sleep. Section 5 discusses the role of behavioral biases in the sleep domain. Concluding remarks are provided in Section 6.

## 2 Background

### 2.1 Literature

We directly relate to recent studies analyzing the effects of wearable technology on sleep and health behavior (Patel et al., 2015; Jakicic et al., 2016). Handel and Kolstad (2017) exploit a large-scale intervention in a firm to randomize subjects into treatments to improve sleep and exercise through planning. They find small effects of planning tools. Our findings suggest in the presence of persistent behavioral biases the introduction of monetary incentives and commitment devices may be more effective than using planning tools alone. Bessone et al. (2021) randomize incentives to sleep longer to analyze the effects of sleep on labor market productivity and health in a developing country, finding little evidence of an impact of sleep on short-run economic outcomes, but significant effects of naps on attention and well-being. While their main goal was to induce exogenous variation in sleep (and naps) to assess its effects on human capital and productivity, our study focuses directly on the mechanisms behind sleep choice. Furthermore, by using Fitbit data on individuals beyond the immediate treatment period, we are the first to examine persistence and habit formation effects in the context of sleep decisions.[4] Our findings regarding habit formation are twofold. First, we find that sleep changes were persistent within the duration of our experiment even after treatment ended, suggesting that temporary incentives could lead to long-run lifestyle changes in the sleep domain. Second, treated subjects' adjustment of screen time and evening routines could imply longer-term habit formation effects.

Our study also contributes to the literature analyzing demand for commitment and the effectiveness of commitment devices (see, e.g., Bryan et al., 2010; Kremer et al., 2019; Schilbach, 2019, for a review). To date, little evidence exists in the sleep domain regarding the effectiveness of commitment devices in improving sleep habits. The evidence on the effectiveness of commitment devices in general is mixed (Laibson, 2015, 2018). Some studies support the idea that commitment devices may help sophisticated agents with present bias mitigate their self-control problems (Ashraf et al., 2006; Kaur et al., 2015; Schilbach, 2019). Others argue that uncertainty could undermine the demand for commitment (Laibson, 2015) and that, unless subjects are sufficiently sophisticated, commitment devices may be welfare reducing (John, 2020; Bai et al., 2021). Furthermore, in a recent work, Carrera et al. (2019b) show that commitment contract take-up may reflect, at least in part, demand effects or "noisy valuation" when there is substantial uncertainty about the desirability of an activity, even if subjects are time consistent. However, the continuous experience and frequent feedback that characterize sleep behavior suggest commitment devices may be more effective in this domain. Our experiment provides a relatively soft commitment device in the form of setting bedtime and sleep duration targets at the cost of forgone rewards.[5] Additionally, we elicit time preferences and measure subjects' belief about own future performance in incentivized tasks.

[4] Breig et al. (2018) also consider sleep in a study using wearable devices. However, their main focus is task allocation. In a 2-week experiment, they randomize feedback on subjects' time allocation and explore how that affects their time use in the following week. Their findings show the role for overoptimism in time allocation decisions. Our focus is instead on sleep, and we conduct an 8-week (15-week in wave 6) field experiment to analyze the effects of randomized incentives to sleep, their effects on time use, and shed light on the role of demand for commitment, overconfidence, and habit formation in the sleep domain.

[5] Previous evidence also suggests that softer commitments may work better than hard commitments (Dupas and Robinson, 2013).

### 2.2 Students: An Interesting Population to Study Sleep Deprivation

While sleep deprivation is a problem for many age groups, there are several reasons for sleep deprivation and sleep choice among college students being of particular interest. First, time management is a major challenge among college students transitioning from high school and home habits to campus life (Misra and McKean, 2000). Second, sleep deprivation among college students is increasingly becoming a reason for concern. According to recent statistics published in a report of the National Institute of Health (Hershner and Chervin, 2014), more than 70% of college students sleep less than eight hours a day, 60% say they are "dragging, tired, or sleepy" at least three days a week, and more than 80% say loss of sleep affects their academic performance. Third, sleep deprivation and poor sleep quality have been associated with various aspects of undergraduate mental health (Milojevich and Lukowski, 2016), including symptoms of psychological distress, anxiety, attention deficit, and depression problems (McEwen, 2006; Kahn-Greene et al., 2007).[6] Fourth, college is also a crucial phase to shape one's lifestyle and habits (Buboltz et al., 2001). Fifth, college students are a group that is physically healthier, with fewer social and familial constraints and with more time flexibility, suggesting that this is an appropriate group for our experimental study of sleep choice.

## 3 Experimental Procedure, Design, and Data

### 3.1 Experimental Procedure

The results reported in this paper were derived from six waves of experimental sessions. The first three waves were conducted in Oxford. The last three were conducted at the University of Pittsburgh. The experiment was first advertised in the University of Oxford and the Oxford Brookes University in the Oxford waves (1–3) and on the University of Pittsburgh campus in the Pittsburgh waves (4–6).[7]

[6] This is of particular concern, given that depression, anxiety, and suicide rates are rising among US college students (Liu et al., 2019).

[7] The experiment was conducted at the Centre for Experimental Social Sciences (CESS) in Nuffield College, Oxford, UK and at the Pittsburgh Experimental Economics Laboratory (PEEL) at the University of Pittsburgh. The experimental procedure was approved by the respective research ethics committees. All subjects provided informed consent before participating, and could withdraw at any point of the study. Interested participants then signed up on our recruiting website. To mitigate experimenter demand effect and selection effect, we advertised our experiment as a study about the use of wearable devices and general health behavior. In all our estimates, we include dummies for

Waves 1-5 were conducted over 8 weeks (which in Oxford coincided with the length of the academic term), with 2 weeks before and 3 weeks after the 3 treatment weeks. Wave 6 was conducted over 15 weeks, with 2 weeks before and 10 weeks after the 3 treatment weeks.[8] Recruitment occurred a week before the beginning of the experiment (Week 0). All weeks were during school terms except weeks 12-15 in wave 6. For waves 1 through 5, subjects were invited to the lab in week 1 for an experimental session and were given a Fitbit Charge HR device. During wave 6, because of COVID-19 restrictions, subjects participated in a virtual lab session (Danz et al., 2021) and received their device, a Fitbit Inspire II, via mail.

We collected baseline data from Fitbit devices for the first two weeks. Experimental surveys and treatments started on Monday morning of Week 3, and all participants' Fitbit data were monitored until the end of the 8 or 15-week period. On Friday of the last week, participants returned the devices and received final payments. A show-up fee of GBP 4 ($\approx$ USD 5.3 at the time of our sessions) in Oxford or USD 6 in Pittsburgh was given both in the Week 1 lab session and when they returned the Fitbit, regardless of their performance. Among subjects who synced at least 80% of their data and answered more than 50% of the surveys, a lottery was then drawn, and 3% of them each won a reward of GBP 150 ($\approx$ USD 199 at the time of our sessions) in Oxford or USD 200 in Pittsburgh.

The first lab session was divided into three parts. The first part was an incentivized elicitation of risk and time preferences (see Appendix B for details of the elicitation methods). The second part of the lab session involved several survey items, eliciting details on subjects' demographics, health conditions, cognition, lifestyle, health behaviors, and physical activity. Additionally, a survey measure of domain-specific risk attitudes (Weber et al., 2002) was implemented, which included a health domain. To further minimize the experimenter demand effect, in all our surveys we included questions on physical exercise, health and mental health. In the third part of the lab session, each participant was given a Fitbit device.

whether the subject was recruited in Oxford or in Pittsburgh.

[8] We thank anonymous referees for suggesting to extend the post-treatment period to better assess the persistence of the treatment effects.

### 3.2 Experimental Design

The first 2 weeks after the (virtual) lab session were for baseline data collection for all subjects. In waves 1, 5, and 6, subjects were randomly assigned to a control condition or an incentive treatment at the end of week 2, while in waves 2-4, subjects were only assigned to one of our treatment conditions. Details of waves and treatments are in Table A.1. In the control group, participants were asked only to wear the device, allowing their Fitbit data to be recorded, and to answer some surveys. All subjects received Weekly Surveys on their health, activities and sleep in the previous week, and completed Time Use Surveys on two randomly chosen days per week. Additionally, subjects in the treatment group also completed questions related to sleep incentives in the treatment weeks as part of the Weekly Survey. Treated participants were also asked to choose a bedtime target (between 10 pm and 1 am) and a sleep duration target (between 7 and 9 hours) for Monday through Thursday nights of the current week and received incentives for achieving the targets. The sleep duration targets were set between 7 and 9 hours to reflect the recommended number of hours of sleep (see Cappuccio et al., 2010). A target was met if someone fell asleep by the bedtime target and slept at least as long as the duration target.[9] [10]

In our main treatment (Incentive-Weekly, waves 2, 3, 5, and 6), the treated weeks were Weeks 3, 4, and 5. Figure A.1 and A.2 illustrate the timeline of our main intervention. We used gain/loss framing: each week, these subjects were told that they would be rewarded GBP 10 (USD 15 in Pittsburgh) for participation in the following week. Rewards and punishments were added to this amount. Each reward was GBP 2.5 (USD 3.75) and each punishment was also GBP 2.5 (USD 3.75), so that the largest gain for achieving targets on all four nights was GBP 20 (USD 30). Subjects would achieve their target by complying with both bedtime and sleep duration targets, measured by Fitbit data, on each given day. A failure was to miss either target on a given day. We also provided feedback on performance in the previous week and asked subjects to predict their own performance for the current and all remaining treated weeks.

[9]One may be concerned that the treatment may lead subjects to sleep more than recommended, given the mixed evidence on the health effects of sleeping longer than 9 hours (Watson et al., 2015). In practice, we find the share of subjects sleeping more than 9 hours actually marginally declined during treatment from 6.9% to 6.7% (p-value=0.24).

[10]A standard concern in field experiments is that subjects in different treatment arms might communicate with each other, compromising the estimated treatment effect. Subjects in our control group did not know others were paid for meeting sleep targets in our experiment. By including individual fixed effects, we also controlled for term and university fixed effects. The stability of our main results to the inclusion of individual fixed effects is reassuring.

We tested two variants to this treatment to see how subjects would respond to differences in frequency and structure of the incentives (Carrera et al., 2019a) (Incentive-Biweekly, waves 1 and 4), and in the size of the monetary incentive (Small Incentive-Biweekly, wave 1). The Incentive-Biweekly was the same as 1, except that the sleep incentives were given biweekly, in weeks 3, 5, and 7 (see Figure A.3). In the Small Incentive-Biweekly, the incentives were given biweekly as in 2, but the amount they could gain for each achievement was smaller.

Furthermore, in waves 1 and 6 a subset of treated subjects were offered two contracts. There was no initial endowment in a treated week. And the first contract was a reward of GBP 2.5 (USD 3.75) for each night the target was met, and there was no punishment. Therefore, meeting the target on all nights of a week could lead to a total reward of GBP 10 (USD 15). The alternative contract would not only involve the same reward for meeting the targets but also penalize unmet targets. The punishment for each failed night was GBP 2 (USD 3).[11] In wave 6, we also included a treatment where subjects did everything the same as the main treatment, including setting sleep targets, except that they did not receive any monetary incentives.

In all treatment groups, subjects only received their monetary payoffs on the day they returned the device. At the end of the experiment, one of the 3 treated weeks was selected for each subject to determine payment for their sleep performance.

### 3.3 Empirical Specification

To analyze the effects of our incentives, we tested the following model

$$Y_{it} = \beta_1 * Treatment_{it} + \beta_2(Post\text{-}Treatment_{it}) + X_{it} + \eta_i + \epsilon_{it} \quad (1)$$

where $Y_{it}$ is a sleep outcome of individual $i$, at time $t$; $Treatment_{it}$ is a dichotomous variable equal to 1 during the weeks of treatment if individual $i$ in treatment group; $Post\text{-}Treatment_{it}$ is a dichotomous variable equal to 1 during the weeks following the last week of treatment if individual $i$ in treatment group; $X_{it}$ a vector of dummies for each day of the week and month fixed effects. We also include a linear weekly trend. Our preferred specification accounts for time-

[11] One treated week was randomly chosen and any loss was deducted from their payoffs to be received in the end of the experiment when they returned the device.

invariant heterogeneity by accounting for individual fixed effects ($\eta_i$). In our OLS regressions without individual fixed effects we include controls for gender and age as dummies. Standard errors were adjusted for clustering at the individual level. In our baseline specifications we exclude weekends, but we show results hold when we include weekends in the analysis.

### 3.4 Data

Summary statistics are reported in Table A.2. A total of 564 subjects participated in the experiment. Among these 508 generated usable data; 56 subjects (9.9%) either felt uncomfortable wearing a Fitbit or withdrew from the study due to other reasons. However, there is only limited selection into the final sample (Table A.3). Individuals who withdrew were more likely to report other ethnicity, were older, and had a shorter ideal sleep. However, there is no selective attrition on age and ideal sleep when restricting to waves 5 and 6. We also note that while most subjects synced regularly, sleep data were missing for some of the nights. This could happen if subjects did not wear the Fitbit at night or the device was not charged. We find no evidence of significant differences in syncing by treatment status (see Table A.4). In the next section, we discuss concerns related to missing data and irregular syncing behavior of participants as well as the stability of our results to the exclusion of individuals who synced the least or the imputation of missing data.

The sample is overall representative of the university population in the two universities with respect to gender and age. Women comprise 52% of undergraduate students at Oxford University, and 54% at the University of Pittsburgh. The median age in the sample is 20. Asian students were over-represented in both pools. They accounted for approximately 27% of our sample, while they comprise 7% of undergraduate enrollment in the United States, 11.4% of the enrollment at the University of Pittsburgh, and approximately 10% of students at the University of Oxford. At the same time, white students were under-represented with respect to the population at both campuses.[12] To address the concern that this may affect the external validity of our results, we replicate the main intervention analysis excluding Asians from the sample. Alternatively, we employed survey weighting to improve our sample representativeness with respect to race. To determine weights, we use data from the 2018 American Community Survey to make the

[12]See https://educationdata.org/college-enrollment-statistics.

sample representative by race with respect to the US college population. Overall, these analyses suggest that the results we document in our sample do not seem to be specific to the particular composition of the sample we recruited (see Section 4.1).

#### 3.4.1 Measuring Sleep

Measuring sleep is challenging. Previous studies have shown that self-reported measures of sleep, whether based on time-use diaries or survey questions, are prone to severe measurement errors. Self-reports tend to overestimate sleep duration compared to objective measures (Lauderdale et al., 2008b). Time-use diaries may also be subject to overestimation bias as, often, the activity lexicon associated with sleeping includes transition states (e.g., falling asleep) (Basner et al., 2007). Personal wearable devices (such as Fitbits) have been increasingly used to study health behavior (e.g., Handel and Kolstad, 2017). Concerns have also been raised regarding the ability of Fitbit devices to provide an accurate measurement of sleep. However recent studies (e.g., Lee et al., 2017) find that wearable activity trackers that detect heart rate perform fairly well in terms of tracking sleep compared to actigraphy, the medical professional grade sleep-tracking device. In Appendix C, we discuss in further detail the current evidence on the reliability and specificity of Fitbits, as well as their limitations.[13] Fitbits also offer limited but useful information about sleep quality. In particular, we have information on sleep efficiency and the number of sleep episodes per night in each wave, and information on the REM cycle collected through Fitbit Inspire II in wave 6.

While other papers have used these various methods to measure activity, ours is one of the few studies integrating time-use data from these three different sources: wearable devices, time-use diaries, and surveys. Table A.5 compares the different measures of sleep obtained using Fitbit devices, survey data, and time-use diaries. On average, subjects reported 7.96 hours of sleep for the previous week in time-use diaries and 6.81 hours of sleep in self-reported surveys. Thus, time-use data tend to significantly overestimate time allocated to sleep, while self-reported sleep duration was very close to the average sleep duration measured by Fitbit devices (6.85

[13] We cannot rule out that subjects may give the Fitbit to a friend for a few nights and game the incentive. However, it has been shown that, while daily resting heart rate can differ substantially across individuals, within individuals it tends to be more consistent over time (Quer et al., 2020). We do not find evidence of a significant correlation between abnormal changes in resting heart rate (above or below 2 standard deviations with respect to baseline) and treatment status (Figure A.4).

hours during the week). Further, according to time-use data, only 6.5% of the subjects reported sleeping less than 6 hours, while the survey-based measure indicated 10% of the subjects slept less than 6 hours—closer to but still significantly smaller than the 23% recorded by Fitbit devices during the school week. These results suggest that sleep-deprived individuals were more likely to overestimate their sleep duration in both time-use diaries and self-reported sleep duration. Fitbit data of sleep before, during, and after the intervention are plotted in Figure A.5. On an average night of the week at baseline, 87% of the time individuals slept less than 8 hours, 53% less than 7 hours, 17% less than 6 hours, and 6% less than 5 hours (see Figure A.6). Sleep duration was highly irregular—the standard deviation was approximately 1.8 hours. Subjects compensated during the weekend for some of their lost sleep hours during the week: approximately 52% of the subjects slept less than 7 hours in the first 2 weeks, while during the weekend the fraction of individuals sleeping less than 7 hours declined to 44% (Figure A.7). Most subjects (94%) considered it ideal to sleep more than 7 hours, with 67% reporting an ideal sleep of at least 8 hours (Figure A.8).[14]

#### 3.4.2 Covariate Balance between Treatment and Control Groups

Table 1 reports the differences between control and treatment groups with respect to baseline characteristics. Overall, subjects are well-balanced with respect to most variables. We report differences between treatment and control for all the waves (columns 1-6), for the waves in which we randomized subjects to either a treatment or a control condition (columns 3-6). Finally, in the Appendix (Table A.6), we report separate tests by wave of the experiment.[15].

## 4 Incentives, Sleep Behavior, and Habit Formation

### 4.1 Financial Incentives and Sleep Behavior

Table 2 shows our main regression results. In Panel A, we report estimates obtained from pooling all the waves, which include the different incentives used and waves 2-4 that had no

[14] In wave 6, we specifically ask subjects to report their ideal sleep duration during a typical week of the term. 91% of the subjects indicated at least 7 hours.

[15] For waves 2-4, for which we did not have a control group, we used the control group from waves 1, 5, and 6 when reporting separate balancing for each wave.

control group. The unit of observation is subject-night.

We find that the treatment increased the likelihood of sleeping between 7 and 9 hours by 6.6 percentage points (column 1), equivalent to a 15% increase with respect to the mean. The effect holds to the inclusion of individual fixed effects (column 2, +13% with respect to mean). Similarly, we show that the monetary incentive reduced the likelihood of sleeping less than 6 hours by 3.5 percentage points (column 3, -14% with respect to the mean). The effect is robust to the inclusion of individual fixed effects (column 4).

In Panel B, we focus on waves 1, 5, and 6 where subjects were randomly assigned to a control or treatment condition. The results are substantially unchanged. Relative to control, we find that subjects receiving monetary incentives were 5.9 percentage points more likely to sleep the recommended number of hours (between 7 and 9 hours) (column 1). This is equivalent to a 14% increase with respect to the mean. This result holds with the inclusion of individual fixed effects (column 2): accounting for persistent individual heterogeneity, the coefficient reduces only marginally and the effect remains economically and statistically significant (+13% with respect to the mean). When receiving the monetary incentive, individuals were 4.2 percentage points less likely to sleep less than 6 hours, equivalent to a 16% reduction with respect to the mean (column 3). This effect holds even with the inclusion of individual fixed effects (column 4). In Panel C, we restrict the analysis to waves 5 and 6, which were both conducted in Pittsburgh and focused on the weekly incentive. Overall, we find very similar and, if anything, larger effects.[16]

In our main estimates, we focus on sleeping the recommended amount of hours (7-9) and on a metric of insufficient sleep as both short and long sleep durations are considered behavioral risk factors for mortality and morbidity (Bin et al., 2013; Cappuccio et al., 2010) and our incentives aimed at promoting the recommended amount of sleep and reducing sleep deprivation. Figure A.5 visualizes the shift in sleep duration (upper figure) and bedtime (lower figure) induced by our intervention.[17] When receiving the monetary incentive, subjects' bedtimes were moved

[16]Table A.7 reports results only for wave 6 in which we used a different wearable device (Fitbit Inspire II). Results are substantially similar, albeit the post-treatment effects are less precisely estimated (see also Section 4.2).

[17]In Table A.8 we report results using alternative metrics of sleep. The findings are qualitatively similar. When we use continuous measures of sleep, the effects are smaller and less precisely estimated. Using a continuous measure of sleep duration, we find an increase of average sleep duration of approximately 5-7 minutes (columns 1-2). Using the logarithm of sleep, which reduces the influence of outliers, we find that the incentives led to a 1.1%-2.4% change in sleep (columns 3-4, Panels A-C of Table A.8). Results are more precisely estimated when restricting the sample to sleeping at least 4 hours and less than 10 hours (Table A.9). Using alternative metrics of short sleep duration, $< 7$ or $< 5$ (columns 5-8 of Table A.8), results tend in the same direction, although they are less precisely estimated when

earlier by approximately 25 minutes, while the average wake-up time was approximately 9-12 minutes earlier (see Table A.10).[18]

Results are also very similar to our baseline estimates when excluding from the sample individuals who missed more than 20%, 25%, and 50% of their data (Tables A.13-A.14). Replacing missing data using the average sleep of the individual at baseline (Table A.15), the estimates on the likelihood of sleeping between 7 and 9 hours (columns 1 and 3) during and after the intervention are somewhat smaller and less precisely estimated, pointing respectively to a 6%-8% and 2%-5% increase with respect to the mean; while the effects are overall more robust and sizeable (10%-20% decrease with respect to the mean) when focusing on the likelihood of sleeping less than 6 hours (columns 2 and 4).

As Asian students are over-represented in our sample, we show that our results are robust when we exclude them from the analysis (see Table A.16). Furthermore, results are very similar when using inverse probability weighting to ensure the representativeness of our sample with respect to race (Table A.17). We build weights using data from the 2018 American Community Survey on college enrollment by race and ethnicity. Table A.18 shows our results are robust to the exclusion of individuals who never slept less than 7 hours at baseline. In wave 6, we included a goal-setting treatment without monetary incentives (Table A.19). The goal-setting intervention did not have significant effects. The coefficient declines substantially when including individual fixed effects. Interestingly, there is some evidence that the goal-setting intervention may have some persistence. While we are under-powered to investigate these effects, future research could shed more light on the role of setting goals in the sleep domain. Even when compared to subjects in the goal-setting treatment, subjects in the monetary incentive treatment slept significantly longer, although effects become less precisely estimated in the post-treatment period (p-value=0.12) than when compared to the control group (see Table 2).

### 4.2 Habit Formation and Sleep

In Table 2, we find evidence that the effects of monetary incentives on sleep persisted to some extent in the weeks following the termination of treatment. Panel A includes all waves. After

using sleeping less than 5 hours.

[18] We find no evidence that sleeping more on incentivized nights crowded out sleep on non-incentivized nights during weekends of the treated weeks (Tables A.11 and Table A.12).

removing the monetary incentive, subjects in the treatment group were 3.9 percentage points more likely to sleep between 7 and 9 hours (+9%, column 1, row 2). Results are only slightly smaller including individual fixed effects (+8%, column 2, row 2). Effects are qualitatively similar when focusing on the left tail of the sleep distribution (column 3-4). While OLS estimates are not precisely estimated, the point-estimate is sizeable (-6%), with a larger and significant effect when including individual fixed effects (-10%). If anything, the effects are larger and more precisely estimated when restricting to the waves in which subjects were randomly assigned to control or treatment conditions (waves 1, 5, and 6), with an increase in the likelihood of sleeping between 7 and 9 hours ranging between 9% (Panel B, column 2) and 15% (Panel B, column 1) and a reduction in the likelihood of sleeping less than 6 hours ranging between 13% (Panel C, column 4) and 16% (Panel C, column 3).

The difference in magnitude between the treatment and post-treatment effect is comparable with recent evidence on habit formation effects when using financial incentives to promote exercising (Carrera et al., 2019a). In waves 1-5, we observed subjects for only three weeks after the end of treatment, and thus this only provides suggestive evidence consistent with a habit formation effect. In wave 6, we extended the post-treatment period to analyze the persistence of the effects up to 10 weeks after the end of treatment. Overall, we find that the increase in likelihood of sleeping between 7 and 9 hours persisted up to the $11^{th}$ week of the experiment and thus for 6 weeks after the removal of the incentive (Table A.20). There is no evidence of a significant difference between treatment and control after week 11. However, we also note that during week 11 of the experiment (6 weeks after the removal of the incentives) classes ended. Final exams took place in week 12 of the experiment with the term ending at the end of the same week on May 1st. Overall, the academic calendar limits our ability to investigate the persistence of habit formation effects in the presence of the typical term constraints (i.e. classes, deadlines etc.) and the expected attrition rate.[19] [20]

[19]To gauge a sense of the role of term in shaping students' sleep schedule, when restricting to the control group, average sleep duration increased by approximately 30 minutes after the end of the term.

[20]Table A.21 shows that our results are overall very similar if excluding the linear weekly trend or including week fixed effects.

### 4.3 Effects on Sleep Quality

Examining other outcomes drawn from Fitbit data (Table A.22), we find no evidence of significant effects of treatment on the efficiency of sleep (columns 1–2), measured as the ratio between sleep duration and time spent in bed (including time awake). The fact that our treatment did not reduce sleep efficiency also confirms that the treated subjects were not just lying in bed more, but their actual sleep duration increased.

In wave 6, we were also able to measure sleep stages as we used more recent wearable devices (i.e., Fitbit Inspire II). There is some evidence of an increase in the average number of minutes in deep and REM sleep. These effects are not precisely estimated, in part because of the smaller sample size for which this data was available. Furthermore, this evidence should be interpreted with caution. While recent models have shown significant improvements, their performance in identifying sleep stages is still relatively inaccurate compared with standard PSG (Haghayegh et al., 2019b; de Zambotti et al., 2018).

### 4.4 Incentives to Sleep and Time Allocation

A natural question is whether and how the allocation of time changed in response to our intervention. Individuals may compensate insufficient sleep at night by napping during the day or by sleeping longer during weekends. Other studies find significant effects of naps on productivity and well-being (Bessone et al., 2021). We investigated whether our intervention affected the time allocated to naps. Less than 1% of the sleep episodes recorded by the Fitbits were classified as naps, defined as any sleep lasting less than 2 hours between 7 am and 7 pm during weekdays. 76% of the subjects never recorded a nap throughout the period of the experiment. Although nap duration is negatively correlated with sleep duration at night, we find no evidence that our intervention systematically affected the likelihood of taking a nap or the nap duration (see Table A.23, columns 1–2). Thus, unsurprisingly, the results are substantially unchanged when we include controls for napping behavior (columns 3–6). We also find no evidence of subjects changing their weekend sleep duration during the intervention in response to the longer sleep duration induced by the incentives during the week (Table A.12).

Subjects may also reallocate their time devoted to other activities when receiving incentives to go to bed earlier and to sleep longer. Using time-use diaries, we directly examine the effects of our incentives on individual time allocation. Time-use data are available for approximately 80% of the participants, and thus, results should be interpreted with some caution. The subjects not responding to the time-use surveys were slightly older, less likely to be White, and more likely to be obese (see Table A.24). Despite some attrition, we believe these results provide interesting insights on the overall effects of our intervention on time allocation.

As mentioned above, consistent with previous evidence (Lauderdale et al., 2008a), we find that individuals tend to overestimate sleep when using time-use diaries. While qualitatively we do observe a mild increase in sleep, the coefficients are only precisely estimated when focusing on the likelihood of sleeping less than 6 hours (Table 3, column 3). When examining other activities, we find no significant evidence that the increase in sleep duration was associated with a change in time spent on studying, working, personal care activities, or exercising. However, there is a significant decline in time spent relaxing, which includes screen time and time spent with friends. When breaking down these results (Panel B, columns 2-7), we find that a significant decline in overall screen time driven in particular by a reduction in time spent on videos and online games. After the incentive was removed, we still observe a decline in time spent watching videos with respect to baseline but the on overall screen time is smaller than during treatment and it is not precisely estimated. While this analysis has several limitations (i.e., attrition, small sample size, measurement error), we believe these results are overall consistent with recent research suggesting that screen time near bedtime is associated with lower sleep duration (Twenge et al., 2017; Billari et al., 2018). Consistent with the evidence that repetition of behavior, such as following fixed routines increases habit formation (e.g., Wood and Neal, 2007; Lally et al., 2010), adjusting activities before bedtime may help develop better sleep habits.

### 4.5 Additional Findings: Sleep Regularity, Structure and Size of the Incentives

This section reports some additional findings regarding the effect of our intervention. Recent studies suggest that sleep variability is an important predictor of well-being and learning (Fang et al., 2021). Interestingly, the monetary incentives affected the regularity of sleep and bedtime, reducing their weekly variance (Table A.25). We find significant and sizable reductions (13-15%)

in the variability of sleep duration and bedtime during and after treatment. There is instead no evidence of a significant change in the variability of wake-up time during treatment, although there is some evidence of a decline after the end of treatment period (8%).

We do not find statistically significant differences when examining the role of the frequency and the structure of the incentives on the treatment effects on the likelihood of sleeping between 7 and 9 hours (Table A.26). However, if anything the biweekly incentive seems to have stronger post-treatment effects, although these differences are not precisely estimated, and thus, should be interpreted with caution.[21]

Finally, we explore the role of incentive size. Using a smaller monetary incentive and eliminating loss framing lead to effects that are smaller in magnitude and less precisely estimated (Table A.27).

# 5 Behavioral Biases and Sleep Choice

## 5.1 Time Inconsistency and Demand for Commitment

Several aspects of our participants' behavior were consistent with partially sophisticated time inconsistency. We correlated our measures of subjects' time preference with baseline sleep patterns and performance in the experiment.[22] While estimates are not precise due to the small sample size, we find that, before intervention, present-biased subjects were more likely to be sleep deprived (Table A.28).[23] In particular, the most present-biased quartile of participants in our sample were 16% less likely to report sleeping between 7 and 9 hours at baseline (column 3). Although these results are largely imprecise and should be interpreted with caution, they suggest present-bias may contribute to explaining sleep deprivation.

Our experiment included two features that allowed us to directly observe the demand for commitment devices. First, in all intervention groups, we asked subjects to choose bedtime targets between 10 pm and 1 am and sleep duration targets between 7 and 9 hours. An agent

[21]In the biweekly treatment, we regard as post-treatment any week after the first week of treatment during which subjects did not receive a monetary incentive. Using an alternative definition and focusing only on the last week of the experiment (week 8), we find similar results.

[22]In Appendix B, we describe in detail how we built our measure of present bias and impatience. We acknowledge that our measure of time preference may be potentially confounded by liquidity constraints (Cohen and Story, 2014).

[23]The sample is restricted to individuals with non-missing information on time-preferences. We couldn't estimate time preferences for approximately 4% of the subjects. They were therefore excluded from this part of analysis.

with standard preferences who maximizes expected rewards from the experiment would choose the least binding targets, namely 1 am and 7 hours. For instance, for a subject who normally sleeps for 6.5 hours each night, any sleep duration target in the given range (7-9 hours) would be restrictive. If the subject is also maximizing rewards, 7 hours would be optimal. However, if they want to motivate themselves to sleep longer, they could choose a more restrictive target, because failing to achieve would lead to forgone payoffs. Furthermore, the choice of more flexible targets would be consistent with demand for flexibility as there may be unanticipated reasons impeding participants to comply with targets (see Section 5.2). The demand for commitment may be influenced by both internal and external uncertainties (Bai et al., 2021). Individuals may be uncertain about their self-control or uncertain about their schedule and thus demand more flexibility.

We uncover some interesting evidence of demand for restrictive bedtime and sleep duration targets. Despite 1 am and 7 hours of sleep being dominant bedtime and sleep duration targets, approximately 60% of the subjects in the treatment group chose more restrictive targets at least once during the three weeks of the experiment (see Figure A.9).[24] 20% of the subjects demanded for a more restrictive target every week, and 41-43% did so for at least two of the three weeks of treatment.

Figure 1 reports the share of participants choosing a bedtime (sleep duration) target earlier (longer) than their average bedtime (sleep duration) at baseline.[25] A recent literature (Bai et al., 2021; John, 2020; Carrera et al., 2019b) has argued that people often fail when choosing commitments, and that offering commitment devices is thus potentially welfare-reducing. We find that individuals choosing restrictive bedtime and sleep duration targets failed on their commitment choices more than 50% of the time (55% for bedtime and 56% for sleep duration targets). On 32.6% of the nights subjects who chose a restrictive bedtime or sleep duration target would have obtained their reward if they had set 1am or 7 hours as their targets. Thus subjects lost approximately $10 per week by choosing more restrictive targets.

It is possible that subjects were simply not in the mindset of maximizing rewards when choosing targets, but they just trying to pick targets that were more reasonable based on their

[24]Recall that each treated subject chose targets once in each of the 3 treated weeks.

[25]Figure A.10 reports the same share among participants sleeping more than 7 hours and going to bed on average before 1am at baseline).

own sleep habits.[26] For instance, for subjects who normally sleeps for 7.5 hours each night, if they simply want to maximize rewards, 7 hours would still be optimal. However, if they think it is unlikely for themselves to sleep less than 7.5 hours, then targets between 7 and 7.5 hours would be indifferent for them. If they would really want to motivate themselves to sleep longer, they would choose a target longer than 7.5 hours. Therefore, we also look at their targets compared with their normal sleep patterns, either at the baseline (first two weeks without incentive), or during the previous week. When compared with baseline sleep pattern, the targets were in fact restrictive for a large majority of subjects. At baseline, approximately 93% of the participants slept less than 7 hours (the minimum sleep duration target) at least 1 day out of the weekdays of a week during term (at baseline); while 87% of individuals went to bed at least one day after 1 am at baseline. Among those choosing more restrictive sleep targets 61% of the time participants chose a target longer than their average sleep duration at baseline and 67% of the time they chose a bedtime earlier than their average baseline bedtime (see Figure A.11). It is also possible that the baseline weeks were too distant to use for comparison, but subjects rather thought about their own behavior in the previous week when setting targets. Our feedback during the experiment also made this easier. When compared with previous week sleep pattern, the targets were also restrictive for a large majority of subjects (Figures A.12-A.13).

Although it is possible that menu or experimenter demand effect (Carrera et al., 2019b) could drive the choice of restrictive targets, we have several reasons to believe that the chance of this should be small. First, the least restrictive targets (1 am and 7 hours) are at opposite extremes on their respective lists, reducing the chance of menu effect. Second, the correlation between pre-treatment behaviour and choice of commitment cannot be explained by menu or demand effect. Third, experimenter demand effect means that subjects could be choosing restrictive targets because they felt this is what is good to want, or to please the experimenter. To gauge a sense of whether this effect played a role in participants' choices, in our end-of-experiment survey in wave 6, we also asked subjects who chose a more restrictive target during the intervention a) whether they were aware that 7 hours and 1 am were easier targets to achieve; b) whether they were aware that they were potentially leaving money on the table; c) whether they did so to motivate themselves; d) whether they did so to please the researcher. 71% of the participants

[26] Thanks to an anonymous referee who pointed out this possibility.

who had chosen at least one time a more restrictive target answered the survey. 69% of them reported to be aware that a sleep duration of 7 hours and a bedtime target of 1 am were easier targets to achieve; 60% reported to be aware that by choosing more restrictive targets they were potentially leaving money on the table; 92% reported they chose more restrictive targets to motivate themselves. Only 2% of the participants mentioned they chose these targets to please the researchers. When asked to explain in their own words why they chose more restrictive targets, many mentioned they did so to get more sleep, get on a schedule, have a more consistent schedule, get better habits, or improve their lifestyle. 77% of the subjects thought that choosing these targets helped them sleep longer and go to bed earlier at least sometimes during the study.

Correlational evidence suggests that our commitment devices were effective in improving sleep duration. When considering the sample of intervention weeks and including controls for demographics and baseline sleep, we find that subjects who chose restrictive targets were more likely to sleep longer (Table 4). Furthermore, participants were 8 percentage points more likely to sleep between 7 and 9 hours. This is despite the fact that those setting restrictive targets only achieved them about 45% of the time, resulting in the loss of approximately 10$ per week.[27]

Exploring the heterogeneity in the propensity to demand more restrictive targets, we find some suggestive evidence that both present-biased and impatient individuals were more likely than other subjects to choose an earlier bedtime at least once during the experiment (Table A.29). Present-biased subjects were 20% more likely to choose a bedtime target before 1 am than the rest of the sample, although this result is not precisely estimated (p-value=0.112, column 1, Panel A). Impatient subjects were 27% more likely to choose a bedtime target before 1 am than the rest of the sample (Column 1 Panel B). Finally risk-averse individuals were 31% less likely to choose early bedtime targets.[28]

[27] Among the individuals choosing restrictive sleep duration (bedtime) targets approximately 46% (48%) achieved their targets. Individuals choosing restrictive bedtime and sleep duration targets failed on their commitment choices more than 50% of the times (55% for bedtime and 56% for sleep duration target). On 32.6% of the nights subjects who chose a restrictive bedtime or sleep duration target would have achieved their reward if they had set 1 am or 7 hours as their targets.

[28] Risk attitudes are found to be domain specific (see e.g. Weber et al. (2002)). Therefore, we measured general risk attitudes using the Multiple Price List (Holt and Laury, 2002) and the domain-specific risk attitudes specifically in the health domain (Weber et al., 2002). There is evidence that more risk averse people are less likely to engage in unhealthy behavior, such as smoking and drug use (e.g. Bartky and Harrison (1979)). Additionally, more risk averse people are also more likely to take preventive measures such as vaccination (e.g. Nuscheler and Roeder (2016); Lutter et al. (2019)). Barsky et al. (1997) show that risk tolerance and impulsivity are correlated, which is also consistent with personality theory in psychology (Mitchell, 1999).

We also included another feature in our experiment to check the influence of time inconsistency on the relationship between time of answering the survey and choice of sleep targets. In particular, time-inconsistent subjects may be more likely to choose more demanding sleep targets earlier in the day, when the cost of last night's bad sleep choice is still felt, and the next sleep is still considered as in the future. Later in the day, when for instance the desire to watch another episode of a TV series sets in, they may be more likely to choose less restrictive targets. To test this we deliberately randomized the time surveys were sent out throughout the experiment (ranging from 6 am to 3 pm), although we could not fully control the timing of the answers. While only 42% of the subject chose a more restrictive bedtime (earlier than 1 am) when responding to the survey after 3 pm, among those responding earlier in the day, 51% did so (see also Figure A.14). Individuals responding before 3 pm to the survey were 8 percentage points more likely to choose a more restrictive target (earlier than 1am), equivalent to a 17% increase with respect to the mean (Table A.30, column 1). We also show a strong relationship between the time at which the email was sent to the subjects and the choice of the target. Those receiving an early email (before noon) were 8.6 percentage point more likely to choose a bedtime target earlier than 1 am. Instrumenting the response to the survey with the time at which the survey was sent (column 3), we find that individuals who answered the survey before 3 pm were 17 percentage points more likely to select a bedtime target earlier than 1am, equivalent to a 39% increase with respect to the mean. Once we include controls for age, gender, race, and wave fixed effects the coefficient declines only marginally (+34%, column 4).

Our experiment included another feature to check demand for commitment. In waves 1 and 6, we asked a sub-sample of treated subjects to choose between a contract that only rewards successes and a contract that not only rewards successes to the same extent but also punishes failures. To maximize monetary payoff, an agent with standard preferences would choose the former, whereas an agent who demands commitment would choose the latter (e.g., Kaur et al., 2015).[29] An agent with naive time-inconsistent preferences may predict that their future self will achieve all targets and thus be indifferent between having and not having a commitment device, whereas a sophisticated agent may anticipate their future time-inconsistent behavior and would

[29]Offering these alternative contracts did not generate significant different effects from our baseline treatment (p-value=0.58 during treatment; p-value= 0.47 after the intervention).

actively demand for a commitment device even at some cost. The "cost" of the commitment device in this setting is the explicit monetary punishments associated with failures. A small fraction of subjects chose the dominated contract (7% of the subjects who were offered the contract with punishment opted for it). However, the small sample size does not allow us to identify any significant relationship between the choice of the dominated contract and sleep outcomes during the intervention. We believe this could be an interesting avenue for future research, and a candidate for alternative format of commitment devices in the sleep setting.

### 5.2 Biased Beliefs

As mentioned earlier, more than 90% of the subjects reported an ideal sleep of longer than 7 hours during term. In wave 6 we asked participants who went to bed later than their ideal bedtime at least once over the first two weeks of the experiment on whether they had any regrets. It is worth noting that while we specifically asked about ideal sleep duration and bedtime during the term, that is with the regular schedule and work loads during an academic term, these might still be interpreted as unconstrained optima by subjects rather than constrained optima. That said, 82% of the participants who went to bed later than their ideal bedtime reported some regrets. 55% of the participants said they had regrets most of the time, while only 17.6% said they never had regrets. In this section, we assess subjects' biased beliefs about their sleep, which could also contribute to explaining the distance between ideal sleep and actual sleep duration.

First, we show that subjects overestimated the duration of own sleep (see Table A.5), consistent with previous evidence (Lauderdale et al., 2008a) that time-use data were particularly prone to this bias. Comparing individual self-reported average sleep on weekday nights from the first day survey with actual sleep as measured by Fitbits, we define as overconfident those participants who reported an average sleep of at least 30 minutes longer than the actual sleep duration as measured by the Fitbits. We find a correlation between insufficient sleep at baseline and this type of overconfidence: individuals sleeping less than 7 hours were significantly more likely to overestimate their sleep duration, suggesting that overconfidence may be an important factor behind insufficient sleep. In particular, 47% of those who averaged less than 7 hours of sleep in the first two weeks of the experiment reported a sleep duration that was at least 30 minutes longer than their average actual sleep. Among those sleeping at least 7 hours, only 11% of the subjects reported a sleep duration 30 minutes longer than their average actual sleep. Participants who reported significantly longer sleep duration than that measured by Fitbits were less likely to report insufficient sleep at baseline based on self-reported data, but more likely to be sleep deprived based on Fitbit data before treatment (Table 5, Panel A). In other words, while individuals who were overconfident about sleep reported longer sleep duration at baseline, these subjects were also sleeping significantly less than the rest of the sample based on Fitbit data.

In wave 6, we incentivized subjects in their first day survey to predict their sleep in the following two weeks (before the beginning of the intervention), and compared these predictions with their actual sleep during those weeks based on Fitbit data. The average prediction was 7.5 hours, while the actual sleep in the first two weeks was 6.9 hours. Approximately 55% of our participants over-predicted sleep duration for the first two weeks of the experiment by more than 30 minutes. Interestingly, individuals who were sleeping less than 7 hours at baseline were significantly more likely to over-predict their sleep in the first two weeks of the experiment (see Figure A.15). Panel B of Table 5 confirms qualitatively that individuals who were defined as overconfident by comparing their first day prediction on sleep during the first two weeks with their actual sleep, over-reported sleep duration in the first day survey, and were sleeping significantly less than the rest of the sample based on Fitbit data.[30]

As an alternative measure of overconfidence, we compare participants' beliefs about their own sleep duration, sleep quality, and risks associated with sleep deprivation with beliefs they had on others' sleep behavior and risks. The data drawn from the first day survey reveal that subjects systematically reported longer sleep duration, better sleep quality, and lower risks associated with sleep for themselves than what they considered the average for persons of the same age (see Table A.32). The majority of subjects (58%) believed they sleep longer than an average person of the same age. They reported sleeping 25 minutes longer than an average person of their age (see Figure A.16).[31]

[30] In Table A.31, we show that defining as overconfident participants who self-reported to sleep between 7 and 9 hours in the first day survey, but slept less than 7 or more than 9 hours according to Fitbit data at baseline, results tend in the same direction.

[31] It is possible that participants may not predict sleep correctly because of awakening in the nights. This may be more of an issue in developing contexts where noise and sleeping conditions may severely disrupt sleep. However, we find no evidence of a significant correlation between over-placement and sleep efficiency (p-value=0.26) or minutes awake (p-value=0.37).

Similarly, 62% of the subjects thought that their sleep quality was better than that of the average person of their age. These statistics are consistent with self-serving belief and an overplacement of own sleep relative to others. We also find evidence of overconfidence with respect to the perceived health and cognitive risks associated with sleep deprivation (see Table A.32). We find similar results when analyzing the correlation between overconfidence and insufficient sleep using an index of overconfidence which included beliefs about sleep quality and sleep risks (Table 5, Panel C).

Overconfidence was not significantly correlated with demand for restrictive targets (Table A.33), but it was correlated with compliance. Interestingly, while choosing dominated targets generally significantly increased the chance of sleeping the recommended amount of hours during the intervention, it did not have such effect among overconfident individuals (Figure 2).[32] Overconfident subjects were less likely to meet targets, and commitment devices were not effective (possibly even welfare diminishing) for them, consistent with Bai et al. (2021).

Participants were also asked to predict the likelihood that they would achieve their chosen target in each of the following treated weeks. Correct predictions were rewarded. The prediction reward was $1 in waves 2 and 4 and $3 in waves 3, 5, and 6. We find no evidence that the increasing prediction incentive affected prediction accuracy.[33] Individuals tended to over-predict their likelihood of achieving the targets (Figure A.18). Among treated subjects, predictions did not seem to be improving over time: while subjects were revising their predictions down from week to week, they kept falling short of their targets as the study proceeded. In the first treated week, 70% of the subjects were too optimistic about the number of nights they could achieve; in the second (third) week of treatment 64% (70%) of the subjects were too optimistic (see Figure A.19).

Our evidence appears consistent with motivated beliefs and sustained overconfidence in the face of repeated feedback. We analyzed subjects' self-reported sleep duration when they are provided with information from the Fitbit. If subjects did not update their self-reported sleep duration when feedback is available, this may further rule out alternative explanations, such as misinformation or uncertainty, and suggest motivated beliefs may play an important role. Indeed,

[32]Figure A.17 replicates Figure 2 using the incentivized measure of overconfidence obtained in wave 6.

[33]We note that incentivizing predictions may be problematic as partially-sophisticated present-biased participants may reinforce their incentives to meet desired goals (Acland and Levy, 2015).

when looking at subjects in the control group, we find no evidence of significant differences in self-reported sleep duration before and after being provided with the Fitbit (Figure A.20), suggesting that they were not fully incorporating information into their beliefs. Furthermore, among participants in the control group, there is no evidence of a significant difference between self-reported and actual sleep between the beginning and the end of the experiment (Figure A.21).

In addition to overconfidence, planning fallacy may be consistent with some of our findings (Buehler et al., 1994; Brunnermeier et al., 2008). This means individuals may underestimate the time it will take to complete tasks, particularly as the academic term progresses and students may increasingly need to work in the evening. We sent a follow-up survey to all subjects after the experiment and particularly asked subjects who did not meet their targets during the treated weeks what were the factors that prevented them from achieving their targets (Figure A.22). 65% of them mentioned they could not resist watching Netflix, playing video games or doing something they wanted to do more than going to bed. 60% of them mentioned some external conditions (i.e. homework, cleaning, caring for someone) that they could only do at that time, while perfectly being aware of these constraints.[34] 32% of participants mentioned that they were finishing something they did not realize it would take so long to complete, which would indicate planning fallacy. When asked to identify the single most important reason for not meeting their sleep targets, only 5% of the participants referred to planning errors, while 47.5% said they chose to do something they preferred and 47.5% referred to some external conditions (Figure A.23). However, given the nature of the experiment and the potential for different interpretations of the factors in the survey, we cannot rule out that planning fallacy may contribute to explaining the behavior of some subjects.

## 6 Conclusion

Statistics reveal that many individuals sleep less than the recommended number of hours. There are several factors affecting individuals' sleep choices. Understanding how to improve

[34] Note that on Mondays of treated weeks we asked subjects to set targets for the next 4 days in the week. Subjects might choose a target even if they knew that there were some known constraints that could prevent them from achieving the targets on some nights, as long as they expected to achieve on more days than they expected to fail.

health habits is crucial in designing policies aimed at promoting healthier behavior. As pointed out by Charness and Gneezy (2009), people tend to underestimate the impact of current actions on future utility and discount the future too much. Our evidence suggests that this tendency also characterizes sleep behavior. The prevalence and persistence of behavioral biases in the sleep domain is particularly interesting given the repeated feedback individuals receive on sleep throughout their lives (Huffman et al., 2018).

We studied sleep choice, and whether commitment devices as well as monetary incentives can improve sleep behavior among students. We find supportive evidence for partially sophisticated time-inconsistent preferences in sleep choice. The subjects in our experiment chose commitment devices even if this meant a lower monetary reward in expectation. Present-biased subjects were more likely to be sleep deprived at baseline, but many of them committed to dominated bedtime or sleep duration targets. Subjects choosing more demanding targets were also more likely to achieve them, with the exception of those who were classified as overconfident. Indeed, many subjects tended to be overconfident in their own sleep duration and quality and were more optimistic about themselves than about others when assessing the risks associated with insufficient sleep.

Our incentives improved sleep behavior and led to some habit formation effects, with subjects in the treatment groups sleeping longer even after the incentives were removed. Furthermore, monetary incentives increased sleep regularity, reducing the variance of bedtime, wake-up time, and sleep duration. Finally, we show that incentives to sleep may also have positive effects on academic outcomes, although these results are at best suggestive and further research is needed to establish this finding. When receiving incentives to sleep longer, individuals significantly reduced screen time (watching TV, YouTube videos, or surfing the Internet), while time spent with friends, working, or studying were not affected. Overall, these results give us a more nuanced understanding of sleep choice. Despite many economic models regarding sleep as an exogenous and homogeneous constraint on time, we provide evidence that behavioral biases play an important role and affect the heterogeneity of choice.

Our findings suggest that time inconsistency and biased beliefs can persist in the face of extensive experience and feedback. Thus, interventions based only on information (i.e. educational programs on sleep hygiene or fatigue management) may not be effective in the presence of these behavioral biases. Self-control problems may lead to procrastination with subjects repeatedly placing higher weight on immediate outcomes, and constantly delaying the start of good sleep habits. Also, people with motivated beliefs may be able to suppress the recall of objective feedback that could challenge their self-image, so that the simple provision of information may be ineffective in correcting such misperceptions. Yet, to the extent subjects become more aware of their time inconsistent preferences due to the repeated feedback, sleep is also a domain where demand for commitment may be relevant and commitment devices effective. We show that appropriate incentives can be used to improve an individual's sleep behavior, while we find no evidence of subjects updating their beliefs with the additional information provided by the wearable devices, suggesting they are not fully incorporating information into their beliefs. Incentives to go to bed earlier and to sleep longer were effective, suggesting that there is a cost to sleep, either in effort or in alternative uses of time, which can be compensated with a monetary payment.

Our findings also suggest that commitment devices and incentive structures may be more effective than planning tools at improving sleep behavior (Handel and Kolstad, 2017), and that temporary interventions, as those adopted by some companies, may have persistent effects, particularly when individuals lack a commitment device in natural settings. Providing incentives and commitment devices may help time inconsistent and overconfident individuals improve their sleep habits. Incentives and commitment devices may promote better sleep behaviors among subjects with self-control problems in the form of a time-inconsistent taste for immediate gratification (O'Donoghue et al., 2006), and among subjects with overconfidence as a result of motivated beliefs because they are resilient to repeated information (Bénabou and Tirole, 2016). Incentives may also mitigate the role of motivated reasoning (Zimmermann, 2019). At the same time, our results imply that interventions that help individuals form routines conducive to healthy sleep habits (i.e., reduced screen time) may have longer-lasting effect.

Future research efforts exploiting larger samples may shed further light on the human capital and health effects of interventions aimed at improving sleep duration and quality. Future studies could also explore the relative effectiveness of non-monetary incentives and alternative commitment devices in nudging individuals into healthier and persistent sleep habits.

## References


Acland, Dan, and Matthew R Levy (2015) 'Naiveté, projection bias, and habit formation in gym attendance.' *Management Science* 61(1), 146–160

Andreoni, James, and Charles Sprenger (2012) 'Estimating time preferences from convex budgets.' *American Economic Review* 102(7), 3333–56

Ashraf, Nava, Dean Karlan, and Wesley Yin (2006) 'Tying odysseus to the mast: Evidence from a commitment savings product in the philippines.' *Quarterly Journal of Economics* 121(2), 635–672

Bai, Liang, Benjamin Handel, Edward Miguel, and Gautam Rao (2021) 'Self-control and demand for preventive health: Evidence from hypertension in india.' *The Review of Economics and Statistics* pp. 1–55

Barsky, Robert B, F Thomas Juster, Miles S Kimball, and Matthew D Shapiro (1997) 'Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study.' *The Quarterly Journal of Economics* 112(2), 537–579

Bartky, Ian R., and Elizabeth Harrison (1979) 'Standard and daylight-saving time.' *Scientific American* 240(5), 46–53

Basner, Mathias, Kenneth M Fomberstein, Farid M Razavi, Siobhan Banks, Jeffrey H William, Roger R Rosa, and David F Dinges (2007) 'American time use survey: sleep time and its relationship to waking activities.' *Sleep* 30(9), 1085–1095

Becker, Gary S (1965) 'A theory of the allocation of time.' *The Economic Journal* pp. 493–517

Bénabou, Roland, and Jean Tirole (2016) 'Mindful economics: The production, consumption, and value of beliefs.' *Journal of Economic Perspectives* 30(3), 141–64

Bessone, Pedro, Gautam Rao, Frank Schilbach, Heather Schofield, and Mattie Toma (2021) 'The economic consequences of increasing sleep among the urban poor.' *The Quarterly Journal of Economics* 136(3), 1887–1941

Biddle, Jeff E, and Daniel S Hamermesh (1990) 'Sleep and the allocation of time.' *Journal of Political Economy* 98(5, Part 1), 922–943

Billari, Francesco C, Osea Giuntella, and Luca Stella (2018) 'Broadband internet, digital temptations, and sleep.' *Journal of Economic Behavior & Organization* 153, 58–76

Bin, Yu Sun, Nathaniel S Marshall, and Nick Glozier (2013) 'Sleeping at the limits: the changing prevalence of short and long sleep durations in 10 countries.' *American Journal of Epidemiology* 177(8), 826–833

Breig, Zachary, Matthew Gibson, and Jeffrey Shrader (2018) 'Why do we procrastinate? present bias and optimism.' Technical Report, Mimeo

Brunnermeier, Markus K, Filippos Papakonstantinou, and Jonathan A Parker (2008) 'An economic model of the planning fallacy.' Technical Report, National Bureau of Economic Research

Bryan, Gharad, Dean Karlan, and Scott Nelson (2010) 'Commitment devices.' *Annual Review of Economics* 2(1), 671–698

Buboltz, Walter C, Franklin Brown, and Barlow Soper (2001) 'Sleep habits and patterns of college students: a preliminary study.' *Journal of American College Health* 50(3), 131–135

Buehler, Roger, Dale Griffin, and Michael Ross (1994) 'Exploring the" planning fallacy": Why people underestimate their task completion times.' *Journal of personality and social psychology* 67(3), 366

Cappuccio, Francesco P, Lanfranco D'Elia, Pasquale Strazzullo, and Michelle A Miller (2010) 'Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies.' *Sleep* 33(5), 585

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor (2019a) 'The structure of health incentives: evidence from a field experiment.' *Management Science*

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky (2019b) 'How are preferences for commitment revealed?' Technical Report, National Bureau of Economic Research

Charness, Gary, and Uri Gneezy (2009) 'Incentives to exercise.' *Econometrica* 77(3), 909–931

Cohen, Deborah A, and Mary Story (2014) 'Mitigating the health risks of dining out: the need for standardized portion sizes in restaurants.' *American Journal of Public Health* 104(4), 586–590

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White (2020) 'Measuring time preferences.' *Journal of Economic Literature* 58(2), 299–347

Danz, David, Neeraja Gupta, Marissa Lepper, Lise Vesterlund, and K Pun Winichakul (2021) 'Going virtual: A step-by-step guide to taking the in-person experimental lab online.' *Available at SSRN 3931028*

de Zambotti, Massimiliano, Aimee Goldstone, Stephanie Claudatos, Ian M Colrain, and Fiona C Baker (2018) 'A validation study of fitbit charge 2™ compared with polysomnography in adults.' *Chronobiology International* 35(4), 465–476

DellaVigna, Stefano, and Ulrike Malmendier (2006) 'Paying not to go to the gym.' *American Economic Review* 96(3), 694–719

Dupas, Pascaline, and Jonathan Robinson (2013) 'Why don't the poor save more? evidence from health savings experiments.' *American Economic Review* 103(4), 1138–71

Falk, Armin, Anke Becker, Thomas J Dohmen, David Huffman, and Uwe Sunde (2016) 'The preference survey module: A validated instrument for measuring risk, time, and social preferences.' *Working Paper*

Fang, Yu, Daniel B Forger, Elena Frank, Srijan Sen, and Cathy Goldstein (2021) 'Day-to-day variability in sleep parameters and depression risk: A prospective cohort study of training physicians.' *NPJ digital medicine* 4(1), 1–9

Frederick, Shane, George Loewenstein, and Ted O'donoghue (2002) 'Time discounting and time preference: A critical review.' *Journal of Economic Literature* 40(2), 351–401

Gibson, Matthew, and Jeffrey Shrader (2018) 'Time use and labor productivity: The returns to sleep.' *Review of Economics and Statistics* 100(5), 783–798

Giuntella, Osea, and Fabrizio Mazzonna (2019) 'Sunset time and the economic effects of social jetlag. evidence from us time zone borders.' *Journal of Health Economics*

Godino, Job G, David Wing, Massimiliano de Zambotti, Fiona C Baker, Kara Bagot, Sarah Inkelis, Carina Pautz, Michael Higgins, Jeanne Nichols, Ty Brumback et al. (2020) 'Performance of a commercial multi-sensor wearable (fitbit charge hr) in measuring physical activity and sleep in healthy children.' *PLOS ONE* 15(9), e0237719

Haghayegh, Shahab, Sepideh Khoshnevis, Michael H Smolensky, and Kenneth R Diller (2019a) 'Accuracy of purepulse photoplethysmography technology of fitbit charge 2 for assessment of heart rate during sleep.' *Chronobiology International* 36(7), 927–933

Haghayegh, Shahab, Sepideh Khoshnevis, Michael H Smolensky, Kenneth R Diller, and Richard J Castriotta (2019b) 'Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis.' *Journal of Medical Internet research* 21(11), e16273

Hamermesh, Daniel S (2019) *Spending Time: The Most Valuable Resource* (Oxford University Press)

Handel, Benjamin, and Jonathan Kolstad (2017) 'Wearable technologies and health behaviors: new data and new methods to understand population health.' *American Economic Review: Papers and Proceedings* 107(5), 481–85

Harrison, Glenn W, Morten I Lau, and Melonie B Williams (2002) 'Estimating individual discount rates in denmark: A field experiment.' *American Economic Review* 92(5), 1606–1617

Heissel, Jennifer A, and Samuel Norris (2018) 'Rise and shine the effect of school start times on academic performance from childhood through puberty.' *Journal of Human Resources* 53(4), 957–992

Hershner, Shelley D, and Ronald D Chervin (2014) 'Causes and consequences of sleepiness among college students.' *Nature and Science of Sleep* 6, 73

Holt, Charles A, and Susan K Laury (2002) 'Risk aversion and incentive effects.' *American Economic Review* 92(5), 1644–1655

Huffman, David, Collin Raymond, and Julia Shvets (2018) 'Persistent overconfidence and biased memory: Evidence from managers'

Jakicic, John M, Kelliann K Davis, Renee J Rogers, Wendy C King, Marsha D Marcus, Diane Helsel, Amy D Rickman, Abdus S Wahed, and Steven H Belle (2016) 'Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the idea randomized clinical trial.' *JAMA* 316(11), 1161–1171

John, Anett (2020) 'When commitment fails: evidence from a field experiment.' *Management Science* 66(2), 503–529

Just, David R, and Joseph Price (2013) 'Using incentives to encourage healthy eating in children.' *Journal of Human Resources* 48(4), 855–872

Kahn-Greene, Ellen T, Desiree B Killgore, Gary H Kamimori, Thomas J Balkin, and William DS Killgore (2007) 'The effects of sleep deprivation on symptoms of psychopathology in healthy adults.' *Sleep Medicine* 8(3), 215–221

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan (2015) 'Self-control at work.' *Journal of Political Economy* 123(6), 1227–1277

Kremer, Michael, Gautam Rao, and Frank Schilbach (2019) 'Behavioral development economics.' *Handbook of Behavioral Economics*

Laibson, David (1997) 'Golden eggs and hyperbolic discounting.' *The Quarterly Journal of Economics* 112(2), 443–478

_ (2015) 'Why don't present-biased agents make commitments?' *American Economic Review* 105(5), 267–72

_ (2018) 'Private paternalism, the commitment puzzle, and model-free equilibrium.' In 'AEA Papers and Proceedings,' vol. 108 pp. 1–21

Lally, Phillippa, Cornelia HM Van Jaarsveld, Henry WW Potts, and Jane Wardle (2010) 'How are habits formed: Modelling habit formation in the real world.' *European Journal of Social Psychology* 40(6), 998–1009

Lauderdale, Diane S, Kristen L Knutson, Lijing L Yan, Kiang Liu, and Paul J Rathouz (2008a) 'Self-reported and measured sleep duration: how similar are they?' *Epidemiology* pp. 838–845

— (2008b) 'Sleep duration: how well do self-reports reflect objective measures? The CARDIA sleep study.' *Epidemiology* 19(6), 838

Lee, Hyun-Ah, Heon-Jeong Lee, Joung-Ho Moon, Taek Lee, Min-Gwan Kim, Hoh In, Chul-Hyun Cho, and Leen Kim (2017) 'Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults.' *Psychiatry Investigation* 14(2), 179

Liu, Cindy H, Courtney Stevens, Sylvia HM Wong, Miwa Yasui, and Justin A Chen (2019) 'The prevalence and predictors of mental health diagnoses and suicide among us college students: Implications for addressing disparities in service use.' *Depression and Anxiety* 36(1), 8–17

Liu, Yong (2016) 'Prevalence of healthy sleep duration among adults—united states, 2014.' *MMWR. Morbidity and mortality weekly report*

Lutter, Johanna I, Boglárka Szentes, Margarethe E Wacker, Joachim Winter, Sebastian Wichert, Annette Peters, Rolf Holle, and Reiner Leidl (2019) 'Are health risk attitude and general risk attitude associated with healthcare utilization, costs and working ability? results from the german kora ff4 cohort study.' *Health economics review* 9(1), 1–11

McEwen, Bruce S (2006) 'Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load.' *Metabolism* 55, S20–S23

Milojevich, Helen M, and Angela F Lukowski (2016) 'Sleep and mental health in undergraduate students with generally healthy sleep habits.' *PLOS ONE* 11(6), e0156372

Misra, Ranjita, and Michelle McKean (2000) 'College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction.' *American Journal of Health studies* 16(1), 41

Mitchell, Vincent-Wayne (1999) 'Consumer perceived risk: conceptualisations and models.' *European Journal of marketing*

Moreno, CRC, FM Louzada, LR Teixeira, F Borges, and G Lorenzi-Filho (2006) 'Short sleep is associated with obesity among truck drivers.' *Chronobiology International* 23(6), 1295–1303

Mullainathan, S (2014) 'Get some sleep, and wake up the gdp the new york times.' *The New York Times*

Nuscheler, Robert, and Kerstin Roeder (2016) 'To vaccinate or to procrastinate? that is the prevention question.' *Health economics* 25(12), 1560–1581

O'Donoghue, Ted, and Matthew Rabin (1999) 'Doing it now or later.' *American Economic Review* 89(1), 103–124

O'Donoghue, Ted, Matthew Rabin et al. (2006) 'Incentives and self-control.' *Econometric Society Monographs* 42, 215

Patel, Mitesh S, David A Asch, and Kevin G Volpp (2015) 'Wearable devices as facilitators, not drivers, of health behavior change.' *JAMA* 313(5), 459–460

Quer, Giorgio, Pishoy Gouda, Michael Galarnyk, Eric J Topol, and Steven R Steinhubl (2020) 'Inter-and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, bmi, and time of year: Retrospective, longitudinal cohort study of 92,457 adults.' *PLOS ONE* 15(2), e0227709

Rabin, Matthew, Ted O'Donoghue et al. (1999) 'Doing it now or later.' *American Economic Review* 89(1), 103–124

Roenneberg, Till (2013) 'Chronobiology: the human sleep project.' *Nature* 498(7455), 427–428

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

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

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen (2010) 'Risk and time preferences: Linking experimental and household survey data from Vietnam.' *American Economic Review* 100(1), 557–71

Twenge, Jean M, Zlatan Krizan, and Garrett Hisler (2017) 'Decreases in self-reported sleep duration among us adolescents 2009–2015 and association with new media screen time.' *Sleep Medicine* 39, 47–53

Volpp, Kevin G, Andrea B Troxel, Mark V Pauly, Henry A Glick, Andrea Puig, David A Asch, Robert Galvin, Jingsan Zhu, Fei Wan, Jill DeGuzman et al. (2009) 'A randomized, controlled trial of financial incentives for smoking cessation.' *New England Journal of Medicine* 360(7), 699–709

Watson, Nathaniel F, M Safwan Badr, Gregory Belenky, Donald L Bliwise, Orfeu M Buxton, Daniel Buysse, David F Dinges, James Gangwisch, Michael A Grandner et al. (2015) 'Recommended amount of sleep for a healthy adult: a joint consensus statement of the american academy of sleep medicine and sleep research society.' *Sleep* 38(6), 843–844

Weber, Elke U, Ann-Renee Blais, and Nancy E Betz (2002) 'A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors.' *Journal of Behavioral Decision Making* 15(4), 263–290

Wood, Wendy, and David T Neal (2007) 'A new look at habits and the habit-goal interface.' *Psychological Review* 114(4), 843

Zimmermann, Florian (2019) 'The dynamics of motivated beliefs.' *American Economic Review*

Table 1: Differences between Treatment and Control at Baseline

| | Waves 1-6 | | Waves 1, 5, and 6 | | Waves 5 and 6 | |
|---|---|---|---|---|---|---|
| Variables: | Coeff. | Std.err | Coeff. | Std.err | Coeff. | Std.err |
| Female | 0.037 | (0.047) | -0.000 | (0.053) | 0.020 | (0.059) |
| Age | -0.181 | (0.343) | 0.281 | (0.380) | 0.149 | (0.392) |
| White | 0.050 | (0.051) | 0.035 | (0.055) | 0.032 | (0.062) |
| Black | -0.013 | (0.027) | -0.027 | (0.031) | -0.027 | (0.036) |
| Asian | -0.033 | (0.048) | -0.009 | (0.052) | -0.000 | (0.059) |
| Other | -0.005 | (0.027) | 0.001 | (0.030) | -0.005 | (0.030) |
| Poor health | 0.025 | (0.029) | 0.009 | (0.031) | 0.032 | (0.036) |
| Satisfied with life | 0.028 | (0.040) | -0.006 | (0.048) | -0.105** | (0.045) |
| Depressed | -0.048 | (0.051) | -0.045 | (0.055) | -0.051 | (0.062) |
| Ever smoked | -0.026 | (0.040) | -0.002 | (0.041) | 0.028 | (0.045) |
| Drinks | -0.056 | (0.042) | -0.037 | (0.045) | -0.043 | (0.046) |
| BMI | 1.010 | (0.666) | 1.080 | (0.850) | 0.593 | (0.597) |
| Overweight | 0.039 | (0.046) | 0.042 | (0.051) | 0.049 | (0.058) |
| Obese | -0.014 | (0.027) | -0.021 | (0.032) | -0.012 | (0.036) |
| Ideal sleep | 0.111 | (0.086) | 0.049 | (0.097) | 0.009 | (0.111) |
| Self-reported sleep | -0.023 | (0.122) | -0.069 | (0.147) | -0.130 | (0.176) |
| Sleep quality | -0.252 | (0.155) | -0.183 | (0.170) | -0.314 | (0.196) |
| Sleep hours (Fitbit) | -0.348 | (0.218) | -0.089 | (0.124) | -0.075 | (0.121) |

*Notes -* Data are drawn from the first-day survey and from the first two weeks of Fitbit data (baseline period). Each cell reports the coefficient of a univariate regression of each covariate on an indicator for whether the individual was assigned to any incentive treatment or control. Robust standard errors are in parentheses.

Table 2: Incentives and Sleep Duration

| | (1) 7≤Sleep≤9 | (2) 7≤Sleep≤9 | (3) Sleep<6 | (4) Sleep<6 |
|---|---|---|---|---|
| Panel A | Waves 1-6 | | | |
| Treatment | 0.0664*** | 0.0546*** | -0.0353** | -0.0328*** |
| | (0.017) | (0.012) | (0.015) | (0.010) |
| Post-Treatment | 0.0395* | 0.0348** | -0.0146 | -0.0251** |
| | (0.022) | (0.015) | (0.021) | (0.012) |
| Individual FE | NO | YES | NO | YES |
| Observations | 18,413 | 18,413 | 18,413 | 18,413 |
| R-squared | 0.021 | 0.004 | 0.027 | 0.005 |
| Mean of Dep. Var. | 0.429 | 0.429 | 0.258 | 0.258 |
| Std.Dev. of Dep. Var. | 0.495 | 0.495 | 0.438 | 0.438 |
| Individuals | 460 | 460 | 460 | 460 |
| Panel B | Waves 1, 5, and 6 | | | |
| Treatment | 0.0596** | 0.0533*** | -0.0426* | -0.0421*** |
| | (0.025) | (0.017) | (0.022) | (0.015) |
| Post-Treatment | 0.0517* | 0.0389** | -0.0340 | -0.0356** |
| | (0.030) | (0.019) | (0.030) | (0.016) |
| Individual FE | NO | YES | NO | YES |
| Observations | 13,460 | 13,460 | 13,460 | 13,460 |
| R-squared | 0.020 | 0.004 | 0.032 | 0.005 |
| Mean of Dep. Var. | 0.419 | 0.419 | 0.264 | 0.264 |
| Std.Dev. of Dep. Var. | 0.493 | 0.493 | 0.441 | 0.441 |
| Individuals | 285 | 285 | 285 | 285 |
| Panel C | Waves 5 and 6 | | | |
| Treatment | 0.0799*** | 0.0576*** | -0.0622** | -0.0455*** |
| | (0.029) | (0.018) | (0.027) | (0.017) |
| Post-Treatment | 0.0638** | 0.0406** | -0.0423 | -0.0330* |
| | (0.032) | (0.020) | (0.032) | (0.017) |
| Individual FE | NO | YES | NO | YES |
| Observations | 11,623 | 11,623 | 11,623 | 11,623 |
| R-squared | 0.021 | 0.004 | 0.034 | 0.005 |
| Mean of Dep. Var. | 0.405 | 0.405 | 0.276 | 0.276 |
| Std.Dev. of Dep. Var. | 0.491 | 0.491 | 0.447 | 0.447 |
| Individuals | 216 | 216 | 216 | 216 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 3 and 4 is an indicator for sleeping less than 6 hours. We pooled any monetary incentive in Panel A and B. Panel C focuses on waves 5 and 6 where we only had the strong weekly incentive as the treatment group. The unit of observation is subject-night. All estimates include controls for day of the week, month fixed effects and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

## Table 3: Incentives and Time Allocation

| Panel A | (1) Time Use Sleep (hours) | (2) Time Use 7≤ Sleep ≤ 9 | (3) Time Use <6 Sleep | (4) Study | (5) Work | (6) Care | (7) Exercise |
|---|---|---|---|---|---|---|---|
| Treatment | 0.1465 | 0.0274 | -0.0391*** | 0.4242** | -0.0927 | -0.0594 | 0.0241 |
| | (0.100) | (0.024) | (0.012) | (0.215) | (0.170) | (0.088) | (0.042) |
| Post-Treatment | 0.0648 | 0.0063 | -0.0315* | 0.2999 | 0.0181 | 0.0094 | 0.0382 |
| | (0.128) | (0.030) | (0.016) | (0.282) | (0.224) | (0.119) | (0.058) |
| Observations | 3,053 | 3,053 | 3,053 | 3,053 | 3,053 | 3,053 | 3,053 |
| Mean of Dep. Var. | 8.121 | 0.384 | 0.0740 | 4.926 | 1.807 | 2.821 | 0.419 |
| Std.Dev. of Dep. Var. | 1.974 | 0.486 | 0.262 | 3.556 | 2.784 | 1.450 | 0.798 |
| Individuals | 455 | 455 | 455 | 455 | 455 | 455 | 455 |
| Panel B | Relaxing | Other Activities | Social Activities | Screen Time | TV/Videos | Internet | Gaming |
| Treatment | -0.5335*** | 0.0907 | 0.0637 | -0.5972*** | -0.2030** | -0.1711 | -0.2231** |
| | (0.168) | (0.147) | (0.106) | (0.167) | (0.095) | (0.111) | (0.099) |
| Post-Treatment | -0.1402 | -0.2903 | -0.0151 | -0.1251 | -0.2111* | 0.1612 | -0.0751 |
| | (0.225) | (0.184) | (0.141) | (0.226) | (0.119) | (0.156) | (0.122) |
| Observations | 3,053 | 3,053 | 3,053 | 3,053 | 3,053 | 3,053 | 3,053 |
| Mean of Dep. Var. | 4.442 | 1.465 | 1.568 | 2.874 | 1.067 | 1.360 | 0.447 |
| Std.Dev. of Dep. Var. | 2.846 | 3.078 | 1.963 | 2.666 | 1.551 | 1.824 | 1.329 |
| Individuals | 455 | 455 | 455 | 455 | 455 | 455 | 455 |

*Notes* - All the dependent variables are drawn from the time-use surveys collected throughout the experiment and are measured in minutes. Subjects were surveyed on their time-use on two randomly chosen days each week. All estimates include controls for gender, age, day of the week, month fixed effects, and a weekly linear trend. The analysis includes waves 1-6. Standard errors are clustered at the individual level.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table 4: Sleep & Bedtime Targets and Sleep Duration

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | 7hrs$<\leq$Sleep$\leq$9hrs | 7hrs$\leq$Sleep$\leq$9hrs | 7hrs$\leq$Sleep$\leq$9hrs | Sleep$<$6hrs | Sleep$<$6hrs | Sleep$<$6hrs |
| Bedtime earlier than 1am | 0.0769*** | | | -0.0413 | | |
| & Sleep $>$ 7hrs | (0.029) | | | (0.026) | | |
| Sleep $>$7hrs | | 0.0562** | | | -0.0341 | |
| | | (0.028) | | | (0.027) | |
| Bedtime earlier than 1am | | | 0.0358 | | | -0.0045 |
| | | | (0.027) | | | (0.025) |
| Observations | 3,640 | 3,640 | 3,640 | 3,640 | 3,640 | 3,640 |
| R-squared | 0.083 | 0.081 | 0.080 | 0.100 | 0.100 | 0.098 |
| Mean of Dep. Var. | 0.475 | 0.475 | 0.475 | 0.234 | 0.234 | 0.234 |
| Std.Dev. of Dep. Var. | 0.499 | 0.499 | 0.499 | 0.424 | 0.424 | 0.424 |
| Individuals | 284 | 284 | 284 | 284 | 284 | 284 |

*Notes -* The dependent variable in columns 1-3 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 4-6 is an indicator for sleeping less than 6 hours. The analysis includes waves 1-6. Data are drawn from Fitbit during the weeks of the intervention (weeks 3, 4, and 5 in waves 2, 3, 5 and 6; weeks 3, 5, 7 in waves 1 and 4). This is a correlational analysis restricting the sample to treated subjects. All estimates include controls for age, gender, ethnicity, race, wave fixed effects and baseline sleep duration. The unit of observation is subject-night in the weeks of treatment. Standard errors are clustered at the individual level. * $p<0.10$ ** $p<0.05$ *** $p<0.01$

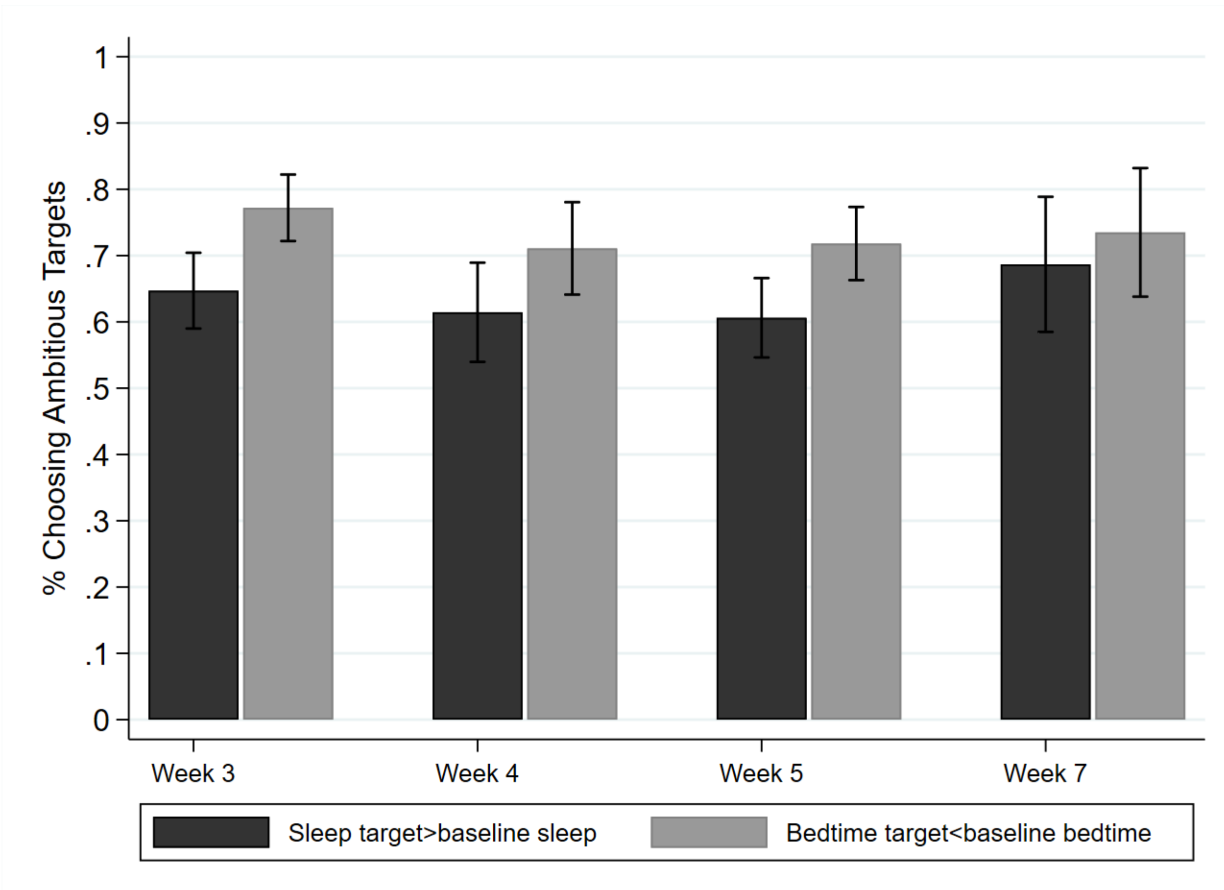
Table 5: Overconfidence and Sleep Duration (Self-reported vs Fitbit data)

| | Self-reported (Day 1 Survey) | | | Actual Sleep (Fitbit) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Sleep hours | 7≤Sleep≤9 | Sleep<7hrs | Sleep Hours | Sleep<7hrs | 7≤Sleep≤9 |
| Panel A: Self-Reported Sleep at Baseline and Actual Sleep | | | | | | |
| Overconfident | 0.9006*** | 0.3397*** | -0.1307*** | -1.3332*** | -0.3217*** | 0.3146*** |
| | (0.089) | (0.044) | (0.023) | (0.131) | (0.044) | (0.043) |
| Observations | 486 | 486 | 486 | 486 | 486 | 486 |
| R-squared | 0.235 | 0.173 | 0.098 | 0.284 | 0.206 | 0.268 |
| Mean of Dep. Var. | 6.810 | 0.572 | 0.103 | 6.902 | 0.428 | 0.173 |
| Std.Dev. of Dep. Var. | 0.999 | 0.495 | 0.304 | 1.591 | 0.495 | 0.378 |
| Panel B: Predicted Sleep Over First Two Weeks And Actual Sleep (Wave 6) | | | | | | |
| Overconfident | 0.3579** | 0.0725 | -0.0991** | -0.5161*** | -0.2320*** | 0.1555*** |
| | (0.156) | (0.075) | (0.050) | (0.103) | (0.073) | (0.053) |
| Observations | 189 | 189 | 189 | 189 | 189 | 189 |
| R-squared | 0.103 | 0.094 | 0.135 | 0.385 | 0.170 | 0.156 |
| Mean of Dep. Var. | 6.760 | 0.635 | 0.0990 | 6.693 | 0.370 | 0.156 |
| Std.Dev. of Dep. Var. | 1.021 | 0.483 | 0.299 | 0.839 | 0.484 | 0.364 |
| Panel C: Overconfidence Index and Actual Sleep | | | | | | |
| Overconfident | 0.8541*** | 0.3237*** | -0.1302*** | -0.7797*** | -0.1786*** | 0.0991 |
| | (0.123) | (0.053) | (0.022) | (0.204) | (0.063) | (0.061) |
| Observations | 486 | 486 | 486 | 486 | 486 | 486 |
| R-squared | 0.153 | 0.126 | 0.081 | 0.173 | 0.133 | 0.141 |
| Mean of Dep. Var. | 6.810 | 0.572 | 0.103 | 6.907 | 0.426 | 0.172 |

*Notes -* Data are drawn from the first-day survey (columns 1-3) and the Fitbit data for the first two weeks of the experiment before intervention (columns 4-6). The analysis includes waves 1-6. The dependent variables are self-reported measures of sleep hours (column 1), sleeping less than 7 hours (column 2), and sleeping between 7 and 9 hours (column 3). In columns 4-6, the dependent variables are the same but measured by Fitbit data. The unit of observation is the subject. Panel A compares self-reported sleep as measured in the first-day survey to average sleep duration at baseline. We define as overconfident individuals those who slept on average 30 minutes less than what reported in the survey. Panel B compares individuals' prediction about their sleep over the following two weeks and their actual sleep as measured by the Fitbits. In Panel B, we restricted the sample to wave 6 as explained in Section 5.2. Again we define as overconfident individuals those who slept 30 minutes less than what they predicted. In Panel C, we define as overconfident individuals those in the upper quartile of an index measuring the difference between own and others' sleep duration, sleep quality, and perceived risks from sleep deprivation.

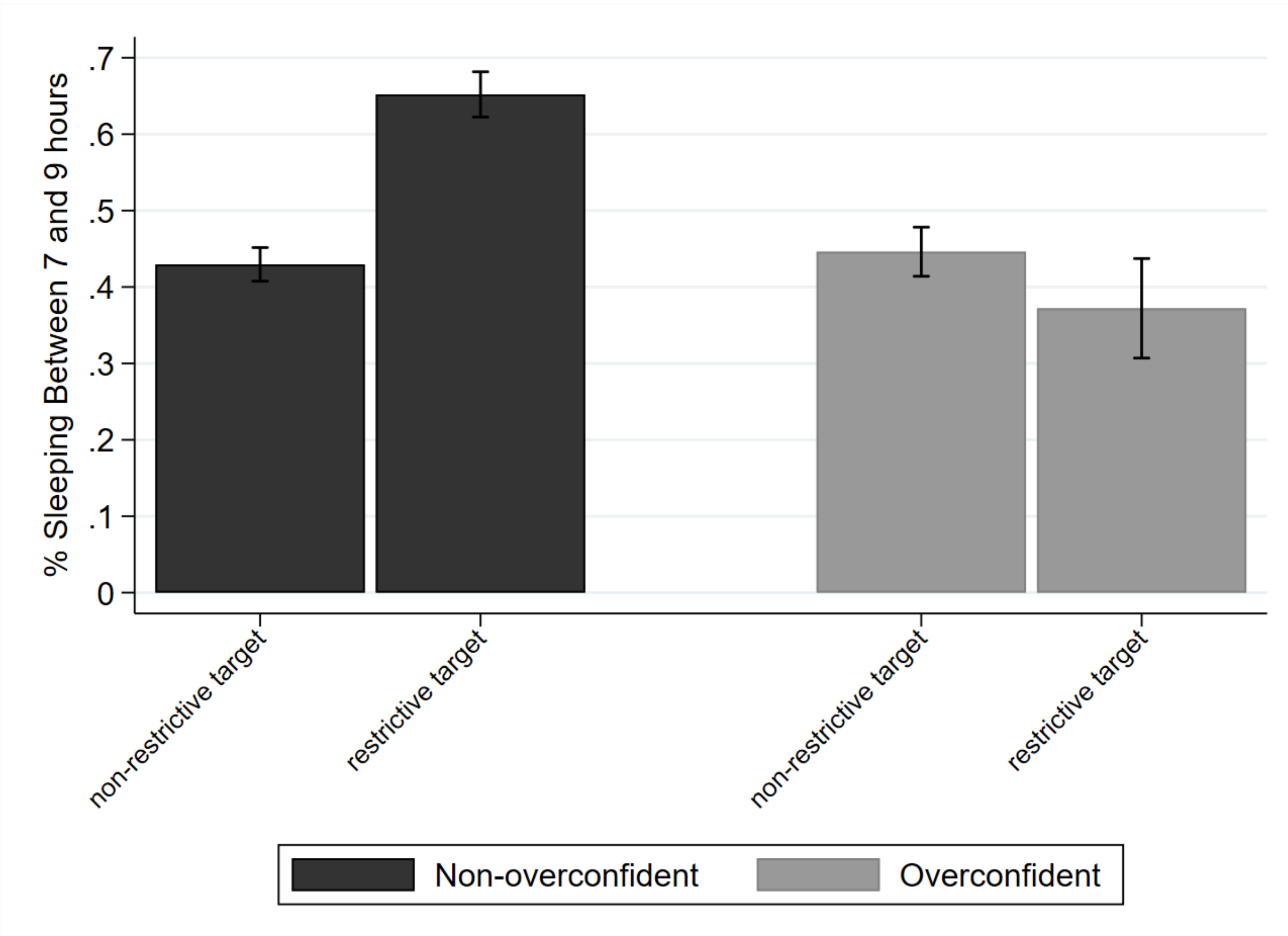
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Figure 1: Share of participants choosing a restrictive target with respect to their baseline sleep duration and bedtime habits




*Notes -* The figure above reports the share of participants choosing a restrictive target with respect to their baseline sleep duration (black) or bedtime target (gray) with 95% confidence intervals. In waves 1 and 4, treated weeks were week 3, 5, and 7. In waves 2, 3, 5, and 6 treated weeks were week 3, 4, and 5. This analysis includes waves 1-6.

Figure 2: Overconfidence, Restrictive Targets, and Sleep




*Notes* - The figure demonstrates the share of individuals sleeping between 7 and 9 hours across our measure of overconfidence as discussed in Section 5.2, and between choosing restrictive sleep targets or not with 95% confidence intervals. This analysis includes waves 1-6.

# APPENDIX

## A Additional Tables and Figures

Table A.1: Summary of Treatments

| **Treatment** | **Wave** | N | **Location** | **Time** | **Incentive** | **Prediction Reward** |
|---|---|---|---|---|---|---|
| **Control** | 1 | N=30 | Oxford | Oct-Dec 2016 | None | No |
| **Treatment** | 1 | N=20 | Oxford | Oct-Dec 2016 | Biweekly, Weak | No |
| **Dominated contract** | 1 | N=19 | Oxford | Oct-Dec 2016 | Biweekly, Weak | No |
| **Treatment** | 2 | N=51 | Oxford | Apr-Jun 2017 | Weekly, Strong | Yes, 1 |
| **Treatment** | 3 | N=43 | Oxford | Oct-Dec 2017 | Weekly, Strong | Yes, 3 |
| **Treatment** | 4 | N=81 | Pittsburgh | Jan-Mar 2018 | Biweekly, Strong | Yes, 1 |
| **Control** | 5 | N=52 | Pittsburgh | Sep-Nov 2018 | None | No |
| **Treatment** | 5 | N=23 | Pittsburgh | Sep-Nov 2018 | Weekly, Strong | Yes, 3 |
| **Control** | 6 | N=48 | Pittsburgh | Jan-May 2021 | Weekly, None | No |
| **Treatment** | 6 | N=46 | Pittsburgh | Jan-May 2021 | Weekly, Strong | Yes, 3 |
| **Dominated contract** | 6 | N=47 | Pittsburgh | Jan-May 2021 | Weekly, Strong | Yes, 3 |
| **Target-No incentive** | 6 | N=48 | Pittsburgh | Jan-May 2021 | Weekly, None | No |

*Notes -* The table above describes the location, timing, incentive structure and sample size used in the different waves of the experiment.

Table A.2: Summary Statistics

| | Waves 1-6 | | Waves 1, 5, and 6 | | Waves 5 and 6 | |
|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| Female | 0.52 | 0.50 | 0.63 | 0.48 | 0.65 | 0.48 |
| Age | 20.94 | 3.94 | 20.62 | 4.16 | 19.77 | 3.24 |
| White | 0.54 | 0.50 | 0.50 | 0.50 | 0.48 | 0.50 |
| Black | 0.08 | 0.27 | 0.09 | 0.29 | 0.09 | 0.29 |
| Asian | 0.27 | 0.45 | 0.31 | 0.46 | 0.34 | 0.48 |
| Other Race Category | 0.10 | 0.31 | 0.10 | 0.29 | 0.08 | 0.27 |
| Poor Health | 0.10 | 0.30 | 0.09 | 0.28 | 0.09 | 0.29 |
| Satisfied with life | 0.34 | 0.47 | 0.27 | 0.44 | 0.18 | 0.39 |
| Depressed | 0.51 | 0.50 | 0.48 | 0.50 | 0.46 | 0.50 |
| Ever smoked | 0.19 | 0.39 | 0.18 | 0.39 | 0.16 | 0.37 |
| Drinks | 0.23 | 0.42 | 0.21 | 0.41 | 0.17 | 0.37 |
| BMI | 23.93 | 8.33 | 24.09 | 7.30 | 23.82 | 4.72 |
| Overweight | 0.27 | 0.44 | 0.28 | 0.45 | 0.29 | 0.46 |
| Obese | 0.07 | 0.26 | 0.09 | 0.29 | 0.09 | 0.29 |
| Ideal Sleep | 7.87 | 0.85 | 7.81 | 0.86 | 7.77 | 0.87 |
| Sleep hours (self-reported) | 6.85 | 1.23 | 6.84 | 1.34 | 6.81 | 1.47 |
| Sleep Quality (self-reported) | 6.63 | 1.60 | 6.67 | 1.59 | 6.59 | 1.60 |
| Sleep hours (Fitbit) | 6.93 | 1.64 | 6.81 | 1.09 | 6.69 | 0.95 |

*Notes -* Data are drawn from the Day 1 Survey and from Fitbit data at baseline. The first-day survey was filled in by 564 individuals when considering all waves (columns 1 and 2); 376 individuals in waves 1, 5, and 6 (coulumns 3-4); 289 individuals in waves 5 and 6 (columns 5 and 6)).

Table A.3: Attrition

| | Waves 1-6 | | Waves 1, 5, and 6 | | Waves 5 and 6 | |
|---|---|---|---|---|---|---|
| Variables: | Coeff. | std.err | Coeff. | std.err | Coeff. | std.err |
| Female | 0.029 | (0.072) | 0.006 | (0.086) | 0.006 | (0.113) |
| Age | 1.380 | (0.863) | 1.808* | (1.083) | 0.420 | (0.493) |
| White | -0.104 | (0.073) | -0.065 | (0.081) | -0.060 | (0.103) |
| Black | 0.018 | (0.044) | 0.005 | (0.050) | -0.059 | (0.055) |
| Asian | -0.080 | (0.056) | -0.091 | (0.067) | -0.070 | (0.092) |
| Other Race Category | 0.167*** | (0.061) | 0.152** | (0.067) | 0.189** | (0.084) |
| Poor health | -0.052 | (0.034) | -0.029 | (0.040) | 0.006 | (0.062) |
| Satisfied with life | -0.052 | (0.066) | -0.043 | (0.072) | -0.072 | (0.082) |
| Depressed | 0.010 | (0.072) | -0.022 | (0.081) | -0.048 | (0.102) |
| Ever smoked | 0.053 | (0.062) | 0.072 | (0.070) | 0.011 | (0.079) |
| Drinks more than once a week | -0.005 | (0.062) | -0.010 | (0.067) | -0.007 | (0.076) |
| BMI | -0.147 | (1.045) | -0.179 | (1.292) | -0.299 | (1.073) |
| Overweight | -0.071 | (0.062) | -0.106 | (0.066) | -0.123 | (0.082) |
| Obese | 0.014 | (0.044) | -0.003 | (0.049) | 0.017 | (0.065) |
| Ideal sleep | -0.199* | (0.109) | -0.242* | (0.129) | -0.014 | (0.144) |
| Sleep duration (self-reported) | 0.033 | (0.153) | 0.118 | (0.183) | 0.340 | (0.283) |
| Sleep quality (self-reported) | -0.120 | (0.281) | -0.074 | (0.342) | -0.058 | (0.450) |
| Sleep duration (Fitbit) | 0.278 | (0.286) | 0.278 | (0.285) | 0.278 | (0.285) |
| Observations | 564 | | 376 | | 283 | |

*Notes -* We report estimates of the univariate regression of the likelihood of withdrawing from the experiment on baseline characteristics.

* p<0.10 ** p<0.05 *** p<0.01

Table A.4: Attrition by Treatment Status (Waves 1, 5, and 6)

| | Weeks 1-15 | Weeks 1-8 |
|---|---|---|
| Treatment | -0.0146 | -0.0160 |
| | (0.017) | (0.017) |
| Post-Treatment | 0.0195 | -0.0271 |
| | (0.023) | (0.021) |
| Observations | 15,827 | 11,289 |
| Individuals | 285 | 285 |
| Mean of Dep. Var. | 0.150 | 0.140 |
| Std.Dev. of Dep. Var. | 0.357 | 0.347 |

*Notes* - The dependent variable is a dummy for whether the subject had synced data for a given night. All estimates include individual fixed effects and day of the week fixed effects. Column 2 restricts the analysis to weeks 1-8.

Table A.5: Comparisons of Sleep Measurements

| | Sleep Duration | 7≤Sleep≤9 | Sleep<6 |
|---|---|---|---|
| Fitbit | 6.92 | 0.40 | 0.27 |
| | (1.64) | (0.26) | (0.26) |
| Self-Reported | 6.85 | 0.57 | 0.10 |
| | (1.22) | (0.49) | (0.30) |
| Time Use | 7.95 | 0.66 | 0.064 |
| | (1.69) | (0.47) | (0.24) |

*Notes* - This table compares averages of sleep collected before our intervention started for three distinct measures. The analysis includes waves 1-6. We restricted the sample of Fitbit and time-use data to the baseline period (first two weeks of data) and excluded weekends. Standard deviations are reported in parentheses. The first row (Fitbit) reports the sleep measures derived from the Fitbit data. The second row (Self-Reported) reports the sleep measures elicited in first-day survey. The third row (Time Use) reports the sleep measures based on the time use surveys.

Table A.6: Differences between Treatment and Control at Baseline, by Wave of the Experiment

| Variables | Wave 1 | | Wave 2 | | Wave 3 | | Wave 4 | | Wave 5 | | Wave 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Female | -0.028 | (0.122) | 0.161** | (0.080) | 0.051 | (0.091) | -0.544*** | (0.041) | -0.091 | (0.130) | 0.004 | (0.065) |
| Age | -0.472 | (0.823) | 1.873*** | (0.590) | 1.047* | (0.562) | -0.190 | (0.497) | 0.033 | (0.390) | 0.134 | (0.536) |
| White | 0.012 | (0.123) | 0.080 | (0.085) | 0.196** | (0.084) | 0.106 | (0.069) | 0.043 | (0.131) | 0.031 | (0.073) |
| Black | -0.014 | (0.059) | -0.092*** | (0.035) | -0.115*** | (0.026) | -0.022 | (0.043) | -0.097 | (0.085) | 0.015 | (0.037) |
| Asian | 0.004 | (0.105) | 0.041 | (0.080) | -0.277*** | (0.037) | -0.077 | (0.059) | 0.131 | (0.124) | -0.060 | (0.070) |
| Other Race Category | -0.005 | (0.083) | -0.028 | (0.038) | 0.196** | (0.076) | -0.007 | (0.036) | -0.077** | (0.037) | 0.015 | (0.037) |
| Poor health | -0.082 | (0.072) | 0.026 | (0.053) | 0.102 | (0.069) | 0.046 | (0.046) | -0.020 | (0.079) | 0.057 | (0.039) |
| Satisfied with life | 0.208* | (0.120) | 0.132 | (0.085) | 0.164* | (0.091) | 0.225*** | (0.069) | 0.015 | (0.131) | -0.059* | (0.034) |
| Depressed | -0.054 | (0.123) | 0.054 | (0.085) | 0.031 | (0.092) | 0.010 | (0.071) | -0.158 | (0.128) | -0.013 | (0.072) |
| Ever smoked | -0.146 | (0.103) | -0.060 | (0.068) | 0.002 | (0.079) | -0.082 | (0.055) | -0.012 | (0.112) | 0.070 | (0.046) |
| Drinks | -0.092 | (0.118) | 0.070 | (0.079) | 0.049 | (0.084) | -0.008 | (0.061) | -0.011 | (0.093) | -0.062 | (0.056) |
| BMI | 2.530 | (2.884) | 1.622 | (3.032) | -1.293* | (0.776) | -0.257 | (0.538) | -0.644 | (1.017) | 1.026 | (0.687) |
| Overweight | 0.041 | (0.109) | -0.130* | (0.068) | -0.070 | (0.079) | -0.013 | (0.064) | -0.022 | (0.119) | 0.071 | (0.067) |
| Obese | -0.061 | (0.078) | -0.072** | (0.034) | -0.067* | (0.037) | -0.069** | (0.031) | -0.135*** | (0.048) | 0.027 | (0.042) |
| Ideal sleep | 0.091 | (0.193) | 0.172 | (0.127) | 0.251* | (0.152) | 0.230* | (0.124) | -0.123 | (0.159) | 0.101 | (0.135) |
| Term sleep | 0.090 | (0.185) | 0.071 | (0.202) | 0.062 | (0.211) | 0.174 | (0.176) | -0.324 | (0.519) | -0.057 | (0.147) |
| Sleep quality (self-reported) | 0.167 | (0.343) | 0.048 | (0.278) | -0.554* | (0.300) | -0.264 | (0.226) | -0.177 | (0.430) | -0.398* | (0.227) |
| Sleep hours (self-reported) | -0.406 | (0.335) | 0.081 | (0.245) | 0.131 | (0.305) | -0.535* | (0.299) | 0.028 | (0.341) | -0.106 | (0.122) |

*Notes* - Data are drawn from the first-day survey and from the first two weeks of Fitbit data (baseline period). Each cell reports the coefficient of a univariate regression of each covariate on an indicator for whether the individual was assigned to any incentive treatment or control. Robust standard errors are in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.7: Incentives and Sleep, Waves 1-5; Wave 6

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 7≤Sleep ≤9 | 7≤Sleep ≤9 | Sleep<6 | Sleep<6 |
| Panel A | | Waves 1-5 | | |
| Treatment | 0.0761*** | 0.0495*** | -0.0518*** | -0.0332** |
| | (0.020) | (0.016) | (0.017) | (0.013) |
| Post Treatment | 0.0618* | 0.0249 | -0.0645** | -0.0486** |
| | (0.033) | (0.022) | (0.029) | (0.019) |
| Individual FE | NO | YES | NO | YES |
| Observations | 8,738 | 8,738 | 8,738 | 8,738 |
| R-squared | 0.026 | 0.004 | 0.030 | 0.006 |
| Mean of Dep. Var. | 0.456 | 0.456 | 0.245 | 0.245 |
| Std.Dev. of Dep. Var. | 0.498 | 0.498 | 0.430 | 0.430 |
| Individuals | 319 | 319 | 319 | 319 |
| Panel B | | Wave 6 | | |
| Treatment | 0.0635** | 0.0598*** | -0.0445 | -0.0479** |
| | (0.031) | (0.020) | (0.030) | (0.019) |
| Post Treatment | 0.0525 | 0.0453** | -0.0242 | -0.0294 |
| | (0.035) | (0.023) | (0.036) | (0.019) |
| Individual FE | NO | YES | NO | YES |
| Observations | 9,675 | 9,675 | 9,675 | 9,675 |
| R-squared | 0.022 | 0.004 | 0.035 | 0.005 |
| Mean of Dep. Var. | 0.405 | 0.405 | 0.270 | 0.270 |
| Std.Dev. of Dep. Var. | 0.491 | 0.491 | 0.444 | 0.444 |
| Individuals | 141 | 141 | 141 | 141 |

*Notes -* The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 3 and 4 is an indicator for sleeping less than 6 hours. The unit of observation is subject-night. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

## Table A.8: Incentives and Sleep Duration (Alternative Metrics)

| | (1) Sleep hrs | (2) Sleep hrs | (3) log (Sleep hrs) | (4) log (Sleep hrs) | (5) Sleep<7 | (6) Sleep<7 | (7) Sleep<5 | (8) Sleep<5 |
|---|---|---|---|---|---|---|---|---|
| Panel A | | | | Waves 1-6 | | | | |
| Treatment | 0.1219* | 0.0383 | 0.0193* | 0.0107 | -0.0703*** | -0.0431*** | 0.0003 | -0.0076 |
| | (0.062) | (0.041) | (0.011) | (0.008) | (0.018) | (0.012) | (0.011) | (0.008) |
| Post-Treatment | 0.0551 | 0.0243 | 0.0126 | 0.0112 | -0.0321 | -0.0193 | 0.0015 | -0.0105 |
| | (0.084) | (0.046) | (0.015) | (0.008) | (0.025) | (0.015) | (0.015) | (0.009) |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES |
| Observations | 18,413 | 18,413 | 18,412 | 18,412 | 18,413 | 18,413 | 18,413 | 18,413 |
| R-squared | 0.041 | 0.005 | 0.056 | 0.005 | 0.028 | 0.006 | 0.031 | 0.003 |
| Mean of Dep. Var. | 6.820 | 6.820 | 1.877 | 1.877 | 0.503 | 0.503 | 0.124 | 0.124 |
| Std.Dev. of Dep. Var. | 1.739 | 1.739 | 0.324 | 0.324 | 0.500 | 0.500 | 0.330 | 0.330 |
| Individuals | 460 | 460 | 460 | 460 | | 460 | 460 | 460 |
| Panel B | | | | Waves 1, 5, and 6 | | | | |
| Treatment | 0.0984 | 0.1102** | 0.0148 | 0.0169 | -0.0568** | -0.0545*** | -0.0114 | -0.0155 |
| | (0.085) | (0.053) | (0.016) | (0.010) | (0.026) | (0.017) | (0.016) | (0.011) |
| | (0.058) | (0.053) | (0.011) | (0.010) | (0.019) | (0.017) | (0.011) | (0.011) |
| Post-Treatment | 0.1624 | 0.1511*** | 0.0264 | 0.0275*** | -0.0538 | -0.0403** | -0.0138 | -0.0209* |
| | (0.118) | (0.055) | (0.021) | (0.010) | (0.034) | (0.020) | (0.022) | (0.011) |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES |
| Observations | 13,460 | 13,460 | 13,459 | 13,459 | 13,460 | 13,460 | 13,460 | 13,460 |
| R-squared | 0.053 | 0.008 | 0.073 | 0.006 | 0.029 | 0.008 | 0.039 | 0.004 |
| Mean of Dep. Var. | 6.767 | 6.767 | 1.872 | 1.872 | 0.521 | 0.521 | 0.121 | 0.121 |
| Std.Dev. of Dep. Var. | 1.666 | 1.666 | 0.317 | 0.317 | 0.500 | 0.500 | 0.326 | 0.326 |
| Individuals | 285 | 285 | 285 | 285 | 285 | 285 | 285 | 285 |
| Panel C | | | | Waves 5 and 6 | | | | |
| Treatment Treatment | 0.1767* | 0.0953 | 0.0247 | 0.0123 | -0.0795*** | -0.0522*** | -0.0206 | -0.0179 |
| | (0.101) | (0.059) | (0.019) | (0.012) | (0.030) | (0.020) | (0.020) | (0.013) |
| Post-Treatment | 0.1799 | 0.1070* | 0.0300 | 0.0204** | -0.0655* | -0.0354* | -0.0165 | -0.0198* |
| | (0.126) | (0.056) | (0.023) | (0.010) | (0.036) | (0.021) | (0.023) | (0.012) |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES |
| Observations | 11,623 | 11,623 | 11,622 | 11,622 | 11,623 | 11,623 | 11,623 | 11,623 |
| R-squared | 0.056 | 0.008 | 0.084 | 0.006 | 0.026 | 0.007 | 0.042 | 0.004 |
| Mean of Dep. Var. | 6.706 | 6.706 | 1.864 | 1.864 | 0.539 | 0.539 | 0.125 | 0.125 |
| Std.Dev. of Dep. Var. | 1.621 | 1.621 | 0.308 | 0.308 | 0.499 | 0.499 | 0.331 | 0.331 |
| Individuals | 216 | 216 | 216 | 216 | 216 | 216 | 216 | 216 |

*Notes* - The dependent variables are sleep hours (columns 1-2), the log of sleep hours (column 3-4), an indicator for sleeping less than 7 hours (columns 5-6), and an indicator for sleeping less than 5 hours (columns 7-8). All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1, 3, 5, 7 include dummies for gender and age. Columns 2, 4, 6, and 8 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.9: Incentives and Sleep - Excluding Sleep<4hrs or Sleep>10hrs

| | (1)<br>Sleep hrs | (2)<br>Sleep hrs | (3)<br>log (Sleep hrs) | (4)<br>log (Sleep hrs) |
|---|---|---|---|---|
| Treatment | 0.1586*** | 0.0884*** | 0.0226*** | 0.0135*** |
| | (0.045) | (0.029) | (0.007) | (0.004) |
| Post-Treatment | 0.0561 | 0.0477 | 0.0079 | 0.0080 |
| | (0.063) | (0.037) | (0.009) | (0.006) |
| Individual FE | NO | YES | NO | YES |
| Observations | 16,882 | 16,882 | 16,882 | 16,882 |
| R-squared | 0.028 | 0.006 | 0.027 | 0.005 |
| Mean of Dep. Var. | 7.004 | 7.004 | 1.930 | 1.930 |
| Std.Dev. of Dep. Var. | 1.235 | 1.235 | 0.185 | 0.185 |

*Notes -* The dependent variables are sleep hours (columns 1-2) and the log of sleep hours (column 3-4). The analysis includes waves 1-6. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.10: Incentives, Time in Bed, Bedtime, Wake-up Time

| | (1)<br>Time in bed | (2)<br>Time in bed | (3)<br>Bedtime | (4)<br>Bedtime | (5)<br>Wake-up Time | (6)<br>Wake-up Time |
|---|---|---|---|---|---|---|
| Treatment | 2.8017 | 3.6762 | -0.6244 | -1.2999*** | -0.1082 | -0.2286*** |
| | (3.690) | (2.421) | (0.454) | (0.342) | (0.088) | (0.058) |
| Post-Treatment | 1.0774 | 2.7471 | 0.2980 | -0.7100 | 0.1521 | -0.0176 |
| | (5.551) | (2.988) | (0.744) | (0.451) | (0.123) | (0.073) |
| Individual FE | NO | YES | NO | YES | NO | YES |
| Observations | 18,413 | 18,413 | 18,413 | 18,413 | 18,413 | 18,413 |
| R-squared | 0.041 | 0.006 | 0.016 | 0.004 | 0.032 | 0.003 |
| Mean of Dep. Var. | 455 | 455 | 27.92 | 27.92 | 8.170 | 8.170 |
| Std.Dev. of Dep. Var. | 111.6 | 111.6 | 14.34 | 14.34 | 2.490 | 2.490 |
| Individuals | 460 | 460 | 460 | 460 | 460 | 460 |

*Notes* - The dependent variable in columns 1 and 2 is time spent in bed (in minutes), including time spent in bed awake. The dependent variable in columns 3 and 4 is bedtime (in hours), where 1am is coded as 25, 2am is 26 and so on. The dependent variable in columns 5 and 6 is the wake-up time (in hours). The analysis includes waves 1-6. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1, 3, and 5 include dummies for gender and age. Columns 2, 4, and 6 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.11: Incentives and Sleep (Spring 2021), Including Weekends

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 7≤Sleep ≤9 | 7≤Sleep ≤9 | Sleep<6 | Sleep<6 |
| Panel A | | Waves 1-6 | | |
| Treatment | 0.0600*** | 0.0472*** | -0.0364*** | -0.0262*** |
| | (0.016) | (0.012) | (0.013) | (0.009) |
| Post-Treatment | 0.0375* | 0.0249** | -0.0193 | -0.0163 |
| | (0.020) | (0.012) | (0.020) | (0.010) |
| Observations | 25,576 | 25,576 | 25,576 | 25,576 |
| R-squared | 0.017 | 0.003 | 0.023 | 0.003 |
| Mean of Dep. Var. | 0.432 | 0.432 | 0.253 | 0.253 |
| Std.Dev. of Dep. Var. | 0.495 | 0.495 | 0.435 | 0.435 |
| Individuals | 460 | 460 | 460 | 460 |
| Panel B | | Waves 1, 5, and 6 | | |
| Treatment | 0.0545** | 0.0440** | -0.0407** | -0.0343*** |
| | (0.023) | (0.017) | (0.020) | (0.013) |
| Post-Treatment | 0.0474* | 0.0295* | -0.0365 | -0.0287** |
| | (0.026) | (0.016) | (0.027) | (0.014) |
| Observations | 18,751 | 18,751 | 18,751 | 18,751 |
| R-squared | 0.017 | 0.003 | 0.027 | 0.003 |
| Mean of Dep. Var. | 0.422 | 0.422 | 0.259 | 0.259 |
| Std.Dev. of Dep. Var. | 0.494 | 0.494 | 0.438 | 0.438 |
| Individuals | 285 | 285 | 285 | 285 |
| Panel C | | Waves 5 and 6 | | |
| Treatment | 0.0793*** | 0.0596*** | -0.0573** | -0.0406*** |
| | (0.026) | (0.019) | (0.024) | (0.015) |
| Post-Treatment | 0.0570** | 0.0342** | -0.0454 | -0.0306** |
| | (0.027) | (0.017) | (0.029) | (0.015) |
| Observations | 16,152 | 16,152 | 16,152 | 16,152 |
| R-squared | 0.017 | 0.003 | 0.027 | 0.003 |
| Mean of Dep. Var. | 0.411 | 0.411 | 0.271 | 0.271 |
| Std.Dev. of Dep. Var. | 0.492 | 0.492 | 0.445 | 0.445 |
| Individuals | 216 | 216 | 216 | 216 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in column 3 and 4 is an indicator for sleeping less than 6 hours. We pooled any monetary incentive in Panel A and B, while Panel C focuses on waves 5 and 6 where we only had the strong weekly incentive. The unit of observation is subject-night. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.12: Incentives and Sleep, Restricting to Weekends

| Outcomes | Sleep hrs | Sleep hrs | 7≤Sleep≤9 | 7≤Sleep≤9 | Sleep<6 | Sleep<6 |
|---|---|---|---|---|---|---|
| Treated | 0.1186 | 0.0289 | 0.0225 | -0.0003 | -0.0316 | -0.0103 |
| | (0.089) | (0.069) | (0.023) | (0.022) | (0.021) | (0.018) |
| Individual FE | NO | YES | NO | YES | NO | YES |
| Observations | 7,163 | 7,163 | 7,163 | 7,163 | 7,163 | 7,163 |
| R-squared | 0.036 | 0.006 | 0.016 | 0.003 | 0.019 | 0.006 |
| Mean of Dep. Var. | 7.038 | 7.038 | 0.437 | 0.437 | 0.240 | 0.240 |
| Std.Dev. of Dep. Var. | 1.816 | 1.816 | 0.496 | 0.496 | 0.427 | 0.427 |

*Notes -* The dependent variable in columns 1 and 2 is sleep duration (in hours). The dependent variable in columns 3 and 4 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 5 and 6 is an indicator for sleeping less than 6 hours. The analysis includes waves 1-6. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2, 4, and 6 include individual fixed effects. Standard errors are clustered at the individual level. The sample is restricted to weekends.

* p<0.10 ** p<0.05 *** p<0.01

Table A.13: Incentives and Sleep, Excluding Individuals with Low Syncing Rate (All Waves)

| | (1)<br>7≤Sleep≤9 | (2)<br>7≤Sleep≤9 | (3)<br>Sleep<6 | (4)<br>Sleep<6 |
|---|---|---|---|---|
| Panel A: Excluding individuals reporting less than 50% of the data | | | | |
| Treatment | 0.0501*** | 0.0492*** | -0.0370** | -0.0361*** |
| | (0.017) | (0.013) | (0.015) | (0.010) |
| Post-Treatment | 0.0485* | 0.0428** | -0.0412* | -0.0361*** |
| | (0.026) | (0.017) | (0.025) | (0.014) |
| Individual FE | NO | YES | NO | YES |
| Observations | 16,305 | 16,305 | 16,305 | 16,305 |
| R-squared | 0.030 | 0.003 | 0.034 | 0.003 |
| Mean of Dep. Var. | 0.433 | 0.433 | 0.248 | 0.248 |
| Std.Dev. of Dep. Var. | 0.496 | 0.496 | 0.432 | 0.432 |
| Individuals | 367 | 367 | 367 | 367 |
| Panel B: Excluding individuals reporting less than 75% of the data | | | | |
| Treatment | 0.0541*** | 0.0458*** | -0.0446** | -0.0450*** |
| | (0.021) | (0.015) | (0.017) | (0.011) |
| Post-Treatment | 0.0471 | 0.0386* | -0.0353 | -0.0397*** |
| | (0.030) | (0.020) | (0.029) | (0.015) |
| Individual FE | NO | YES | NO | YES |
| Observations | 12,432 | 12,432 | 12,432 | 12,432 |
| R-squared | 0.024 | 0.004 | 0.027 | 0.004 |
| Mean of Dep. Var. | 0.439 | 0.439 | 0.240 | 0.240 |
| Std.Dev. of Dep. Var. | 0.496 | 0.496 | 0.427 | 0.427 |
| Individuals | 271 | 271 | 271 | 271 |
| Panel C: Excluding individuals reporting less than 80% of the data | | | | |
| Treatment | 0.0569** | 0.0508*** | -0.0423** | -0.0428*** |
| | (0.022) | (0.015) | (0.018) | (0.010) |
| Post-Treatment | 0.0488 | 0.0398* | -0.0416 | -0.0405** |
| | (0.031) | (0.021) | (0.030) | (0.016) |
| Individual FE | NO | YES | NO | YES |
| Observations | 11,372 | 11,372 | 11,372 | 11,372 |
| R-squared | 0.024 | 0.003 | 0.028 | 0.004 |
| Mean of Dep. Var. | 0.436 | 0.436 | 0.243 | 0.243 |
| Std.Dev. of Dep. Var. | 0.496 | 0.496 | 0.429 | 0.429 |
| Individuals | 245 | 245 | 245 | 245 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columsn 3 and 4 is an indicator for sleeping less than 6 hours. The analysis includes waves 1-6. Panel A excludes individuals reporting less than 50% of the data. Panel B excludes individuals reporting less than 75% of the data. Panel C excludes individuals reporting less than 80% of the data. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.14: Incentives and Sleep, Excluding Individuals with Low Syncing Rate (Waves 1, 5, 6)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 7≤Sleep≤9 | 7≤Sleep≤9 | Sleep<6 | Sleep<6 |
| Panel A: Excluding individuals reporting less than 50% of the data | | | | |
| Treatment | 0.0444* | 0.0453** | -0.0493** | -0.0590*** |
| | (0.025) | (0.018) | (0.022) | (0.014) |
| Post-Treatment | 0.0536* | 0.0499** | -0.0464 | -0.0569*** |
| | (0.032) | (0.023) | (0.031) | (0.018) |
| Individual FE | NO | YES | NO | YES |
| Observations | 10,725 | 10,725 | 10,725 | 10,725 |
| R-squared | 0.028 | 0.004 | 0.039 | 0.004 |
| Mean of Dep. Var. | 0.417 | 0.417 | 0.257 | 0.257 |
| Std.Dev. of Dep. Var. | 0.493 | 0.493 | 0.437 | 0.437 |
| Individuals | 234 | 234 | 234 | 234 |
| Panel B: Excluding individuals reporting less than 75% of the data | | | | |
| Treatment | 0.0469* | 0.0367* | -0.0512** | -0.0594*** |
| | (0.028) | (0.019) | (0.024) | (0.014) |
| Post-Treatment | 0.0537 | 0.0386 | -0.0366 | -0.0488*** |
| | (0.034) | (0.023) | (0.034) | (0.019) |
| Individual FE | NO | YES | NO | YES |
| Observations | 9,104 | 9,104 | 9,104 | 9,104 |
| R-squared | 0.019 | 0.003 | 0.028 | 0.005 |
| Mean of Dep. Var. | 0.427 | 0.427 | 0.248 | 0.248 |
| Std.Dev. of Dep. Var. | 0.495 | 0.495 | 0.432 | 0.432 |
| Individuals | 191 | 191 | 191 | 191 |
| Panel C: Excluding individuals reporting less than 80% of the data | | | | |
| Treatment | 0.0489* | 0.0360* | -0.0534** | -0.0561*** |
| | (0.029) | (0.019) | (0.024) | (0.014) |
| Post-Treatment | 0.0605* | 0.0375 | -0.0461 | -0.0476** |
| | (0.035) | (0.024) | (0.034) | (0.019) |
| Individual FE | NO | YES | NO | YES |
| Observations | 8,679 | 8,679 | 8,679 | 8,679 |
| R-squared | 0.020 | 0.003 | 0.032 | 0.005 |
| Mean of Dep. Var. | 0.430 | 0.430 | 0.246 | 0.246 |
| Std.Dev. of Dep. Var. | 0.495 | 0.495 | 0.431 | 0.431 |
| Individuals | 180 | 180 | 180 | 180 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columsn 3 and 4 is an indicator for sleeping less than 6 hours. The analysis includes waves 1, 5, and 6. Panel A excludes individuals reporting less than 50% of the data. Panel B excludes individuals reporting less than 75% of the data. Panel C excludes individuals reporting less than 80% of the data. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.15: Incentives and Sleep, Replacing Missing Observations with Baseline Data

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Waves 1-6<br>$7 \leq$Sleep$\leq 9$ | Waves 1-6<br>Sleep$<6$ | Waves 1, 5, 6<br>$7 \leq$Sleep$\leq 9$ | Waves 1, 5, 6<br>Sleep$<6$ |
| Treatment | 0.021* | -0.030*** | 0.029* | -0.054*** |
| | (0.011) | (0.009) | (0.015) | (0.012) |
| Post-Treatment | 0.009 | -0.029** | 0.017 | -0.051*** |
| | (0.015) | (0.012) | (0.020) | (0.015) |
| Individual FE | YES | YES | YES | YES |
| Observations | 22,405 | 22,405 | 15,759 | 15,759 |
| R-squared | 0.003 | 0.003 | 0.003 | 0.005 |
| Mean of Dep. Var. | 0.418 | 0.252 | 0.403 | 0.257 |
| Std.Dev. of Dep. Var. | 0.493 | 0.434 | 0.491 | 0.437 |
| Individuals | 460 | 460 | 284 | 284 |

*Notes* - The dependent variable is an indicator for sleeping between 7 and 9 hours. We replace missing observations with the individual average sleep at baseline. Columns 1 and 2 include waves 1-6; columns 3 and 4 include waves 1, 5, and 6. All estimates include controls for day of the week, month fixed effects, individual fixed effects, and a weekly linear trend. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.16: Incentives and Sleep Duration, Excluding Asians

| | (1)<br>7≤Sleep≤9 | (2)<br>7≤Sleep≤9 | (3)<br>Sleep<6 | (4)<br>Sleep<6 |
|---|---|---|---|---|
| Treatment | 0.0578*** | 0.0524*** | -0.0395** | -0.0396*** |
| | (0.020) | (0.014) | (0.016) | (0.012) |
| Post-Treatment | 0.0366 | 0.0465*** | -0.0168 | -0.0362*** |
| | (0.026) | (0.017) | (0.023) | (0.014) |
| Individual FE | NO | YES | NO | YES |
| Observations | 13,261 | 13,261 | 13,261 | 13,261 |
| R-squared | 0.023 | 0.003 | 0.034 | 0.003 |
| Mean of Dep. Var. | 0.449 | 0.449 | 0.241 | 0.241 |
| Std.Dev. of Dep. Var. | 0.497 | 0.497 | 0.427 | 0.427 |
| Individuals | 336 | 336 | 336 | 336 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 3 and 4 is an indicator for sleeping less than 6 hours. We exclude Asians from this analysis. The analysis includes waves 1-6. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.
* p<0.10 ** p<0.05 *** p<0.01

Table A.17: Incentives and Sleep Duration, Weighting by Race

| | (1)<br>7≤Sleep≤9 | (2)<br>7≤Sleep≤9 | (3)<br>Sleep<6 | (4)<br>Sleep<6 |
|---|---|---|---|---|
| Treatment | 0.0566*** | 0.0508*** | -0.0367** | -0.0390*** |
| | (0.019) | (0.013) | (0.016) | (0.012) |
| Post-Treatment | 0.0330 | 0.0381** | -0.0174 | -0.0342*** |
| | (0.024) | (0.016) | (0.022) | (0.013) |
| Individual FE | NO | YES | NO | YES |
| Observations | 18,413 | 18,413 | 18,413 | 18,413 |
| R-squared | 0.023 | 0.003 | 0.032 | 0.003 |
| Mean of Dep. Var. | 0.429 | 0.429 | 0.258 | 0.258 |
| Std.Dev. of Dep. Var. | 0.495 | 0.495 | 0.438 | 0.438 |
| Individuals | 460 | 460 | 460 | 460 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 3 and 4 is an indicator for sleeping less than 6 hours. The analysis includes waves 1-6. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. All estimates include weights to ensure representativeness by race with respect to the US college population. Standard errors clustered at the individual level are reported in parentheses.
* p<0.10 ** p<0.05 *** p<0.01

Table A.18: Incentives and Sleep Duration, Excluding Never Sleep Deprived at Baseline

| | (1)<br>7≤Sleep≤9 | (2)<br>7≤Sleep≤9 | (3)<br>Sleep<6 | (4)<br>Sleep<6 |
|---|---|---|---|---|
| Treatment | 0.0692*** | 0.0553*** | -0.0354** | -0.0301*** |
| | (0.018) | (0.012) | (0.015) | (0.010) |
| Post-Treatment | 0.0448** | 0.0368** | -0.0214 | -0.0267** |
| | (0.023) | (0.015) | (0.021) | (0.012) |
| Individual FE | NO | YES | NO | YES |
| Observations | 17,843 | 17,843 | 17,843 | 17,843 |
| R-squared | 0.020 | 0.004 | 0.020 | 0.005 |
| Mean of Dep. Var. | 0.437 | 0.437 | 0.248 | 0.248 |
| Std.Dev. of Dep. Var. | 0.496 | 0.496 | 0.432 | 0.432 |
| Individuals | 418 | 418 | 418 | 418 |

*Notes* - The dependent variable in columns 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 3 and 4 is an indicator for sleeping less than 6 hours. We exclude those who were never sleep-deprived at baseline in this analysis. The analysis includes waves 1-6. All estimates include controls for gender, dummies of the week, month fixed effects and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.19: Incentives and Sleep Duration, Goal-Setting (Wave 6)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $7 \leq Sleep \leq 9$ | $7 \leq Sleep \leq 9$ | $7 \leq Sleep \leq 9$ | $7 \leq Sleep \leq 9$ |
| Goal Setting vs Control | 0.0498 | 0.0114 | | |
| | (0.042) | (0.030) | | |
| Post-Goal Setting vs Control | 0.0779* | 0.0363 | | |
| | (0.045) | (0.034) | | |
| Treatment vs Goal Setting | | | 0.0205 | 0.0533*** |
| | | | (0.032) | (0.020) |
| Post-Treatment vs Post-Goal Setting | | | -0.0148 | 0.0248 |
| | | | (0.038) | (0.022) |
| Individual FE | NO | YES | NO | YES |
| Observations | 6,721 | 6,721 | 9,746 | 9,746 |
| R-squared | 0.017 | 0.003 | 0.019 | 0.003 |
| Mean of Dep. Var. | 0.413 | 0.413 | 0.422 | 0.422 |
| Std.Dev. of Dep. Var. | 0.492 | 0.492 | 0.494 | 0.494 |
| Individuals | 96 | 96 | 141 | 141 |

*Notes* - The dependent variable is an indicator for sleeping between 7 and 9 hours. Columns 1 and 2 compare participants who were invited to set a sleep duration and bedtime goal without monetary incentives to participants in the control group. Columns 3 and 4 compare participants in the monetary incentive treatment to those invited to set bedtime and sleep duration goals without incentive. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.20: Incentive and Sleep: Persistence of the Intervention (Wave 6)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 7≤Sleep≤9 | 7≤Sleep≤9 | 7≤Sleep≤9 | 7≤Sleep≤9 |
| Treatment | 0.0635** | 0.0598*** | 0.0693** | 0.0657*** |
| (weeks 3-5) | (0.031) | (0.020) | (0.031) | (0.020) |
| Post-treatment | 0.0525 | 0.0453** | | |
| (weeks 6-15) | (0.035) | (0.023) | | |
| Post-Treatment | | | 0.0610* | 0.0556** |
| (weeks 6-8) | | | (0.034) | (0.022) |
| Post-Treatment | | | 0.0562* | 0.0495** |
| (weeks 9-11) | | | (0.031) | (0.022) |
| Post-Treatment | | | -0.0166 | -0.0190 |
| (weeks 12-15) | | | (0.035) | (0.023) |
| Individual FE | NO | YES | NO | YES |
| Observations | 9,675 | 9,675 | 9,675 | 9,675 |
| R-squared | 0.022 | 0.004 | 0.023 | 0.005 |
| Mean of Dep. Var. | 0.405 | 0.405 | 0.405 | 0.405 |
| Std.Dev. of Dep. Var. | 0.491 | 0.491 | 0.491 | 0.491 |
| Individuals | 141 | 141 | 141 | 141 |

*Notes* - The dependent variable is an indicator for sleeping between 7 and 9 hours. The results document the persistence of the effect in wave 6 after the removal of the monetary incentive. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.21: Incentives and Sleep Duration, Including Trends

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | 7≤Sleep≤9 | | | Sleep<6 | |
| Treatment | 0.0547*** | 0.0546*** | 0.0528*** | -0.0330*** | -0.0328*** | -0.0398*** |
| | (0.012) | (0.012) | (0.013) | (0.010) | (0.010) | (0.012) |
| Post-Treatment | 0.0268* | 0.0348** | 0.0317** | -0.0166 | -0.0251** | -0.0221* |
| | (0.014) | (0.015) | (0.016) | (0.012) | (0.012) | (0.013) |
| Linear Weekly Trend | NO | YES | NO | NO | YES | NO |
| Week FE | NO | NO | YES | NO | NO | YES |
| Observations | 18,413 | 18,413 | 18,413 | 18,413 | 18,413 | 18,413 |
| R-squared | 0.004 | 0.004 | 0.004 | 0.004 | 0.005 | 0.005 |
| Mean of Dep. Var. | 0.429 | 0.429 | 0.429 | 0.258 | 0.258 | 0.258 |
| Std.Dev. of Dep. Var. | 0.495 | 0.495 | 0.495 | 0.438 | 0.438 | 0.438 |
| Individuals | 460 | 460 | 460 | 460 | 460 | 460 |

*Notes* - The dependent variable in columns 1-3 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 4-6 is an indicator for sleeping less than 6 hours. The analysis includes waves 1-6. The unit of observation is subject-night. All estimates include month, day of the week, and individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.
* p<0.10 ** p<0.05 *** p<0.01

Table A.22: Incentives and Sleep Quality

| | (1) Efficiency | (2) Efficiency | (3) Deep sleep | (4) Deep sleep | (5) REM sleep | (6) REM sleep |
|---|---|---|---|---|---|---|
| Waves: | 1-6 | 1-6 | 6 (minutes) | 6 (minutes) | 6 (minutes) | 6 (minutes) |
| Treatment | -0.0419 | -0.1537 | 0.5576 | 1.4523 | 0.6462 | 1.6500 |
| | (0.425) | (0.245) | (2.024) | (1.124) | (2.780) | (1.583) |
| Post-Treatment | 0.1330 | 0.1330 | 0.5880 | 2.2856** | 1.3713 | 2.7488* |
| | (0.663) | (0.353) | (2.308) | (1.051) | (3.229) | (1.390) |
| Individual FE | NO | YES | NO | YES | NO | YES |
| Observations | 18,413 | 18,413 | 9,243 | 9,243 | 9,243 | 9,243 |
| R-squared | 0.046 | 0.017 | 0.026 | 0.007 | 0.024 | 0.004 |
| Mean of Dep. Var. | 92.73 | 92.73 | 76.62 | 76.62 | 89.82 | 89.82 |
| Std.Dev. of Dep. Var. | 8.153 | 8.153 | 27.54 | 27.54 | 35.46 | 35.46 |
| Individuals | 460 | 460 | 141 | 141 | 141 | 141 |

*Notes* - The dependent variables are sleep efficiency measured as the ratio between sleep duration and time spent in bed (columns 1 and 2); minutes of deep sleep (columns 3 and 4); and minutes of REM sleep (columns 5 and 6). Columns 1-2 include the entire sample; in columns 3-6, the sample is restricted to wave 6 for which these data are available. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1, 3, and 5 include dummies for gender and age. Columns 2, 4, and 6 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

## Table A.23: Incentives to Sleep and Naps

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Nap | Nap duration | $7 \leq$ Sleep $\leq 9$ | $7 \leq$ Sleep $\leq 9$ | Sleep $<6$ | Sleep $<6$ |
| **Panel A: Waves 1-6** | | | | | | |
| Treatment | -0.0017 | -0.0009 | 0.0541*** | 0.0544*** | -0.0318*** | -0.0325*** |
| | (0.002) | (0.002) | (0.012) | (0.012) | (0.010) | (0.010) |
| Post-Treatment | -0.0023 | -0.0028 | 0.0341** | 0.0342** | -0.0238** | -0.0240** |
| | (0.002) | (0.003) | (0.015) | (0.015) | (0.012) | (0.012) |
| Nap | | | YES | | YES | |
| Nap duration | | | | YES | | YES |
| Individual FE | YES | YES | YES | YES | YES | YES |
| Observations | 18,413 | 18,413 | 18,413 | 18,413 | 18,413 | 18,413 |
| R-squared | 0.005 | 0.004 | 0.006 | 0.006 | 0.017 | 0.016 |
| Mean of Dep. Var. | 0.006 | 0.008 | 0.429 | 0.429 | 0.258 | 0.258 |
| Std.Dev. of Dep. Var. | 0.078 | 0.109 | 0.495 | 0.495 | 0.438 | 0.438 |
| Individuals | 460 | 460 | 460 | | 460 | 460 |
| **Panel B: Waves 1,5, and 6** | | | | | | |
| Treatment | 0.0009 | 0.0023 | 0.0536*** | 0.0538*** | -0.0426*** | -0.0429*** |
| | (0.003) | (0.004) | (0.017) | (0.017) | (0.015) | (0.015) |
| Post-Treatment | -0.0018 | -0.0030 | 0.0384** | 0.0384** | -0.0346** | -0.0345** |
| | (0.003) | (0.004) | (0.019) | (0.019) | (0.016) | (0.016) |
| Nap | | | YES | | YES | |
| Nap duration | | | | YES | | YES |
| Individual FE | YES | YES | YES | YES | YES | YES |
| Observations | 13,460 | 13,460 | 13,460 | 13,460 | 13,460 | 13,460 |
| R-squared | 0.002 | 0.002 | 0.006 | 0.006 | 0.015 | 0.014 |
| Mean of Dep. Var. | 0.005 | 0.007 | 0.419 | 0.419 | 0.264 | 0.264 |
| Std.Dev. of Dep. Var. | 0.073 | 0.103 | 0.493 | 0.493 | 0.441 | 0.441 |
| Individuals | 285 | 285 | 285 | 285 | 285 | 285 |
| **Panel C: Waves 5 and 6** | | | | | | |
| Treatment | 0.0004 | 0.0002 | 0.0577*** | 0.0576*** | -0.0457*** | -0.0456*** |
| | (0.003) | (0.004) | (0.018) | (0.018) | (0.017) | (0.017) |
| Post-Treatment | -0.0014 | -0.0028 | 0.0402* | 0.0401* | -0.0323* | -0.0320* |
| | (0.003) | (0.004) | (0.020) | (0.020) | (0.017) | (0.017) |
| Nap | | | YES | | YES | |
| Nap duration | | | | YES | | YES |
| Individual FE | YES | YES | YES | YES | YES | YES |
| Observations | 11,623 | 11,623 | 11,623 | 11,623 | 11,623 | 11,623 |
| R-squared | 0.002 | 0.002 | 0.005 | 0.005 | 0.013 | 0.012 |
| Mean of Dep. Var. | 0.005 | 0.007 | 0.405 | 0.405 | 0.276 | 0.276 |
| Std.Dev. of Dep. Var. | 0.072 | 0.102 | 0.491 | 0.491 | 0.447 | 0.447 |
| Individuals s | 216 | 216 | 216 | 216 | 216 | 216 |

*Notes -* The dependent variable in column 1 is an indicator for whether a participant took a nap (any sleep shorter than 2 hours between 7am and 7pm). The dependent variable in column 2 is nap duration. The dependent variable in columns 3 and 4 is an indicator for sleeping between 7 and 9 hours. The dependent variable in columns 5 and 6 is an indicator for sleeping less than 6 hours. This analysis includes weekends. All estimates include controls for day of the week fixed effects, a linear weekly trend, and individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.24: Baseline Characteristics and Sample Attrition in Time-Use Survey

| | | |
|---|---|---|
| Female | -0.036 | (0.055) |
| Age | 1.202** | (0.507) |
| White | -0.178*** | (0.056) |
| Black | 0.046 | (0.036) |
| Asian | 0.008 | (0.048) |
| Other Race Category | 0.125*** | (0.045) |
| Poor health | -0.030 | (0.030) |
| Satisfied with life | -0.062 | (0.054) |
| Depressed | 0.082 | (0.057) |
| Ever smoked | 0.099** | (0.050) |
| Drinks more than once a week | -0.010 | (0.050) |
| BMI | -0.091 | (0.770) |
| Overweight | 0.009 | (0.052) |
| Obese | 0.056* | (0.033) |
| Ideal sleep | -0.194** | (0.091) |
| Sleep duration (self-repoted) | 0.052 | (0.182) |
| Sleep quality (self-reported) | -0.055 | (0.200) |
| Sleep hours | 0.069 | (0.538) |

*Notes -* Data are drawn from the Day 1 Survey. The analysis includes waves 1-6. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual did not respond to the time-use survey.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.25: Incentives and Sleep Regularity (Weekly-Level Analysis)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Std.Dev. Sleep Hours | | Std.Dev. Bedtime | | Std.Dev. Wake up time | |
| Treatment | -0.0753 | -0.2154*** | -0.0727 | -0.1780*** | 0.0346 | -0.0368 |
| | (0.052) | (0.045) | (0.056) | (0.053) | (0.049) | (0.038) |
| Post-Treatment | -0.0754 | -0.1867*** | -0.1258** | -0.2163*** | -0.0271 | -0.1074*** |
| | (0.051) | (0.044) | (0.052) | (0.048) | (0.046) | (0.037) |
| Individual FE | NO | YES | NO | YES | NO | YES |
| Observations | 4,215 | 4,215 | 4,215 | 4,215 | 4,215 | 4,215 |
| R-squared | 0.024 | 0.015 | 0.035 | 0.013 | 0.032 | 0.006 |
| Mean of Dep. Var. | 1.389 | 1.389 | 1.340 | 1.340 | 1.205 | 1.205 |
| Std.Dev. of Dep. Var. | 0.787 | 0.787 | 0.816 | 0.816 | 0.674 | 0.674 |
| Individuals | 460 | 460 | 460 | 460 | 460 | 460 |

*Notes -* The dependent variable in column 1 and 2 is the weekly standard deviation of sleep hours. The analysis includes waves 1-6. The dependent variable in column 3 and 4 is the weekly standard deviation of bedtime. The dependent variable in column 5 and 6 is the weekly standard deviation of wake-up time. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1, 3, and 5 include dummies for gender and age.Columns 2, 4, and 6 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.26: Incentive Structure and Sleep

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 7≤Sleep≤9 | 7≤Sleep≤9 | Sleep<6 | Sleep<6 |
| Panel A: Weekly, Waves 2, 3, 5 and 6 | | | | |
| Treatment | 0.0796*** | 0.0513*** | -0.0448** | -0.0207 |
| | (0.020) | (0.014) | (0.019) | (0.013) |
| Post-Treatment | 0.0477* | 0.0296* | -0.0234 | -0.0155 |
| | (0.025) | (0.017) | (0.024) | (0.013) |
| Individual FE | NO | YES | NO | YES |
| Observations | 14,307 | 14,307 | 14,307 | 14,307 |
| R-squared | 0.025 | 0.004 | 0.033 | 0.004 |
| Mean of Dep. Var. | 0.422 | 0.422 | 0.266 | 0.266 |
| Std.Dev. of Dep. Var. | 0.494 | 0.494 | 0.442 | 0.442 |
| Individuals | 310 | 310 | 310 | 310 |
| Panel B: Biweekly, Waves 1 and 4 | | | | |
| Treatment | 0.0470* | 0.0579*** | -0.0510*** | -0.0585*** |
| | (0.024) | (0.021) | (0.018) | (0.017) |
| Post-Treatment | 0.0144 | 0.0426 | -0.0433 | -0.0673** |
| | (0.041) | (0.038) | (0.033) | (0.031) |
| Individual FE | NO | YES | NO | YES |
| Observations | 4,106 | 4,106 | 4,106 | 4,106 |
| R-squared | 0.035 | 0.008 | 0.043 | 0.011 |
| Mean of Dep. Var. | 0.454 | 0.454 | 0.232 | 0.232 |
| Std.Dev. of Dep. Var. | 0.498 | 0.498 | 0.422 | 0.422 |
| Individuals | 150 | 150 | 150 | 150 |

*Notes -* The dependent variable in column 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in column 3 and 4 is an indicator for sleeping less than 6 hours. Panel A restricts the analysis to waves in which the monetary incentive was weekly. Panel B restricts the analysis to waves in which the incentive was biweekly. All estimates include controls for day of the week, month fixed effects, and a weekly linear trend. Columns 1 and 3 include dummies for gender and age. Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.27: Incentives and Sleep: the Role of the Size of the Financial Incentive

| | Weak Incentive (Wave 1) | | Strong-Incentive (Waves 2-6) | |
|---|---|---|---|---|
| | $7 \leq$ Sleep $\leq 9$ | Sleep $< 6$ | $7 \leq$ Sleep $\leq 9$ | Sleep $< 6$ |
| Treatment | 0.0247 | -0.0375 | 0.0569*** | -0.0311*** |
| | (0.041) | (0.034) | (0.012) | (0.011) |
| Post Treatment | 0.0135 | -0.0681 | 0.0330** | -0.0211* |
| | (0.066) | (0.049) | (0.015) | (0.012) |
| Individual FE | YES | YES | YES | YES |
| Observations | 1,837 | 1,837 | 16,576 | 16,576 |
| R-squared | 0.008 | 0.017 | 0.004 | 0.004 |
| Mean of Dep. Var. | 0.502 | 0.193 | 0.494 | 0.265 |
| Std.Dev. of Dep. Var. | 0.500 | 0.395 | 0.421 | 0.442 |
| Individuals | 69 | 69 | 391 | 391 |

*Notes -* Columns 1 and 2 restrict the analysis to the first wave where we use the weak monetary incentive. Columns 3 and 4 restrict the analysis to the other waves in which we used a stronger monetary incentive. The dependent variable in column 1 and 3 is an indicator for sleeping between 7 and 9 hours. The dependent variable in column 2 and 4 is an indicator for sleeping less than 6 hours. All estimates include day of the week, month fixed effects, individual fixed effects and a weekly linear trend. Standard errors clustered at the individual level are reported in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.28: Time Preferences and Sleep Duration at Baseline

| | (1) Sleep hrs | (2) Sleep hrs | (3) $7 \leq$ Sleep $\leq 9$ | (4) $7 \leq$ Sleep $\leq 9$ |
|---|---|---|---|---|
| Present-biased | -0.1036 | | -0.0692 | |
| | (0.171) | | (0.053) | |
| Impatient | | 0.1048 | | 0.0122 |
| | | (0.174) | | (0.054) |
| Observations | 487 | 487 | 487 | 487 |
| R-squared | 0.149 | 0.149 | 0.123 | 0.120 |
| Mean of Dep. Var. | 6.906 | 6.906 | 0.425 | 0.425 |
| Std.Dev. of Dep. Var. | 1.595 | 1.595 | 0.495 | 0.495 |

*Notes* - The dependent variable in columns 1 and 2 is sleep duration (in hours). The dependent variable in columns 3 and 4 is an indicator for whether the participant slept between 7 and 9 hours. The analysis includes waves 1-6. Data for time and risk preferences are drawn from the first-day survey. Data on sleep are drawn from fitbit data collected in first two weeks of the experiment and before the intervention. All estimates include controls for gender, age, race, and wave fixed effects. Robust standard errors are in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.29: Present-Bias, Impatience, Overconfidence and Commitment

| | (1) Bedtime target: earlier than 1am | (2) Sleep target>7hrs | (3) Both |
|---|---|---|---|
| Panel A: | | | |
| Present-biased | 0.1147 | -0.0412 | -0.0434 |
| | (0.072) | (0.071) | (0.072) |
| Observations | 284 | 284 | 284 |
| R-squared | 0.114 | 0.135 | 0.125 |
| Mean of Dep. Var. | 0.581 | 0.595 | 0.419 |
| Std.Dev. of Dep. Var. | 0.494 | 0.492 | 0.494 |
| Panel B: | | | |
| Impatient | 0.1601** | 0.0228 | 0.1337* |
| | (0.069) | (0.071) | (0.072) |
| Observations | 284 | 284 | 284 |
| R-squared | 0.122 | 0.134 | 0.136 |
| Mean of Dep. Var. | 0.581 | 0.595 | 0.419 |
| Std.Dev. of Dep. Var. | 0.494 | 0.492 | 0.494 |
| Panel C: | | | |
| Risk-averse | -0.1798** | -0.0928 | -0.1676** |
| | (0.083) | (0.087) | (0.079) |
| Observations | 284 | 284 | 284 |
| R-squared | 0.121 | 0.138 | 0.137 |
| Mean of Dep. Var. | 0.581 | 0.595 | 0.419 |
| Std.Dev. of Dep. Var. | 0.494 | 0.492 | 0.494 |

*Notes -* The dependent variable in column 1 is an indicator for whether the participants in the incentive treatment selected a bedtime target earlier than 1 am. The dependent variable in column 2 is an indicator for whether the participants in the incentive treatment selected a sleep duration target longer than 7 hours. The dependent variable in column 3 is an indicator for whether the participant selected restrictive targets for both bedtime and sleep. The analysis includes waves 1-6. Data are drawn from Fitbit data, weekly surveys collected during the weeks of the intervention, and the first-day survey. The sample is restricted to subjects in the treatment group. All estimates include controls for age, gender, ethnicity, race, wave fixed effects and baseline sleep duration. Robust standard errors are in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A.30: Survey Time and Early Bedtime Target

| | (1) OLS | (2) Reduced-Form | (3) 2SLS | (4) 2SLS |
|---|---|---|---|---|
| Response before 3pm | 0.0800** | | 0.1848** | 0.1589* |
| | (0.038) | | (0.089) | (0.088) |
| Early email | | 0.0859** | | |
| | | (0.042) | | |
| Controls | NO | NO | NO | YES |
| Observations | 853 | 853 | 853 | 853 |
| R-squared | 0.028 | 0.028 | 0.018 | 0.057 |
| Mean of Dep. Var. | 0.464 | 0.464 | 0.464 | 0.464 |
| Std.Dev. of Dep. Var. | 0.499 | 0.499 | 0.499 | 0.499 |

*Notes -* The dependent variable is whether a subject selected a bedtime target earlier than 1am. The analysis includes waves 1-6. Data are drawn from the weeks of treatment. In column 4, we include controls for age, gender, race, and wave fixed effects. The unit of observation is subject-week during the weeks of treatment. The unit of analysis is subject-week. Standard errors clustered at the individual level are reported in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.31: Overconfidence and Sleep Duration (Self-reported vs Fitbit, 7≤Sleep≤9)

| | Self-reported (Day 1 Survey) | | | Actual Sleep (Fitbit) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Sleep hours | 7≤Sleep≤9 | Sleep<6 | Sleep hours | 7≤Sleep≤9 | Sleep<6 |
| Overconfident | 0.7947*** | 0.5940*** | -0.1411*** | -0.5661*** | -0.5729*** | 0.0790* |
| | (0.079) | (0.031) | (0.022) | (0.177) | (0.031) | (0.044) |
| Observations | 486 | 486 | 486 | 486 | 486 | 486 |
| R-squared | 0.191 | 0.342 | 0.101 | 0.170 | 0.366 | 0.141 |
| Mean of Dep. Var. | 6.810 | 0.572 | 0.103 | 6.902 | 0.428 | 0.173 |
| Std.Dev. of Dep. Var. | 0.999 | 0.495 | 0.304 | 1.591 | 0.495 | 0.378 |

*Notes -* Data are drawn from the first-day survey (columns 1-3) and the Fitbit data for the first two weeks of the experiment before intervention (columns 4-6). The dependent variables are self-reported measures of sleep hours (column 1), sleeping less than 7 hours (column 2), and sleeping between 7 and 9 hours (column 3). In columns 4-6, the dependent variables are the the same but as measured by Fitbit data. The analysis includes waves 1-6. We defined overconfident subjects who reported to sleep more than 7 hours and less than 9, but did not. All estimates include controls for gender, age, race and ethnicity, and wave fixed effects. Robust standard errors are in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A.32: Perceived own and other's sleep quality and sleep deprivation risks

| | Own | | Others | | Difference | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev |
| Sleep quality | 6.64 | 1.58 | 5.86 | 1.31 | 0.78 | 1.77 |
| Sleep duration | 6.84 | 1.26 | 6.45 | 1.15 | 0.40 | 1.28 |
| Risks associated with sleep deprivation: | | | | | | |
| Mental alertness (1-100) | 26.64 | 12.81 | 59.98 | 23.28 | -33.42 | 24.11 |
| Weight gain (1-100) | 42.62 | 24.19 | 52.86 | 21.93 | -10.27 | 22.31 |
| Insomnia (1-100) | 26.37 | 19.23 | 37.85 | 21.83 | -11.74 | 19.98 |
| Getting a cold (1-100) | 38.17 | 22.92 | 45.77 | 24.52 | -7.84 | 21.23 |
| Arterial (1-100) | 33.45 | 22.06 | 36.31 | 22.22 | -2.53 | 19.15 |
| Average risk | 33.34 | 13.81 | 46.54 | 16.96 | -13.15 | 13.93 |
| Observations | 508 | 508 | 508 | 508 | 508 | 508 |

*Notes -* These variables are all measured in the baseline survey collected on the first day of the experiment upon enrollment. The analysis includes waves 1-6. Columns 1-2 report summary statistics for own sleep duration, quality, and risks associated with sleep. Columns 3-4 report summary statistics for others' sleep duration, quality, and risks associated with sleep, where others are defined as individuals in the same age and gender group. Columns 5-6 report the difference between perceptions about own and others' sleep duration, sleep quality, and risks associated with sleep deprivation.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

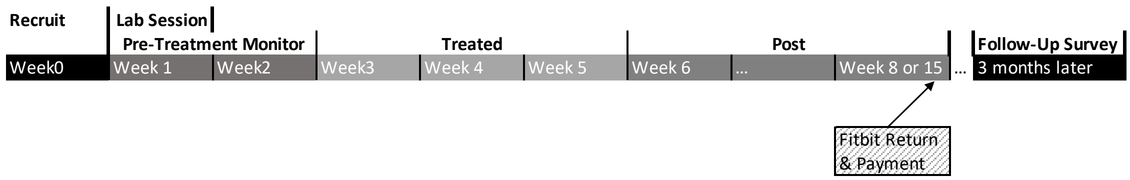
Table A.33: Demand for Committment and Overconfidence

| | (1) Bedtime target: earlier than 1am | (2) Sleep target>7hrs | (3) Both |
|---|---|---|---|
| Panel A: Reported Sleep vs Actual Sleep | | | |
| Overconfident | -0.0892 | -0.1012 | -0.0992 |
| (sleep duration) | (0.070) | (0.069) | (0.068) |
| Observations | 284 | 284 | 284 |
| R-squared | 0.097 | 0.089 | 0.077 |
| Mean of Dep. Var. | 0.580 | 0.594 | 0.417 |
| Std.Dev. of Dep. Var. | 0.495 | 0.492 | 0.494 |
| Panel B: Index of Overconfidence | | | |
| Overconfident | -0.1129 | 0.0551 | -0.0330 |
| (sleep duration, quality, and risk) | (0.098) | (0.090) | (0.097) |
| Observations | 284 | 284 | 284 |
| R-squared | 0.097 | 0.083 | 0.070 |
| Mean of Dep. Var. | 0.581 | 0.595 | 0.419 |
| Std.Dev. of Dep. Var. | 0.494 | 0.492 | 0.494 |

*Notes* - The dependent variable in column 1 is an indicator for whether the participants in the incentive treatment selected a bedtime target earlier than 1 am. The dependent variable in column 2 is an indicator for whether the participants in the incentive treatment selected a sleep duration target longer than 7 hours. The dependent variable in column 3 is an indicator for whether the participant selected restrictive targets for both bedtime and sleep. The analysis includes waves 1-6. In Panel A, we defined as overconfident all participants who in the first two weeks of the experiment (baseline) slept at least 30 minutes less than what they self-reported in the first-day survey. In Panel B, we defined as overconfident those in the upper quartile of an index built on the over-placement of own sleep duration, sleep quality, and sleep risks with respect to others in the same age and gender group. All estimates include controls for gender, age, race and ethnicity, and wave fixed effects. Robust standard errors are in parentheses.

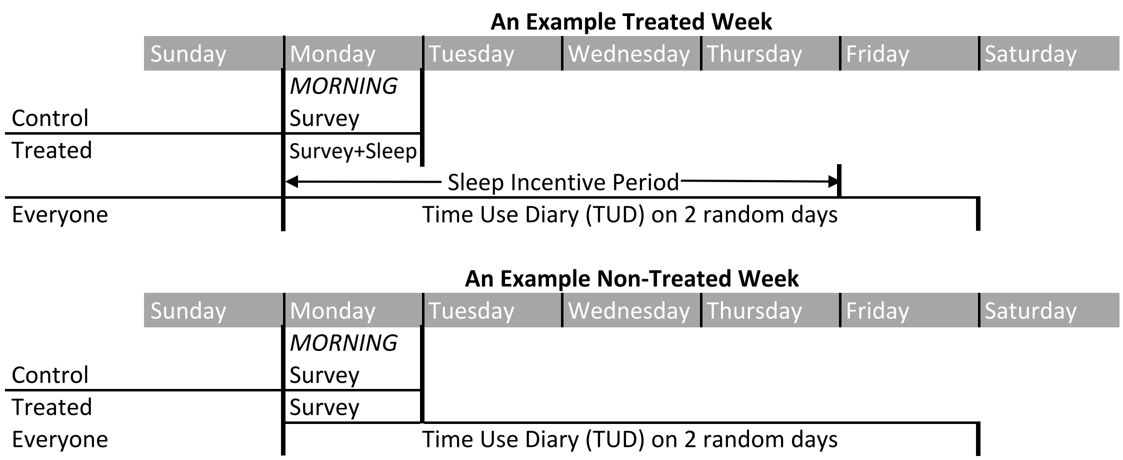
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Figure A.1: Design Illustration



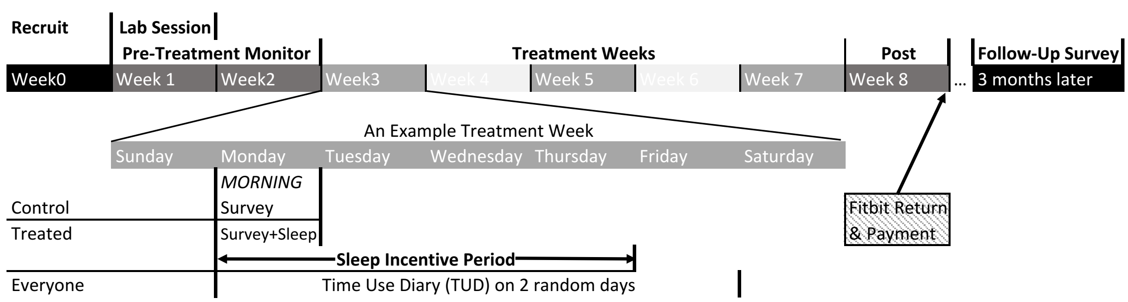

*Notes -* The above figure describes the timeline of our experiment for individuals in the Weekly-Incentive treatment.

Figure A.2: Treatment Week



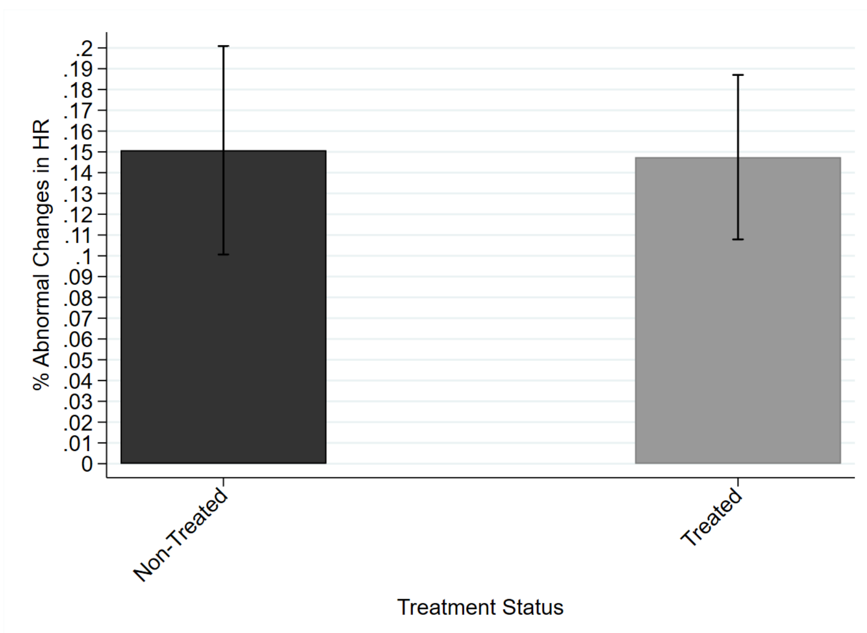

*Notes -* The above figure describes the typical week timeline during the experiment.

Figure A.3: Design Illustration: Biweekly Intervention



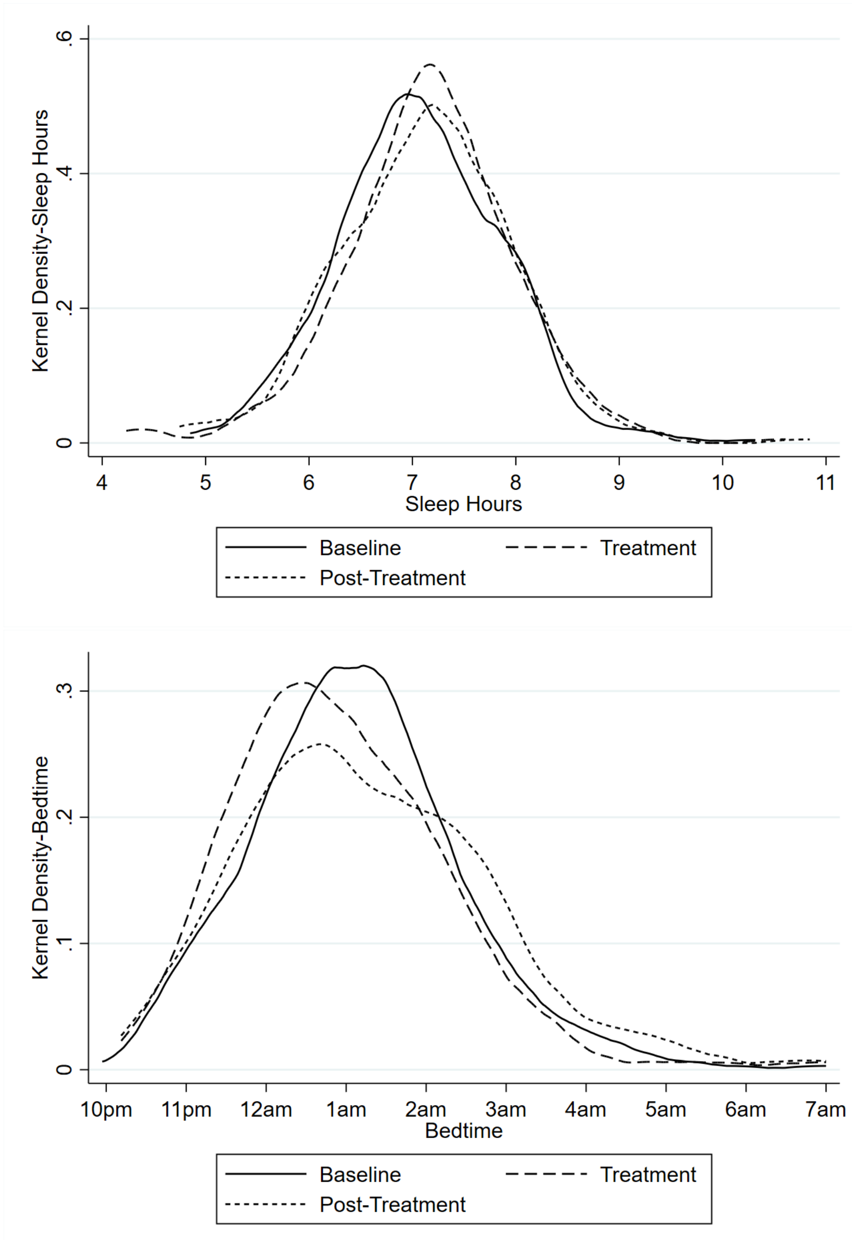

*Notes -* The above figure describes the timeline of our experiment for individuals in the Biweekly-Incentive treatment.

Figure A.4: % of participants with abnormal changes in heart rate between baseline and treatment



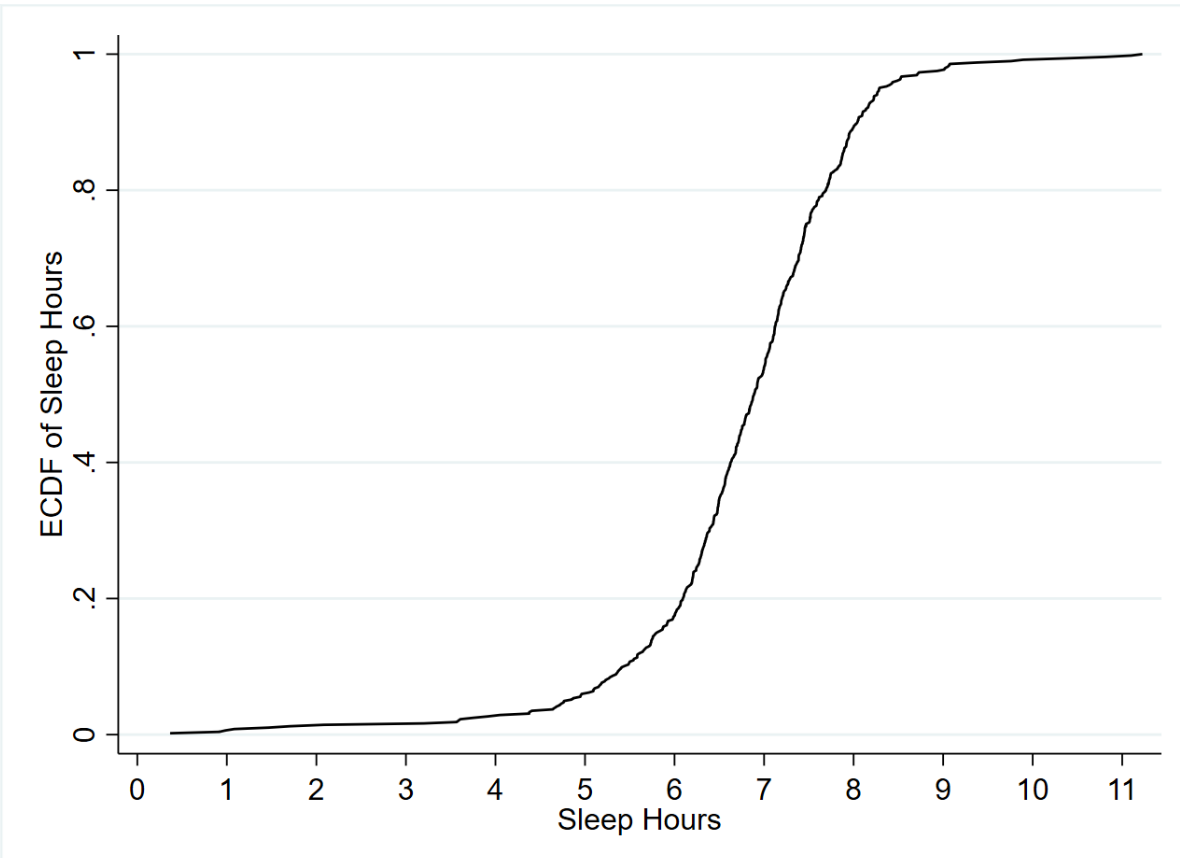

*Notes* - The figure above reports the share of individuals experiencing abnormal changes in heart rate between treatment period and baseline with 95% confidence intervals. This analysis includes waves 1-6.

Figure A.5: Sleep duration and bedtime distribution before and after treatment



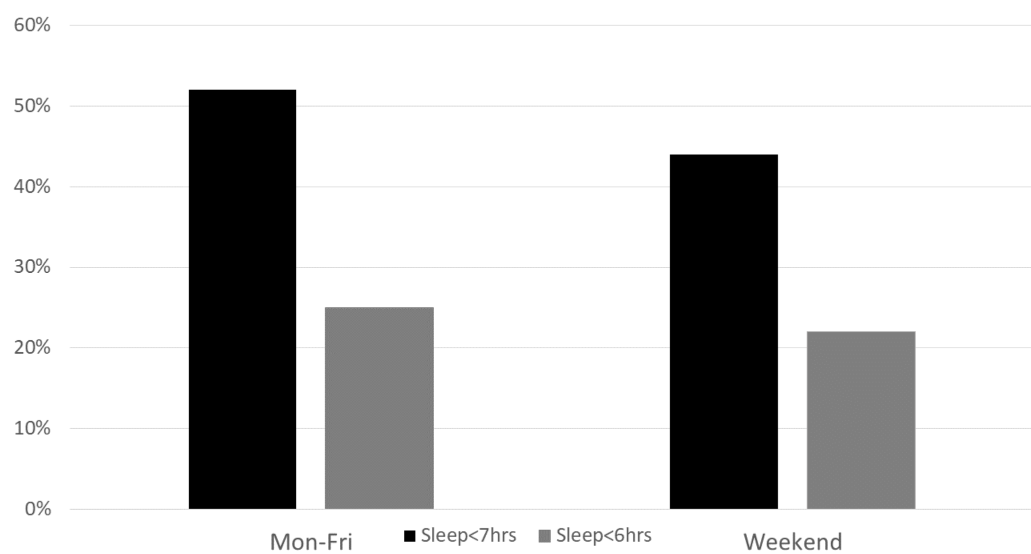

*Notes -* Data are drawn from the Fitbit devices. The top (bottom) figure plots the distribution of sleep duration (bedtime) before, during, and after treatment. In the figures, we pooled all incentive treatments in waves 1-6. Weekends were excluded from this figure.

Figure A.6: Sleep duration before intervention



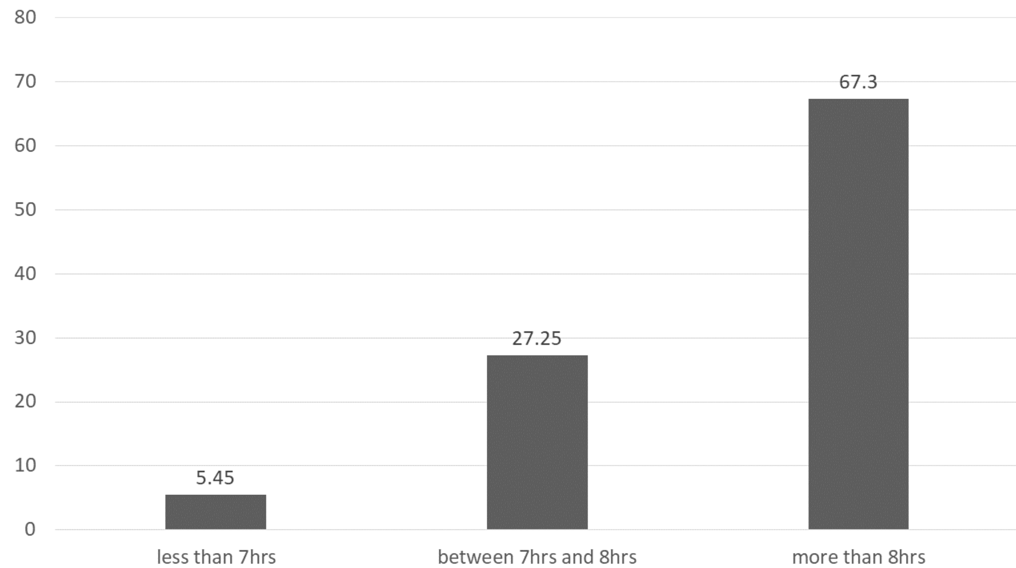

*Notes* - The figure plots the cumulative distribution function of sleep hours in our sample at baseline. Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention. Weekends were excluded from this figure. This analysis includes waves 1-6.

Figure A.7: Sleep Duration over the Week



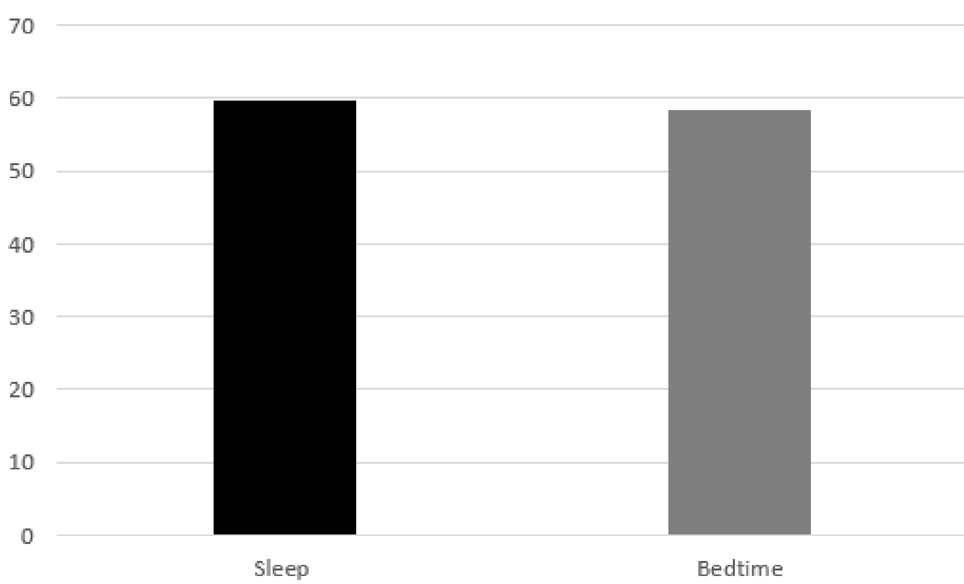

*Notes* - The figure describes the proportion of subjects sleeping less than 7(6) hours between Monday-Friday and the weekend. Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.This analysis includes waves 1-6.

Figure A.8: Ideal Sleep Duration at Baseline



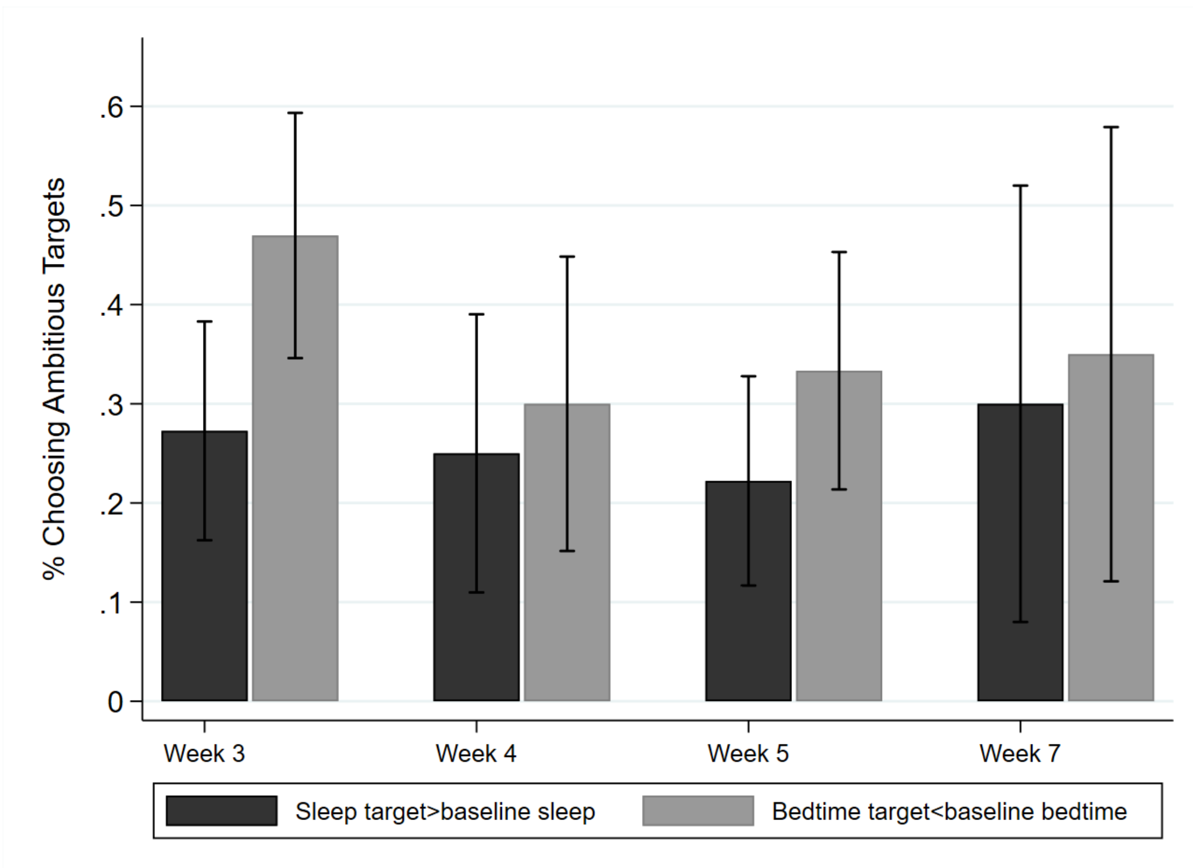

*Notes -* The figure reports the share of reporting less than 7 hours, between 7and 8 hours, and more than 8 hours as ideal sleep duration. Data were collected during the first-day session. This analysis includes waves 1-6.

Figure A.9: Demand for a More Restrictive Target



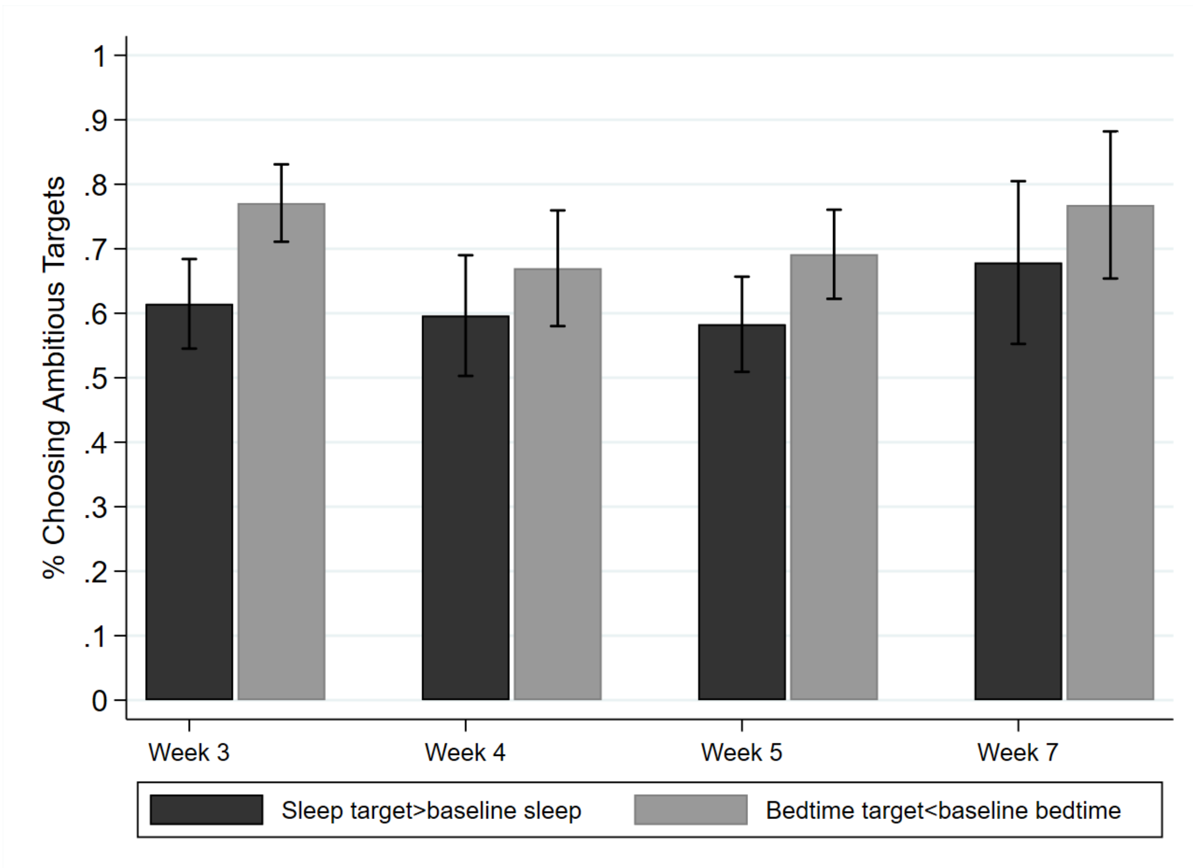

*Notes -* The figure reports the proportion of subjects choosing a restrictive sleep duration (longer than 7 hours) or bedtime (earlier than 1am) target in at least one of the weeks of the intervention.This analysis includes waves 1-6.

Figure A.10: Share of participants choosing a restrictive target with respect to their baseline sleep duration and bedtime habits



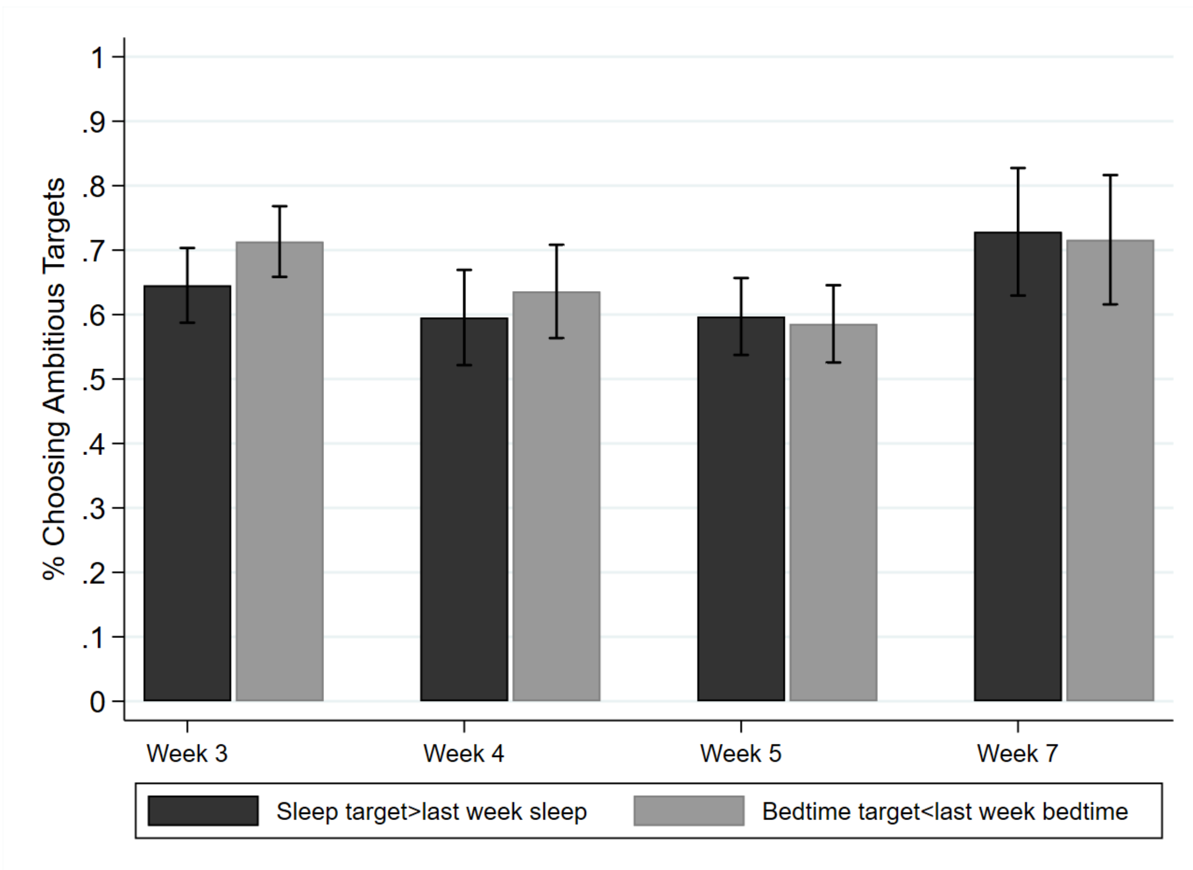

*Notes -* The figure reports the share of participants choosing a restrictive target with respect to their baseline bedtime and sleep habits with 95% confidence intervals. This analysis includes waves 1-6. In waves 1 and 4, treated weeks were week 3, 5, and 7. In waves 2, 3, 5, and 6 treated weeks were week 3, 4, and 5.

Figure A.11: Share of participants choosing a restrictive target with respect to their average sleep duration and bedtime at baseline among those choosing bedtime earlier than 1am and sleep duration longer than 7 hours



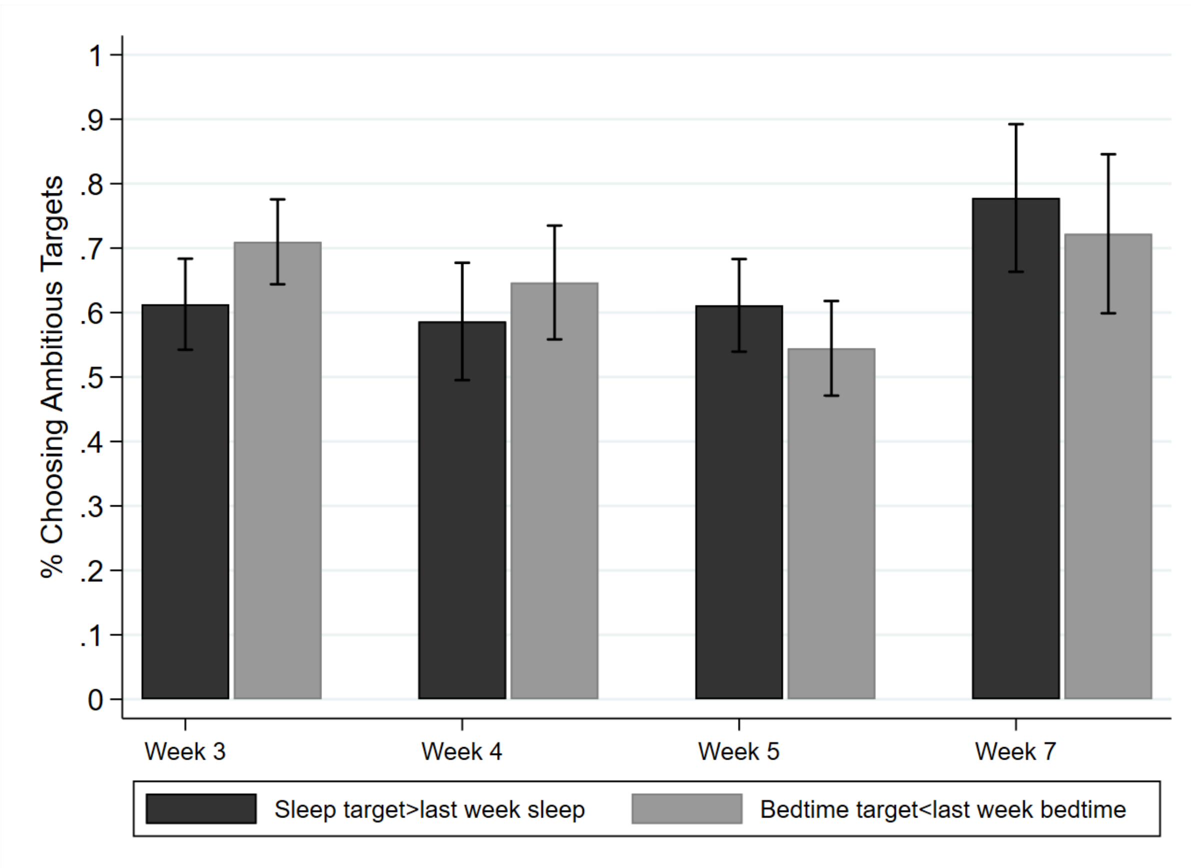

*Notes* - The figure reports the share of participants choosing a restrictive target with respect to their baseline bedtime and sleep habits with 95% confidence intervals. This analysis includes waves 1-6. In waves 1 and 4, treated weeks were week 3, 5, and 7. In waves 2, 3, 5, and 6 treated weeks were week 3, 4, and 5.

Figure A.12: Share of participants choosing a restrictive target with respect to their previous week bedtime and sleep habits



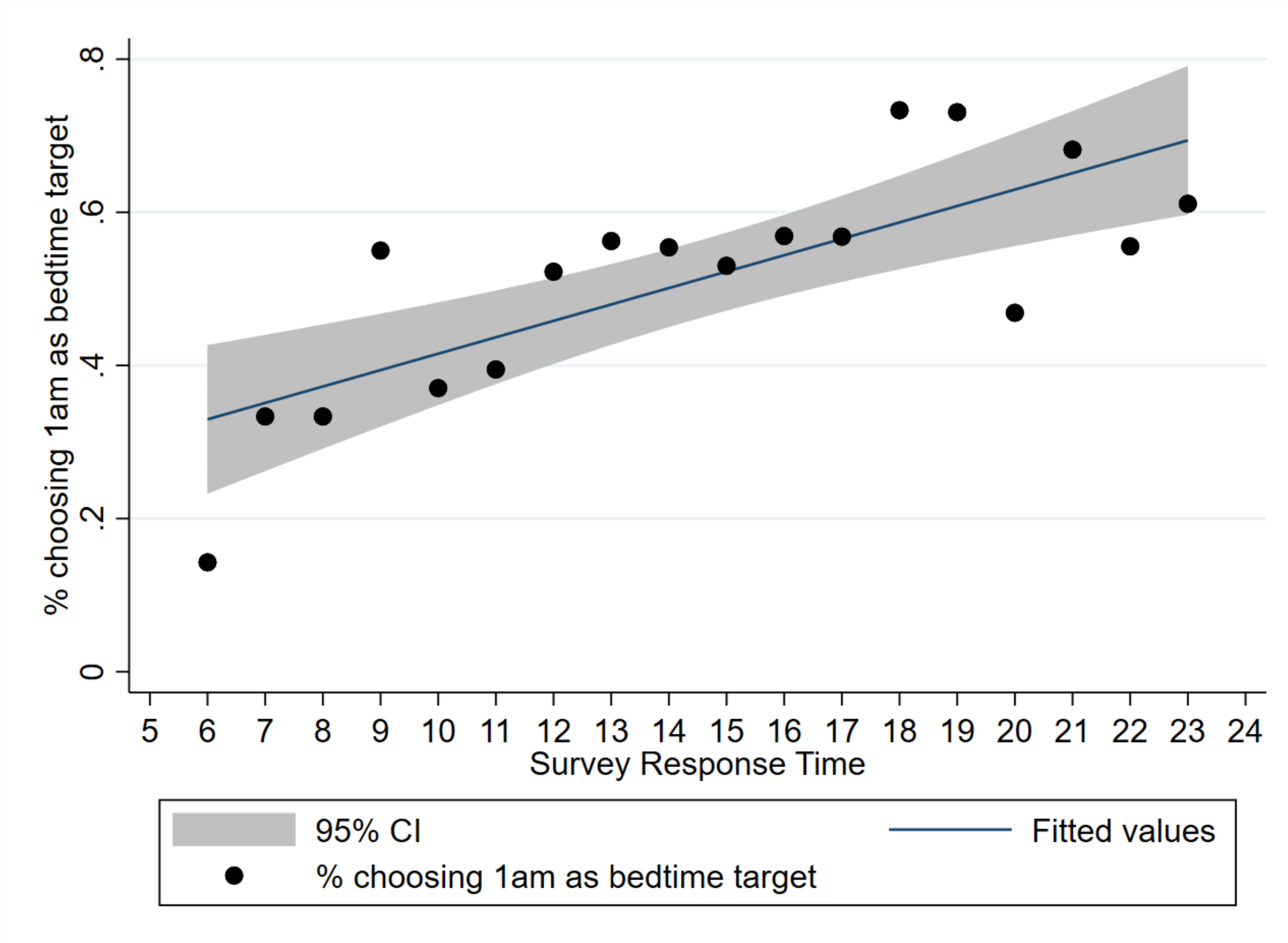

*Notes* - The figure reports the share of participants choosing a restrictive target with respect to their previous week average sleep duration and bedtime at baseline with 95% confidence intervals. This analysis includes waves 1-6. In waves 1 and 4, treated weeks were week 3, 5, and 7. In waves 2, 3, 5, and 6 treated weeks were week 3, 4, and 5.

Figure A.13: Share of participants choosing a restrictive target with respect to their previous week bedtime and sleep habits among those choosing bedtime earlier than 1am and sleep duration longer than 7 hours



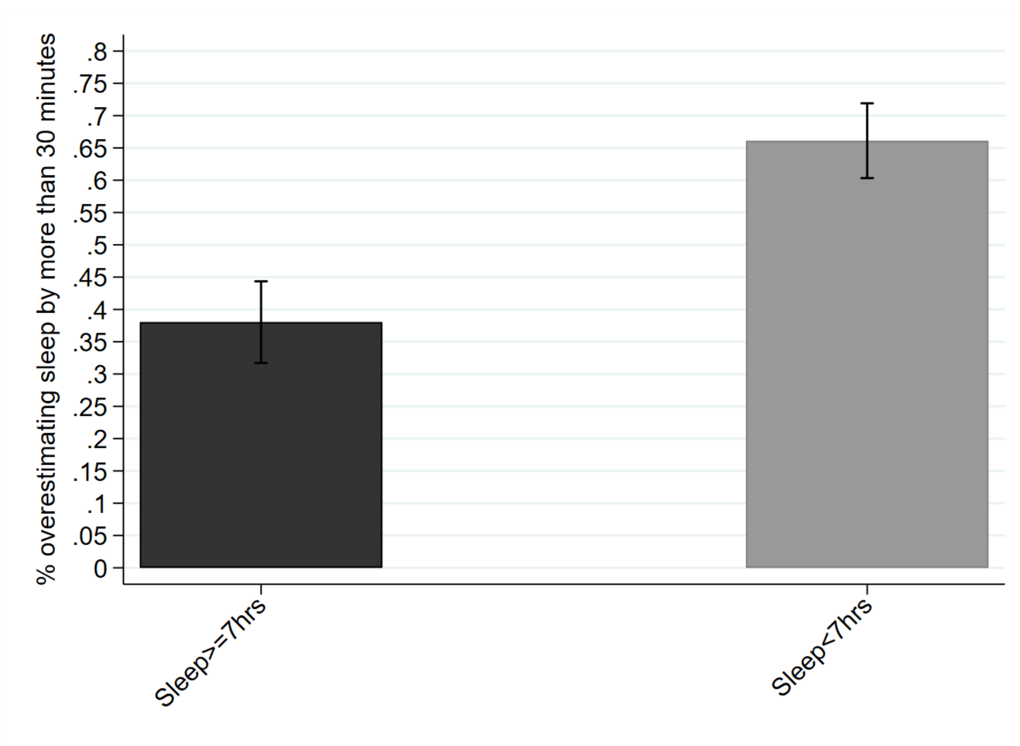

*Notes -* The figure reports the share of participants choosing a restrictive target with respect to their previous week average sleep duration and bedtime at baseline with 95% confidence intervals. This analysis includes waves 1-6. The sample is restricted to participants who chose a bedtime earlier than 1am and a sleep duration target longer than 7 hours.

Figure A.14: Timing of Survey Response and Bedtime Target Choice



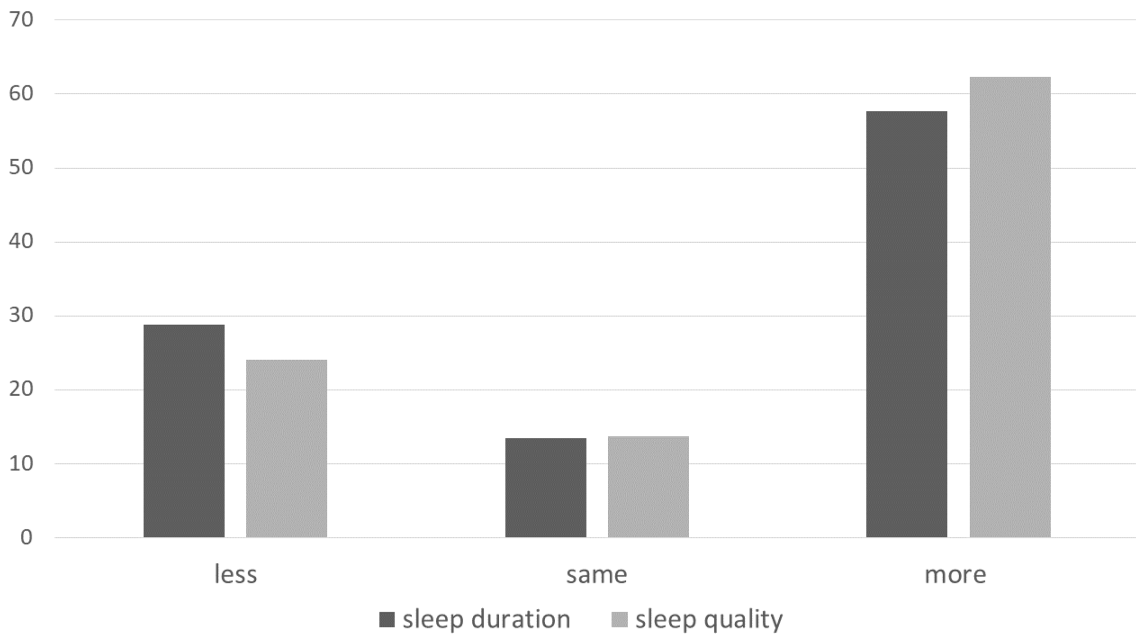

*Notes* - The figure reports the share of individuals choosing the least binding bedtime target (1 am) by the survey response time (6 am-8 pm). This analysis includes waves 1-6. The data are drawn from the weekly surveys during intervention weeks.

Figure A.15: Self-reported sleep and sleep deprivation: share of individuals overestimating their own sleep (>30minutes)



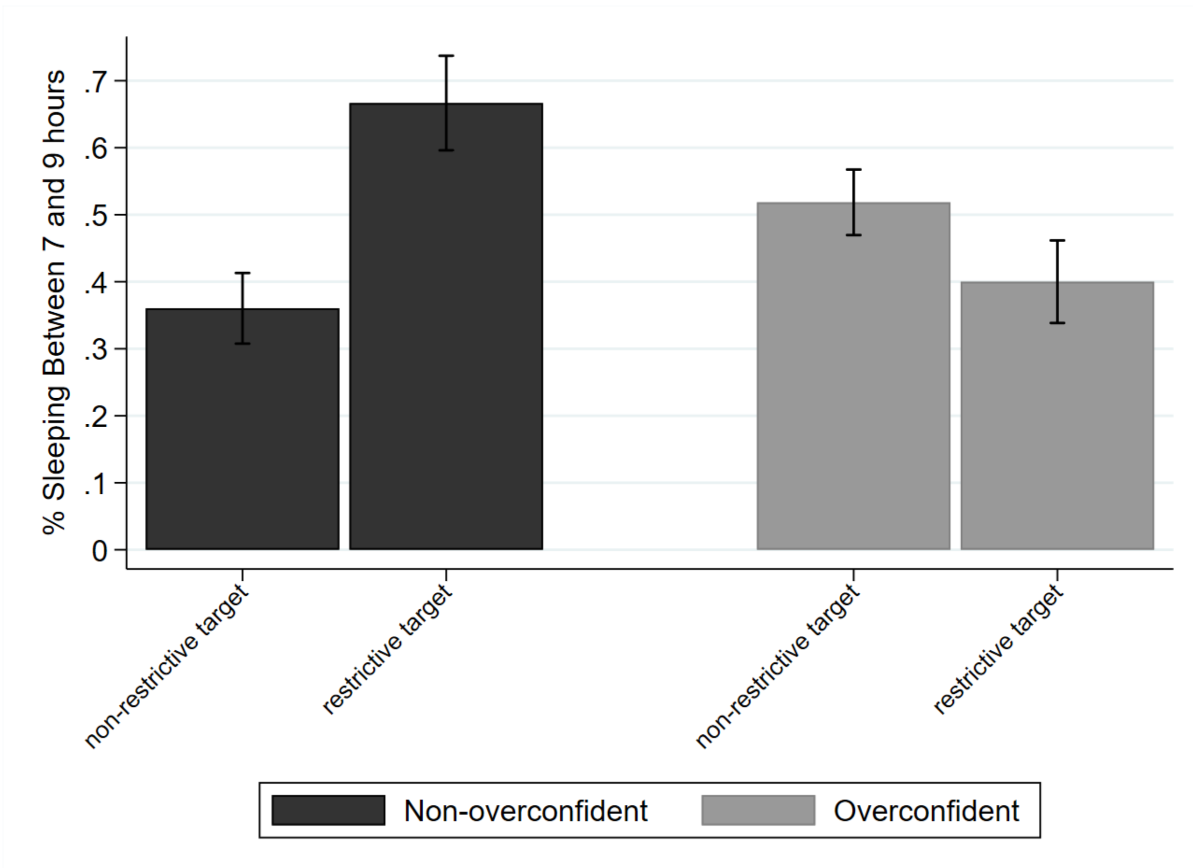

*Notes -* The figure documents the share of individuals overpredicting their sleep duration for the first two weeks of the experiment by more than 30 minutes with 95% confidence intervals. This analysis includes waves 1-6.

Figure A.16: Self-reported Sleep Duration and Quality and Beliefs about Others' Sleep Duration and Quality



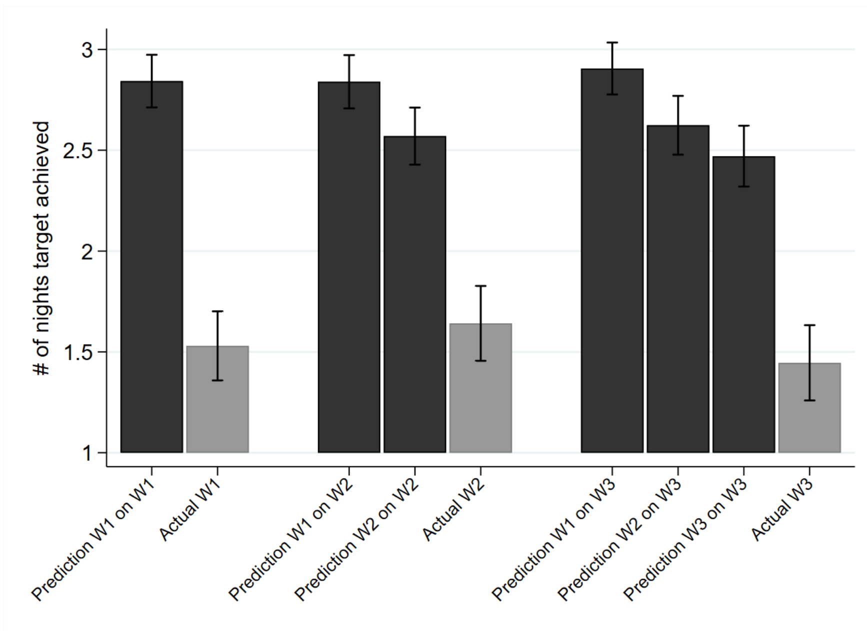

*Notes -* The figure reports the share of individuals reporting lower, the same, or higher sleep duration (in black) or higher sleep quality (in gray) than what reported when asked about the typical sleep duration and quality of individual of their same age and gender. This analysis includes waves 1-6.

Figure A.17: Overconfidence (based on Wave 6-incentivated predictions), Commitment and Sleep



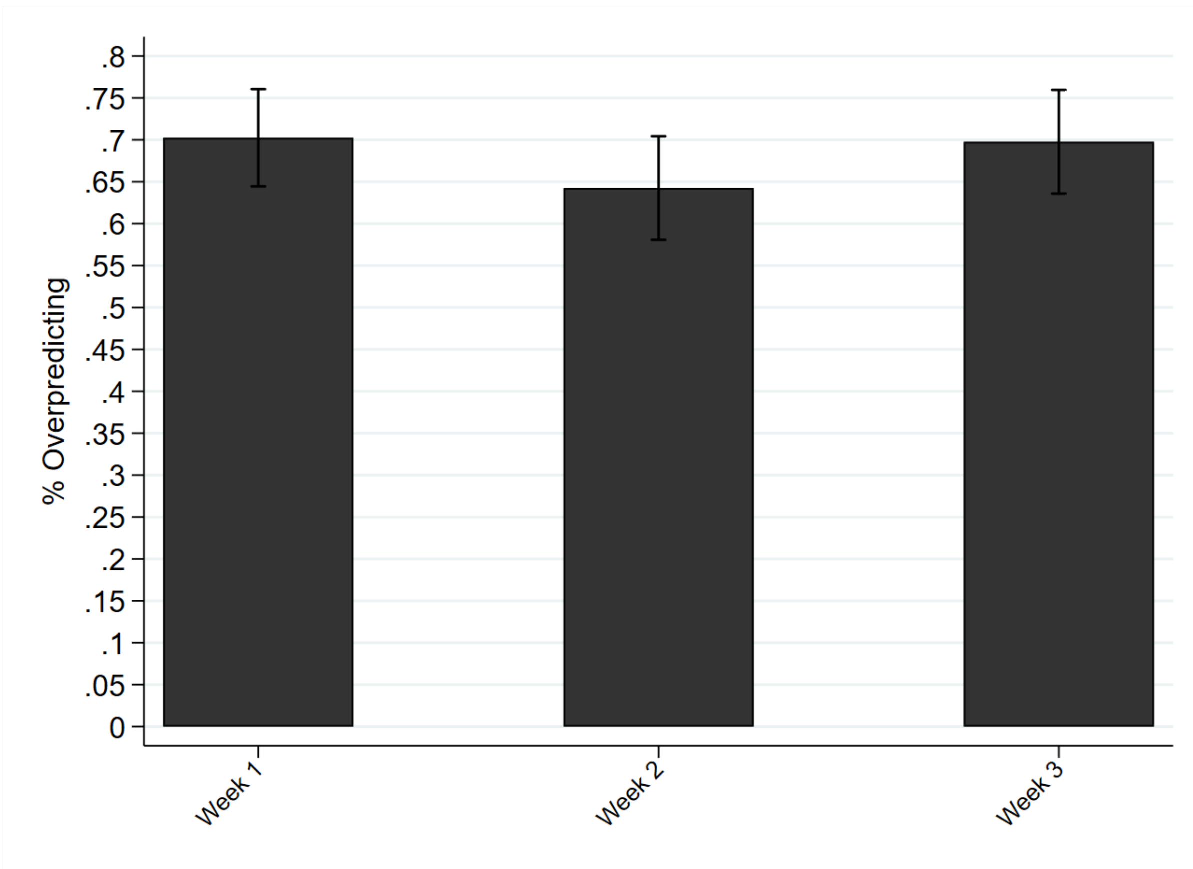

*Notes -* The figure documents the share of individual sleeping between 7 and 9 hours by whether they selected or not a restricted target (either sleep duration or bedtime) and by overconfidence defined as overpredicting own sleep in the first two weeks of the experiment by more than 30 minutes with 95% confidence intervals.

Figure A.18: Prediction vs Actual Behavior



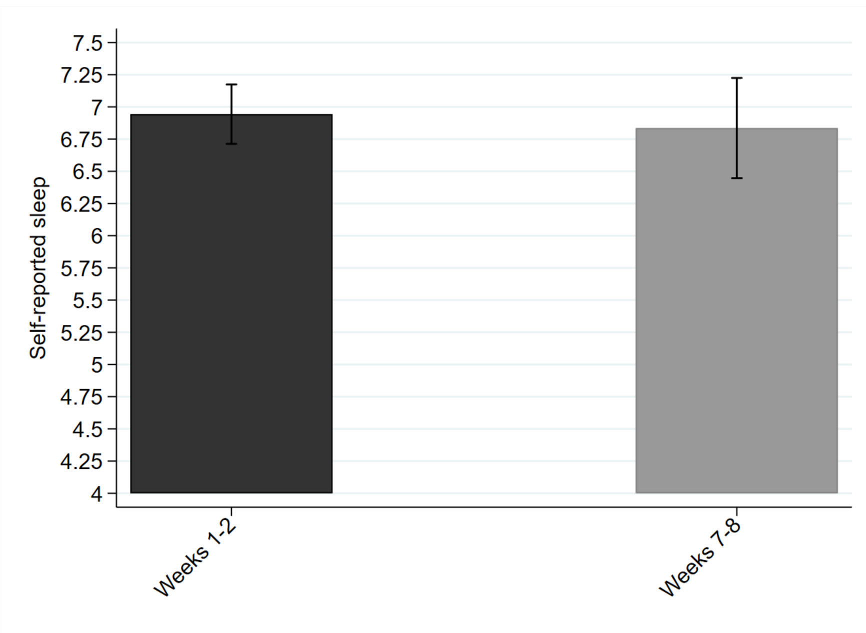

*Notes -* The figure above reports the average number of nights participants predicted to meet their sleep and bedtime targets versus the average number of nights they actually met their targets with 95% confidence intervals. At the beginining of the first week of treatment they were asked to predict for the following three weeks of treatment; at the beginining of the second week of treatment they were asked to predict for the following two weeks of treatment; at the beginining of the third week of treatment they were asked to predict for the following week of treatment. This analysis includes waves 1-6.

Figure A.19: Overpredicting Sleep Duration during Treatment



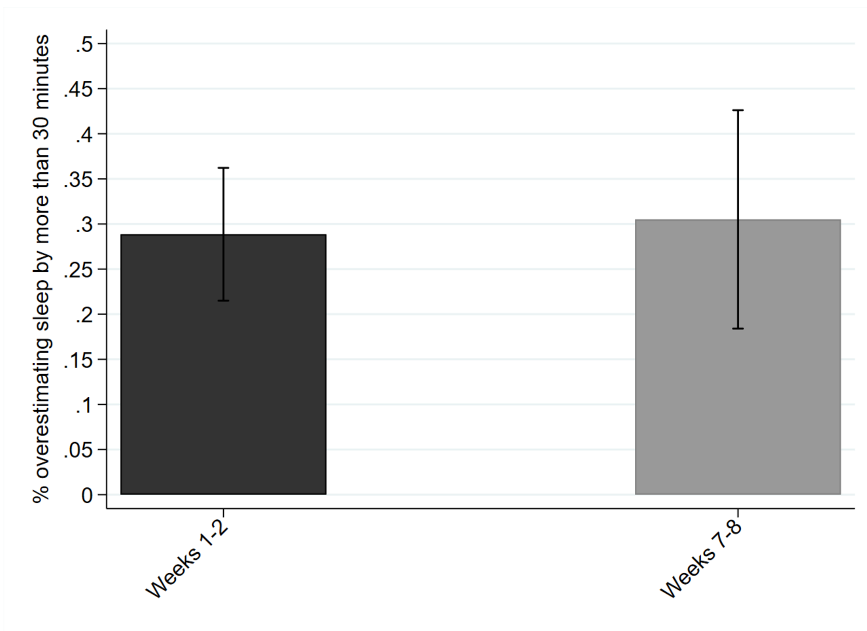

*Notes -* The figure documents the share of individuals overpredicting their sleep duration for the following week during the three weeks of treatments with 95% confidence intervals. This analysis includes waves 1-6.

Figure A.20: Self-reported sleep duration at baseline and endline among subjects in control group.



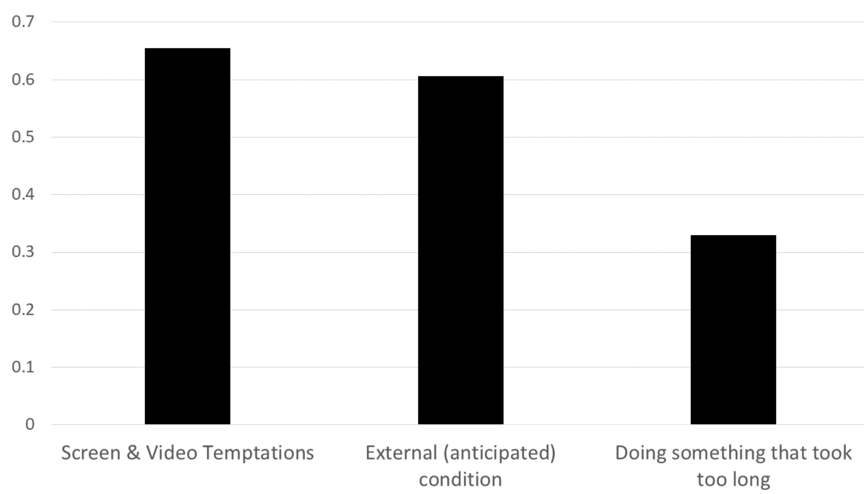

*Notes -* The figure above reports self-reported sleep duration of participants in control group at baseline and endline with 95% confidence intervals. This analysis includes waves 1-6.

Figure A.21: Share of participants overestimating sleep by more than 30 minutes among subjects in control group



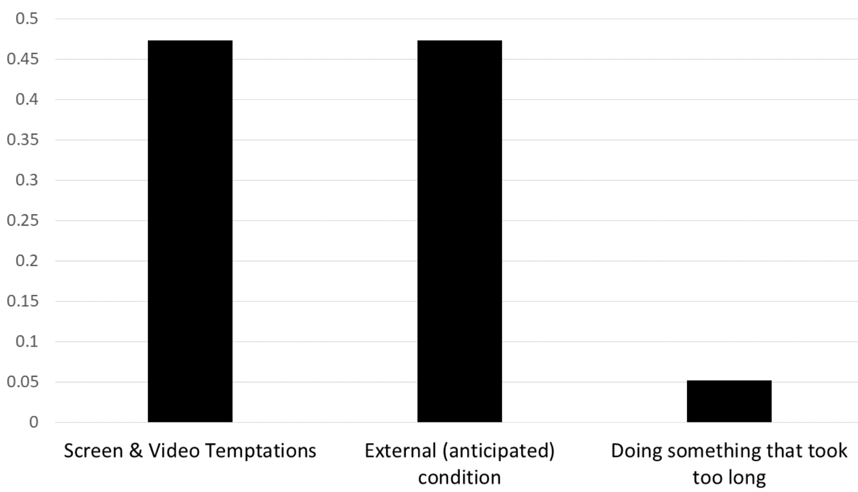

*Notes -* The figure above reports the share of participants in control group overestimating sleep by more than 30 minutes at baseline and endline with 95% confidence intervals.

Figure A.22: Reasons for not meeting target




*Notes -* The figure above reports the share of participants mentioning screen or other digital temptations, an external (anticipated) conditions, or that they were doing something that took too long as the reason for not meeting their target during the intervention.

Figure A.23: Main reason for not meeting target




*Notes -* The figure above reports the share of participants mentioning screen or other digital temptations, an external (anticipated) conditions, or that they were doing something that took too long as the reason for not meeting their bedtime target during the intervention.

## B Elicitation of Risk and Time Preferences

We elicited subjects' risk and time preferences during the first session (in the lab for waves 1-5, and online for wave 6), mainly based on the Double Multiple Price List (DMPL) method in Andreoni and Sprenger (2012). These are then based on risk preference elicitation in Holt and Laury (2002) and time preference elicitation in Harrison et al. (2002), with applications in many settings (see e.g. Tanaka et al., 2010; Falk et al., 2016). We did this by using 2 multiple price lists for risk preference, and 4 multiple price lists for time preference. In each list, there were two columns, representing Option A and Option B. On each list, one of the two options was fixed, and the other option changed from one row to the next. In each row, subjects had to indicate their preferred option: Option A or Option B. To avoid multiple switching points on a single list, they only had to click once on a list at the point where their preferred option switched from Option A to Option B. And then all other rows are automatically filled taking monotonicity into account. This is also different from the implementation in which subjects only need to state their indifference point, which was shown to be problematic (Frederick et al., 2002). Subjects were given examples and the opportunity to practice before making decisions that counted for payment.

To elicit the risk preference parameter, we used two lists. On each list, Option A was a fixed lottery: a 50% chance of getting GBP 6 and a 50% chance of getting GBP 0. Option B was always a sure amount. The lists we used are illustrated in Figures B.1 and B.2. Payments associated with the risk preference task was made at the end of the first session.

To elicit the time preference parameters, we used four lists each comparing a fixed lottery with various certainty amounts. We elicited time preferences using four price lists, each comparing different sooner payments with a fixed future payment. We varied both the size and timing (immediately or in 4 weeks) of the sooner payments as well as the gap between the sooner and later payments (4 or 8 weeks). On each list, Option A was associated with a monetary payment at a sooner time and Option B implied some monetary payment at a later time. The amount to be gained at the later time is fixed at GBP 6, and the amount to be gained at the sooner time varied on each list. Among the lists, the sooner time is either today or in 4 weeks, and the delay between the later and the sooner time is either 4 weeks or 8 weeks. The lists we used are illustrated in

Figures B.3, B.4, B.5 and B.6. All payments (earlier and later) associated with the time preference task was made digitally (Amazon Gift Cards in waves 1-5, and Venmo in wave 6).

We admit that our measure of time preference is not perfect, because for instance it can be confounded by liquidity constraints (Cohen et al., 2020). While there are more sophisticated ways to elicit time preferences, we chose a relatively simple and easy task to reduce burden on subjects and shorten the duration of the lab session.

Figure B.1: Choice List for Risk Preference 1

| Option A | Option B |
|---|---|
| 50% Chance of £6 and 50% Chance of £0 | £0.00 |
| 50% Chance of £6 and 50% Chance of £0 | £0.30 |
| 50% Chance of £6 and 50% Chance of £0 | £0.60 |
| 50% Chance of £6 and 50% Chance of £0 | £0.90 |
| 50% Chance of £6 and 50% Chance of £0 | £1.20 |
| 50% Chance of £6 and 50% Chance of £0 | £1.50 |
| 50% Chance of £6 and 50% Chance of £0 | £1.80 |
| 50% Chance of £6 and 50% Chance of £0 | £2.10 |
| 50% Chance of £6 and 50% Chance of £0 | £2.40 |
| 50% Chance of £6 and 50% Chance of £0 | £2.70 |
| 50% Chance of £6 and 50% Chance of £0 | £3.00 |
| 50% Chance of £6 and 50% Chance of £0 | £3.30 |
| 50% Chance of £6 and 50% Chance of £0 | £3.60 |
| 50% Chance of £6 and 50% Chance of £0 | £3.90 |
| 50% Chance of £6 and 50% Chance of £0 | £4.20 |
| 50% Chance of £6 and 50% Chance of £0 | £4.50 |
| 50% Chance of £6 and 50% Chance of £0 | £4.80 |
| 50% Chance of £6 and 50% Chance of £0 | £5.10 |
| 50% Chance of £6 and 50% Chance of £0 | £5.40 |
| 50% Chance of £6 and 50% Chance of £0 | £5.70 |
| 50% Chance of £6 and 50% Chance of £0 | £6.00 |

Figure B.2: Choice List for Risk Preference 2

| Option A | Option B |
|---|---|
| 50% Chance of £6 and 50% Chance of £0 | £0.00 |
| 50% Chance of £6 and 50% Chance of £0 | £0.30 |
| 50% Chance of £6 and 50% Chance of £0 | £0.60 |
| 50% Chance of £6 and 50% Chance of £0 | £0.90 |
| 50% Chance of £6 and 50% Chance of £0 | £1.20 |
| 50% Chance of £6 and 50% Chance of £0 | £1.50 |
| 50% Chance of £6 and 50% Chance of £0 | £1.80 |
| 50% Chance of £6 and 50% Chance of £0 | £2.10 |
| 50% Chance of £6 and 50% Chance of £0 | £2.40 |
| 50% Chance of £6 and 50% Chance of £0 | £2.70 |
| 50% Chance of £6 and 50% Chance of £0 | £3.00 |
| 50% Chance of £6 and 50% Chance of £0 | £3.30 |
| 50% Chance of £6 and 50% Chance of £0 | £3.60 |
| 50% Chance of £6 and 50% Chance of £0 | £3.90 |
| 50% Chance of £6 and 50% Chance of £0 | £4.20 |
| 50% Chance of £6 and 50% Chance of £0 | £4.50 |
| 50% Chance of £6 and 50% Chance of £0 | £4.80 |
| 50% Chance of £6 and 50% Chance of £0 | £5.10 |
| 50% Chance of £6 and 50% Chance of £0 | £5.40 |
| 50% Chance of £6 and 50% Chance of £0 | £5.70 |
| 50% Chance of £6 and 50% Chance of £0 | £6.00 |

Figure B.3: Choice List for Time Preference 1

| Option A | Option B |
|---|---|
| Receive £5.80 today | Receive £6 in 4 weeks |
| Receive £5.60 today | Receive £6 in 4 weeks |
| Receive £5.40 today | Receive £6 in 4 weeks |
| Receive £5.20 today | Receive £6 in 4 weeks |
| Receive £5.00 today | Receive £6 in 4 weeks |
| Receive £4.80 today | Receive £6 in 4 weeks |
| Receive £4.60 today | Receive £6 in 4 weeks |
| Receive £4.40 today | Receive £6 in 4 weeks |
| Receive £4.20 today | Receive £6 in 4 weeks |
| Receive £4.00 today | Receive £6 in 4 weeks |
| Receive £3.80 today | Receive £6 in 4 weeks |
| Receive £3.60 today | Receive £6 in 4 weeks |
| Receive £3.40 today | Receive £6 in 4 weeks |
| Receive £3.20 today | Receive £6 in 4 weeks |
| Receive £3.00 today | Receive £6 in 4 weeks |
| Receive £2.80 today | Receive £6 in 4 weeks |
| Receive £2.60 today | Receive £6 in 4 weeks |
| Receive £2.40 today | Receive £6 in 4 weeks |
| Receive £2.20 today | Receive £6 in 4 weeks |
| Receive £2.00 today | Receive £6 in 4 weeks |
| Receive £1.80 today | Receive £6 in 4 weeks |
| Receive £1.60 today | Receive £6 in 4 weeks |
| Receive £1.40 today | Receive £6 in 4 weeks |
| Receive £1.20 today | Receive £6 in 4 weeks |
| Receive £1.00 today | Receive £6 in 4 weeks |
| Receive £0.80 today | Receive £6 in 4 weeks |
| Receive £0.60 today | Receive £6 in 4 weeks |
| Receive £0.40 today | Receive £6 in 4 weeks |
| Receive £0.20 today | Receive £6 in 4 weeks |

Figure B.4: Choice List for Time Preference 2

| Option A | Option B |
|---|---|
| Receive £5.80 today | Receive £6 in 8 weeks |
| Receive £5.60 today | Receive £6 in 8 weeks |
| Receive £5.40 today | Receive £6 in 8 weeks |
| Receive £5.20 today | Receive £6 in 8 weeks |
| Receive £5.00 today | Receive £6 in 8 weeks |
| Receive £4.80 today | Receive £6 in 8 weeks |
| Receive £4.60 today | Receive £6 in 8 weeks |
| Receive £4.40 today | Receive £6 in 8 weeks |
| Receive £4.20 today | Receive £6 in 8 weeks |
| Receive £4.00 today | Receive £6 in 8 weeks |
| Receive £3.80 today | Receive £6 in 8 weeks |
| Receive £3.60 today | Receive £6 in 8 weeks |
| Receive £3.40 today | Receive £6 in 8 weeks |
| Receive £3.20 today | Receive £6 in 8 weeks |
| Receive £3.00 today | Receive £6 in 8 weeks |
| Receive £2.80 today | Receive £6 in 8 weeks |
| Receive £2.60 today | Receive £6 in 8 weeks |
| Receive £2.40 today | Receive £6 in 8 weeks |
| Receive £2.20 today | Receive £6 in 8 weeks |
| Receive £2.00 today | Receive £6 in 8 weeks |
| Receive £1.80 today | Receive £6 in 8 weeks |
| Receive £1.60 today | Receive £6 in 8 weeks |
| Receive £1.40 today | Receive £6 in 8 weeks |
| Receive £1.20 today | Receive £6 in 8 weeks |
| Receive £1.00 today | Receive £6 in 8 weeks |
| Receive £0.80 today | Receive £6 in 8 weeks |
| Receive £0.60 today | Receive £6 in 8 weeks |
| Receive £0.40 today | Receive £6 in 8 weeks |
| Receive £0.20 today | Receive £6 in 8 weeks |

Figure B.5: Choice List for Time Preference 3

| Option A | Option B |
|---|---|
| Receive £5.80 in 4 weeks | Receive £6 in 8 weeks |
| Receive £5.60 in 4 weeks | Receive £6 in 8 weeks |
| Receive £5.40 in 4 weeks | Receive £6 in 8 weeks |
| Receive £5.20 in 4 weeks | Receive £6 in 8 weeks |
| Receive £5.00 in 4 weeks | Receive £6 in 8 weeks |
| Receive £4.80 in 4 weeks | Receive £6 in 8 weeks |
| Receive £4.60 in 4 weeks | Receive £6 in 8 weeks |
| Receive £4.40 in 4 weeks | Receive £6 in 8 weeks |
| Receive £4.20 in 4 weeks | Receive £6 in 8 weeks |
| Receive £4.00 in 4 weeks | Receive £6 in 8 weeks |
| Receive £3.80 in 4 weeks | Receive £6 in 8 weeks |
| Receive £3.60 in 4 weeks | Receive £6 in 8 weeks |
| Receive £3.40 in 4 weeks | Receive £6 in 8 weeks |
| Receive £3.20 in 4 weeks | Receive £6 in 8 weeks |
| Receive £3.00 in 4 weeks | Receive £6 in 8 weeks |
| Receive £2.80 in 4 weeks | Receive £6 in 8 weeks |
| Receive £2.60 in 4 weeks | Receive £6 in 8 weeks |
| Receive £2.40 in 4 weeks | Receive £6 in 8 weeks |
| Receive £2.20 in 4 weeks | Receive £6 in 8 weeks |
| Receive £2.00 in 4 weeks | Receive £6 in 8 weeks |
| Receive £1.80 in 4 weeks | Receive £6 in 8 weeks |
| Receive £1.60 in 4 weeks | Receive £6 in 8 weeks |
| Receive £1.40 in 4 weeks | Receive £6 in 8 weeks |
| Receive £1.20 in 4 weeks | Receive £6 in 8 weeks |
| Receive £1.00 in 4 weeks | Receive £6 in 8 weeks |
| Receive £0.80 in 4 weeks | Receive £6 in 8 weeks |
| Receive £0.60 in 4 weeks | Receive £6 in 8 weeks |
| Receive £0.40 in 4 weeks | Receive £6 in 8 weeks |
| Receive £0.20 in 4 weeks | Receive £6 in 8 weeks |

Figure B.6: Choice List for Time Preference 4

| Option A | Option B |
|---|---|
| Receive £5.80 in 4 weeks | Receive £6 in 12 weeks |
| Receive £5.60 in 4 weeks | Receive £6 in 12 weeks |
| Receive £5.40 in 4 weeks | Receive £6 in 12 weeks |
| Receive £5.20 in 4 weeks | Receive £6 in 12 weeks |
| Receive £5.00 in 4 weeks | Receive £6 in 12 weeks |
| Receive £4.80 in 4 weeks | Receive £6 in 12 weeks |
| Receive £4.60 in 4 weeks | Receive £6 in 12 weeks |
| Receive £4.40 in 4 weeks | Receive £6 in 12 weeks |
| Receive £4.20 in 4 weeks | Receive £6 in 12 weeks |
| Receive £4.00 in 4 weeks | Receive £6 in 12 weeks |
| Receive £3.80 in 4 weeks | Receive £6 in 12 weeks |
| Receive £3.60 in 4 weeks | Receive £6 in 12 weeks |
| Receive £3.40 in 4 weeks | Receive £6 in 12 weeks |
| Receive £3.20 in 4 weeks | Receive £6 in 12 weeks |
| Receive £3.00 in 4 weeks | Receive £6 in 12 weeks |
| Receive £2.80 in 4 weeks | Receive £6 in 12 weeks |
| Receive £2.60 in 4 weeks | Receive £6 in 12 weeks |
| Receive £2.40 in 4 weeks | Receive £6 in 12 weeks |
| Receive £2.20 in 4 weeks | Receive £6 in 12 weeks |
| Receive £2.00 in 4 weeks | Receive £6 in 12 weeks |
| Receive £1.80 in 4 weeks | Receive £6 in 12 weeks |
| Receive £1.60 in 4 weeks | Receive £6 in 12 weeks |
| Receive £1.40 in 4 weeks | Receive £6 in 12 weeks |
| Receive £1.20 in 4 weeks | Receive £6 in 12 weeks |
| Receive £1.00 in 4 weeks | Receive £6 in 12 weeks |
| Receive £0.80 in 4 weeks | Receive £6 in 12 weeks |
| Receive £0.60 in 4 weeks | Receive £6 in 12 weeks |
| Receive £0.40 in 4 weeks | Receive £6 in 12 weeks |
| Receive £0.20 in 4 weeks | Receive £6 in 12 weeks |

# C Fitbit Reliability and Specificity

Recently, due to the popularity of consumer-grade wearable devices, there has been medical studies checking and comparing the validity of their measures. The Fitbit model (Fitbit Charge HR) we used in waves 1-5 of the experiment were found to have adequate sensitivity in detecting sleep, a significant improvement over self-reported sleep measures, and comparable with PSG and other actigraphy devices (Lee et al., 2017; Haghayegh et al., 2019b; Godino et al., 2020). Admittedly, earlier models of Fitbit were also found to have certain issues, such as relatively low sensitivity in detecting wake (Godino et al., 2020; Moreno et al., 2006) We note that recent generation Fitbit models (including Fitbit Inspire 2) rely not only on body movement but also on the HRV sensor and a new software algorithm technology, and therefore can detect wake epochs during intended sleep much more accurately than earlier models. While there is only limited evidence on these new models, the amount of bias in estimating total sleep time compared to PSG was found to be clinically negligible. Meta-analysis of the published data also substantiated the lack of a statistically significant difference between sleep-staging Fitbit models and PSG in measuring WASO, TST, and SE, and with effect sizes of differences in the range of small, even for SOL (Haghayegh et al., 2019b,a). Epoch-by-epoch (EB) analyses conducted on the data from the same four comparisons also revealed high sensitivity (i.e., between 0.95 and 0.96) and specificity (i.e., between 0.58 and 0.69) in detecting sleep. Comparison trials involving sleep-staging Fitbit models disclosed much higher specificity in detecting sleep than all of the nonsleep-staging Fitbit model comparison trials that reported specificity in the range of 0.10-0.52. This recent evidence also suggests that their performance in differentiating wake from sleep epochs is better than that reported in the literature for actigraphy.