Using online panels to collect self-reported and device-based sleep duration data: Examining sources of variation

EsCHER Working Paper No. 2026001
Date: March 2026

Stefan A. Lipman, Ph.D.
Andrea Pogliano, M.Sc.
Dorien Beeres, M.Sc.
Lili L. Kokai, Ph.D.
Georg Granic, Ph.D.
Joost Oude Groeniger, Ph.D.
Kirsten Rohde, Ph.D.
Hans van Kippersluis, Ph.D.

Erasmus University Rotterdam
Making Minds Matter

Title
Using online panels to collect self-reported and device-based sleep duration data: Examining sources of variation

Authors
See page 2

Keywords
sleep duration measurement; activity tracker; online panels; screenshots of device data; participant attrition; selection effects

JEL classification
I10; I12; C83; C81; D91

Cite as
Lipman, S.A., Pogliano, A., Beeres, D., Kokai, L.L., Granic, G.D., Oude Groeniger, J., Rohde, K.I.M, & van Kippersluis, H. (2026). Using online panels to collect self-reported and device-based sleep duration data: Examining sources of variation. EsCHER Working Paper Series No. 2026001, Erasmus University Rotterdam. Available from: https://www.eur.nl/en/research/research-groups-initiatives/escher/research/working-papers

Erasmus Centre for Health Economics Rotterdam (EsCHER) is part of Erasmus University Rotterdam.
Want to know more about EsCHER? Visit www.eur.nl/escher
Want to contact EsCHER? E-mail escher@eur.nl

Interested in more EsCHER Working Papers? Download from www.eur.nl/escher/research/workingpapers

==

Using online panels to collect self-reported and device-based sleep duration data: Examining sources of variation

Stefan A. Lipman[1,2,*], Andrea Pogliano[3], Dorien Beeres[4], Lili L. Kókai[4,5], Georg Granic[3], Joost Oude Groeniger[4,6], Kirsten Rohde[7], and Hans van Kippersluis[3]

[1] *Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands*
[2] *Erasmus Centre for Health Economics Rotterdam, Erasmus University Rotterdam, Rotterdam, the Netherlands*
[3] *Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands*
[4] *Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands*
[5] *Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, the Netherlands*
[6] *Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, the Netherlands*
[7] *School of Business & Economics, Maastricht University, Maastricht, the Netherlands*
[*] *Corresponding author, E: lipman@eshpm.eur.nl*

Abstract

Objective: Given the increasing need for accurate yet cost-efficient sleep measurement tools, this study explores whether recruiting device owners from online panels can produce high-quality data on sleep duration. Specifically, we investigate the feasibility and benefits of collecting device-based data on sleep duration from individuals who already own an activity tracker, as well as two potential drawbacks: selection and attrition effects.

Methods: We conducted a three-Wave survey experiment with 885 UK-based respondents recruited through the online panel Prolific. Participants were allocated to either an unrestricted (UR) sample or a sample restricted to device owners (RDO). Demographics were collected in both groups, as well as self-reports of sleep duration in every Wave. RDO respondents also uploaded data on sleep duration captured by their devices. We examined the frequency and quality of uploaded device data, differences between self- and device-reported sleep, selection effects and the effects of mandatory upload requirements on participant drop-out.

Results: Nearly all screenshots allowed extraction of sleep duration, although verifying date accuracy and retaining participants though multiple Waves was challenging. Device-based sleep duration aligned closely with self-reports among device owners, with little to no systematic differences. Selection and attrition effects clearly apply: restricting recruitment to device owners changed sample composition (older, higher income, more employed, fewer children), and mandatory uploads causally increased attrition.

Conclusions: Collecting device-reported sleep data via online panels is feasible and cost-efficient, but introduces trade-offs in representativeness and retention. Researchers should carefully weigh this trade-off between cost-efficiency and external validity in the design of future sleep studies.

Keywords: sleep duration measurement; activity tracker; online panels; screenshots of device data; participant attrition; selection effects

1 Introduction

Insufficient sleep is increasingly recognized as a major public health concern [The Lancet Diabetes & Endocrinology, 2024], for several reasons. First, it negatively impacts health, cognitive skills, decision-making, and productivity [Gibson and Shrader, 2018, Giuntella and Mazzonna, 2019, Heissel and Norris, 2018, Osborne-Christenson, 2022]. A recent RAND study, for example, estimates the productivity loss in 2025 due to insufficient sleep at $718 billion in the U.S., U.K, Japan, Germany, and Canada combined [Chattu et al., 2018]. Second, insufficient sleep is widespread. One third of the U.S. population report that they usually sleep less than the recommended minimum of seven hours per night [Liu et al., 2016]. A systematic review and meta-analysis of over 1 million observations from the U.S., U.K., and the Netherlands found that one in four individuals sleep less than age-specific recommendations [Kocevska et al., 2021]. Despite the apparent need for initiatives or policies to alleviate the impact of insufficient sleep, field studies aiming to increase sleep duration remain rare [Rao et al., 2021, Giuntella et al., 2024, Bessone et al., 2021]. One reason may be the difficulty of objectively and affordably measuring sleep duration in field settings.

Polysomnography (PSG) is the gold standard for objectively measuring sleep duration (and quality) in laboratories [Marino et al., 2013]. While PSG is highly accurate in capturing sleep and wake times, it is intrusive and costly, requiring study respondents to wear multiple sensors and stay overnight in a specialized environment [Hussain et al., 2022]. This makes measuring sleep duration with PSG in field settings challenging. As an alternative, studies have relied on subjective measures of sleep, i.e.,self-reports [Jin and Ziebarth, 2020], though these are found to poorly correlate with objective measures of sleep duration [Santos et al., 2021, Girschik et al., 2012]. Individuals often self-report time spent in bed rather than actual time asleep, generally leading to an overestimation of sleep duration [Lauderdale et al., 2008, Jackson et al., 2018, Benz et al., 2023].

Wearable activity trackers, e.g., wrist-worn devices, offer a less intrusive method to measure sleep duration (and quality). These wearable trackers infer sleep from actigraphy (body movements) and/or heart rate

fluctuations [Ancoli-Israel et al., 2003]. Typically, researchers studying sleep provide these trackers to study respondents [Avery et al., 2022, Bessone et al., 2021, Ong et al., 2023]. Unlike PSG, they enable the measurement of sleep duration in natural environments, such as at home. Although wearable trackers are less accurate than PSG, they are considered a valid alternative for measuring sleep duration [Marino et al., 2013, Lau et al., 2022]. Additionally, wearable trackers can be worn for extended periods, providing insight into long-term sleep patterns. This makes them particularly useful for studying sleep in populations with unpredictable or changing sleep patterns, such as shift workers, individuals with circadian rhythm disruption, insomnia, children, and the elderly [Lujan et al., 2021, Ancoli-Israel et al., 2015].

However, providing wearable activity trackers to study respondents is costly, given that wrist-worn devices alone typically cost at least $50, alongside the time required to set up the data collection infrastructure and the environmental costs associated with manufacturing and potentially later disposing of electronic devices. A potential cost-saving approach is recruiting respondents who already own an activity tracker and ask them to share/upload their data. Commercial online panels typically allow to screen for activity tracker ownership, offering access to large pools of potential respondents. If feasible, recruiting device owners through online panels could provide high-quality, device-reported behavioral data from a large population at a lower cost.

In this study, our main objective was to examine the feasibility, benefits and drawbacks of recruiting device owners through online panels. We recruited a sample of online respondents via Prolific [Palan and Schitter, 2018], an online platform that in 2023 (when this study was launched) provided access to a sample of more than 28,000 activity tracker device owners. We we recruited half the sample by including only respondents who already owned an activity tracker (hereafter, the restricted-to-device-owners [RDO] sample). The remaining half was drawn from the full Prolific pool on a first-come, first-served basis (the unrestricted [UR] sample). This design allows us to contribute to the literature in three ways. Our first contribution is that we study the feasibility of collecting sleep data from an online sample of device owners by examining whether respondents are willing and able to upload screenshots from their devices that are interpretable (i.e., provide information on

sleep duration). Importantly, we assess respondents willingness to upload such data repeatedly by collecting sleep data across three Waves. A second contribution of our study is that we assess the potential benefits of this approach, i.e., we study how these data compare to self-report measures of sleep duration (both within the RDO sample and compared to the UR sample).

Our third contribution is to examine two potential drawbacks of recruiting device owners through online panels, as compared to collecting only self-reported sleep data. First, restricting the sample to device owners may introduce *selection effects*, since device owners may differ from non-owners in demographics and health behaviors, including sleep patterns [Yen et al., 2022]. If such differences exist, studies that limit participation to device owners may suffer from reduced external validity. Second, we assess *attrition related to uploading*: respondents may be more likely to drop out when asked to upload device data, whether due to technical barriers, privacy concerns, inexperience, or the greater burden of uploading compared to self-reporting [Dunn and Hazzard, 2019, Smirnova, 2022]. If only a subset of device owners is willing or able to upload data, this may further undermine the external validity of relying on a RDO sample. To test for these issues, our survey experiment compares the characteristics of respondents in the RDO and UR samples across Waves and introduces random variation in whether uploading is mandatory or optional. This allows us to identify the causal impact of requiring device uploads. Both groups were asked to upload device data (on sleep for RDO and phone usage for UR), enabling us to explore whether attrition depends on the type of data requested.

Our key findings are that it is possible to obtain high-quality device-reported data through online panels, enabling affordable and timely access to objectively measured data on sleep duration. However, in our study, differences with self-reported sleep duration were small to non-existent. It is, furthermore, important to realize that potential drawbacks apply: we find that limiting the sample to device-owners reduces representativeness, and mandatory uploading increases (the already substantial) drop-out during the study. In the remainder of the paper we describe our experiment in detail in section 2, present our results in section 3, and critically discuss our findings in section 4.

2 Methods

2.1 Recruitment

The design and analysis plan for this survey experiment was pre-registered on AsPredicted.org (registration number 144053). We conducted a three-Wave study with UK-based respondents recruited via Prolific, with one week between each Wave.

Our target sample size was based on an ex-ante power analysis for independent samples t-tests. Assuming a small-to-medium between-group effect size (Cohen's $d = 0.25$), with 80% power ($\beta = 0.80$) and a 5% significance level ($\alpha = 0.05$), we required a total sample of 506 respondents (253 per group). This effect size was based on prior studies comparing measures of sleep duration, and allows us to detect a between-group difference in mean sleep duration of approximately 17–23 minutes, assuming pooled standard deviations comparable to those reported in prior studies comparing these measures. [Lauderdale et al., 2008, Jackson et al., 2018]. This target was set to ensure adequate statistical power for between-group comparisons in each Wave of data collection.

To account for attrition—estimated at 30% between Waves 1 and 2, and 5% between Waves 2 and 3—we oversampled in Wave 1, targeting 810 respondents in total. Respondents were evenly divided between two samples. The *restricted to device owners* (RDO) sample included those who had indicated in their Prolific registration that they owned an activity tracker. The *unrestricted* (UR) sample was drawn without such restrictions (excluding those in the RDO group), mimicking a typical convenience sample for self-reported sleep studies. Because some UR participants might also own activity trackers, both groups completed survey questions assessing activity tracker ownership. This allowed us to examine whether sleep duration differed as a function of tracker ownership, including within the UR sample itself.

To test the design, we conducted a pilot with 40 respondents per group. The pilot and full study data were pooled, as no changes were made to the survey post-piloting, yielding a total Wave 1 sample of 883 (RDO: $n = 441$; UR: $n = 442$). Waves 1, 2, and 3 of the main study launched on September 20/21, 28/29, and October 5/6 2023, respectively. Each Wave was first released to the RDO sample and released to the UR

sample on the following day, which enabled the application of Prolific's participant-level exclusion feature (excluding respondents based on completed studies) to ensure no respondent overlap. Median completion time was approximately 3 minutes for Waves 1 and 3, and 12 minutes for Wave 2. RDO and UR respondents exhibited small differences in survey completion times across Waves; the RDO-UR differences in mean completion time were approximately 80 seconds in Wave 1, 30 seconds in Wave 2, and -12 seconds in Wave 3. Respondents received £0.50 for completing Waves 1 and another £0.50 for Wave 3, and £4.00 for Wave 2. Those completing all three Waves earned an additional £2.50 bonus.

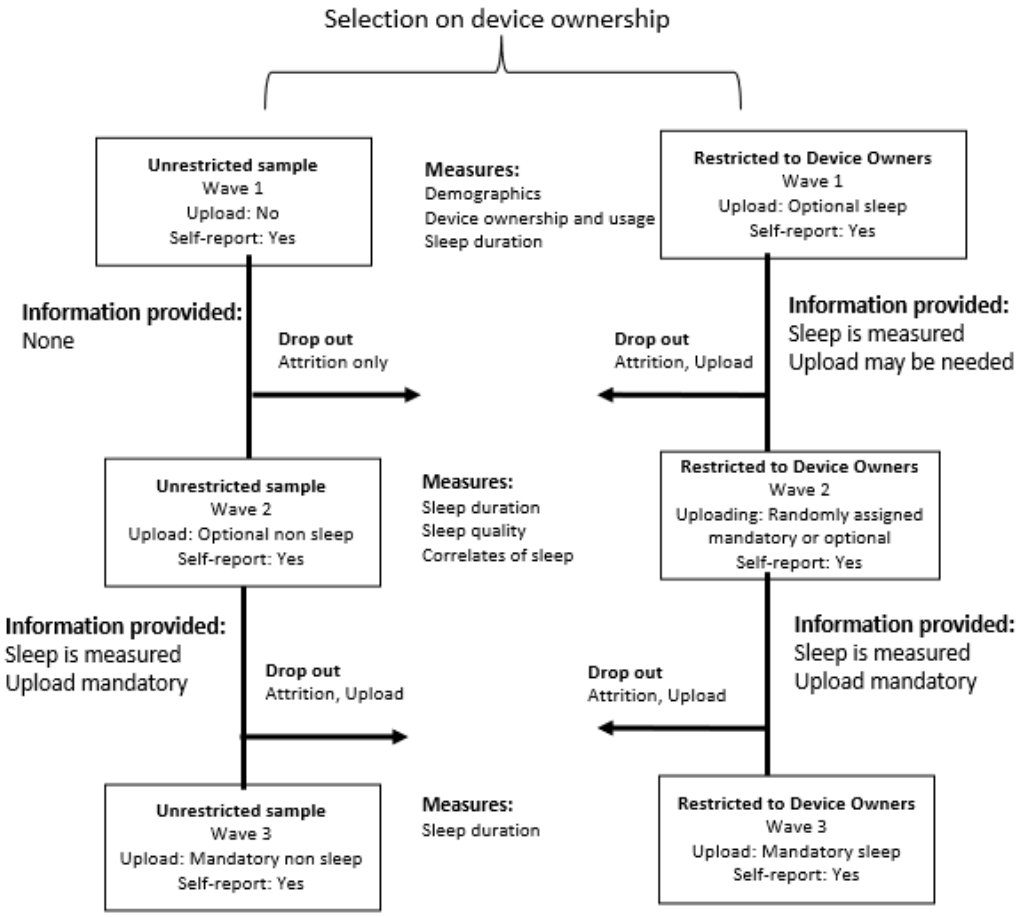

Figure 1: Survey Design

2.2 Study design

Figure 1 summarizes our study design. The Online Appendix to this paper contains a transcript of the surveys. The surveys were delivered via Qualtrics.

2.2.1 Wave 1

In Wave 1, respondents from both samples provided demographic information (age, employment status, gender, education and income). They were asked if they owned an activity tracker, and, if so, which type, and how often they wore it. All respondents were informed of the goal of the study and self-reported their sleep duration of the previous night by answering *'How long did you sleep last night? This time might be shorter than you have spent in bed. Try to be as precise as possible.'* All respondents reported the hours and minutes they slept. RDO respondents were given the option to voluntarily upload a screenshot of their activity tracker showing their sleep data from the previous night. To facilitate the upload, the survey included URLs to pages explaining how to view sleep data for popular activity trackers (i.e., Fitbit, Xiaomi and Apple Watch), as well as instructions on taking screenshots on iOS and Android phones. At the end of Wave 1 , RDO respondents were informed that they might be asked in a subsequent Wave to upload a screenshot of their sleep data from their activity tracker, without specifying the exact timing. They were shown the following message: "We may ask you to upload a screenshot of your sleep data from your activity tracker, so please be sure to wear your activity tracker at night during the next 7 days." In contrast, respondents in the UR sample were not given the option to upload a screenshot in Wave 1 or asked to do so in the next Wave.

2.2.2 Wave 2

In Wave 2, all respondents first self-reported their sleep from the previous night, as in Wave 1. The RDO sample was then randomized into two subsamples. One group was required to upload a screenshot of their device data (*mandatory*: "Please provide a screenshot of your activity tracker's sleep data from last night", with survey item set on 'optional'), while the other group was given the option to do so (*optional*: "Would you like to provide a screenshot of your activity tracker's sleep data from last night?", with survey item set on 'optional'). They received the same instructions as in Wave 1 for viewing sleep data and taking screenshots. Respondents in the UR sample could optionally choose to upload a screenshot of their phone's tracking of their screen time

the previous day, with instructions provided on how to obtain phone usage data and take screenshots for both iOS and Android devices. Subsequently, all respondents answered additional demographic questions (household composition, parental status, body mass index) and completed additional measures on sleep quality and correlates of sleep. The latter data are intended for follow-up studies and are not discussed further in this paper (see preregistration for details). At the end of the Wave 2 survey, all respondents were notified that they would be required to upload a screenshot (of their activity tracker or phone usage data for the RDO and UR sample, respectively) in the next Wave.

2.2.3 Wave 3

In Wave 3, all respondents self-reported their sleep from the previous night one more time. For respondents in the RDO sample, uploading a screenshot of their sleep data from the activity tracker was mandatory. Similarly, for the UR sample, it was mandatory to upload a screenshot of their phone apps usage. Both samples received instructions on how to upload, as in Wave 2.

2.3 Data Analysis

Our survey experiment was designed to assess the feasibility, benefits and drawbacks of collecting sleep data from online samples of device owners. First, to assess the feasibility, we analyze respondents' willingness to upload screenshots of device-reported sleep duration, and the quality of these screenshots. Second, we evaluate the benefits of requiring uploads of these screenshots compared to asking respondents to self-report their sleep duration. Finally, we evaluate two potential drawbacks of collecting device-reported sleep data in online studies: (i) *selection effects* arising from recruiting only device owners, which may reduce external validity, and (ii) *attrition related to uploading*, where respondents may drop out when asked to submit screenshots, thus compromising both sample representativeness and data completeness.

2.3.1 Feasibility of recruiting device owners through online panels to obtain device-reported data on sleep duration

We begin by analyzing the frequency and quality of uploaded device data across samples and Waves. All uploaded screenshots were manually reviewed and scored on 3 criteria using binary indicators. First, we scored if key information — specifically, sleep or phone usage duration — was directly visible from the screenshots. Second, we scored if a detailed summary of sleep or phone usage was provided. Specifically, for the RDO sample we scored if a detailed breakdown into sleep stages was available, whereas for the UR sample we scored if distribution of phone usage across different applications was provided. Third, we scored if it was possible to verify that the data was generated on the correct date or at least the correct day of the week. Screenshots did not meet this criteria when, for example, no day or date information was visible at all, or the screenshot referred to the wrong day of the week or date (compared to survey completion date). Importantly, respondents were not explicitly requested to submit screenshots with detailed information or showing date information. We also identify how many screenshots meet all three criteria.[1]

2.3.2 Benefits of recruiting device owners through online panels to obtain device-reported data on sleep duration

Next, we compared device-reported sleep duration with self-report measures across samples and Waves. Within the RDO sample, we used paired t-tests to assess individual-level discrepancies between device- and self-reported sleep in each Wave. To evaluate sample-level differences, we compared both device- and self-reported sleep in the RDO sample with self-reported sleep in the UR sample using Probit regressions, controlling for demographic variables (which were split into 2 or 3 categories to facilitate interpretation). The following demographic variables were included as covariates in subsequent analyses, with categories defined post hoc and reference groups indicated in parentheses: 1) Gender (3 categories): males (reference), females, and other genders; 2) Age (3 categories): $\leq$ 34 years (reference), 35–44 years, and

[1]Respondents were not explicitly requested to include screenshots that satisfied these criteria, we asked for a screenshot of sleep data from the previous night without any further quality requirements

45+ years; 3) Education (3 categories): below Bachelor's degree (reference), Bachelor's degree, and postgraduate education (Master's or doctorate); 4) Income (3 categories): low (£40,000, reference), medium (£40,000–59,999), and high (£60,000+); 5) Employment (3 categories): other, including students, retired, looking for work, and stay-at-home parents (reference), versus part-time and full-time.

2.3.3 Drawbacks of recruiting device owners through online panels to obtain device-reported data on sleep duration

To identify *selection effects*, we examined how sample composition varied between the RDO and UR samples. Specifically, we compared demographic characteristics using Chi-squared tests on demographic data collected, allowing us to detect systematic differences related to device ownership.

To assess *attrition related to uploading*, we analyzed dropout rates across Waves in relation to when uploading screenshots became mandatory (Wave 2 for half of the RDO sample and Wave 3 for the other half). Using Probit regressions with demographic controls, we tested whether attrition was random or concentrated within specific subgroups, or related to the content of the requested data (sleep vs. phone usage data). We further explored the nature of dropout by distinguishing between respondents who did not start the next Wave (potentially indicating unintentional dropout) and those who started but did not complete it (potentially indicating reluctance to upload data). Finally, within the RDO sample, we identified the *causal* effect of requiring device sleep data uploads by comparing dropout rates between respondents randomly assigned to mandatory versus optional uploading conditions.[2].

[2]In our preregistered analysis plan, we examined whether differences in sleep across Waves and samples could be explained by (i) information effects or (ii) feedback effects from wearing a device (see Online Appendix). To test information effects, we compared changes in sleep duration between Waves 1 and 2 in the RDO and UR samples using Probit regressions controlling for demographics, as only RDO respondents were informed that their sleep would be tracked. To test feedback effects, we classified RDO respondents who wore their device at least four nights per week in Wave 1 as nightwearers and compared their Wave 1–2 changes in sleep duration to those of non-nightwearers using Probit regressions. We find no evidence for either mechanism.

3 Results

3.1 Feasibility of recruiting device owners through online panels to obtain device-reported data on sleep duration

The most frequently owned activity tracker brands were Fitbit, Apple Watch, and Garmin, with 227/76 (RDO/UR) respondents owning Fitbit, 112/78 owning an Apple Watch, and 93/59 owning a Garmin device. A minority (14) owned a Xiaomi device, while 19% of respondents owned a device of another brand[3]. Table 1 shows the number and quality of screenshots uploaded across the Waves and samples, and clearly shows that obtaining device-reported information on sleep (or phone usage) duration is feasible. That is, the proportion of screenshots satisfying criterion 1 is considerable across all samples and Waves (with only 61 out of 949, i.e., 6.4%, screenshots not providing information on sleep or phone usage duration). In total, we obtained 508 screenshots from which sleep duration data could be extracted across three Waves (i.e., 94.6% of all screenshots uploaded by the RDO sample, and 380 screenshots from which phone usage data could be extracted (i.e., 94.6% of all UR uploads). Detailed information on sleep stages or the applications used is provided in a large number of screenshots uploaded across samples and Waves (ranging from 32% to 53.6% of screenshots uploaded). The number of screenshots meeting all 3 criteria simultaneously, however, is low across the Waves and samples, which is mainly driven by the low number of screenshots for which it could be verified that sleep or phone usage data was provided for the correct day (ranging from 9.3% to 33.3% of the screenshots uploaded). In most cases in which criterion 3 was not met, screenshots showed data for 'today', without any date visible on screen, which means the screenshots could have been valid, but could also, in principle, have been taken on a different day.

[3]Looking at differences between devices' brands was not an aim of this study, yet we do find suggestive evidence that respondents who own different devices report different sleep durations. Such differences are absent in Wave 1 and appear in the subsequent Waves, when the device was most likely used for sleep measurement. This suggests differences in how different devices measure sleep, which might limit the comparability of the data among respondents, see more evidence of this in table A1 in the Appendix

A total of 82 RDO (36.6% of Wave 3) and 140 UR (54.3% of Wave 3) respondents uploaded screenshots at each opportunity (regardless of whether it was optional or mandatory). Interestingly, in Waves 1 and 2, 22.4% and 41.5% of screenshots in the RDO were uploaded when uploads were not mandatory. Similarly, in Wave 2, 150 screenshots, i.e., 48% of the sample, were uploaded voluntarily. Screenshots uploaded voluntarily also seemed more likely to meet quality criteria 1. For criterion 1 (i.e., sleep or screen time duration visible), between approximately 94–98% of optional uploads met the criterion, compared with about 91–94% of mandatory uploads (Chi squared, p value =0.03). We also examined whether respondents who uploaded data when uploads were voluntary declined to upload when uploads later became mandatory. This pattern was observed for 9 respondents in the UR sample and 2 respondents in the RDO sample, indicating that such behavior exists but is relatively uncommon. In Appendix Table A2, we show that voluntary uploading is affected by individuals' socioeconomic position (i.e., higher income and education). Yet, Table 1 also shows that willingness to upload screenshots repeatedly is limited: there is substantial attrition between Waves, especially between Waves 1 and 2 where about 25% of the sample drops out (we explore the causes of this attrition further in subsequent sections).

3.2 Benefits of recruiting device owners through online panels to obtain device-reported data on sleep duration

3.2.1 Comparing self-reported and device-reported sleep duration within RDO respondents

There was little evidence for within-individual differences between measures of self-reported sleep duration and device-reported sleep duration. In RDO respondents who provided both, measures of self-reported sleep duration and device-reported sleep duration were strongly positively correlated (Pearson correlations > 0.77), and not significantly different from each other (paired T-tests: all p-values > 0.20). For the majority of pairs of self-reported and device-reported sleep duration, we find that the reports are identical (353 out of 508, i.e., 69.4%). When re-

	Wave 1		Wave 2		Wave 3	
	RDO	**UR**	**RDO**	**UR**	**RDO**	**UR**
Sample size	441	442	253	310	224	258
N (%) uploads	99 (100%)	–	209 (100%)	150 (100%)	229 (100%)	262 (100%)
i. Duration present	96 (97.0%)	–	198 (94.7%)	141 (94.0%)	214 (93.5%)	239 (91.2%)
ii. Detailed breakdown	47 (47.5%)	–	112 (53.6%)	48 (32.0%)	112 (48.9%)	96 (36.6%)
iii. Day/date verifiable	33 (33.3%)	–	69 (33.0%)	14 (9.3%)	75 (32.7%)	39 (14.9%)
% satisfying all 3 criteria	19 (19.2%)	–	36 (17.2%)	0 (0%)	32 (14.0%)	0 (0%)
N (%) voluntary	99 (100%)	–	105 (100%)	150 (100%)	–	–
i. Duration present	96 (97.0%)	–	103 (98.1%)	141 (94.0%)	–	–
ii. Detailed breakdown	47 (47.5%)	–	59 (56.2%)	48 (32.0%)	–	–
iii. Day/date verifiable	33 (33.3%)	–	29 (27.6%)	14 (9.3%)	–	–
% satisfying all 3 criteria	19 (19.2%)	–	21 (20.0%)	0 (0%)	–	–
N (%) mandatory	–	–	104 (100%)	–	229 (100%)	262 (100%)
i. Duration present	–	–	95 (91.3%)	–	214 (93.5%)	239 (91.2%)
ii. Detailed breakdown	–	–	53 (51.0%)	–	112 (48.9%)	96 (36.6%)
iii. Day/date verifiable	–	–	40 (38.5%)	–	75 (32.7%)	39 (14.9%)
% satisfying all 3 criteria	–	–	15 (14.4%)	–	32 (14.0%)	0 (0%)

Table 1: Frequency and quality of uploaded screenshots

Notes: The number of uploads exceeds the number of respondents in some cases because respondents could upload multiple files.

RDO = restricted to device owners sample; UR = unrestricted sample.

ports were not identical, individuals were slightly more likely to under-report sleep with self-report (98 out of 508) than over-report (60 out of 508, i.e., 11.8%) compared to device-reported sleep. These results are also shown in Figure 2, in a Bland-Altman plot, which illustrates the non-systematic degree of over- and under-reporting. The differences between self-reported and device-reported sleep duration were not associated with any RDO respondent characteristics (Appendix Table A3).

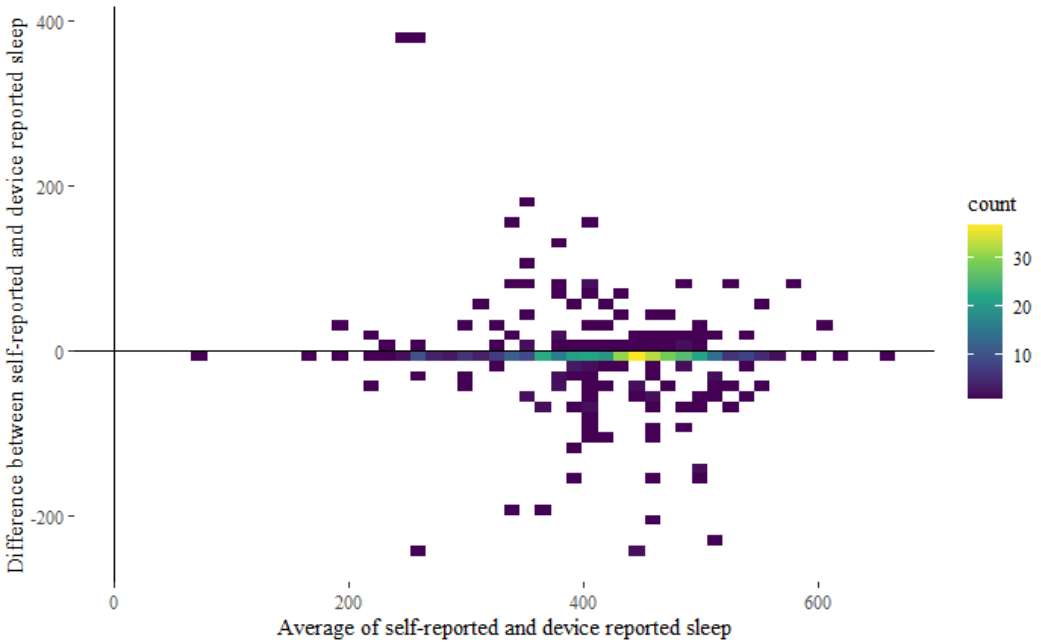

Figure 2: Bland-Altman plot comparing self- and device-reported sleep duration in minutes for the RDO sample

3.2.2 Comparing sleep duration across Waves and samples

We find only small differences across Waves and samples in sleep duration. Figure 3 illustrates the distribution and central tendencies of self-reported sleep across the samples and Waves. Appendix Figure A1 shows a similar distribution plot comparing self-reported sleep in the UR sample with device-reported sleep in the RDO sample. Mean (standard deviation in brackets) self-reported sleep per night in RDO

respondents was 408 (69), 423 (79), 418 (83) minutes in Wave 1, 2 and 3, respectively. In UR respondents, mean self-reported sleep per night was 425 (72), 425 (84), and 438 (79) minutes in Waves 1, 2 and 3, respectively. Within samples, paired comparisons suggest that in the RDO sample, self-reported sleep significantly increased from Wave 1 to Wave 2 (p=0.015), but was not significantly different between Wave 2 and 3 (p=0.36). In the UR sample self-reported sleep increased from Waves 2 to 3 (p=0.011) but did not change significantly between Waves 1 and 2 (p=0.71). These differences may be related to attrition.

Comparing between samples, within Waves 1 and 3 the UR sample self-reports to sleep significantly longer (17-20 minutes) than the RDO sample (t-tests, p-values < 0.01), a result that was only replicated for Wave 3 when comparing device-reported sleep in RDO respondents with self-reported sleep in UR respondents. We also see that in Waves 1 and 3 the proportion of RDO vs. UR respondents with self-reported sleep durations within the recommended ranges of 7 to 9 hours appeared to differ somewhat between the Waves (55% vs 60% in Wave 1, 64% and 62% in Wave 2 and 59% vs 65% in Wave 3). These differences were not statistically significant (Chi-squared test p's > 0.13).

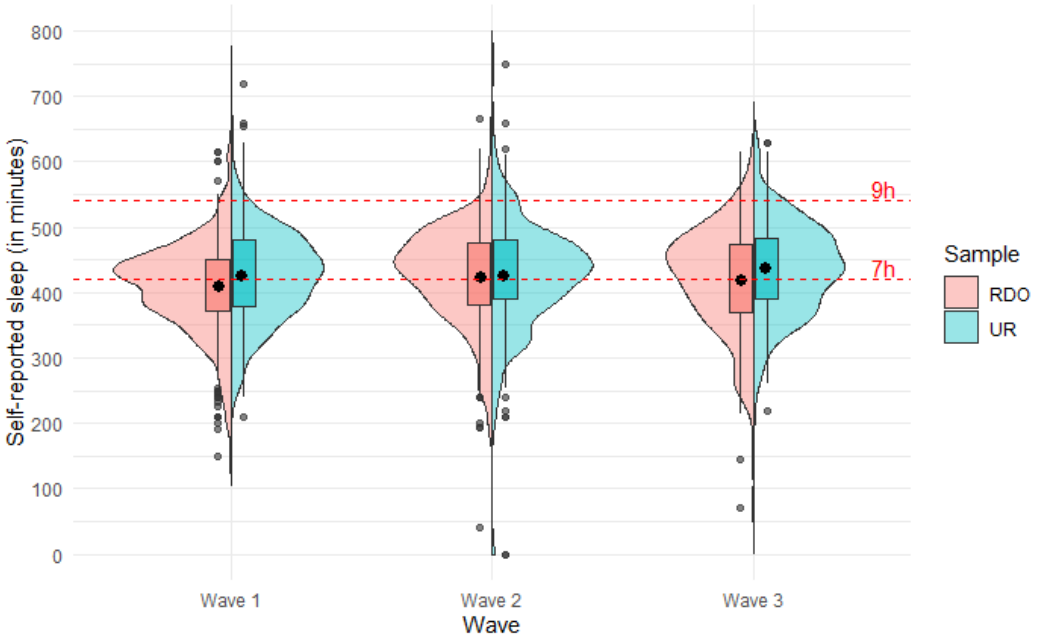

Figure 3: Distribution of self-reported sleep duration by sample and Wave

We substantiated these results with regressions controlling for demographics: respondents in the RDO sample sleep significantly shorter than respondents in the UR sample in Wave 1 and Wave 3, see Table 2. The absence of a significant difference between samples in Wave 2 does not seem to be caused by attrition: we find the same trend both when controlling for demographics and when only including respondents who stay in the sample for all three Waves (models (2) and (4)). In model (6) of Table 2 we use a modified dependent variable to capture sleep, which consists of device-reported sleep for the RDO sample and self-reported sleep for the UR sample, and find that the sign and significance of difference in sleep between the samples in Wave 3 does not depend on the mode of reporting. Furthermore, we find that sleep was associated with age (Wave 1 and 3), income (Wave 1 and 2), employment status (Wave 2) and education (Wave 1). We also looked at whether sleep differed within the UR sample between those that reported owning a device and those who did not (Table A4) and we do not find any significant difference between them across any of the three Waves.

	(1) Wave 1 (Full) Self-report	(2) Wave 1 (Restricted) Self-report	(3) Wave 2 (Full) Self-report	(4) Wave 2 (Restricted) Self-report	(5) Wave 3 Self-report	(6) Wave 3 Self-report (UR) or Device (RDO)
RDO	-17.08***	-15.36**	-4.555	-0.843	-19.68***	-14.78*
	(4.750)	(6.308)	(6.958)	(7.521)	(7.526)	(7.860)
Gender: Female	-0.887	1.976	6.610	9.334	4.960	6.219
	(5.086)	(6.727)	(7.403)	(8.021)	(8.027)	(8.379)
Gender: Other Genders	-1.687	20.70	-43.47	-16.58	28.94	31.78
	(26.54)	(48.03)	(47.17)	(57.28)	(57.32)	(59.23)
Age: 35-44 years	-12.74**	-11.32	-14.61*	-14.94	-28.53***	-28.91***
	(5.762)	(7.619)	(8.427)	(9.085)	(9.091)	(9.474)
Age: 45+ years	-21.58***	-19.78**	-5.969	-6.780	-7.680	-11.88
	(5.805)	(7.813)	(8.574)	(9.316)	(9.323)	(9.717)
Education: Bachelor	9.576*	13.27*	3.564	3.304	15.47*	18.71**
	(5.418)	(7.132)	(7.881)	(8.504)	(8.510)	(8.914)
Education: Postgraduate education	13.11**	11.35	17.88*	17.39*	-6.530	2.047
	(6.347)	(8.506)	(9.373)	(10.14)	(10.15)	(10.59)
Income: Middle	11.52*	17.13**	18.62**	18.10*	2.158	0.455
	(6.156)	(7.924)	(8.884)	(9.449)	(9.455)	(9.861)
Income: High	7.734	13.95*	18.29**	22.47**	6.574	-2.447
	(5.895)	(7.740)	(8.515)	(9.229)	(9.236)	(9.609)
Employment: Part-time	-11.18	-12.87	-32.28***	-28.43**	-17.41	-19.15
	(7.851)	(10.18)	(11.31)	(12.14)	(12.15)	(12.65)
Employment: Full-time	4.714	-0.0380	-14.62	-14.29	-2.160	-8.114
	(6.661)	(8.552)	(9.493)	(10.20)	(10.20)	(10.63)
Constant	423.0***	421.3***	427.3***	421.9***	443.9***	448.3***
	(8.268)	(10.91)	(12.02)	(13.00)	(13.01)	(13.59)
Observations	883	482	565	482	482	472
R-squared	0.059	0.065	0.044	0.043	0.059	0.050

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Linear Regression Results for Self-reported Sleep Duration by Sample and Wave, Controlling for Demographics.

Note: Columns (1) and (2) present analyses for Wave 1, where (1) uses the full sample, and (2) is restricted to those who completed all Waves. Columns (3) and (4) present analyses for Wave 2, with (3) using the full sample and (4) restricted to those who completed all Waves.

3.3 Drawbacks of recruiting device owners through online panels to obtain device-reported data on sleep duration

3.3.1 Selection Effects

Table 3 shows that, besides showing differences in sleep duration at baseline in Wave 1, the two samples differ substantially in terms of their demographic composition. In Wave 1, individuals in the RDO sample are older, have a higher income, and are more likely to be full-time employed than those in the UR sample. In Wave 2, the two samples differ only in terms of age. Noticeably, however, individuals in the RDO sample are more likely to have children. In Wave 3, some differences between the two samples reappear for income and employment. These findings are supported by multivariate regressions, as shown in Table A5 in the Appendix. These selection effects appear to be driven by (selection on) device-ownership: when dividing the UR sample between those owning a device and those who do not, we find that the subsample of UR owning a device differs from the UR subsample that does not own a device, but UR device owners do not differ in observables from the RDO sample, see Table A6.

3.3.2 Attrition related to uploading

To identify the potential causes of the substantial attrition between Waaves, in Tables 4 and A7 we distinguish between respondents who drop out from the study before the start of a new Wave (i.e., not starting the survey at all) and during the Wave itself (i.e., not completing the survey after starting it).

We find some differences between the samples (Table A7). In particular, drop-out between completing Wave 1 and starting Wave 2 is slightly, but not significantly, larger in the RDO sample (who are informed at the end of Wave 1 that they may be required to upload) compared to the UR sample. Additionally, in the UR sample only 21 (6%) respondents who started Wave 2 did not complete it, while in the RDO sample 63 (20%) respondents (of which 47 were randomly assigned to be required to upload device data) dropped out while completing Wave 2.

	UR			RDO			Difference		
	W1	W2	W3	W1	W2	W3	W1	W2	W3
Total	442	310	258	441	253	224			
Gender									
Male	159 (35.97%)	114 (36.77%)	91 (35.27%)	145 (32.88%)	83 (32.81%)	74 (33.04%)	*	ns	ns
Female	277 (62.67%)	193 (62.26%)	165 (63.95%)	295 (66.89%)	170 (67.19%)	150 (66.96%)			
Other Genders	6 (1.36%)	3 (0.97%)	2 (0.78%)	1 (0.23%)	0 (0.00%)	0 (0.00%)			
Age									
$\leq$ 34 years	175 (39.59%)	120 (38.71%)	104 (40.31%)	143 (32.43%)	76 (30.04%)	64 (28.57%)	**	**	***
35-44 years	140 (31.67%)	101 (32.58%)	86 (33.33%)	135 (30.61%)	80 (31.62%)	71 (31.70%)			
45+ years	127 (28.73%)	89 (28.71%)	68 (26.36%)	163 (36.96%)	97 (38.34%)	89 (39.73%)			
Education									
Below Bachelor's degree	171 (38.69%)	114 (36.77%)	91 (35.27%)	172 (39.00%)	95 (37.55%)	84 (37.50%)	ns	ns	ns
Bachelor's degree	171 (38.69%)	130 (41.94%)	111 (43.02%)	172 (39.00%)	99 (39.13%)	89 (39.73%)			
Postgraduate education	100 (22.62%)	66 (21.29%)	56 (21.71%)	97 (22.00%)	59 (23.32%)	51 (22.77%)			
Income									
Low	188 (42.53%)	127 (40.97%)	108 (41.86%)	145 (32.88%)	86 (33.99%)	78 (34.82%)	**	ns	*
Medium	109 (24.66%)	75 (24.19%)	64 (24.81%)	124 (28.12%)	72 (28.46%)	68 (30.36%)			
High	145 (32.81%)	108 (34.84%)	86 (33.33%)	172 (39.00%)	95 (37.55%)	78 (34.82%)			
Employment									
Other	96 (21.72%)	64 (20.65%)	53 (20.54%)	69 (15.65%)	46 (18.18%)	40 (17.86%)	***	ns	*
Part-time	90 (20.36%)	66 (21.29%)	57 (22.09%)	71 (16.10%)	39 (15.42%)	35 (15.62%)			
Full-time	256 (57.92%)	180 (58.06%)	148 (57.36%)	301 (68.25%)	168 (66.40%)	149 (66.52%)			
Has children		173 (55.81%)	144 (55.81%)		161 (63.64%)	141 (62.95%)		**	*

*** p<0.01, ** p<0.05, * p<0.1, ns=non significant χ^2 test

Table 3: Demographics by Sample and Wave

Taken together, moving from Wave 1 to Wave 2 leads to a much higher drop out for the RDO sample than for the UR sample. Something similar happens from Wave 2 to Wave 3: the proportion of respondents dropping out is higher for the UR than the RDO sample, both before starting the Wave (6.8% and 3.6% for UR and RDO respectively) and during the Wave itself (10.7% and 8.2%). The fact that dropout is relatively higher when mandatory uploading is introduced (between Wave 1 and 2 for RDO and between Wave 2 and 3 for UR) suggests that attrition might be due to mandatory uploading.

These results are substantiated with Probit regressions reported in Table 4, in which we also control for demographics. Being in the RDO group increases the probability of dropping out during Wave 2, when it is mandatory for some respondents to upload screenshots (model 2). Furthermore, part-time and full-time workers, and older individuals were more likely to drop out before Wave 2. We also find that respondents with higher self-reported sleep in the previous Wave (i.e., SR Sleep W1 and SR Sleep W2) are less likely to drop out before a Wave. Finally, Table A8 in the Appendix shows that in the RDO sample, females and those with a Bachelor's degree are less likely to drop out when uploading is mandatory compared to those with another education level and other genders. Finally, model (5) in Table 4 confirms the causality of drop-out related to mandatory uploads. Here, drop out is significantly higher for respondents randomly allocated to the mandatory group compared to respondents allocated to the non-mandatory group. Specifically, being in the mandatory group increases the probability of dropping out by 22 percentage points compared to being in the non-mandatory group, holding all other variables constant. This shows clear *causal* evidence that mandatory upload leads to attrition.

4 Discussion

Recently, experiments on improving sleep, rather than relying on potentially biased self-reported measures of sleep, have employed the costly strategy of supplying study respondents with activity tracking devices [Avery et al., 2022, Bessone et al., 2021, Ong et al., 2023]. This paper reported an analysis of the feasibility, benefits, and drawbacks of collecting sleep data from an online sample of respondents who already own

	(1) Drop out before W2	(2) Not finishing W2	(3) Drop out before W3	(4) Not finishing W3	(5) Not finishing W2 (RDO)
RDO	0.0867	0.679***	-0.306	-0.180	
	(0.0940)	(0.140)	(0.194)	(0.158)	
Mandatory					0.883***
					(0.179)
Gender: Female	0.00147	0.0828	-0.207	-0.0637	-0.0155
	(0.100)	(0.142)	(0.193)	(0.165)	(0.184)
Gender: Other Genders	0.467	0.918	0.779		
	(0.487)	(0.709)	(0.790)		
Age: 35-44 years	-0.127	-0.0901	-0.142	-0.0221	-0.382
	(0.112)	(0.172)	(0.224)	(0.197)	(0.236)
Age: 45+ years	-0.233**	0.123	-0.173	0.200	-0.0256
	(0.115)	(0.164)	(0.228)	(0.192)	(0.216)
Education: Bachelor's degree	-0.153	-0.102	-0.201	-0.196	-0.136
	(0.107)	(0.152)	(0.213)	(0.179)	(0.198)
Education: Postgraduate education	-0.0348	-0.0824	-0.203	-0.111	-0.192
	(0.124)	(0.178)	(0.252)	(0.205)	(0.243)
Income: Middle	-0.0363	0.0990	-0.0312	-0.0912	0.00850
	(0.122)	(0.169)	(0.267)	(0.211)	(0.227)
Income: High	-0.0437	0.000807	0.420*	0.249	0.162
	(0.116)	(0.171)	(0.227)	(0.190)	(0.233)
Employment: Part-time	0.322**	-0.510**	-0.256	-0.0144	-0.518
	(0.157)	(0.241)	(0.322)	(0.256)	(0.334)
Employment: Full-time	0.245*	-0.120	-0.113	-0.0158	-0.0385
	(0.136)	(0.177)	(0.252)	(0.217)	(0.257)
SR Sleep W1	-0.00140**	0.000155			
	(0.000670)	(0.000939)			
SR Sleep W2			-0.00224**	0.00155	
			(0.00107)	(0.00101)	
Constant	-0.104	-1.497***	-0.291	-1.884***	-1.095***
	(0.325)	(0.469)	(0.533)	(0.535)	(0.343)
Observations	883	647	563	531	316
Pseudo R-squared	0.017	0.076	0.070	0.029	0.106

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Probit regressions, Likelihood of Dropping out at different stages, dependent on Demographics and Treatment Group

Note: SR Sleep W1 and SR Sleep W2 refer to self-reported sleep in Wave 1 and Wave 2 respectively.

such devices. Our results, collectively, suggest that obtaining device-reported data on sleep duration through online panels is feasible, with the large majority of screenshots uploaded from devices allowing interpretation of sleep duration. Obtaining such data longitudinally appears possible, but drop-out from between Waves in our survey experiment was considerable. Differences with self-reported sleep are small to non-existent, both within device owners that both self-report and uploads device data but also when comparing to a sample that only self-reports sleep. We find evidence that two potential drawbacks of collecting sleep data from an online sample of device owners apply. First, restricting the sample to people who own devices may lead to considerable sample selection effects. Second, our survey experiment provides causal evidence that shows that requiring the uploading of device-reported data *increases* drop-out.

Respondents in our study uploaded screenshots from their mobile or wearable devices, from which we aimed to extract estimates of sleep duration or phone usage. In most cases, we could do so, with minimal differences in extraction success between sleep and phone usage screenshots. Screenshots often included breakdowns into sleep stages or app usage, suggesting that researchers studying, for example, sleep quality defined as the number of awakenings or duration of deep sleep stages [De Zambotti et al., 2019] may benefit from collecting such data from online samples. However, most screenshots did not indicate whether the data reflected the correct day (instead in most cases simply referring to today). In principle, respondents could have submitted screenshots from other days or different users' devices. Future studies may consider implementing strategies to confirm that data corresponds to the correct user and date (e.g., requiring uploading of video-based screen-captures that also show the correct date). Interestingly, data quality was highest among the many voluntarily submitted screenshots. We also find little evidence for differences in feasibility between uploading screenshots showing sleep and screentime duration, suggesting that our results may also inform about the feasibility of uploading screenshots of other device-bases measures (e.g., step count). Overall, our findings suggest that asking online respondents to upload screenshots on sleep duration is a feasible data collection method.

Our survey experiment was also designed to assess the feasibility of col-

lecting device-reported data in longitudinally. While previous research has reported high retention rates on Prolific for long-term surveys (61% after 19 months), comparable to conventional cohort studies [Kothe and Ling, 2019], we observed lower retention over just 14–15 days: 51% in the RDO sample and 59% in the UR sample. Two main factors likely contributed to this drop-out. First, to optimize recall in self-reports, we asked respondents to report sleep duration from the previous night. Each survey Wave launched on a Wednesday morning and closed after approximately 24 hours, allowing us to capture weekday-specific variation in sleep patterns (e.g., longer sleep on weekends [Wittmann et al., 2006]). This short response window contrasts with the longer timeframes used in Kothe and Ling [2019]. Second, non-response appeared selective: respondents who were employed, had a higher income, or reported longer sleep in the prior Wave were less likely to start subsequent surveys. This suggests that job demands may interfere with survey participation, completion, and/or reduce sensitivity to financial incentives [Singer and Ye, 2013].

Overall, our estimates of mean self-reported sleep duration (6.8 to 7.3 hours, depending on Wave and sample) align with prior research [Kocevska et al., 2021]. However, unlike earlier large-scale studies [Kocevska et al., 2021, Liu et al., 2016], only a slight majority of respondents (51–61%) reported sleep durations within the recommended 7–9 hour range. Prior work has documented discrepancies of 0.77 to 1.5 hours between self-reported and actigraphy-based sleep measures [Lauderdale et al., 2008, Jackson et al., 2018], but we find little to no systematic difference between self-reported and device-reported durations. Among respondents who provided both self-reports and sleep screenshots, the two measures closely aligned, with no consistent pattern of over- or under-reporting. This suggests that within a sample of device owners, self-reports are reasonably accurate—even without prompting respondents to consult their devices. This accuracy may reflect either direct use of device feedback when self-reporting or improved estimation skills due to long-term device use. Notably, we did find that sleep duration in the RDO sample was significantly lower in Wave 1 (self-report only) and Wave 3 (both self- and device-reported) than self-reported sleep in the UR sample. Although the direction of this difference is consistent with prior findings [Lauderdale et al., 2008, Jackson et al., 2018], in our study it is likely related to selection, as it

is already observed at baseline in Wave 1.

The (lack of) differences observed in sleep duration between the samples may not be solely attributable to measurement mode, as restricting the sample to device owners introduces two potential drawbacks that complicate interpretation. First, selection effects apply and are the primary explanation for pre-existing differences in sleep duration in Wave 1. Compared to the UR sample, RDO respondents were more likely to be middle-aged, employed full-time, have higher income, and to have children. Even among self-reports in Wave 1, sleep duration was lower in the RDO sample—by approximately 17 minutes—also after controlling for demographic differences. This aligns with earlier research showing that wearable device users differ in health-related behaviors such as physical activity [Yen et al., 2022]. We also observed that respondents with higher sleep durations in Waves 1 and 2 were less likely to complete later Waves, further raising concerns about potential upward bias in sleep estimates due to selective retention. One possible explanation is that those who missed their opportunity to complete the survey were catching up on sleep.

Second, our survey experiment provides causal evidence showing that attrition increases when uploading screenshots is required. We randomized respondents in the RDO sample to either voluntarily or mandatorily upload screenshots in Wave 2: and drop-out increases substantially in those randomly required to upload. This selective attrition may explain why no differences in sleep duration were observed between the samples in Wave 2. That is, because in Wave 2, half of the RDO respondents were required to upload screenshots a considerable proportion of the sample drops out. This may suggest that attrition due to uploading obscures pre-existing differences between the two samples. In Wave 3, the UR sample is required for the first time in the current study to upload, and a large proportion of that sample drops out, and we observe a significant difference in measures of sleep duration, similar in size to that found in Wave 1. There are several reasons why mandatory uploads could lead to attrition, including technical difficulties or inexperience with uploading, and privacy concerns [Keusch et al., 2025]. Our work provides little insight which of these causes drives our result, but we do find that drop-out related to mandatory uploading is higher in males and lower-educated respondents. This finding highlights a signifi-

cant challenge in collecting such data longitudinally, requiring repeated uploading. Given these challenges, researchers interested in using this method should explore solutions to mitigate attrition, such as offering additional incentives for uploads, simplifying the upload procedure, and emphasizing the importance of contributing this data. Alternative strategies to explore could be providing even more detailed step-by-step explanations (e.g. making 'wizard' videos showing the uploading procedure to counteract inexperience), and/or providing material that could be reassuring to respondents with privacy concerns

The findings of this study should be interpreted in light of several limitations. First, even without restricting the sample to device-owners, our UR sample is not representative of the UK population. We find over-representation of female respondents. Our sample also appears skewed towards individuals of middle-age (between 25-44). Whereas in 2021 18.6% of the population in the UK was 65+, our unrestricted sample only included about 4% of respondents in this age category. Without restrictions, our sample was also highly educated, with 61% having attained a Bachelors/Graduate degree. This may be a consequence of Prolific's high standards for survey completion and reponse quality as shown in previous work comparing panels [Stagnaro et al., 2024], and/or the general tendency for older individual to engage less in online activities [Turner et al., 2021].

Second, although our sample size was based on a priori power analysis and we oversampled, some of the small and statistically insignificant differences in sleep duration we observed may reflect true differences that we were not powered to detect. Although it is questionable if differences in sleep duration smaller than 15 minutes are clinically meaningful, future work may aim to use larger sample sizes. Third, our study compared a sample restricted to device-owners to an unrestricted sample, of whom the slight majority also owned activity trackers. The unrestricted sample was included to identify what the effect was of screening online surveys to include device owners *only*. It may also be of interest to explore differences in sleep and sleep-related characteristics between individuals who own or do not own wearable devices. Yet, our study was not (pre-registered or) powered for that comparison. Fourth, our study design aimed to identify effects of making uploading of device-reported data mandatory. As such, the only incentive for uploading this data

was the fee paid out on study completion, which was not contingent on the quality of the uploaded device data. Respondents could (and in a minority of cases did) upload data that was irrelevant, and in many cases it is not straightforward to identify whether a screenshot of sleep duration for the correct night was uploaded. If this approach would be extended to future work, additional checks on the veracity of the screenshots would be needed.

Fifth, our study did not restrict the RDO sample to people owning devices of a specific brand or type. Although we find little differences in reported sleep duration in Wave 1 across people owning different device brands, we find suggestive evidence that in subsequent Waves differences exist. This suggests that different device types or brands may record sleep duration differently [Lee et al., 2023], which limits the comparability of sleep measures across subjects. Researchers should consider accounting for device type in their methodology analysis - either by including device fixed effects or restricting comparisons to a single device type when comparability is essential.

Sixth, the online panel Prolific only allows screening of respondents based on a questionnaire they complete when registering. There is no strong incentive for participants keep this information up-to-date afterward. This may have affected our study results. In Wave 1, we observed differences in sleep duration between the RDO and UR samples, but we did not observe differences between individuals who owned a device and those who did not within the UR sample (where device ownership was measured after respondents were included). A likely explanation is that, in the UR sample, some participants may have purchased an activity tracker after registering on Prolific. Consequently, device owners in the UR sample may have had their devices for a relatively short time on average, which could explain the absence of differences within this group. Our data, however, does not enable testing this conjecture.

In conclusion, our study shows that collecting high-quality device-reported sleep data through online panels is feasible, providing affordable and timely access to objective measures. However, differences between device-reported and self-reported sleep duration were small to non-existent, indicating that self-reports among device owners may often be sufficient for studies focused solely on duration. Device data become most valuable when richer measures—such as sleep stages, qual-

ity, or night-to-night variation—are needed, but this value will likely only materialize if studies include strategies for verifying the screenshots. Researchers should also weigh potential drawbacks: restricting recruitment to device owners reduces representativeness, and requiring mandatory uploads increases the already substantial drop-out. Overall, the decision to collect device data should balance the added value of detailed sleep information against the costs in representativeness and retention, and should be accompanied by procedures to ensure that uploaded data are verifiable.

References

Sonia Ancoli-Israel, Roger Cole, Cathy Alessi, Mark Chambers, William Moorcroft, and Charles P Pollak. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26(3):342–392, 2003.

Sonia Ancoli-Israel, Jennifer L Martin, Terri Blackwell, Luis Buenaver, Lianqi Liu, Lisa J Meltzer, Avi Sadeh, Adam P Spira, and Daniel J Taylor. The sbsm guide to actigraphy monitoring: clinical and research applications. *Behavioral sleep medicine*, 13(sup1):S4–S38, 2015.

Mallory Avery, Osea Giuntella, and Peiran Jiao. Why don't we sleep enough? a field experiment among college students. *Review of Economics and Statistics*, pages 1–45, 2022.

Fee Benz, Dieter Riemann, Katharina Domschke, Kai Spiegelhalder, Anna F Johann, Nathaniel S Marshall, and Bernd Feige. How many hours do you sleep? a comparison of subjective and objective sleep duration measures in a sample of insomnia patients and good sleepers. *Journal of Sleep Research*, 32(2):e13802, 2023.

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

Vijay Kumar Chattu, Md Dilshad Manzar, Soosanna Kumary, Deepa Burman, David Warren Spence, and Seithikurippu R Pandi-Perumal. The global problem of insufficient sleep and its serious public health implications. In *Healthcare*, volume 7, page 1. MDPI, 2018.

Massimiliano De Zambotti, Nicola Cellini, Aimee Goldstone, Ian M Colrain, and Fiona C Baker. Wearable sleep technology in clinical and research settings. *Medicine and science in sports and exercise*, 51(7):1538, 2019.

Patrick Dunn and Eric Hazzard. Technology approaches to digital health literacy. *International journal of cardiology*, 293:294–296, 2019.

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

Jennifer Girschik, Lin Fritschi, Jane Heyworth, and Flavie Waters. Validation of self-reported sleep against actigraphy. *Journal of epidemiology*, 22(5):462–468, 2012.

Osea Giuntella and Fabrizio Mazzonna. Sunset time and the economic effects of social jetlag: evidence from us time zone borders. *Journal of health economics*, 65: 210–226, 2019.

Osea Giuntella, Silvia Saccardo, and Sally Sadoff. Sleep: Educational impact and habit formation. Technical report, National Bureau of Economic Research, 2024.

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

Zawar Hussain, Quan Z Sheng, Wei Emma Zhang, Jorge Ortiz, and Seyedamin Pouriyeh. Non-invasive techniques for monitoring different aspects of sleep: A comprehensive review. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(2):1–26, 2022.

Chandra L Jackson, Sanjay R Patel, W Braxton Jackson, Pamela L Lutsey, and Susan Redline. Agreement between self-reported and objectively measured sleep duration among white, black, hispanic, and chinese adults in the united states: Multi-ethnic study of atherosclerosis. *Sleep*, 41(6):zsy057, 2018.

Lawrence Jin and Nicolas R Ziebarth. Sleep, health, and human capital: Evidence from daylight saving time. *Journal of Economic Behavior & Organization*, 170: 174–192, 2020.

Florian Keusch, Bella Struminskaya, Joris Mulder, and Stein Jongerius. Data donation as a method to measure physical activity in older adults: Cross-sectional web survey assessing consent rates, donation success, and bias. *Journal of Medical Internet Research*, 27:e69799, 2025.

Desana Kocevska, Thom S Lysen, Aafje Dotinga, M Elisabeth Koopman-Verhoeff, Maartje PCM Luijk, Niki Antypa, Nienke R Biermasz, Anneke Blokstra, Johannes Brug, Wiliam J Burk, et al. Sleep characteristics across the lifespan in 1.1 million people from the netherlands, united kingdom and united states: a systematic review and meta-analysis. *Nature human behaviour*, 5(1):113–122, 2021.

Emily Kothe and Mathew Ling. Retention of participants recruited to a multi-year longitudinal study via prolific. *Preprint*, 2019.

TeYang Lau, Ju Lynn Ong, Ben KL Ng, Lit Fai Chan, Daphne Koek, Chuen Seng Tan, Falk Müller-Riemenschneider, Karen Cheong, Stijn AA Massar, and

Michael WL Chee. Minimum number of nights for reliable estimation of habitual sleep using a consumer sleep tracker. *Sleep Advances*, 3(1):zpac026, 2022.

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

Taeyoung Lee, Younghoon Cho, Kwang Su Cha, Jinhwan Jung, Jungim Cho, Hyunggug Kim, Daewoo Kim, Joonki Hong, Dongheon Lee, Moonsik Keum, et al. Accuracy of 11 wearable, nearable, and airable consumer sleep trackers: prospective multicenter validation study. *JMIR mHealth and uHealth*, 11(1):e50983, 2023.

Yong Liu, Anne G Wheaton, Daniel P Chapman, Timothy J Cunningham, Hua Lu, and Janet B Croft. Prevalence of healthy sleep duration among adults—united states, 2014. *Morbidity and Mortality Weekly Report*, 65(6):137–141, 2016.

Matthew R Lujan, Ignacio Perez-Pozuelo, and Michael A Grandner. Past, present, and future of multisensory wearable technology to monitor sleep and circadian rhythms. *Frontiers in digital health*, 3:721919, 2021.

Miguel Marino, Yi Li, Michael N Rueschman, John W Winkelman, JM Ellenbogen, Jo M Solet, Hilary Dulin, Lisa F Berkman, and Orfeu M Buxton. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, 36(11):1747–1755, 2013.

Ju Lynn Ong, Stijn AA Massar, TeYang Lau, Ben KL Ng, Lit Fai Chan, Daphne Koek, Karen Cheong, and Michael WL Chee. A randomized-controlled trial of a digital, small incentive-based intervention for working adults with short sleep. *Sleep*, 46(5):zsac315, 2023.

Eric Jonathon Osborne-Christenson. Saving light, losing lives: How daylight saving time impacts deaths from suicide and substance abuse. *Health economics*, 31: 40–68, 2022.

Stefan Palan and Christian Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

Gautam Rao, Susan Redline, Frank Schilbach, Heather Schofield, and Mattie Toma. Informing sleep policy through field experiments. *Science*, 374(6567):530–533, 2021.

Ronaldo B Santos, Soraya Giatti, Aline N Aielo, Wagner A Silva, Barbara K Parise, Lorenna F Cunha, Silvana P Souza, Airlane P Alencar, Paulo A Lotufo, Isabela M Bensenor, et al. Prevalence and predictors of under or overestimation sleep duration in adults: The elsa-brasil study. *Sleep Epidemiology*, 1:100013, 2021.

Eleanor Singer and Cong Ye. The use and effects of incentives in surveys. *The annals of the American academy of political and social science*, 645(1):112–141, 2013.

Svetlana Smirnova. "you don't own a fitbit, the fitbit owns you": A taxonomy of privacy attitudes in the context of self-quantification. *Communication and Health: Media, Marketing and Risk*, pages 253–270, 2022.

Michael Nicholas Stagnaro, James Druckman, Adam J Berinsky, Antonio Alonso Arechar, Robb Willer, and David Rand. Representativeness versus attentiveness: A comparison across nine online survey samples. *PsyArXiv*, 2024.

The Lancet Diabetes & Endocrinology. Sleep: a neglected public health issue. *The Lancet Diabetes Endocrinology*, 12(6):365, 2024.

Anne M Turner, Thomas Engelsma, Jean O Taylor, Rashmi K Sharma, and George Demiris. Recruiting older adult participants through crowdsourcing platforms: Mechanical turk versus prolific academic. In *AMIA Annual Symposium Proceedings*, volume 2020, page 1230, 2021.

Marc Wittmann, Jenny Dinich, Martha Merrow, and Till Roenneberg. Social jetlag: misalignment of biological and social time. *Chronobiology international*, 23(1-2): 497–509, 2006.

Hsin-Yen Yen, Yung Liao, and Hao-Yun Huang. Smart wearable device users' behavior is essential for physical activity improvement. *International Journal of Behavioral Medicine*, 29(3):278–285, 2022.

5 Appendix

Respondents were not required to own any particular device in our study, and consequently, device type or brand may affect measures of sleep duration. To explore this, in Table A1, we report linear regressions of sleep duration measured either through self-reports or device reports as dependent, with the type of tracker the respondents report to own as independent variable (with owning an Apple Watch as the reference). We find that, whereas in Wave 1, when devices are less likely to be used to inform self-reports, only a small significant difference in sleep duration is present for Fitbit, sizable differences in sleep duration appear in Waves 2 and 3. Compared to those owning Apple Watches, respondents owning a Garmin report higher estimates, whereas those owning a Fitbit have lower estimates (for Wave 3, DR). This suggests that different types of devices might report sleep somewhat differently. Finally, notice that, controlling for demographic variables, the differences among device types do not change (results available on request).

Table A1: Sleep Outcomes by Tracker Brand

	(1) Wave 1 Self-report	(2) Wave 2 Self-report	(3) Wave 2 Device-report	(4) Wave 3 Self-report	(5) Wave 3 Device-report
Fitbit	-11.57*	-6.13	1.62	-13.88	-33.11**
	(6.79)	(9.89)	(14.73)	(11.38)	(16.67)
Garmin	9.22	35.77**	62.64***	28.36*	26.78
	(9.45)	(14.18)	(19.29)	(15.86)	(22.10)
Other	-0.67	-0.17	33.87*	4.00	-6.34
	(8.10)	(12.11)	(18.73)	(13.85)	(20.62)
Xiaomi Mi Band	5.97	-15.10	-19.36	59.50**	32.08
	(19.65)	(26.80)	(45.53)	(29.22)	(51.90)
Constant	417.82***	424.10***	417.69***	426.84***	437.26***
	(5.31)	(7.84)	(12.63)	(9.34)	(14.58)
Observations	687	421	196	359	214
R-squared	0.010	0.025	0.085	0.039	0.061

Notes: OLS regressions. The omitted reference category is the Apple Watch tracker type. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

To explore who uploads when it is not strictly required, in Table A2, we restrict to the sample to the respondents for which uploading was not mandatory. We run a Probit regression separately for three groups, in which the outcome variable took a variable of 1 if the respondent uploaded device data and 0 otherwise. Column 1 shows this analysis for all RDO respondents in Wave 1, Column 2 for the RDO respondents in Wave 2 randomized to the non-mandatory uploading, and Column 3 for all respondents in the UR sample in Wave 2 (i.e.,Optional W1 RDO, Optional W2 RDO, and Optional W3 UR). We find some evidence that socio-economic status influences voluntary uploading in RDO respondents, but the effects of education and job status (which increase propensity to upload), and income (which decreases propensity to upload) is in opposite directions. In UR respondents, the only demographic that affects voluntary uploading (negatively) is age, suggesting that older individuals were less likely to upload.

	(1) Optional W1 RDO	(2) Optional W2 RDO	(3) Optional W2 UR
Gender: Female	0.0548	-0.104	-0.199
	(0.146)	(0.226)	(0.154)
Age: 35-44 years	0.187	-0.0639	-0.0294
	(0.176)	(0.263)	(0.168)
Age: 45+	0.306*	-0.0128	-0.343*
	(0.172)	(0.264)	(0.176)
Education: Bachelor's degree	0.269*	0.414*	-0.0508
	(0.156)	(0.235)	(0.164)
Education: Postgraduate education	0.211	0.506*	-0.115
	(0.190)	(0.261)	(0.194)
Income: Middle	-0.195	-0.0266	-0.0180
	(0.176)	(0.269)	(0.183)
Income: High	-0.439**	-0.456*	0.00827
	(0.177)	(0.256)	(0.172)
Employment: Part-time	0.308	0.602	0.274
	(0.251)	(0.385)	(0.223)
Employment: Full-time	0.443**	0.309	0.0650
	(0.215)	(0.290)	(0.188)
Constant	-1.260***	0.00463	0.0527
	(0.269)	(0.394)	(0.231)
Observations	440	169	331
Pseudo R-squared	0.025	0.049	0.016

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2: Probit regression, Likelihood of Uploading Screenshot when Not-Mandatory

In Table A3, we test whether the differences between self-reported and device-reported sleep duration depend on demographics. We restricted our analyses to RDO respondents that contributed both self-reported and device-reported sleep duration data. We run three linear regressions for the three Waves. The outcome variable is defined as the difference in sleep duration between SR and DR sleep duration. We find no association between the outcome and demographics, with the exception of part-time workers in Wave 2, who seem to display a greater difference between SR and DR.

	(1) W1Difference	(2) W2Difference	(3) W3Difference
Gender: Female	-10.34	-2.000	0.437
	(8.290)	(5.875)	(8.551)
Age: 35-44 years	5.260	2.164	0.203
	(10.68)	(7.152)	(10.56)
Age: 45+ years	4.680	-2.764	7.444
	(9.803)	(6.905)	(10.24)
Education: Bachelor's degree	-1.904	7.125	-6.366
	(9.158)	(6.391)	(9.118)
Education: Postgraduate education	1.253	8.425	-13.92
	(11.44)	(7.679)	(10.90)
Income: Middle	1.415	9.457	5.093
	(9.985)	(6.974)	(10.05)
Income: High	6.203	-6.449	20.29*
	(10.52)	(7.270)	(10.37)
Employment: Part-time	18.02	19.63**	9.541
	(15.12)	(9.850)	(13.97)
Employment: Full-time	0.789	13.92*	17.63
	(12.37)	(8.263)	(11.43)
Constant	-4.546	-20.05*	-24.41
	(16.34)	(10.49)	(15.26)
Observations	96	196	214
R-squared	0.056	0.060	0.046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A3: Linear Regression: Differences between SR and DR on Demographics

In our main results we find some evidence for differences in sleep between the RDO and UR samples, which differed on a priori sample selection rules related to device ownership, suggesting that differences between the samples relate to device ownership. The UR sample, however, also contains a sizable group of device-owners. In Table A4, we tested whether sleep differed within the UR sample between those that owned a device and those who did not (i.e.,URowners takes a value of 1 for UR respondents owning a device and 0 otherwise). We run separate regressions for the three Waves, also restricting the sample to those individuals that completed all three Waves (columns 2 and 4). The outcome variable is self-reported sleep duration. We find no significant difference across the three Waves and in any of our models.

	(1) Wave 1 (Full) Self-report	(2) Wave 1 (Restricted) Self-report	(3) Wave 2 (Full) Self-report	(4) Wave 2 (Restricted) Self-report	(5) Wave 3 Self-report
URowners	-1.887	-0.995	8.192	10.06	2.052
	(6.939)	(8.690)	(10.09)	(11.03)	(10.49)
Gender: Female	-3.734	-0.980	7.270	9.835	8.544
	(7.280)	(9.100)	(10.37)	(11.55)	(10.98)
Gender: Other Genders	-15.80	17.38	-34.59	-4.765	39.95
	(29.15)	(47.41)	(49.13)	(60.18)	(57.22)
Age: 35–44 years	-15.92**	-10.12	-17.95	-21.39*	-37.05***
	(8.019)	(9.705)	(11.42)	(12.32)	(11.71)
Age: 45+ years	-36.00***	-27.21**	-8.085	-12.14	-4.320
	(8.313)	(10.48)	(11.89)	(13.31)	(12.65)
Education: Bachelor's degree	4.022	10.67	-6.497	-2.595	19.35*
	(7.779)	(9.638)	(11.09)	(12.23)	(11.63)
Education: Postgraduate education	19.19**	20.64*	21.74	22.70	9.276
	(9.011)	(11.44)	(13.23)	(14.53)	(13.81)
Income: Middle	0.259	1.370	8.241	0.907	-6.104
	(8.769)	(10.94)	(12.73)	(13.88)	(13.20)
Income: High	11.31	23.62**	13.99	17.47	4.302
	(8.392)	(10.43)	(12.00)	(13.24)	(12.59)
Part-time Worker	-5.716	-4.781	-29.36*	-22.50	-13.51
	(10.54)	(12.93)	(15.08)	(16.41)	(15.61)
Full-time Worker	6.328	-2.102	-25.56**	-23.78*	-13.27
	(8.931)	(11.06)	(12.89)	(14.04)	(13.35)
Constant	431.5***	424.5***	437.2***	431.2***	445.6***
	(11.36)	(13.86)	(16.01)	(17.59)	(16.73)
Observations	442	258	310	258	258
R-squared	0.076	0.076	0.051	0.054	0.063

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Linear Regression Results for Self-reported Sleep Duration by Wave comparing between UR respondents (not) owning a device, including Demographic Controls.
Note: Columns (1) and (2) are Wave 1 (full and restricted); Columns (3) and (4) are Wave 2; Column (5) is Wave 3.

In Table A5, we investigate whether there are selection effects when selecting device owners. We run three Probit regressions, one for each Wave, and investigate whether the probability of being in the RDO sample (instead of the UR sample) is associated to demographics. We find that older respondents are more likely to belong to the RDO sample across all Waves. In Wave 1, Employment: Full-times and respondents with higher income are more likely to belong to he RDO sample.

	Wave 1 (RDO=1, UR=0)	Wave 2 (RDO=1, UR=0)	Wave 3 (RDO=1, UR=0)
Gender: Female	0.174*	0.156	0.205*
	(0.0930)	(0.108)	(0.121)
Gender: Other Gender	-0.976*		
	(0.565)		
Age: 35-44 years	0.101	0.113	0.154
	(0.105)	(0.125)	(0.138)
Age: 45+ years	0.373***	0.428***	0.480***
	(0.106)	(0.125)	(0.140)
Education: Bachelor's degree	-0.0469	-0.0704	-0.122
	(0.0994)	(0.116)	(0.129)
Education: Postgraduate education	-0.0582	-0.0155	-0.0278
	(0.116)	(0.137)	(0.152)
Income: Middle	0.220*	0.168	0.240*
	(0.112)	(0.130)	(0.143)
Income: High	0.211**	0.129	0.128
	(0.108)	(0.126)	(0.139)
Employment: Part-time	0.0445	-0.155	-0.155
	(0.144)	(0.168)	(0.185)
Employment: Full-time	0.343***	0.254*	0.287*
	(0.122)	(0.139)	(0.155)
Constant	-0.587***	-0.494***	-0.646***
	(0.151)	(0.176)	(0.198)
Observations	885	644	532
Pseudo R-squared	0.029	0.026	0.032

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A5: Probit regression: Likelihood of Belonging to the RDO sample based on demographics

In Table A6, we look for selection effects when comparing the subsample of UR respondents that report owning a tracker to respondents in UR without tracker and to respondents in RDO. We find that UR device owners are not distinguishable from RDO respondents (i.e.,Chi-

squared statistics comparing column 1 and 3 are small and not significant, while they tend to be richer and more likely to work full time than UR respondents without a tracker.

	UR (tracker = 1)	UR (tracker = 0)	RDO	χ^2: Col1 vs. Col2	χ^2: Col1 vs. Col3
Total	246	196	441		
Gender					
Male	83 (33.74%)	76 (38.78%)	145 (32.88%)	2.66	0.18
Female	161 (65.45%)	116 (59.18%)	295 (66.89%)		
Other genders	2 (0.81%)	4 (2.04%)	1 (0.23%)		
Age					
$\leq$ 34 years	105 (42.68%)	70 (35.71%)	143 (32.43%)	3.78	6.16
35-44 years	79 (32.11%)	61 (31.12%)	135 (30.61%)		
45+ years	62 (25.20%)	65 (33.16%)	163 (36.96%)		
Education					
Less than Bachelor's	93 (37.80%)	78 (39.80%)	172 (39.00%)	2.86	0.004
Bachelor's	103 (41.87%)	68 (34.69%)	172 (39.00%)		
Higher than Bachelor's	50 (20.33%)	50 (25.51%)	97 (22.00%)		
Income					
Low	84 (34.15%)	104 (53.06%)	145 (32.88%)	19.65***	0.16
Medium	62 (25.20%)	47 (23.98%)	124 (28.12%)		
High	100 (40.65%)	45 (22.96%)	172 (39.00%)		
Employment					
Other	41 (16.67%)	55 (28.06%)	69 (15.65%)	8.66**	0.005
Part-time	51 (20.73%)	39 (19.90%)	71 (16.10%)		
Full-time	154 (62.60%)	102 (52.04%)	301 (68.25%)		

*** p<0.01, ** p<0.05, * p<0.1, χ^2 test

Table A6: Demographics by Treatment Group and Wave (RDO Wave 1)

In Table A7, we provide descriptive statistics about the number of respondents in each Wave, specifying when the drop out took place and whether upload was mandatory or optional for the respondents.

	UR	% drop*	**RDO**	% drop*
Completed W1	442		441	
Started W2	331	-25.11%	316	-28.34%
			169 NM ; 147 M	
Completed W2	310	-6.34%	253	-19.94%
			153 NM ; 100 M	-9.47% NM ; -31.97% M
Started W3	289	-6.77%	244	-3.56%
Completed W3	258	-10.73%	224	-8.20%

NM = Non-Mandatory ; M = Mandatory
*relative to the previous observation

Table A7: Attrition By Samples and Wave

In Table A8, we investigate whether mandatory uploading is associated with dropout for specific demographic groups. We run a Probit model focusing on respondents in RDO who completed Wave 2. The outcome variable takes a value of 1 if the respondent belongs to the Mandatory RDO subsample and a value of 0 if they belong to the optional subsample. We find that having achieved postgraduate education and being female is associated with a higher probability of being in the Mandatory subsample, implying that these individuals are less likely to drop out when the upload is mandatory. In column 2, we add self reported sleep in Wave 1 as a predictor (i.e.,SR Sleep W1), finding no evidence that this correlates with attrition.

	Mandatory or Optional (1)	Mandatory or Optional (2)
Gender: Female	-0.291*	-0.294*
	(0.175)	(0.176)
Age: 35–44 years	0.0964	0.0738
	(0.214)	(0.215)
Age: 45+ years	0.0101	-0.0100
	(0.211)	(0.211)
Education: Bachelor's degree	0.372**	0.418**
	(0.187)	(0.191)
Education: Postgraduate education	-0.0435	-0.0358
	(0.228)	(0.229)
Income: Middle	0.0967	0.157
	(0.212)	(0.216)
Income: High	-0.0501	-0.0412
	(0.214)	(0.214)
Employment: Part-time	0.341	0.318
	(0.285)	(0.286)
Employment: Full-time	-0.0172	-0.0135
	(0.237)	(0.239)
SR Sleep W1		-0.00184
		(0.00123)
Constant	-0.298	0.433
	(0.308)	(0.578)
Observations	253	253
Pseudo R-squared	0.032	0.039

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A8: Probit regression, Likelihood of Being in the Mandatory RDO Group After Completing Wave 2

In Table A9 we explore whether differences in sleep between Waves and samples are related to two additional mechanisms: i) information effects and ii) effects related to feedback from wearing a device. For i), the RDO sample was explicitly informed that the study aimed to track their sleep over the coming week, and this information may potentially encourage them to change their sleep. No such information was provided to the UR sample, hence we identify the effect of this information by comparing changes in sleep duration between Waves 1 and 2 in the RDO sample with those in the UR sample. We explore this in column 1 by means of a Probit regression: we find that change in sleep

between Wave 2 and Wave 1 is not associated to being in the RDO. For ii), we assess the impact of device feedback associated with wearing an activity tracker, by identifying respondents in the RDO sample who indicated in Wave 1 that they wore their tracker for four or more nights per week as "nightwearing". We then compare changes in sleep duration from Wave 1 to Wave 2 nightwearing and non-nightwearing respondents. If feedback affected sleep duration, this would predominatly affect those not as used to wearing devices during the night, i.e., non-nightwearing respondents. In column 2, we display a Probit regression: we find that change in sleep between Wave 2 and Wave 1 is not associated to nighwearing of the tracker.

	(1) SleepW2-SleepW1	(2) SleepW2-SleepW1
RDO	11.19	
	(7.287)	
Gender: Female	7.265	1.608
	(7.752)	(11.66)
Gender: Other Gender	-66.35	
	(49.39)	
Age: 35-44 years	-1.973	1.984
	(8.825)	(13.92)
Age: 45+ years	15.11*	11.42
	(8.979)	(13.88)
Education: Bachelor's degree	-7.663	-4.961
	(8.253)	(12.29)
Education: Postgraduate education	6.788	10.74
	(9.815)	(14.69)
Income: Middle	6.357	-1.715
	(9.303)	(13.94)
Income: High	11.87	17.99
	(8.917)	(14.03)
Employment: Part-time	-27.12**	-21.91
	(11.85)	(19.10)
Employment: Full-time	-18.87*	-3.948
	(9.941)	(15.66)
Tracker at Night		3.140
		(12.49)
Constant	3.464	3.590
	(12.58)	(21.68)
Observations	565	255
R-squared	0.039	0.028

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A9: Linear Regression: Difference in Sleep in Wave 2 and Wave 1, on Group and Device Night Use

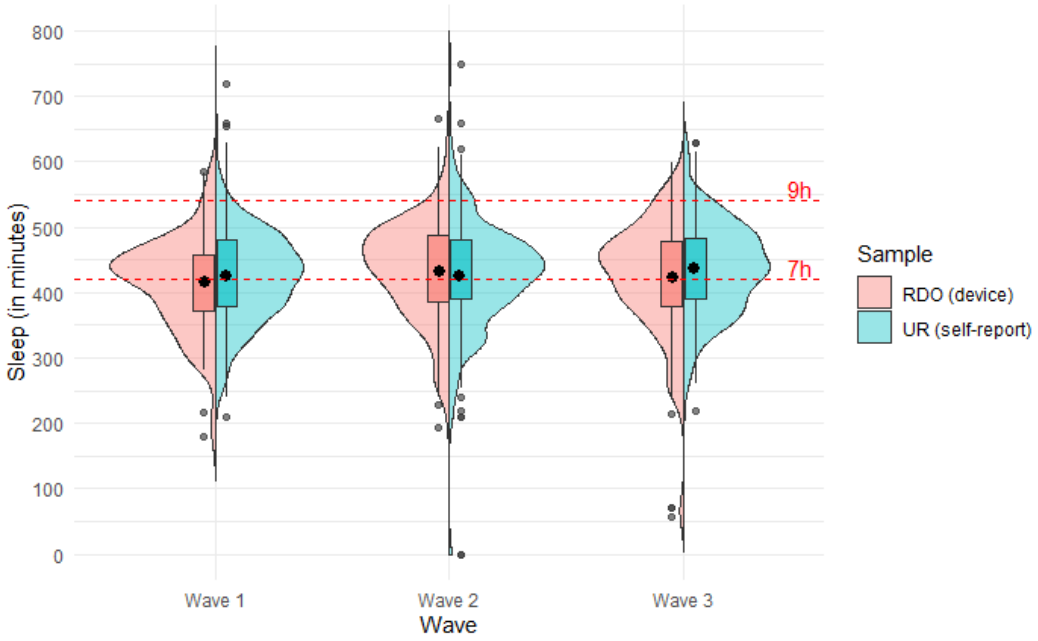

Figure A1: Distribution of self-reported sleep duration (UR sample) and device-reported sleep duration (RDO sample) by Wave

Wave 1 Questionnaire

The demo version of the original questionnaires in qualtrics can be accessed at:

RDO: RDO Qualtrics link Wave 1

UR: UR Qualtrics link Wave 1

Below is the LaTeX version adapted from the original survey.

Prolific Sleep Wave 1

Condition == 1 (Questions posed to device owner)
Condition == 2 (Questions posed to general public)

Informed Consent

Prolific ID: What is your Prolific ID? Please note that this response should auto-fill with the correct ID. We need your ID to verify you completed all parts of the survey.

Confirmation: Please confirm you are not a bot.

Introduction
This project is part of an academic research study run by Dr. Stefan Lipman and Dr. Georg Granic at Erasmus University Rotterdam. This form contains all relevant information necessary for you to consider the invitation to take part in this research. If anything is unclear, please contact Dr. Lipman (lipman@eshpm.eur.nl).

Purpose of the research: Getting sufficient sleep is of great importance for health and well-being. Yet many people sleep less than ideal. Different strategies exist to improve sleep, but evaluating them requires accurate measurement. This study investigates how people experience sleep quality.

Voluntary participation and right to withdraw: Participation is entirely voluntary. You may withdraw at any time without consequences. You may also withdraw consent for use of your data; any data collected after withdrawal will not be used.

Reimbursements: Participants can earn up to **7 GBP** across three questionnaires. This first questionnaire pays **0.50 GBP**. The second questionnaire (after 7 days, $\tilde{3}$0 minutes) pays **4 GBP**. The third questionnaire (after another 7 days) pays **0.50 GBP**. Completing all three yields a **2 GBP bonus**.

Privacy and data handling: Sleep data will be anonymized. Anonymized data may be made publicly available for educational and future research purposes.

Consent

- I voluntarily participate in this study (1)
- I do not wish to participate (2)

Demographics

Age: How old are you?

- Under 18 (1)
- 18-24 years old (2)
- 25-34 years old (3)
- 35-44 years old (4)
- 45-54 years old (5)
- 55-64 years old (6)
- 65+ years old (7)

Employment status: What best describes your employment status over the last three months?

- Working full-time (1)
- Working part-time (2)

- Unemployed/looking for work (3)
- A homemaker/stay-at-home parent (4)
- Student (5)
- Retired (6)
- Other (7)

Gender: How do you describe yourself?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer to self-describe (4) _____
- Prefer not to say (5)

Education: What is the highest level of education you have completed?

- Some Primary (1)
- Completed Primary School (2)
- Some Secondary (3)
- Completed Secondary School (4)
- Vocational or Similar (5)
- Some University but no degree (6)
- University Bachelors Degree (7)
- Graduate or professional degree (masters or doctorate) (8)
- Prefer not to say (9)

Income: What was your total household income before taxes during the past 12 months?

- Less than 20,000 pounds (1)

- 20,000-39,999 pounds (2)
- 40,000-59,999 pounds (3)
- 60,000-99,999 pounds (4)
- More than 100,000 pounds (5)

Study Explainer

Condition 1: Getting sufficient sleep is of great importance for health and well-being. Yet many people sleep less than ideal. Different strategies exist to improve sleeping but finding out if they work requires accurate measurement of sleep. The purpose of this study is to investigate if the measurement of sleep can be improved using activity trackers (e.g. Fitbit, Xiaomi Mi Band, Apple Watch, Garmin). We would like to ask you some questions about your usage of the activity tracker.

Condition 2: Getting sufficient sleep is of great importance for health and well-being. Yet many people sleep less than ideal. Different strategies exist to improve sleep but finding out if they work require accurate measurement of sleep.

Do you own an activity tracker? (Condition 2)

- Yes (1)
- No (2)

Device Questions

Type of tracker: (If owns tracker OR Condition 1)

- Fitbit (please add type) (1) _____
- Xiaomi Mi Band (please add type) (2) _____
- Apple Watch (please add type) (3) _____
- Garmin (please add type) (4) _____
- Other (5) _____

Usage: In an average week, how many days and/or nights do you wear your activity tracker?

Days per week:	0	1	2	3	4	5	6	7
Nights per week:	0	1	2	3	4	5	6	7

Sleep and Upload

Self-reported sleep: How long did you sleep last night?
Hours: ___
Minutes: ___

Optional upload: (Condition 1) Instructions for providing a screenshot of your sleep data, including device-specific links:

- Fitbit: `https://help.fitbit.com/articles/en_US/Help_article/1314.htm`
- Xiaomi Mi Band: `https://www.superwatches.com/sleep-monitoring-on-mi-band-5-6/`
- Apple Watch: `https://support.apple.com/guide/watch/track-your-sleep-apd830528336/watchos`

Go/no-go: (Condition 1)

- I own an activity tracker (1)
- I am ready to use my activity tracker in the next 7 days and its full functionality is available to me (4)
- I am willing to synchronize my activity tracker with the accompanying application (5)
- I am willing and able to make a screenshot of my sleep data and upload it if requested (6)

Offboarding:

- Condition 1: Thank you for taking part in this first questionnaire! So far you have earned 0.50 GBP. Take part in the next questionnaire to earn an additional 4 GBP. Check Prolific next week. Contact: dr. Stefan Lipman (lipman@eshpm.eur.nl)

- Condition 2: Same as above, without tracker instructions.

Wave 2 Questionnaire

The demo version of the original questionnaires in qualtrics can be accessed a:

RDO: Qualtrics Link Wave 2 (RDO)
UR: Qualtrics Link Wave 2 (UR)

Below is the LaTeX version adapted from the original survey.

Prolific Sleep Wave 2

Condition == 1 (Questions posed to device owner)
Condition == 2 (Questions posed to general public)

Landing Page

Prolific ID: What is your Prolific ID? _____

Verification: Please verify you are not a robot.

Welcome:
Welcome to the second questionnaire of this study! This questionnaire will take about 30 minutes to complete, and you will earn **4 GBP**. For each part of this questionnaire you will receive precise instructions. Please read them carefully.

Demographics

Living situation:

- I live by myself (2)
- I live with my partner (3)
- I live with my partner and child(ren) (4)
- I live with other adult non-relatives (5)
- I live with my parents/relatives (6)

- Other: _____ (7)

Children: Do you have children?

- Yes (1)
- No (2)

Night shift work: Do you work night shifts as part of your job?

- No, never (1)
- Once or twice per month (2)
- About once per week (3)
- Multiple times per week (4)

Noise at home: How often do you experience noise (unwanted sounds)?
Scale: 1=Never, 2=Rarely, 3=Some days/nights, 4=Most days/nights, 5=Every day/night

Cause	1	2	3	4	5
Neighbours' activities	☐	☐	☐	☐	☐
Traffic (cars, trains, planes)	☐	☐	☐	☐	☐
Passersby	☐	☐	☐	☐	☐
Industry (factories, farming)	☐	☐	☐	☐	☐
Construction	☐	☐	☐	☐	☐
Nightlife	☐	☐	☐	☐	☐
Own housemates (children, family, roommates, etc.)	☐	☐	☐	☐	☐

Children-specific questions (if you have children):

Age of youngest child:

- Younger than 3 months (1)
- Between 3 months and 1 year (2)
- Between 1 and 6 years old (3)
- Between 6 and 12 years old (4)
- Between 12 and 16 years old (6)

- Older than 16 years (7)

Number of children: _____
Weight: _____ (if unsure, approximate)
Height: _____ (if unsure, approximate)

Sleep Measurement

Self-reported sleep last night:
Hours: _____
Minutes: _____

Phone usage yesterday:
Hours: _____
Minutes: _____

Fatigue

Level of fatigue (0 = None, 10 = Unbearably severe):

In the last week: 0 1 2 3 4 5 6 7 8 9 10
Today: 0 1 2 3 4 5 6 7 8 9 10

Sleep Measurement

Display if Condition == 1 AND Upload == 1
We are interested in the sleep data collected through your activity tracker. Would you like to provide a screenshot of your activity tracker about your sleep of last night? See below the instructions on how to do so. If you prefer not to upload a screenshot, please move on to the next question.

How to access your sleep data?
Please find information on how to see your sleep data per device here:

- Fitbit: Link
- Xiaomi Mi Band: Link

- Apple Watch: Link

How to make a screenshot?
Android: Press Power and Volume Down buttons simultaneously.
iOS/iPhone: Press Side button and Volume up button (or Power and Home button for older models).

Display if Condition == 1 AND Upload == 2
We are interested in the sleep data collected through your activity tracker. Please provide a screenshot of your activity tracker's data about your sleep of last night. [Instructions same as above]

Display if Condition == 2
We are interested in the data your phone collects on your phone usage. If you want, you can provide a screenshot of your phone usage data.

- **Android:** Settings → Digital Wellbeing and Parental Controls.
- **iOS:** Settings → Screen Time.
- **Guide:** How much time you spend on your phone

How long did you sleep last night? This time might be shorter than you have spent in bed. Try to be as precise as possible.
Hours: _____ Minutes: _____

How much have you used your phone today?
Hours: _____ Minutes: _____

Satisfaction with Life Scale

Please indicate the extent to which you agreed with the following statements in the last week.
Scale: 1=Strongly disagree, 2=Disagree, 3=Slightly disagree, 4=Neither, 5=Slightly agree, 6=Agree, 7=Strongly agree

Statement	1	2	3	4	5	6	7
In most ways my life is close to my ideal	☐	☐	☐	☐	☐	☐	☐
The conditions of my life are excellent	☐	☐	☐	☐	☐	☐	☐
I am satisfied with my life	☐	☐	☐	☐	☐	☐	☐
So far I have gotten the important things I want in life	☐	☐	☐	☐	☐	☐	☐
If I could live my life over, I would change nothing	☐	☐	☐	☐	☐	☐	☐

Ideal Sleep

How many hours do you need to sleep on weekdays in order to feel fully rested?
Hours: _____ Minutes: _____

Ideally, how many hours would you like sleep on weekdays (on average, per night)? You may prefer to sleep more than you specified in your answer to the last question, because you enjoy sleeping.
Hours: _____ Minutes: _____

How many hours do you actually sleep on weekdays (on average, per night)?

Hours: _____ Minutes: _____

Bedtime Procrastination Scale

For each of the following statements, please report the extent it generally applies to you.
Scale: 1 (Not at all) to 5 (Always)

Statement	1	2	3	4	5
I go to bed later than I had intended.	☐	☐	☐	☐	☐
I go to bed early if I have to get up early.	☐	☐	☐	☐	☐
If it's time to turn off lights, I do it immediately.	☐	☐	☐	☐	☐
Often I am still doing other things at bedtime.	☐	☐	☐	☐	☐
I easily get distracted when I want to go to bed.	☐	☐	☐	☐	☐
I do not go to bed on time.	☐	☐	☐	☐	☐
I have a regular bedtime which I keep to.	☐	☐	☐	☐	☐
I want to go to bed on time but I just don't.	☐	☐	☐	☐	☐
I can easily stop activities at bedtime.	☐	☐	☐	☐	☐
I regret going to bed late during the next day.	☐	☐	☐	☐	☐

Single Item Sleep Quality Scale (SISQS)

Instructions: The following two questions refer to your overall sleep quality. Please think of your sleep **overall**, such as how many hour of sleep you got, how easily you fell asleep, how often you woke up during the night, how often you woke up earlier than you had to, and how refreshing your sleep was.

During the past 7 days, how would you rate your sleep quality overall?

- 0 Terrible
- 1
- 2 Poor
- 3
- 4
- 5 Fair
- 6
- 7
- 8 Good
- 9
- 10 Excellent

How would you rate your sleep quality overall **last night**?

- 0 Terrible
- 1
- 2 Poor
- 3
- 4
- 5 Fair
- 6
- 7
- 8 Good
- 9
- 10 Excellent

Morningness/Eveningness

At what time would you get up if you were entirely free to plan your own day?

- Earlier than 06:30 (in the morning) (1)
- Between 06:30 and 07:45 (2)
- Between 07:45 and 09:45 (3)
- Between 09:45 and 11:00 (4)
- Between 11:00 and 12:00 (noon) (5)
- Later than 12:00 (noon) (6)

During the first half hour after having woken up, how tired do you feel generally?

- Very tired (1)
- Fairly tired (2)

- Fairly refreshed (3)
- Very refreshed (4)

At approximately what time in the evening do you feel tired, and, as a result, in need of sleep?

- Earlier than 21:00 (1)
- Between 21:00 and 22:15 (2)
- Between 22:15 and 00:45 (3)
- Between 00:45 and 02:00 (4)
- Later than 02:00 (5)

At approximately what time of day do you usually feel your best?

- Between 05:00 and 08:00 (in the morning) (1)
- Between 08:00 and 10:00 (in the morning) (2)
- Between 10:00 and 17:00 (3)
- Between 17:00 and 22:00 (4)
- Between 22:00 and 05:00 (in the morning) (5)

One hears about 'morning' and 'evening' types of people. Which one of these types do you consider yourself to be?

- Definitely a 'morning' type (1)
- Rather a 'morning' type than an 'evening' type (2)
- Rather a 'evening' type than a 'morning' type (3)
- Definitely a 'evening' type (4)

At what time do you actually wake up on weekdays, on average?

- Earlier than 06:30 (in the morning) (1)
- Between 06:30 and 07:45 (2)
- Between 07:45 and 09:45 (3)

- Between 09:45 and 11:00 (4)
- Between 11:00 and 12:00 (noon) (5)
- Later than 12:00 (noon) (6)

At what time do you actually go to bed on weekdays, on average?

- Earlier than 21:00 (1)
- Between 21:00 and 22:15 (2)
- Between 22:15 and 00:45 (3)
- Between 00:45 and 02:00 (4)
- Later than 02:00 (5)

Beliefs

Please indicate the extent to which you agree.
Scale: 1=Strongly disagree to 7=Strongly agree

Statement	1	2	3	4	5	6	7
Brain/body can learn to function on less sleep	☐	☐	☐	☐	☐	☐	☐
If you can get it, more sleep is always better	☐	☐	☐	☐	☐	☐	☐
Falling asleep cannot be willed; must be unintentional	☐	☐	☐	☐	☐	☐	☐

Sleep Norms

Personal: During the past month, how often have you experienced...
Scale: 1=Not in past month, 2=¡1×/week, 3=1×/week, 4=3+×/week

	1	2	3	4
Waking in the middle of the night	☐	☐	☐	☐
Waking up too early	☐	☐	☐	☐
Taking ¿30 mins to fall asleep	☐	☐	☐	☐
Not feeling refreshed after waking	☐	☐	☐	☐
Not feeling energetic throughout the day	☐	☐	☐	☐

symptom.worry
How much do you worry (during the day or night) about:

	Level of Worry									
	1	2	3	4	5	6	7	8	9	10
Symptom	(Not at all)									(Extremely)
waking up in the middle of the night	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
waking up too early	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
taking longer than 30 minutes to fall asleep	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
not feeling refreshed or energized after waking in the morning	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
not feeling energetic and alert throughout the day	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

During the past month, how often do you think most people your age experienced ...

	1	2	3	4
Waking in the middle of the night	☐	☐	☐	☐
Waking up too early	☐	☐	☐	☐
Taking ¿30 mins to fall asleep	☐	☐	☐	☐
Not feeling refreshed after waking	☐	☐	☐	☐
Not feeling energetic throughout the day	☐	☐	☐	☐

How many hours do you think most people your age...?
Should sleep per night: _____ hrs
Actually sleep per night: _____ hrs

To what extent do you think people your age ...

	1	2	3	4	5
forego their sleep in order to meet their obligations	☐	☐	☐	☐	☐
should forego their sleep in order to meet their obligations	☐	☐	☐	☐	☐

Sleep Hygiene Index (SHI)

SHI: Below you will find a list of statements. Please select how true each statement is for you.

Statement	**Never** (1)	**Rarely** (2)	**Frequently** (3)	**Always** (4)
I take daytime naps lasting two or more hours	☐	☐	☐	☐
I go to bed at different times from day to day	☐	☐	☐	☐
I get out of bed at different times from day to day	☐	☐	☐	☐
I exercise to the point of sweating within 1 hr of going to bed	☐	☐	☐	☐
I stay in bed longer than I should two or three times a week	☐	☐	☐	☐
I use alcohol, tobacco, or caffeine within 4hrs of going to bed or after going to bed	☐	☐	☐	☐
I do something that may wake me up before bedtime (e.g., video games, internet, or clean)	☐	☐	☐	☐
I go to bed feeling stressed, angry, upset, or nervous	☐	☐	☐	☐
I use my bed for things other than sleeping or sex (e.g., watch TV, read, eat, or study)	☐	☐	☐	☐
I sleep on an uncomfortable bed (e.g., poor mattress, too many/few blankets)	☐	☐	☐	☐
I sleep in an uncomfortable bedroom (e.g., too bright, stuffy, hot, cold, or noisy)	☐	☐	☐	☐
I do important work before bedtime (e.g., pay bills, schedule, or study)	☐	☐	☐	☐
I think, plan, or worry when I am in bed	☐	☐	☐	☐
Display only if living with partner: I go to bed the same time as my partner	☐	☐	☐	☐

Types of Stressors

How distressed have you felt in the last week?
Scale: 1=No stress, 9=Extreme stress
Types.Stress
For the types of stressors below, please indicate how distressed you have felt about them in the last week (either during the day or night).

	Level of Distress								
	1	2	3	4	5	6	7	8	9
Stressor	(None)								(Extreme)
Financial (e.g. worry about paying the bills)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Occupational – work (e.g. difficulties fulfilling the duties of your job)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Occupational – social (e.g. difficulties with co-workers, boss, union)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Environmental (e.g. noises in the environment from traffic)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Health (e.g. having an illness or health problem)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Relationships (e.g. problems partner, children, family or friends)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Fatigue

Please, indicate the level of fatigue that you have experienced on the line below. (0=None, 10=Unbearably severe)
In the last week: _____
Today: _____

Offboarding

Display if Condition == 1
Thank you for taking part in this second survey! In this questionnaire, you have earned 4 GBP. Would you like to earn more? Take part in the next session of this experiment! In seven days, the third and final questionnaire of this study will become available in Prolific. Check Prolific to see when the new survey becomes available. You will also have to upload a screenshot of your sleep activity from your activity tracker, so be sure to wear your activity tracker at night in the next 7 days. It will be very short, and you will earn 2.50 GBP by completing it. If you have any questions about the study or comments, please reach out to dr. Stefan Lipman (lipman@eshpm.eur.nl).

Display if Condition == 2
Thank you for taking part in this second survey! In this questionnaire, you have earned 4 GBP. Would you like to earn more? Take part in the next session of this experiment! In seven days, the third and final questionnaire of this study will become available in Prolific. Check Prolific to see when the new survey becomes available. It will be very short, and you will earn 2.50 GBP by completing it. If you have any questions about the study or comments, please reach out to dr. Stefan Lipman (lipman@eshpm.eur.nl)

Wave 3 Questionnaire

The demo version of the original questionnaires in qualtrics can be accessed at:

RDO: Qualtrics Link Wave 3 (RDO)
UR: Qualtrics Link Wave 3 (UR)

Below is the LaTeX version adapted from the original survey.

Prolific Sleep Wave 3

Condition == 1 (Questions posed to device owner)
Condition == 2 (Questions posed to general public)

Landing Page

Prolific ID: What is your Prolific ID? _____

Verification: Please verify you are not a robot.

Onboarding Message

Displayed if Condition = 1
Welcome to the third and last questionnaire of this sleep study! This survey is very short as you only need to report your sleep and upload a screenshot from your activity tracker. You will earn **3 GBP** by completing this questionnaire.

Displayed if Condition = 2
Welcome to the third and last questionnaire of this sleep study! This survey is very short as you only need to report your sleep and upload a screenshot of your phone usage data. You will earn **2.50 GBP** by completing this questionnaire.

Sleep Data Upload (Condition 1)

We are interested in the sleep data collected through your activity tracker. Please provide a screenshot of your activity tracker's data about your sleep of last night.

How to access your sleep data:
Find information per device below, or use your device's specific instructions:

- Fitbit Instructions
- Xiaomi Mi Band Instructions
- Apple Watch Instructions

How to make a screenshot:

- **Android:** Press the Power and Volume Down buttons simultaneously.
- **iOS/iPhone:** Press the Side button and the Volume Up button at the same time.

Phone Usage Data Upload (Condition 2)

We are interested in the data your phone collects on your phone usage.
Please provide a screenshot of your phone usage for today.

How to access your phone usage data:

- **Android:** Go to *Settings* → *Digital Wellbeing and Parental Controls*
- **iOS:** Go to *Settings* → *Screen Time*

Reference: Twilio Guide to Screen Time

Self-Report Section

Sleep.SR – How long did you sleep last night?
This time might be shorter than the time you spent in bed. Try to be as precise as possible.

Hours: _____
Minutes: _____

phone.use.SR – How much have you used your phone today?
Hours: _____
Minutes: _____

Closing

Thank You
Thank you for taking part in the last questionnaire of this study. By completing it, you have earned **2.50 GBP**, which results in a total payment of **7 GBP**. If you have any questions about the study or comments, please reach out to Dr. Stefan Lipman.