

Precision Assessment of Real-World Associations Between Stress and Sleep Duration: Individual-Level Modeling of Actigraphy Data Collected Continuously for an Academic Year

Constanza M Vidal Bustamante, Garth Coombs III, Habiballah Rahimi-Eichi, Patrick Mair, Jukka-Pekka Onnela, Justin T Baker, Randy L Buckner

Submitted to: JMIR Formative Research
on: October 26, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	49
Figures	50
Figure 1.....	51
Figure 2.....	53
Figure 3.....	55
Figure 4.....	56
Figure 5.....	57
Figure 6.....	58
Figure 7.....	59
Figure 8.....	60
Multimedia Appendixes	61
Multimedia Appendix 1.....	62
Related publication(s) - for reviewers eyes onlies	63
Related publication(s) - for reviewers eyes only 0.....	63
Related publication(s) - for reviewers eyes only 0.....	63

Precision Assessment of Real-World Associations Between Stress and Sleep Duration: Individual-Level Modeling of Actigraphy Data Collected Continuously for an Academic Year

Constanza M Vidal Bustamante^{1, 2*} BA, MA; Garth Coombs III^{1, 2*} PhD; Habiballah Rahimi-Eichi^{1, 3, 4, 2} PhD; Patrick Mair¹ PhD; Jukka-Pekka Onnela⁵ PhD; Justin T Baker^{3, 4} PhD, MD; Randy L Buckner^{1, 4, 6, 2} PhD

¹Department of Psychology Harvard University Cambridge US

²Center for Brain Science Harvard University Cambridge US

³Institute for Technology in Psychiatry McLean Hospital Belmont US

⁴Department of Psychiatry Harvard Medical School Boston US

⁵Department of Biostatistics Harvard T.H. Chan School of Public Health Harvard University Boston US

⁶Athinoula A. Martinos Center for Biomedical Imaging Massachusetts General Hospital Charlestown US

*these authors contributed equally

Corresponding Author:

Constanza M Vidal Bustamante BA, MA

Department of Psychology

Harvard University

52 Oxford Street

Northwest Building, East Wing, Room 295.06

Cambridge

US

Abstract

Background: Heightened stress and insufficient sleep are common in the transition to college, often co-occur, and have both been linked to negative health outcomes. A challenge concerns disentangling the temporal directionality of their associations in daily life. Prior research has reported inconsistent findings, with changes in perceived stress preceding and/or succeeding changes in sleep duration. One possibility is that these day-to-day associations vary across individuals, but short study periods and group-level analyses may have obscured person-specific phenotypes.

Objective: We present a precision approach to the study of day-to-day associations between perceived stress and objective sleep duration that leverages intensive longitudinal data collected via wearables and smartphones. We developed an individual-level linear model (iLM) that provides stable, individually tailored estimates unbiased by the group. The model was designed to be parsimonious while accounting for the non-independence and temporal structure of the data.

Methods: Fifty-five college students sampled continuously for a full academic year provided daily self-reports of perceived stress via a smartphone app and wore an actigraphy wristband for the estimation of daily sleep duration (median usable daily observations per participant = 178). We developed and validated the iLM framework on a pilot subset of six participants that was fully independent from the target sample. In addition to the predictor of interest, the model included a covariate for day of the week to account for weekly patterns. The model was tested by applying it to variables expected to be strongly correlated with each other: objective sleep duration and self-reported quality of the same sleep event. Adequacy of the model specification was confirmed through several diagnostic tests. The model was then applied to the target sample of 49 first-year college students for the examination of bi-directional associations between daily stress levels and sleep duration.

Results: Proof-of-concept analyses captured expected associations between objective sleep duration and subjective sleep quality and passed all model diagnostics. Target analyses revealed negative associations between sleep duration and perceived stress in most individuals. Critically, our participants represented all the phenotypes of temporal association identified in prior literature: for some, elevated stress in the day was associated with shorter sleep later that night; for others, shorter sleep was associated with elevated stress the next day; others showed both directions of associations, and some showed no association. Of note, when modeled using a group-based, multilevel model, individual estimates were systematically attenuated, and some even sign-reversed.

Conclusions: The dynamic interplay of stress and sleep in daily life is likely person specific. Paired with intensive longitudinal data, our individual-level model provides a precision framework for the estimation of stable real-world behavioral and psychological dynamics, and may support the personalized prioritization of intervention targets for health and wellbeing.

(JMIR Preprints 26/10/2023:53441)

DOI: <https://doi.org/10.2196/preprints.53441>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/53441>

Original Manuscript

Original Paper

Precision Assessment of Real-World Associations Between Stress and Sleep Duration: Individual-Level Modeling of Actigraphy Data Collected Continuously for an Academic Year

Constanza M. Vidal Bustamante^{a,b,*}, Garth Coombs 3rd^{a,b,*}, Habiballah Rahimi-Eichi^{a-d}, Patrick Mair^a, Jukka-Pekka Onnela^e, Justin T. Baker^{c,d}, and Randy L. Buckner^{a-c,f}

^a Department of Psychology, Harvard University, Cambridge, MA 02138

^b Center for Brain Science, Harvard University, Cambridge, MA 02138

^c Department of Psychiatry, Harvard Medical School, Boston, MA 02114

^d Institute for Technology in Psychiatry, McLean Hospital, Belmont, MA 02478

^e Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115

^f Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Charlestown, MA 02129

* Equal contribution

Correspondence should be addressed to Constanza M. Vidal Bustamante, Department of Psychology, Harvard University, Northwest Science Building 295.06, 52 Oxford Street, Cambridge, MA 02138.
Email: cvidal@g.harvard.edu

Abstract

Background: Heightened stress and insufficient sleep are common in the transition to college, often co-occur, and have both been linked to negative health outcomes. A challenge concerns disentangling whether perceived stress precedes and/or succeeds changes in sleep. These day-to-day associations may vary across individuals, but short study periods and group-level analyses in prior research may have obscured person-specific phenotypes.

Objective: To obtain stable estimates of lead-lag associations between perceived stress and objective sleep duration in the individual, unbiased by the group, by developing an individual-level linear model (iLM) that can leverage intensive longitudinal data while remaining parsimonious.

Methods: Fifty-five college students volunteered to provide daily self-reports of perceived stress via a smartphone app and wore an actigraphy wristband for the estimation of daily sleep duration continuously throughout the school year (median usable daily observations per participant = 178). The iLM, developed in a Bayesian framework, included the predictor and outcome of interest and a covariate for day of the week to account for weekly patterns. We validated the model on a subset of six pilot participants by applying it to variables expected to correlate positively within individuals: objective sleep duration and self-reported sleep quality. The model was then applied to a fully independent target sample of 49 first-year college students for the examination of bi-directional associations between daily stress levels and sleep duration.

Results: Proof-of-concept analyses captured expected associations between objective sleep duration and subjective sleep quality in every pilot participant. Target analyses revealed negative associations between sleep duration and perceived stress in most participants, but their temporal association varied. Of the 19 participants (39%) that showed a significant association (probability of direction > 0.975), 42% showed elevated stress in the day associated with shorter sleep later that night; 26% showed shorter sleep associated with elevated stress the next day; and 32% showed both directions of associations. Of note, when modeled using a group-based, multilevel model, individual estimates were systematically attenuated, and some even sign-reversed.

Conclusions: The dynamic interplay of stress and sleep in daily life is likely person specific. Paired with intensive longitudinal data, our individual-level model provides a precision framework for the

estimation of stable real-world behavioral and psychological dynamics, and may support the personalized prioritization of intervention targets for health and wellbeing.

Word count = 375 (max. = 450)

Keywords: deep phenotyping; individualized models; intensive longitudinal data; sleep; stress; actigraphy; accelerometer; wearable; mobile phone; digital health

Introduction

The transition to college is often accompanied by elevated stress and insufficient sleep [1–3], both of which have been found to impact daily functioning and, upon repeated exposure, to be associated with negative health outcomes ranging from internalizing disorders including anxiety and depression to cardiac and metabolic disease [4–9]. A better understanding of the dynamic, day-to-day interplay between perceived stress and sleep duration could inform the mechanisms of their downstream impacts and the development of interventions to prevent or mitigate them [10–14]. Yet only a few studies have collected daily observations to evaluate within-person associations between stress responses and sleep duration, and their results are mixed: some studies report that heightened stress is followed by shorter sleep that night, but not vice versa [15,16], others that shorter sleep is followed by heightened stress the next day, but not vice versa [17], and yet others report bidirectional relationships [18, 19].

Inconsistent findings in stress-sleep associations might be at least partially explained by two interrelated methodological limitations, both of which are addressed by the individual-level modeling approach presented in this study. First, existing longitudinal studies have aggregated individual observations collected over short study periods of 14 days or less, thus limiting the estimation of stable associations that are robust to changing environmental demands (e.g., first week of the semester versus final exams period). Thanks to the adoption of digital phenotyping tools such as wearables and smartphones, research designs that sample individuals over much longer periods of time are increasingly feasible [20–22]. Sleep duration can now be passively tracked through continuous actigraphy sensing over months and even years, while perceived stress levels can be probed via brief daily phone-based surveys, with high compliance rates in student samples and small participant burden [23–25].

Moreover, the prevailing focus on group modeling and sample-level effects obscures the possibility that day-to-day associations between perceived stress and sleep duration may vary across individuals. For instance, it is possible that in certain individuals, getting less sleep than usual has no significant impact on stress levels the following day, but heightened stress during the day leads to shorter sleep duration later that night, while the opposite pattern might be true for others. Even when group-level models allow for individual-level estimates, such as in multilevel models with random effects, the degree to which the individual estimates are pulled toward the group (a “shrinkage” effect) is different for each individual based on the amount of data they provide, thus reducing the individual tailoring in a non-uniform manner [26].

Individual-level models, which are fitted over a single individual's intensive longitudinal observations, may offer a critical alternative to the estimation of stress-sleep associations. Compared to group-level approaches, individual-level models provide estimates of phenotypes that are tailored to each person's data and unbiased by the group [27–30]. They may also be more readily applicable in real-world contexts, where a clinician might use a precision health approach to evaluate and support each individual patient based on their personal data rather than on a hypothetical average patient [31–34]. Of note, individual-level approaches can also contribute to conclusions at the group or population level, such as by estimating the prevalence of each individually-derived phenotype, rather than blurring across individuals to arrive at a central tendency that might not accurately represent many of the included individuals.

In this study, we introduce a novel individual-level linear modeling (iLM) approach that leverages daily observations collected over a full academic year for the assessment of day-to-day associations between actigraphy-derived sleep duration and self-reported stress levels in first-year college students. Our aims were twofold. First, we used a pilot dataset to develop and validate a parsimonious individual-level linear model that estimates concurrent or lagged associations between two daily variables of interest while accounting for the weekly structure in student behavior. We then leveraged this individual-level model for the target examination of bidirectional day-to-day associations between perceived stress and objective sleep duration in an independent dataset of 49 first-year college students. We expected a negative association for most participants—such that higher stress levels are associated with shorter sleep—but we also anticipated that some participants might show positive or null associations. We further hypothesized that the lead-lag relationship between perceived stress and sleep duration (i.e., which of the two temporally precedes the other) would vary across individuals, such that, for some, elevated stress would associate with shorter sleep that night; for others, shorter sleep would associate with elevated stress the next day, and for others, both directions of association would be identified.

Methods

Participants

Pilot Group

Six undergraduates returning for their sophomore year volunteered for a year-long study (all 19 years old; 3 female and 3 male). All had participated in a previous pilot study in our laboratory during their first year of college and were known to be compliant. These six pilot participants provided data to

develop the statistical models that were then applied to the new group of target participants described below. Pilot participants enrolled for the current study during the first two weeks of their Fall semester. Like the target participants, they were required to be taking full-time classes and own a smartphone compatible with the study smartphone application, Beiwe, which is part of the open-source Beiwe platform for digital phenotyping [35]. Given our focus on individual-level models where each person serves as their own baseline and there is no aggregation across participants, students were not excluded for current or past psychiatric disorders or medication use, nor if they began treatment or medication for mental health issues during the study. Interested participants scheduled an in-person consent session where study procedures were explained.

Target Group

Forty-nine undergraduates beginning their first year of college volunteered for a year-long study (18-19 years old; mean age = 18.1; 25 female and 24 male). Data from these participants has been previously reported by this paper's authors in [24]. Participants living on campus were recruited via flyers posted on campus boards and distributed via email lists and were enrolled during the first two weeks concurrently with the pilot participants. Enrollment criteria were the same as for the pilot participants other than the target first-year participants were all new to the university. Initially 68 participants enrolled, with 19 excluded based on issues with data acquisition including early withdrawal from the study (n=7), technical failure of the actigraphy data (n=1), poor quality actigraphy data (n=2), and completion of <50% of the daily surveys (n=9). Of the final sample (N=49), 63% identified as White, 14% Black, 10% Asian, 4% American Indian, and 4% Mixed-Race. Twelve percent reported prior diagnosis of a psychiatric disorder (including anxiety, depression, and/or ADHD); 8% maintained active diagnoses. First-year students have not yet declared their area of study, but participants reported their desired future occupation to be 31% medicine, 14% business or finance, 12% academia or other research, 10% engineering, 10% policy or government, 8% law, and 6% other or undecided. 94% of the participants were iPhone users and 6% were Android users.

Study Design

As previously reported in [24], this intensive longitudinal observational study collected passive and active data as participants engaged in their lives over a full academic year, extending a few days into the summer break. Both pilot and target participants completed smartphone-based daily surveys and a voice-recorded diary, wore an actigraphy wristband for continuous activity and sleep monitoring

for the duration of the study (GENEActiv Original, Activinsights Ltd., Kimbolton, UK), completed a battery of online questionnaires at the beginning, middle, and end of the study, and attended brief in-person check-ins every 3-4 weeks.

Ethical Considerations

Informed consent and all study procedures and methods were approved by the Institutional Review Board of Harvard University. All participants completed an in-person informed consent session where study procedures were explained and any questions were clarified. Participants were informed they could withdraw their study participation at any time. Participants were compensated \$1 per each daily survey they submitted, \$1 per day for continuously wearing the actigraphy wristband, and \$20 per hour for online surveys and attending in-person visits. Milestone bonus payments for completing half of the study (\$100) and the full study (\$300) were also provided to compensate participants' continued compliance.

Study data collected across devices were stored and automatically backed up in a secure data warehouse configured to automatically import data from various collection streams. All data were kept as securely as possible and were only accessible to study staff. Participants' data were labeled with a randomly generated participant ID. Personally identifying information, such as names and contact information, were kept separate from all other collected data in a locked file cabinet, in a locked office, behind an ID-card secured suite during off hours, and in a password-protected database.

Measures and Quality Control

Objective Sleep Duration

Daily sleep duration reflects the number of minutes between the estimated start and end of the day's longest detected sleep episode. As redescribed from the original [24], sleep duration was derived from the accelerometer data collected through the actigraphy wristband (GENEActiv Original, Activinsights Ltd., Kimbolton, UK) and analyzed via the Deep Phenotyping of Sleep (DPSleep) processing pipeline [36]. Participants wore the wristband on their nondominant wrist continuously, including during sleep and when bathing. Tri-axial acceleration was collected with a sampling frequency of 30 Hz during the academic semesters and 10 Hz during the winter break (to extend battery life and memory while participants were away from campus). Participants were instructed to press the wristband's button when they began trying to fall sleep at night and immediately after they awoke in the morning. Individuals exchanged their wristband for a fully charged one with reset memory at the in-person check-ins.

The DPSleep processing pipeline was applied to the raw actigraphy data to detect the major sleep episode for each day [36]. The pipeline first converted the accelerometer data to minute-based activity estimates, removed the minutes when the individual was not wearing the device, and then estimated the major sleep episode based on a sliding window. Days where one of the boundaries of the sleep episode (i.e., rises in relative activity both before and after a period of lower activity) could not be detected due to missing data were labeled as unusable. Two independent trained raters examined the automatically detected start and end times and usability label of each sleep episode against the minute-based activity levels and the participant button presses when available. When necessary, they adjusted the automatic times and labels. Any disagreements between the two raters' assessments were reviewed by the research team and resolved through discussion. A full description of the processing pipelines applied to the actigraphy data, including quality control steps, can be found in [36].

All data that passed quality control was included in analysis, including days with no detected sleep episode (i.e., with no extended periods of lower relative activity).

Daily Phone-Based Surveys

Smartphone surveys were administered via the Beiwe application [35]. Each night before bed participants completed a 46-item self-report survey related to their daily lives. As described originally in [24], questions assessed a range of behaviors and internal states over the past 24 hours, including sleep quality, stress levels and sources, positive and negative affect, general physical health, daily consumption habits, studying behaviors, and sociability and support [24]. This paper reports analyses using two survey questions selected *a priori* that probed daily subjective sleep quality and perceived stress. The sleep quality question asked, "How did you sleep last night?" and was answered on a 5-point scale (1. Terribly: little or no sleep; 2. Not so well: got some sleep but not enough; 3. Sufficient: got enough sleep to function; 4. Good: got a solid night sleep and felt well-rested; 5. Exceptional: one of my best nights of sleep). The perceived stress question asked, "How much did you feel stressed over the past 24 hours?" and was also answered on 5-point scale (1. Very slightly or not at all, 2. A little, 3. Moderately, 4. Quite a bit, 5. Extremely).

Surveys submitted between 5PM (local time) the day the survey opened and 6AM the following day were considered on time. Surveys submitted past 6AM the day after the survey was prompted were marked as missing. A participant was included in analysis if they were compliant with at least 100 daily surveys across the data collection period, and only on-time surveys from those participants were included.

Analytical Approach

Development of the individual-level linear model (iLM)

An iLM was developed on the intensive day-level longitudinal data from the six pilot participants to test person-specific, day-to-day behavioral associations. The iLM framework allows for individually tailored estimates by treating each day as the unit of observation and the individual as the “population,” as opposed to each individual as the unit of observation used to estimate the general population. An individual’s observed time series data can be understood as just one realization of a stochastic process whose data generating process we are trying to model and understand [37], in this case through linear models. The intercept and slope estimates are unique to the individual, and we interpret them as a proxy for the individual’s phenotype. Note that in this framework, each individual model can be interpreted as an independent test of the hypothesized association (e.g., is shorter sleep than usual associated with heightened perceived stress the following day?).

The model was structured to be parsimonious while accounting for the non-independence of the daily measures and the temporal structure imposed by the academic schedule. Day of the week was included as a categorical covariate to account for weekend vs. weekday effects and other weekly structure imposed by the college schedule (e.g., classes that meet on Monday-Wednesday-Friday vs. Tuesday-Thursday). Additionally, since behavioral patterns during the semester vary substantially from the 5-week winter break (when students do not have classes and typically are away from campus), we decided *a priori* to only include in the model compliant data collected during the school semesters.

The final individual-level linear model took the following general form:

$$y_t = \beta_0 + \beta_1 x_{t \vee t-1} + \beta_2 \text{DayOfWeek}_t + \varepsilon_t$$

A daily observation on day_t starts with the nighttime sleep episode and ends with the submission of the daily survey submitted in the evening before the next sleep episode. In the formula above, y_t is a single participant’s outcome variable (e.g., sleep duration in minutes) at daily timepoint t ; β_0 is that participant’s individual intercept; x is the predictor variable (e.g., sleep quality or stress on a 1-5 scale) at daily timepoint t or $t-1$ (depending on the lag of the tested association); β_1 is the participant’s individual slope for x ; DayOfWeek_t is the day of the week in which the outcome observation was acquired (modeled as a categorical variable ordered Saturday to Friday); and ε_t is a normally distributed random error term.

All modeling was carried out in a Bayesian inference framework, which treats unknown parameters (e.g., a slope) as random variables with a probability distribution rather than a discrete

value; this distribution is updated based on the observed data (resulting in a “posterior distribution”) and serves as a measure of uncertainty around the parameter [38-40]. The Bayesian framework was favored for the present analyses due to its flexibility in computing models with varied specifications (including complex random effect specifications in the multilevel models we fitted as part of our model validation process), robustness to sample data characteristics (e.g., dispersion), and intuitive interpretation of the posterior distribution and 95% *uncertainty* interval (i.e., conditional on the data and the model, the probability that the parameter is contained in the interval is 0.95) [39-40]. For comparison, parallel individual-level models fitted in a frequentist inference framework in the pilot validation stage yielded nearly identical point estimates to their Bayesian counterparts (**Figure S2 of Multimedia Appendix 1**), suggesting our model specification is robust across both statistical inference frameworks. All models were estimated in R 4.3.1 [41]. Bayesian models were estimated using the Stan modeling language [42] and the packages *rstanarm* (v2.21.4; [43]), *tidybayes* (v3.0.1; [44]), and *bayestestR* (v0.13.1; [45]). Frequentist individual-level models were estimated using the *stats* package included in base R [41].

Bayesian models were fit with default weakly informative priors specifying a gaussian distribution ($M=0$, $SD=2.5$) to represent our diffuse prior knowledge. We estimated parameters using a Markov chain Monte Carlo (MCMC) approach. For each parameter, we sampled from four stationary Markov chains, each comprising 5,000 sampling iterations, including a burn-in period of 2,500 iterations that were discarded (for a total of 10,000 post-warmup draws). Convergence of the four chains to a single stationary distribution was assessed quantitatively via the R-hat convergence diagnostic [46] (adequate convergence defined as $R\text{-hat} < 1.1$) and qualitatively by visual inspection of trace plots showing the estimated parameter as a function of each chain’s iteration number (adequate convergence defined as the chains overlapping with each other throughout and a lack of structured patterns in each chain). Each model’s Effective Sample Size (ESS) metric is reported. Each estimate in the MCMC process is serially correlated with the previous estimates; the higher the correlation, the more samples are needed to get to a stationary distribution. In the presence of nearly no autocorrelation, the ESS will be equal to the number of posterior draws requested (in this case, 10,000). Generally, the ESS should be at least 1,000 to obtain stable estimates [38, 47].

Adequacy of the model specification was assessed via two methods. First, posterior predictive checks entailed a visual comparison of the distribution of the observed outcome variable to the distribution of 100 simulated outcome datasets generated by applying 100 draws from the model parameters’ posterior distribution to our observed dataset. Similarity in the distributions of the observed and model-generated outcomes suggests that the model specification captured the data

well. Second, we inspected the model residuals against the model's predicted values to confirm homoscedasticity, and against themselves to rule out problematic autocorrelation due to temporal dependencies in longitudinal data.

Point estimates of intercepts and slopes were computed as the median value of their respective posterior distributions. Ninety-five percent uncertainty intervals (UI) were computed as the 2.5% and 97.5% quantiles of the posterior distribution. To provide intuitive parallels to a frequentist inference framework, we interpreted a predictor slope as statistically significant if its 95% UI did not contain zero, or put differently, if the proportion of the posterior distribution falling in the direction (positive or negative) of the point estimate (also known as “probability of direction”, or *pd*) was higher than 0.975 (which approximates a frequentist one-tailed *p*-value < 0.025 ; [48]). Although we focus on 95% as the cutoff for the UI given the widespread use of this number in the literature, it should be noted that this threshold has been criticized as arbitrary [39]. To go beyond testing whether the slope is different from exactly zero, we also report the percentage of the 95% UI that falls within a “region of practical equivalence” (ROPE), defined as parameter values that are sufficiently close to zero to be considered equivalent to the null for practical purposes [49], and mathematically defined as a standardized effect size of under 0.1 (i.e., half of a “small effect” as defined in [50]). Given our individualized approach, ROPEs were estimated separately for each participant based on their observed data.

Validation of the iLM in the pilot data

Data from six pilot participants were used to test the properties of the individual-level model and explore associations among variables of interest to validate the approach. A first sanity check was to assess whether the individual-model captured the expected association between a sleep event's objective duration (estimated from an actigraphy wristband) and the subjective rating (reported at the end of each day as part of a smartphone-based survey) for the same (“concurrent”) sleep event. Participants were expected to rate episodes of shortened sleep as worse quality compared to nights of longer sleep. This contemporaneous model, known as a static model in the time series literature [37], was specified as follows:

$$\text{ObjectiveSleepDuration}_t = \beta_0 + \beta_1 \text{SubjectiveSleepRating}_t + \beta_2 \text{DayOfWeek}_t + \varepsilon_t$$

A second, exploratory lagged model was fit testing associations between sleep duration on day_t and sleep quality reported the day before, on day_{t-1} . We expected this lagged association would be weaker than the concurrent association outlined above (given that the variables tested are no longer referring to the same sleep event), and importantly, that the association would be in the

opposite direction, such that poor quality sleep on one night is associated with longer sleep duration the next night, signaling a compensatory sleep rebound effect. To test this, a similar model was fit to that above, with the difference that the predictor was lagged by one day. This lagged model, known as a finite distributed lag model in time series analysis [37], was specified as follows:

$$\text{ObjectiveSleepDuration}_t = \beta_0 + \beta_1 \text{SubjectiveSleepRating}_{t-1} + \beta_2 \text{DayOfWeek}_t k_t + \varepsilon_t$$

Both models were fit with a gaussian distribution to reflect the observed normal distribution of sleep duration in our sample.

We conducted model checks to evaluate the performance of the iLM framework. Posterior predictive checks (described above) evaluated that a gaussian model specification captured the distribution of the data well. Additionally, given the longitudinal design, we examined whether the model residuals lacked meaningful autocorrelation, which would suggest that our day-of-week covariate sufficiently captured the weekly structure of sleep duration. Finally, we compared the estimates obtained through the iLMs to the estimates obtained when the same data for the six pilot participants was fit within a single, group-based multilevel linear model (MLM), specifying fixed effects for intercepts and slopes and additional participant-level random intercepts and random slopes for the main predictor of interest (in this case, sleep quality). This comparison allowed us to assess our expectation that MLMs would provide roughly comparable individual estimates to the iLM, but with the critical difference that it would systematically attenuate these individual estimates, especially for individuals that deviate from the predominant association phenotype or when there is large heterogeneity in these phenotypes across individuals.

Application of the iLM to the novel target participants

After developing the individual-level linear model and validating it extensively over the pilot dataset, the modeling framework was then carried forward and applied to the independent target dataset of first-year college students. All models were fit with a gaussian distribution. We first applied the same sanity-check models as in the pilot dataset. Subsequently, we used the iLM method to test a priori target hypotheses regarding the association between perceived daily stress and objective sleep duration.

Two target models were fit for each individual to examine bi-directional associations between sleep duration and perceived stress. A daily observation on day_t consists of last night's sleep duration (ObjectiveSleepDuration_t, recorded passively via the actigraphy wristband) and the subjective rating of the present day's overall stress levels (SubjectiveStressRating_t, reported by the participant in the evening). First, to test whether stress level during the day is related to sleep duration later that night

(i.e., a “stress-then-sleep” association), the model used the stress rating the day before sleep (day_{t-1}) as the predictor of subsequent sleep duration (day_t). Thus, the formula for this model was specified as follows:

$$\text{ObjectiveSleepDuration}_t = \beta_0 + \beta_1 \text{SubjectiveStressRating}_{t-1} + \beta_2 \text{DayOfWeek}_t + \varepsilon_t$$

Next, to test whether sleep duration is related to stress levels the day after (i.e., a “sleep-then-stress” association), the model used stress rating the day after sleep (day_t) as predictor of the previous sleep duration (day_{t-1} , sleep duration last night):

$$\text{ObjectiveSleepDuration}_t = \beta_0 + \beta_1 \text{SubjectiveStressRating}_t + \beta_2 \text{DayOfWeek}_t + \varepsilon_t$$

In this model, stress rating (predictor) is back-predicting sleep duration (outcome), which effectively tests the question “is stress today associated with sleep last night?”. Since the objective sleep duration measure occurs temporally before the daily stress rating is submitted, results are interpreted as sleep duration being associated with increased/decreased stress the next day. Implementing the model in this way (rather than using sleep duration as the predictor and stress rating as the outcome) presented the advantage that the slope estimates across the two models are on equivalent units, namely, the number of minutes changed in sleep duration per unit of change in stress rating. This allows us to directly compare the quantitative outputs for the models testing the questions “is stress associated with subsequent sleep?” and “is sleep associated with subsequent stress?”.

As with the pilot dataset, we conducted model checks to further evaluate the specification and performance of the individual-level linear model framework in our target dataset, including posterior predictive checks, inspection of model residuals, and a comparison of the estimates obtained through the iLMs to the estimates obtained when the same data for the target sample was fit in a single, group-level multilevel linear model with random intercepts and random slopes per participant on the stress predictor. Additional inspection of model results against each participant’s raw time series data was conducted to complement our interpretation.

Results

Table 1 presents participant-level available data and summary statistics for sleep and stress variables used in analysis. Pilot participants provided a median of 178 usable observations for modeling, i.e., day-level observations collected during the school semesters with usable actigraphy and survey data (range 119-212 out of a total possible of 223). Target participants provided a median of 178 usable observations (range of 84-214 days). Participants’ total number of usable observations was not

correlated with their mean sleep duration ($r = 0.09$; $p = 0.53$) or mean stress levels ($r = -0.11$; $p = 0.42$). Additionally, participants' mean sleep duration did not differ on days with or without available survey data (paired t -test, $p = 0.59$), and participants' mean stress levels did not differ on days with or without available actigraphy data (paired t -test, $p = 0.70$). These observations suggest that participants' overall sleep and stress metrics did not introduce systematic missingness in the data (more detail on participant-level missing data is presented in **Figure S1 of Multimedia Appendix 1**).

Table 1. Participant-level descriptive statistics of data used in analyses.

Participant	Daily Observations	Mean Sleep Duration (SD)	Mean Sleep Quality (SD)	Mean Stress (SD)
P1	119	493.22 (101.89)	3.59 (0.64)	1.36 (0.74)
P2	212	412.08 (84.57)	2.50 (0.68)	2.95 (0.82)
P3	156	404.03 (68.81)	2.48 (0.55)	2.89 (0.93)
P4	200	425.20 (82.42)	2.64 (0.69)	3.15 (0.42)
P5	200	387.76 (58.32)	3.14 (0.74)	2.23 (1.14)
P6	147	379.29 (139.37)	2.65 (1.12)	3.85 (1.14)
T1	147	464.92 (96.66)	3.81 (0.94)	2.76 (1.06)
T2	212	460.47 (82.68)	3.08 (0.62)	3.16 (0.83)
T3	139	415.09 (114.00)	2.61 (0.71)	3.59 (0.82)
T4	103	349.96 (89.02)	2.81 (0.83)	1.69 (1.09)
T5	189	446.40 (60.47)	3.33 (0.62)	2.51 (0.83)
T6	199	465.13 (76.76)	3.07 (0.63)	2.68 (0.77)
T7	131	389.17 (56.96)	3.06 (0.91)	1.95 (1.19)
T8	173	402.33 (108.91)	2.87 (0.49)	2.69 (0.63)
T9	102	448.53 (144.97)	2.95 (0.79)	1.98 (1.01)
T10	208	432.88 (114.48)	2.71 (0.73)	3.07 (1.11)
T11	173	373.92 (94.30)	2.62 (0.73)	2.77 (0.78)
T12	87	475.03 (75.76)	2.83 (1.00)	3.57 (1.14)
T13	214	441.09 (73.22)	2.49 (0.56)	3.25 (0.70)
T14	84	454.23 (100.68)	3.40 (0.62)	2.52 (0.81)
T15	98	412.93 (64.32)	2.57 (0.66)	3.41 (0.87)
T16	152	414.31 (127.89)	2.70 (0.66)	3.08 (0.83)
T17	191	440.73 (74.20)	3.06 (0.50)	1.37 (0.68)
T18	92	385.57 (70.90)	3.09 (0.51)	2.87 (0.63)
T19	199	464.09 (74.95)	2.87 (0.44)	2.94 (0.59)
T20	192	497.65 (77.78)	3.69 (0.58)	2.30 (0.95)
T21	199	390.89 (82.58)	3.36 (0.76)	2.85 (1.05)
T22	126	464.25 (84.70)	3.33 (0.90)	2.53 (1.04)
T23	84	418.65 (73.60)	2.87 (0.62)	3.39 (0.79)
T24	148	482.32 (81.06)	3.19 (0.87)	2.01 (0.77)
T25	191	406.52 (95.58)	2.70 (0.78)	3.47 (1.00)
T26	205	446.68 (64.03)	3.01 (0.10)	2.94 (1.14)
T27	177	443.10 (114.21)	3.16 (0.82)	3.28 (0.76)

T28	209	409.12 (65.85)	2.92 (0.59)	2.67 (0.91)
T29	197	429.07 (74.85)	3.00 (0.47)	1.68 (0.88)
T30	195	396.16 (62.76)	2.67 (0.59)	2.85 (1.05)
T31	205	388.53 (94.91)	2.82 (0.83)	2.52 (0.71)
T32	117	397.49 (118.98)	2.61 (0.86)	3.73 (1.10)
T33	114	383.05 (135.02)	2.65 (0.89)	2.09 (1.27)
T34	206	436.76 (80.84)	2.88 (0.73)	2.43 (0.89)
T35	207	461.99 (71.50)	3.08 (0.78)	3.16 (1.23)
T36	188	470.90 (86.02)	3.36 (0.80)	2.70 (1.10)
T37	180	423.49 (78.52)	3.41 (0.63)	2.21 (0.85)
T38	139	409.22 (132.79)	2.56 (0.85)	3.39 (0.82)
T39	114	414.25 (99.79)	2.74 (0.67)	2.25 (1.07)
T40	102	483.04 (53.83)	3.37 (0.53)	2.84 (1.24)
T41	136	429.99 (91.77)	2.34 (0.67)	3.54 (0.74)
T42	193	416.79 (95.16)	3.26 (0.80)	1.21 (0.46)
T43	166	464.57 (116.07)	3.04 (0.35)	2.23 (0.82)
T44	179	456.01 (67.63)	3.34 (0.74)	1.96 (1.03)
T45	197	474.08 (64.90)	2.99 (0.60)	1.18 (0.54)
T46	203	478.48 (70.13)	1.99 (0.58)	2.35 (1.01)
T47	178	382.32 (78.80)	3.54 (0.72)	1.79 (0.95)
T48	189	406.65 (103.77)	2.81 (1.00)	2.17 (1.36)
T49	171	444.64 (87.14)	2.91 (0.86)	1.90 (1.11)

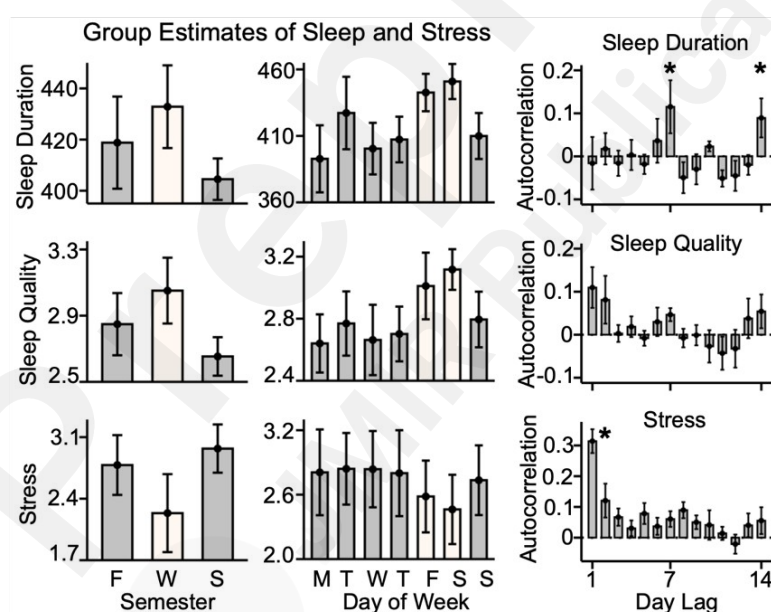
Notes: Sleep Duration is in minutes; Sleep Quality and Stress are on 1-5 Likert scale. Participant IDs starting with "P" indicate Pilot sample participants. IDs starting with "T" indicate Target sample participants. SD = standard deviation.

Sleep and Stress Fluctuate in Relation to the Academic Calendar in the Pilot Participants

There was a pattern of enhanced sleep and lower stress when students were released from the structured academic demands of the in-person school semesters. Participants were enrolled for a full academic year, including a Fall Semester and a Spring Semester (each lasting roughly 16 weeks), as well as a 5-week class-free Winter break bridging the two semesters. During Winter break and weekends, pilot participants had longer objective (actigraphy-derived) Sleep Duration, better subjective Sleep Quality, and felt less Stress relative to school semesters and weekdays (**Figure 1**, left and center columns). The temporal structure of sleep and stress variables was further evidenced by their autocorrelation estimates. Autocorrelations were generally small ($|r| < 0.2$; third column in Figure 1) but were strongest (highlighted with asterisks) at a 7-day lag (and again at a 14-day lag) for Sleep Duration, consistent with a weekly sleep schedule. By contrast, autocorrelations for Stress were strongest at 1- and 2-day lags, suggesting that experiences of stress might come in chains of more than one day.

The school break and week-related changes observed in the data informed the design of the iLM seeking to capture stable person-level associations. Given that Winter break presents different environmental demands from the school semester, we decided *a priori* to exclude from the model observations collected during this period. Additionally, to account for weekly patterns in sleep behavior (outcome variable) within the school semesters, we added day of the week as a covariate in the model.

Figure 1. Group Sleep and Stress metrics reveal structure related to the academic calendar. Pilot participants' group (between-person) means of Sleep Duration (in minutes, top row), Sleep Quality (5-point Likert scale, middle row), and perceived Stress (5-point Likert scale, bottom row), aggregated by school semester (Fall and Spring) and Winter Break (first column) and by day of the week (second column). The third column shows between-person means of autocorrelation estimates over a 14-day window. Asterisks highlight that autocorrelations were strongest at a 7-day lag (and again at 14-day lag) for Sleep Duration and at 1-day and 2-day lags for perceived Stress. Error bars show standard error of the mean. F = Fall semester, W = Winter break, S = Spring semester; M = Monday, T = Tuesday, W = Wednesday, T = Thursday, F = Friday, S = Saturday, S = Sunday.



Individual-Level Linear Models Capture Day-To-Day Associations in the Pilot Participants

To test the viability of the iLM approach, a proof-of-concept model examined the association between objective Sleep Duration and subjective Sleep Quality in the pilot participants (**Table 2 and Figure 2**). This model allowed for a test of construct validity given the tested association was intuitive and expected. The model tested the association between Sleep Duration and Sleep Quality for the same sleep episode (i.e., a “Concurrent” association). All pilot participants showed the

expected positive association: when participants slept for a shorter period than usual, they also subjectively rated those same sleep events as worse quality (shown in orange in Figure 2). The effect size was large for all individuals, and the 95% uncertainty interval (UI) lay outside of the region of practical equivalence (ROPE) for all models. These results provided preliminary evidence that the iLM is a valid framework to reliably detect relations between psychological variables (in this case, subjective Sleep Quality rating) and objective behaviors (Sleep Duration) at the individual participant level.

A second, exploratory iLM tested the association between Sleep Duration (on day_t) and Sleep Quality the Day Before (on day_{t-1}). We expected this model to show significant, but weaker, effects relative to the first model, given that the variables were referring to lagged sleep events. Moreover, we expected negative associations, such that worse sleep quality on one night would associate with longer sleep duration the next night, i.e., a compensatory sleep rebound effect. Half the participants (P1, P4, and P6) showed this expected pattern of association ($pd > 0.975$; shown in blue coloring in **Figure 2**). One participant (P2) showed a positive association, such that worse subjective sleep quality experienced on the preceding night was associated with shorter sleep durations on the subsequent night. This participant might experience chains of poor sleep over multiple days (e.g., reduced sleep days ahead of a deadline to accommodate increased workload), rather than a sleep rebound effect immediately the following night. Two additional participants showed a positive slope estimate but without reaching statistical significance ($pd < 0.975$; P3 and P5).

We did not observe structured patterns of association between individual slope estimates and individuals' Mean Sleep Quality, Mean Sleep Duration or total number of Daily Observations (**Figure 2b**). This suggests that the estimated slopes were not systematically influenced by person-level characteristics of the psychological and behavioral phenomena of interest.

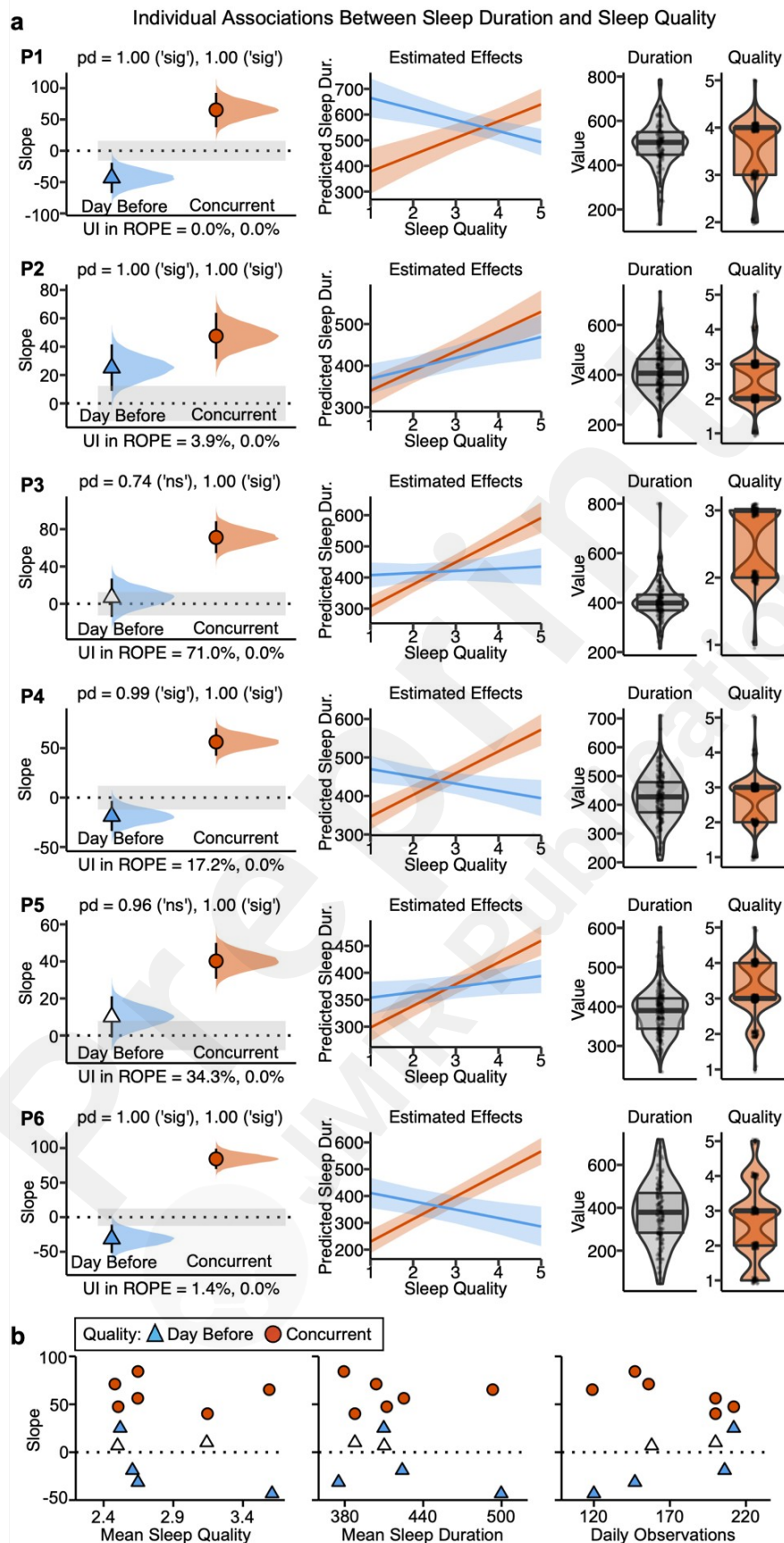
Table 2. Pilot sample results of individual-level linear models assessing Sleep Duration associated with Concurrent Sleep Quality and assessing Sleep Duration associated with Sleep Quality the Day Before.

Model and Participant	Slope (Med.)	95% UI Low	95% UI High	pd	ROPE (+/-)	%UI in ROPE	R-hat	ESS
Sleep Duration ~ Concurrent Sleep Quality								
P1	65.31*	37.73	92.42	1	15.84	0.00	1.00	7291
P2	47.48*	31.42	63.87	1	12.35	0.00	1.00	8177
P3	71.10*	54.30	88.46	1	12.50	0.00	1.00	7484
P4	56.32*	42.31	70.24	1	12.00	0.00	1.00	7649
P5	40.22*	30.65	49.95	1	7.89	0.00	1.00	7030

	P6	84.28*	69.58	99.15	1	12.50	0.00	1.00	6578
Sleep Duration ~ Sleep Quality the Day Before									
	P1	-43.28*	-67.61	-18.99	1	15.84	0.00	1.00	7726
	P2	25.05*	8.84	41.57	1	12.35	3.93	1.00	8244
	P3	6.80	-14.01	27.05	0.74	12.50	71.04	1.00	7478
	P4	-18.92*	-33.78	-3.38	0.99	12.00	17.20	1.00	7823
	P5	10.07	-1.05	21.18	0.96	7.89	34.29	1.00	8030
	P6	-31.39*	-51.90	-10.44	1	12.50	1.36	1.00	6911

Notes: All models included day of week as a covariate, but table only shows model parameters for the main predictor, Sleep Quality. P1-P6: Pilot participants 1-6. Med. = Median of the posterior distribution; 95% UI = 95% uncertainty interval; Low = low bound; High = high bound; pd = probability of direction; ROPE = region of practical equivalence; ESS = effective sample size. Tilde (~) denotes “regressed on”. Asterisk (*) next to Slope value denotes statistical significance.

Figure 2. Within-individual linear models capture real-world associations between objective Sleep Duration and subjective Sleep Quality. (a) Each row (P1-P6) shows model results for an individual pilot participant. Individual-level models testing Sleep Duration associated with Sleep Quality the Day Before are displayed with blue; models testing Sleep Duration associated with Concurrent Sleep Quality are displayed with orange. The first column shows the models’ estimated slopes (computed as the median of the posterior distribution) and uncertainty metrics. Symbol shading signifies statistically significant slopes. Error bars show 95% uncertainty intervals (UI). Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPE); pd = probability of direction. The second column shows the models’ predicted Sleep Duration as a function of Sleep Quality; shading around the lines indicates 95% UI. The third column shows the participant-level distributions of Sleep Duration (in minutes) and Sleep Quality (1-5 scale) across all daily observations used in analysis. **(b)** Slope estimates from the first column in (a) are plotted against participant-level estimates across the study period.

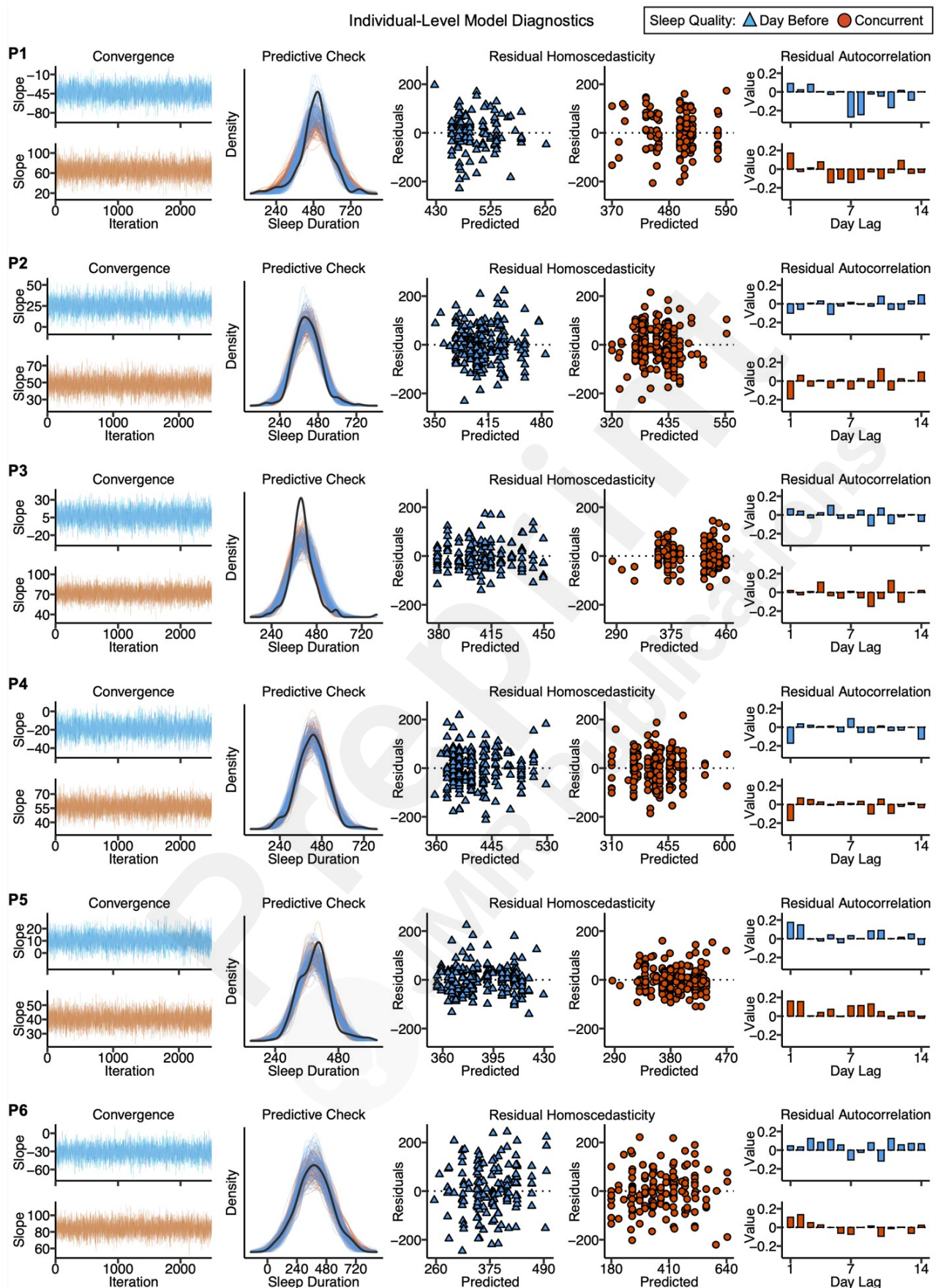


Model Diagnostics Confirm Convergence and Adequate Specification

Both visual and quantitative MCMC diagnostic checks revealed that all iLMs converged successfully (**Figure 3**). Trace plots (Figure 3, first column) revealed no structured pattern in the estimated slopes across sampling iterations. R-hat values were < 1.1 and Effective Sample Sizes were $> 1,000$ for the estimated slopes of all models (Table 2). Posterior predictive checks comparing the observed distribution of the outcome variable (Sleep Duration) to 100 randomly sampled simulated datasets from the posterior predictive distribution confirmed that a gaussian model specification captured the observed data well (Figure 3, second column). Inspection of the model residuals against the model's predicted values confirmed homoscedasticity, with no structured pattern in the plots (Figure 3, third and fourth columns). Finally, residual autocorrelation was generally low ($|r| < 0.2$) for all models (Figure 3, fifth column), suggesting that the model specification was able to account for the temporal structure in the data.

Figure 3. Within-individual model diagnostics suggest adequate convergence and specification.

Each row (P1-P6) shows model diagnostics for an individual pilot participant. Diagnostics for the models assessing the association between Sleep Duration and Sleep Quality the Day Before are displayed with blue; models assessing the association between Sleep Duration and Concurrent Sleep Quality are displayed with orange. The first column's trace plots display time series of each Markov chain's estimated Slope (y-axis) as a function of post-warmup Iterations (x-axis). The second column shows posterior Predictive Checks; black Density lines show the observed distributions of the outcome (Sleep Duration), and the thin colored Density lines show 100 replicated outcome distributions generated based on random samples from the models' parameters' posterior distributions. The third and fourth columns show the models' Predicted values (x-axis) against the models' Residuals (y-axis). The fifth column shows Autocorrelation plots of model residuals.

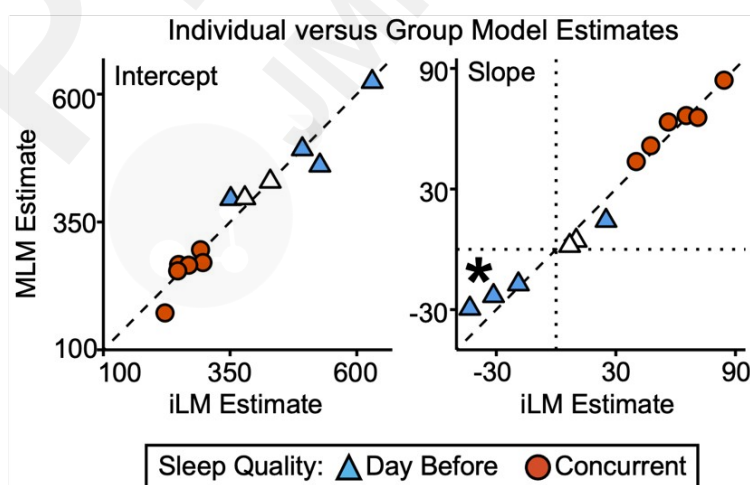


Individual-Level Linear Models Yield Similar, but Not Identical, Estimates to a Group-Based Multilevel Model

The slopes estimated from the iLM are similar to those obtained through a hierarchical or multilevel linear model (MLM) testing the same associations between concurrent objective Sleep Duration and subjective Sleep Quality (**Figure 4**). This suggests the individually tailored slopes obtained from our iLM are comparable to more traditional group-based approaches. Of note, although the estimates are similar, they are not identical.

Comparison of the two models suggested attenuation of the individual-level estimates in the group models (see part of Figure 4 highlighted with an asterisk). Even with random effects by participant, in the MLM these estimates are, by design, biased by the group and may underestimate individual-level effects, especially for uncommon phenotypes [20]. As will be revealed later in the results from the larger sample of target participants, group-based estimation can even result in the reversal of the sign of the association for some individuals.

Figure 4. Within-individual linear models of the association between Sleep Duration and Sleep Quality yield similar, but not identical, estimates to a group-based linear model. Estimates of the intercepts and slopes from the individual-level linear models (iLM; x-axis) for each participant compared to the random effects estimated from a group-based multilevel linear model (MLM; y-axis). Triangles show estimates for the models examining the association between Sleep Duration and Sleep Quality the Day Before; circles show estimates for the models examining the association between Sleep Duration and Concurrent Sleep Quality. Symbol shading signifies statistically significant slopes in the iLMs. Dashed diagonal lines represent identical estimates between the individual- and group-based approaches. Asterisk highlights attenuation of individual estimates in the MLM relative to the iLMs.



The iLM is a Valid Framework to Identify Individual-Level Associations

Proof-of-concept analyses of the pilot data demonstrated that the iLM successfully captures expected associations between objective behavior and psychological variables, has an adequate model

specification, and captures individual-level estimates unbiased by the group. Moreover, point estimates obtained with our Bayesian iLMs are virtually identical to those obtained through iLMs fitted with a frequentist inference framework (**Figure S2 of Multimedia Appendix 1**), suggesting that our individual-level modeling approach is robust across both major statistical inference frameworks. In all, the iLM is a valid framework to identify individual-level associations, which justified carrying over the model to test target hypotheses regarding the association between perceived stress and objective sleep duration in the independent, target dataset of N=49 first-year students.

Sleep and Stress Fluctuate with the Academic Calendar in the Target Participants

As with the pilot participants, during Winter Break and weekends, participants had longer objective Sleep Duration, better subjective Sleep Quality, and felt less Stressed relative to the Fall and Spring semesters (first and second columns in **Figure S3 of Multimedia Appendix 1**). We replicated the autocorrelation pattern seen in the pilot dataset: Sleep Duration showed the greatest autocorrelation at a 7-day lag (and then again at a 14-day lag), indicating sleep patterns are tied to a weekly schedule, while Stress showed the greatest autocorrelation at a 1-day and 2-day lags, indicating that experiences of stress might come in chains of a few days (third column in **Figure S3 of Multimedia Appendix 1**). These school break and week-related dynamics confirm that our target dataset captured real-life dynamics associated with college life and reinforce our individual-level modeling decisions regarding the exclusion of Winter Break data and addition of day of the week as a covariate.

Shorter Objective Sleep Duration is Associated with Worse Subjective Sleep Quality in the Target Participants

All participants showed a positive slope estimate for the association between Concurrent Sleep Duration and Sleep Quality, and this effect was statistically significant in 46 out of 49 target participants ($p < 0.05$; **Table 3** and **Figure 5a**, in orange). In other words, when participants slept for a shorter period than usual, they also rated those same sleep events as worse quality, demonstrating that the iLM can capture expected real-world relations between behavioral and psychological phenomena. Among participants that showed a significant association, a 1-point increase in a 5-point sleep quality scale was associated with a sleep episode that was also longer by a range of 16 to 119 minutes (median = 59 minutes), a substantial increase in sleep duration for most

participants considering that the average sleep duration in the sample was 432 minutes (7.2 hours).

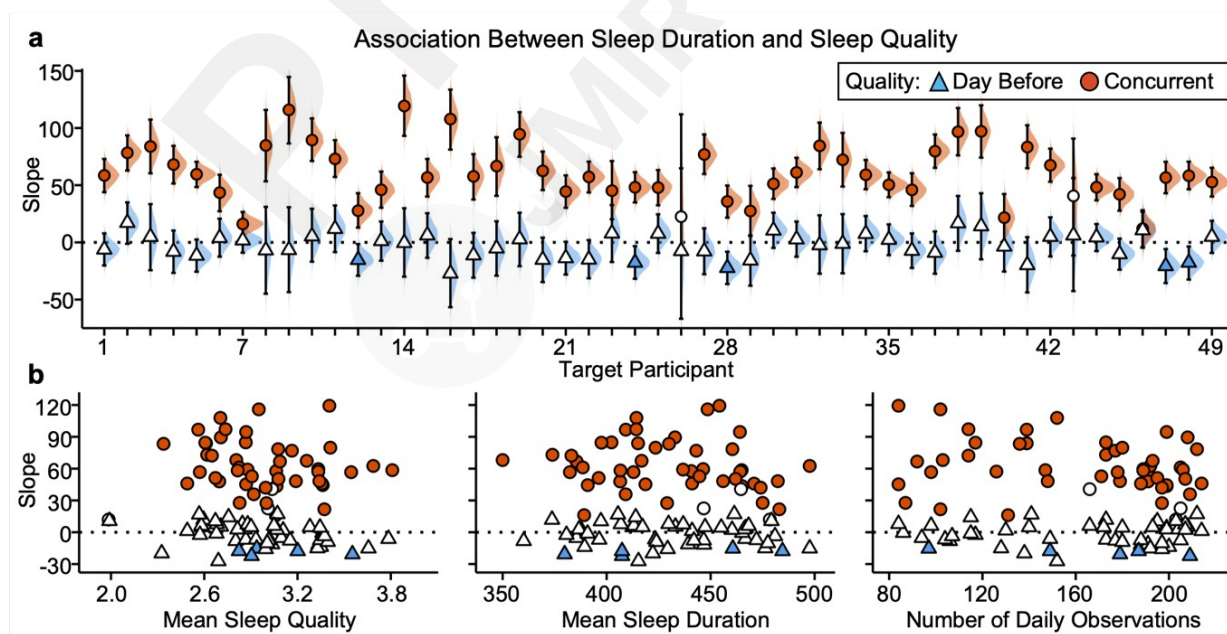
Table 3. Target sample results of individual-level linear models assessing Sleep Duration associated with Concurrent Sleep Quality.

Participant	Slope (Med)	95% UI Low	95% UI High	pd	ROPE (+/-)	%UI in ROPE	R-hat	ESS
T1	58.57*	43.90	72.99	1	10.30	0.00	1.00	7523
T2	78.35*	62.80	93.58	1	13.36	0.00	1.00	8984
T3	83.88*	60.42	107.40	1	16.12	0.00	1.00	7211
T4	68.08*	51.48	84.48	1	10.74	0.00	1.00	7738
T5	59.57*	48.38	70.46	1	9.77	0.00	1.00	8168
T6	43.51*	27.26	59.18	1	12.22	0.00	1.00	8029
T7	16.10*	5.49	26.60	1	6.26	0.82	1.00	7772
T8	84.78*	53.43	115.84	1	22.05	0.00	1.00	8598
T9	115.91*	86.49	144.67	1	18.39	0.00	1.00	6519
T10	89.47*	71.20	108.46	1	15.77	0.00	1.00	7813
T11	73.07*	57.17	89.42	1	12.98	0.00	1.00	7347
T12	27.69*	12.94	43.04	1	7.56	0.00	1.00	7765
T13	45.99*	30.31	61.92	1	13.01	0.00	1.00	7855
T14	119.30*	93.16	145.87	1	16.15	0.00	1.00	6735
T15	56.74*	40.07	72.96	1	9.77	0.00	1.00	8048
T16	107.86*	81.17	133.61	1	19.39	0.00	1.00	8373
T17	57.72*	37.47	77.20	1	14.88	0.00	1.00	8767
T18	66.72*	40.75	91.96	1	14.01	0.00	1.00	7313
T19	94.56*	74.90	113.90	1	16.98	0.00	1.00	7965
T20	62.54*	45.72	79.42	1	13.31	0.00	1.00	9058
T21	44.44*	30.24	58.59	1	10.88	0.00	1.00	8560
T22	57.24*	43.62	70.66	1	9.38	0.00	1.00	7436
T23	45.24*	20.23	71.17	1	11.94	0.00	1.00	7503
T24	48.20*	34.92	61.39	1	9.34	0.00	1.00	8482
T25	47.93*	32.46	63.25	1	12.21	0.00	1.00	8742
T26	22.51	-66.79	112.05	0.70	64.98	84.15	1.00	9231
T27	76.92*	59.80	94.43	1	13.93	0.00	1.00	6829
T28	35.80*	21.59	49.73	1	11.07	0.00	1.00	9366
T29	27.47*	5.63	49.43	0.99	15.80	13.09	1.00	9333
T30	51.20*	37.94	64.63	1	10.68	0.00	1.00	10463
T31	61.13*	48.35	74.01	1	11.39	0.00	1.00	8859
T32	84.47*	63.97	104.78	1	13.82	0.00	1.00	8314
T33	72.26*	48.83	95.74	1	15.13	0.00	1.00	7530
T34	59.12*	46.46	72.07	1	11.04	0.00	1.00	8186
T35	50.34*	39.48	61.15	1	9.11	0.00	1.00	9152
T36	45.89*	31.88	60.45	1	10.69	0.00	1.00	8367
T37	79.76*	65.18	94.36	1	12.42	0.00	1.00	8393
T38	96.75*	76.15	117.53	1	15.58	0.00	1.00	6420

T39	97.12*	74.12	119.88	1	14.99	0.00	1.00	7432
T40	21.66*	1.58	42.28	0.98	10.25	11.05	1.00	7083
T41	83.39*	64.82	102.49	1	13.72	0.00	1.00	7958
T42	67.41*	53.03	81.99	1	11.89	0.00	1.00	8166
T43	40.65	-11.52	90.81	0.94	33.52	38.51	1.00	7499
T44	48.21*	36.79	59.67	1	9.10	0.00	1.00	8509
T45	42.05*	27.71	56.25	1	10.78	0.00	1.00	9139
T46	11.97	-4.23	28.15	0.93	12.00	50.13	1.00	9021
T47	56.68*	42.98	70.50	1	10.92	0.00	1.00	8050
T48	58.27*	46.41	70.59	1	10.36	0.00	1.00	7819
T49	52.76*	40.08	65.24	1	10.18	0.00	1.00	7410

Notes: All individual-level models included day of week as a covariate, but table only shows model parameters for the main predictor, Sleep Quality. T1-T49: Target participants 1-49. Med. = Median of the posterior distribution; 95% UI = 95% uncertainty interval; Low = low bound; High = high bound; pd = probability of direction; ROPE = region of practical equivalence; ESS = effective sample size. Asterisk (*) next to Slope value denotes statistical significance.

Figure 5. Longer objective Sleep Duration associates with Higher subjective Sleep Quality in most individuals. (a) Estimated Slopes (corresponding to the median of the posterior distribution) from the individual-level models assessing Sleep Duration associated with Sleep Quality the Day Before (triangles) and with Concurrent Sleep Quality (circles) are plotted along the y-axis ordered by participant (x-axis). Statistically significant slope estimates are shaded blue (Day Before) or orange (Concurrent). Error bars show 95% uncertainty intervals and shaded density plots show the full posterior distributions of the slopes. **(b)** Slope estimates from panel (a) are plotted against participant-level estimates across the study period.



For the exploratory lagged model, the expected negative association between Sleep Duration

and Sleep Quality the Day Before reached statistical significance in only five participants ($pd > 0.975$; **Table 4** and **Figure 5a**, in blue). In other words, sleep rated as worse quality was consistently followed by longer sleep the following night in just a small subset of participants, suggesting that sleep rebound effects are perhaps less common or reliable than anticipated. Among participants that showed a significant association, a 1-point decrease in a 5-point sleep quality scale was associated with a subsequent sleep episode that was longer by a range of 15 to 22 minutes across participants (median = 18 minutes), a much smaller effect compared to the concurrent model.

We did not observe structured patterns of association between individual slope estimates and individuals' Mean Sleep Quality, Mean Sleep Duration, or total Number of Daily Observations, suggesting that the model results were not systematically influenced by person-level aggregates of the variables that went into the model (**Figure 5b**).

Table 4. Target sample results of individual-level linear models assessing Sleep Duration associated with Sleep Quality the Day Before.

Participant	Slope (Med.)	95% UI Low	95% UI High	pd	ROPE (+/-)	%UI in ROPE	R-hat	ESS
T1	-6.18	-20.18	7.90	0.81	10.30	72.79	1.00	7331
T2	17.18	-1.23	35.01	0.97	13.36	33.68	1.00	7898
T3	4.86	-24.38	33.47	0.62	16.12	73.18	1.00	7339
T4	-8.14	-26.68	10.37	0.81	10.74	61.51	1.00	7692
T5	-11.59	-25.67	2.53	0.95	9.77	38.92	1.00	8057
T6	3.99	-12.52	20.73	0.68	12.22	84.16	1.00	8124
T7	1.79	-9.04	12.40	0.63	6.26	76.76	1.00	7715
T8	-6.92	-44.90	31.06	0.64	22.05	75.05	1.00	8294
T9	-6.64	-43.52	30.76	0.63	18.39	68.00	1.00	6542
T10	5.38	-16.89	29.41	0.68	15.77	82.23	1.00	8276
T11	11.88	-8.52	32.29	0.88	12.98	54.38	1.00	7592
T12	-15.31*	-29.24	-1.46	0.98	7.56	11.92	1.00	6981
T13	1.30	-15.99	18.29	0.56	13.01	90.63	1.00	8803
T14	-0.44	-30.03	29.90	0.51	16.15	76.59	1.00	7033
T15	6.21	-13.53	25.51	0.73	9.77	60.83	1.00	7607
T16	-27.16	-56.68	2.91	0.96	19.39	29.00	1.00	7630
T17	-11.26	-30.90	8.36	0.87	14.88	64.72	1.00	8548

T18	-5.27	-29.15	18.45	0.67	14.01	74.54	1.00	7075
T19	2.84	-20.85	25.94	0.59	16.98	87.60	1.00	7754
T20	-15.15	-34.71	4.14	0.94	13.31	42.17	1.00	7930
T21	-13.72	-28.15	0.59	0.97	10.88	34.11	1.00	8292
T22	-14.97	-31.39	2.04	0.96	9.38	24.14	1.00	7859
T23	7.89	-17.17	32.62	0.73	11.94	59.32	1.00	7688
T24	-17.72*	-31.83	-3.24	0.99	9.34	10.94	1.00	8332
T25	7.69	-9.26	24.56	0.82	12.21	71.23	1.00	7155
T26	-7.54	-80.37	64.79	0.58	64.98	96.56	1.00	8465
T27	-7.78	-27.82	12.37	0.78	13.93	74.72	1.00	8543
T28	-22.29*	-36.43	-8.12	1	11.07	3.94	1.00	7913
T29	-15.76	-37.87	6.16	0.92	15.80	50.18	1.00	7036
T30	10.45	-4.82	25.99	0.91	10.68	51.31	1.00	6695
T31	2.76	-12.58	18.24	0.64	11.39	86.79	1.00	6494
T32	-2.39	-27.37	23.67	0.57	13.82	74.03	1.00	7438
T33	-1.18	-27.00	24.74	0.54	15.13	79.09	1.00	6429
T34	7.50	-7.80	23.05	0.83	11.04	68.41	1.00	8185
T35	2.35	-11.15	15.76	0.63	9.11	83.18	1.00	7304
T36	-7.26	-22.38	7.85	0.83	10.69	67.71	1.00	7446
T37	-8.99	-27.43	9.56	0.83	12.42	65.23	1.00	7425
T38	16.81	-7.39	40.61	0.92	15.58	45.69	1.00	7448
T39	14.22	-14.82	42.98	0.84	14.99	52.11	1.00	7901
T40	-3.62	-25.53	18.36	0.63	10.25	64.99	1.00	7758
T41	-19.81	-43.76	4.64	0.94	13.72	30.36	1.00	8205
T42	4.94	-12.29	21.91	0.72	11.89	80.34	1.00	6956
T43	6.17	-42.60	56.27	0.60	33.52	84.63	1.00	7107
T44	4.48	-7.61	16.09	0.76	9.10	78.73	1.00	7931
T45	-10.23	-23.89	3.28	0.93	10.78	53.33	1.00	8686
T46	10.73	-4.71	26.80	0.91	12.00	56.45	1.00	8086
T47	-20.67*	-35.59	-5.83	1	10.92	8.02	1.00	9068
T48	-17.71*	-32.53	-3.60	0.99	10.36	13.20	1.00	8895
T49	4.96	-9.25	18.87	0.76	10.18	78.56	1.00	8084

Notes: All individual-level models included day of week as a covariate, but table only shows model parameters for the main predictor, Sleep Quality. T1-T49: Target participants 1-49. Med. = Median of the posterior distribution; 95% UI = 95% uncertainty interval; Low = low bound; High = high bound; pd = probability of direction; ROPE = region of practical equivalence; ESS = effective sample size. Asterisk (*) next to Slope value denotes statistical significance.

Higher Subjective Stress is Associated with Shorter Objective Sleep Duration in Most Target Participants, but the Direction of the Temporal Association Varies

The slope estimate for the association between Stress and Sleep duration was negative in 86% of all individual-level linear models fitted, such that an increase in Stress was associated with a decrease in Sleep Duration (**Tables 5-6** and **Figure 6a**). This association between Stress and Sleep Duration reached statistical significance in 19 out of 49 participants ($p > 0.975$), and all but one showed a negative association. Of note, the only participant that showed a significant positive association (T15, for the association between Sleep Duration and Stress the Day After) also had no usable actigraphy data over the full Spring semester due to a technical issue with their wristband, and their estimated positive slope should therefore be interpreted with caution considering the structured missingness in their data. No other participant in the target sample showed this kind of systematic missingness in either the actigraphy or survey data streams (see **Figure S1 of Multimedia Appendix 1**).

Table 5. Target sample results of individual-level linear models assessing Sleep Duration associated with Stress the Day Before.

Participant	Slope (Med.)	95% UI Low	95% UI High	pd	ROPE (+/-)	%UI in ROPE	R-hat	ESS
T1	-17.39*	-30.84	-4.23	0.99	9.09	9.64	1.00	7920
T2	-25.47*	-38.71	-12.52	1	9.94	0.00	1.00	7532
T3	-11.75	-36.19	12.93	0.83	13.68	56.25	1.00	7497
T4	-17.88*	-31.13	-4.48	0.99	8.25	5.45	1.00	7531
T5	-6.59	-17.46	4.17	0.89	7.29	55.09	1.00	8189
T6	-12.27	-26.09	1.72	0.96	9.98	36.82	1.00	8310
T7	0.37	-7.96	8.40	0.53	4.82	78.46	1.00	7568
T8	-3.97	-31.88	24.48	0.60	16.89	78.29	1.00	9005
T9	-12.75	-39.38	13.84	0.83	14.68	56.53	1.00	6773
T10	-16.42*	-29.99	-2.96	0.99	10.19	16.55	1.00	8429
T11	-16.79	-34.71	1.64	0.96	11.91	29.51	1.00	8051
T12	-4.08	-16.42	8.22	0.74	6.73	65.81	1.00	7427
T13	-9.18	-22.63	4.30	0.91	10.44	57.36	1.00	8887
T14	-10.01	-34.82	15.03	0.79	12.84	58.66	1.00	6569
T15	6.40	-8.58	21.04	0.80	7.33	54.53	1.00	8457
T16	-4.81	-28.89	19.49	0.65	15.46	78.60	1.00	7154
T17	0.47	-15.28	15.70	0.52	11.19	88.72	1.00	7480
T18	2.55	-17.09	22.65	0.60	11.74	78.35	1.00	8255

T19	-33.47*	-50.42	-16.44	1	12.28	0.00	1.00	7885
T20	0.66	-12.06	13.07	0.55	8.29	84.76	1.00	7970
T21	-16.43*	-27.28	-5.51	1	7.88	3.93	1.00	8275
T22	-8.13	-22.65	6.56	0.86	8.08	49.73	1.00	8495
T23	-16.85	-36.46	2.27	0.96	9.30	19.97	1.00	6432
T24	-20.04*	-35.95	-4.06	0.99	10.92	11.06	1.00	8168
T25	-25.95*	-39.11	-13.25	1	9.44	0.00	1.00	7895
T26	-7.07	-14.59	0.61	0.97	5.63	34.56	1.00	8812
T27	-7.98	-28.47	12.74	0.78	14.84	76.27	1.00	8484
T28	0.77	-8.97	10.58	0.56	7.28	89.44	1.00	8063
T29	-16.06*	-27.87	-3.96	1	8.52	8.74	1.00	7850
T30	-11.06*	-19.45	-2.63	1	5.92	9.27	1.00	8544
T31	7.99	-9.54	25.47	0.81	13.61	74.60	1.00	8993
T32	-15.49	-33.63	3.45	0.95	10.88	30.23	1.00	9197
T33	-15.60	-33.42	2.04	0.96	10.69	28.00	1.00	7358
T34	-20.70*	-32.87	-8.37	1	9.07	0.79	1.00	7973
T35	0.24	-7.97	8.63	0.52	5.75	87.05	1.00	8332
T36	-11.30*	-22.46	-0.28	0.98	7.70	25.16	1.00	6861
T37	-14.13*	-28.07	-0.10	0.98	9.32	23.84	1.00	7889
T38	-26.90*	-48.84	-5.17	0.99	15.98	14.28	1.00	7520
T39	-13.52	-31.21	4.22	0.93	9.20	29.79	1.00	7380
T40	-3.02	-12.06	5.95	0.75	4.39	59.48	1.00	7465
T41	15.91	-5.36	37.06	0.93	12.45	35.81	1.00	7703
T42	-13.10	-41.31	15.65	0.82	20.77	71.60	1.00	8825
T43	-14.40	-35.39	5.80	0.92	14.06	48.56	1.00	7357
T44	-6.76	-15.22	1.43	0.94	6.47	46.99	1.00	8141
T45	-5.39	-20.66	10.17	0.75	12.13	81.79	1.00	8410
T46	-7.56	-17.10	2.05	0.94	6.98	44.73	1.00	7513
T47	-2.84	-14.49	8.84	0.69	8.30	83.36	1.00	8012
T48	-9.00	-20.33	2.12	0.95	7.59	39.17	1.00	8326
T49	-7.55	-18.29	3.36	0.92	7.85	52.52	1.00	7613

Notes: All individual-level models included day of week as a covariate, but table only shows model parameters for the main predictor, Stress the Day After. T1-T49: Target participants 1-49. Med. = Median of the posterior distribution; 95% UI = 95% uncertainty interval; Low = low bound; High = high bound; pd = probability of direction; ROPE = region of practical equivalence; ESS = effective sample size. Asterisk (*) next to Slope value denotes statistical significance.

Table 6. Target sample results of individual-level linear models assessing Sleep Duration associated with Stress the Day After.

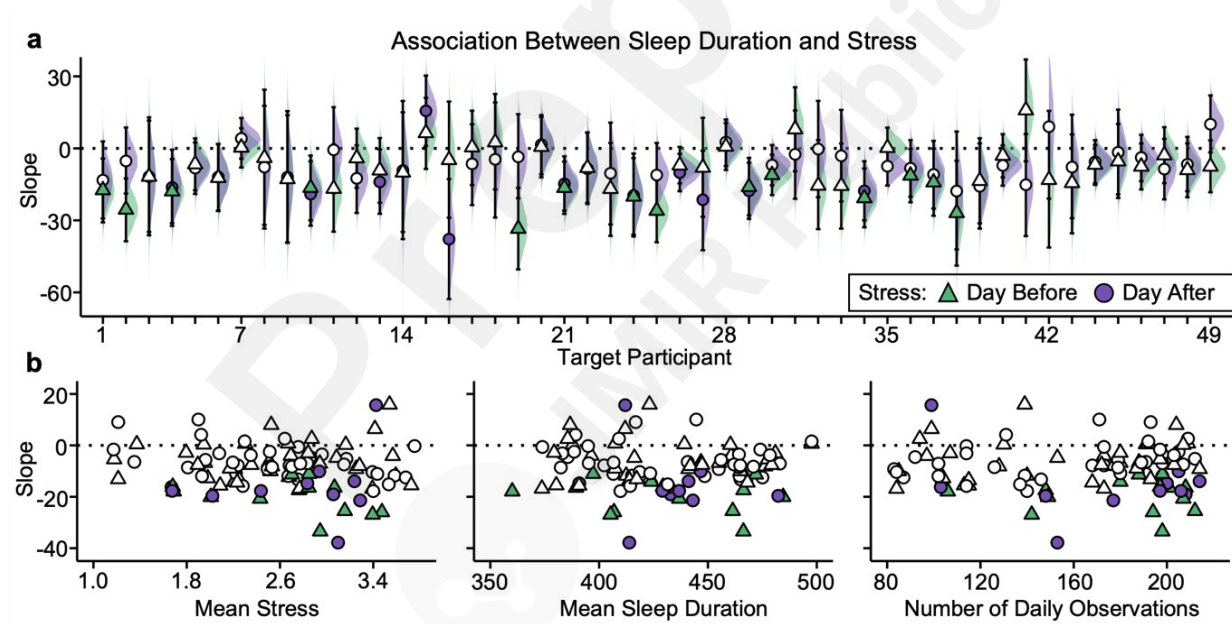
Participant	Slope (Med)	95% UI Low	95% UI High	pd	ROPE (+/-)	%UI in ROPE	R-hat	ESS
T1	-13.23	-29.23	2.91	0.95	9.09	29.61	1.00	5797

T2	-5.24	-18.80	8.68	0.77	9.94	76.74	1.00	8875
T3	-11.77	-34.81	11.56	0.85	13.68	56.47	1.00	7338
T4	-16.22*	-32.32	-0.57	0.98	8.25	15.11	1.00	8021
T5	-8.21	-18.99	2.43	0.94	7.29	42.56	1.00	7377
T6	-11.94	-25.88	1.90	0.96	9.98	37.93	1.00	8926
T7	4.12	-4.28	12.64	0.84	4.82	56.73	1.00	8095
T8	-7.79	-33.38	17.75	0.73	16.89	77.04	1.00	8068
T9	-11.98	-39.16	15.54	0.80	14.68	57.80	1.00	8224
T10	-19.06*	-32.23	-5.39	1	10.19	7.72	1.00	7881
T11	-0.62	-18.60	17.18	0.53	11.91	85.57	1.00	7375
T12	-12.49	-26.89	1.31	0.96	6.73	18.85	1.00	7661
T13	-13.95*	-27.24	-0.47	0.98	10.44	30.02	1.00	7455
T14	-9.24	-37.81	19.76	0.73	12.84	55.80	1.00	5908
T15	15.64*	1.01	30.35	0.98	7.33	11.11	1.00	7561
T16	-37.82*	-62.78	-12.78	1	15.46	1.60	1.00	7345
T17	-6.43	-23.61	10.03	0.78	11.19	72.94	1.00	7712
T18	-4.56	-28.81	19.13	0.65	11.74	66.61	1.00	8545
T19	-3.52	-20.73	14.26	0.66	12.28	83.96	1.00	8844
T20	1.55	-10.79	13.91	0.60	8.29	84.36	1.00	7304
T21	-14.76*	-25.94	-3.31	0.99	7.88	9.48	1.00	7913
T22	-8.48	-23.16	6.68	0.87	8.08	47.68	1.00	8106
T23	-10.36	-31.70	10.72	0.83	9.30	44.67	1.00	7312
T24	-19.63*	-36.73	-2.13	0.99	10.92	14.05	1.00	7158
T25	-11.10	-24.15	2.27	0.95	9.44	39.41	1.00	8495
T26	-10.11*	-17.83	-2.28	1	5.63	10.24	1.00	8767
T27	-21.40*	-42.43	-0.97	0.98	14.84	25.22	1.00	8373
T28	2.60	-7.35	12.04	0.70	7.28	84.95	1.00	8551
T29	-17.68*	-29.37	-5.97	1	8.52	3.66	1.00	8521
T30	-6.95	-15.37	1.40	0.95	5.92	40.05	1.00	8895
T31	-2.47	-20.99	15.76	0.60	13.61	88.88	1.00	9079
T32	-0.28	-20.33	19.80	0.51	10.88	76.02	1.00	8377
T33	-3.11	-22.10	16.05	0.62	10.69	73.59	1.00	7235
T34	-17.71*	-29.99	-5.33	1	9.07	6.43	1.00	8398
T35	-7.43	-15.50	0.75	0.96	5.75	32.79	1.00	8137
T36	-8.27	-20.14	3.43	0.92	7.70	45.84	1.00	6671
T37	-10.86	-25.05	3.13	0.94	9.32	41.21	1.00	7458
T38	-17.78	-42.09	7.02	0.93	15.98	43.95	1.00	8022
T39	-15.79	-33.41	2.06	0.96	9.20	21.49	1.00	7141
T40	-7.11	-15.56	1.26	0.95	4.39	24.88	1.00	7108
T41	-15.15	-36.52	5.96	0.92	12.45	39.48	1.00	8726
T42	9.05	-20.92	38.93	0.72	20.77	79.25	1.00	8862
T43	-7.80	-29.05	14.03	0.76	14.06	72.40	1.00	7545
T44	-5.63	-14.57	3.44	0.89	6.47	57.66	1.00	8151

T45	-1.63	-19.16	16.16	0.58	12.13	86.41	1.00	8594
T46	-3.81	-13.64	5.68	0.79	6.98	75.80	1.00	8400
T47	-8.61	-21.25	3.95	0.91	8.30	47.73	1.00	7590
T48	-6.65	-17.83	4.77	0.87	7.59	56.89	1.00	7442
T49	10.06	-1.66	22.09	0.96	7.85	34.47	1.00	8521

Notes: All individual-level models included day of week as a covariate, but table only shows model parameters for the main predictor, Stress the Day Before. T1-T49: Target participants 1-49. Med. = Median of the posterior distribution; 95% UI = 95% uncertainty interval; Low = low bound; High = high bound; pd = probability of direction; ROPE = region of practical equivalence; ESS = effective sample size. Asterisk (*) next to Slope value denotes statistical significance.

Figure 6. Higher subjective Stress associates with shorter objective Sleep Duration in most individuals. (a) Estimated Slopes (corresponding to the median of the posterior distribution) from the individual-level models assessing Sleep Duration associated with Stress the Day Before (triangles) and with Stress the Day After (circles) are plotted along the y-axis ordered by participant (x-axis). Statistically significant slope estimates are shaded green (Day Before) or purple (Concurrent). Error bars show 95% uncertainty intervals and shaded density plots show the full posterior distributions of the slopes. (b) Slope estimates from panel (a) are plotted against each participant-level estimates across the study period.



Of the 19 participants that showed a statistically significant relationship between Stress and Sleep Duration, eight participants (42%) showed only the “stress-then-sleep” phenotype, i.e., days with shorter Sleep Durations were preceded by higher Stress the Day Before (**Figure 6**, in green), but not vice versa. Five participants (26%) showed only the “sleep-then-stress” phenotype, i.e., nights with shorter Sleep Duration were followed by increased Stress the Day After (**Figure 6**, in purple). Six participants showed bidirectional effects (32%), such that nights with shorter Sleep Duration were preceded by increased Stress the Day Before as well as followed by increased Stress the Day

After. Among participants that showed a significant association between today's stress levels and sleep duration later that night, a 1-point increase in a 5-point perceived stress scale was associated with shorter subsequent sleep duration ranging from 11 to 33 minutes across participants (median = 17 minutes). Among participants that showed a significant association between today's stress levels and last night's sleep duration, a 1-point increase in a 5-point perceived stress scale was associated with shorter previous sleep duration ranging from 10 to 38 minutes across participants (median = 18 minutes). These effects would compound to more substantial reductions in sleep duration with greater increases in daily stress.

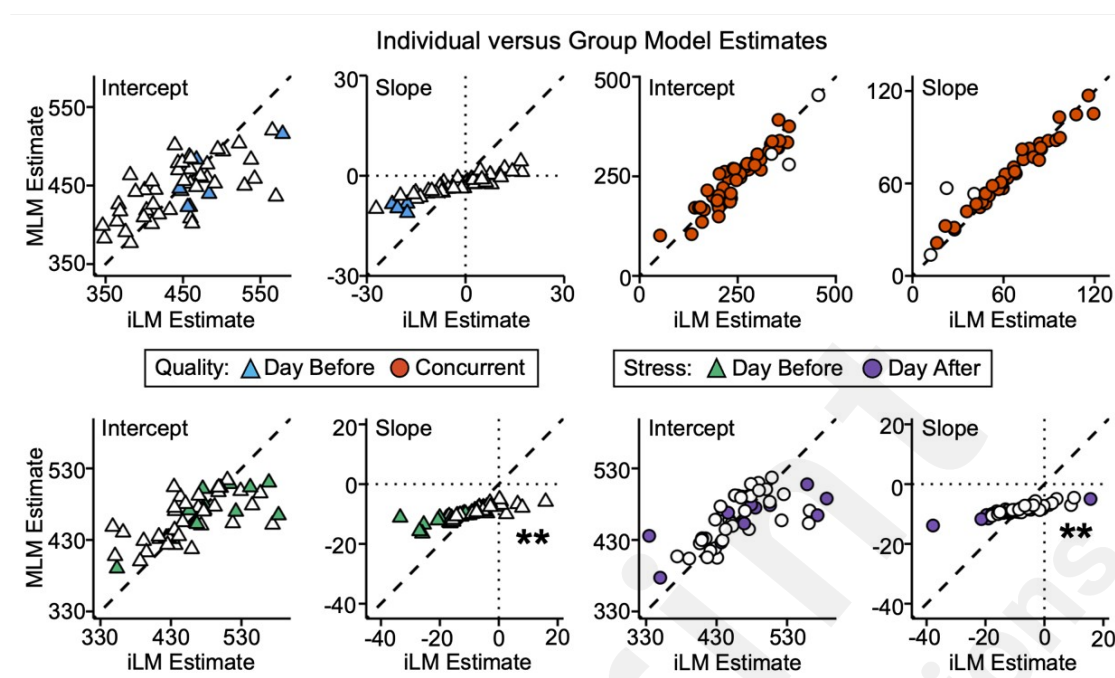
Individual slope estimates of the association between Sleep Duration and Stress showed no clear pattern of association with individuals' Mean Sleep Duration or Number of Daily Observations that went into the model (**Figure 6b**, center and right). There was a slight trend for individuals with higher Mean Stress to have a larger absolute slope estimate (**Figure 6b**, left), perhaps because participants with very low Stress levels have little variance to be modeled.

Person-Specific Estimates Get Attenuated in Group-Based Modelling

A comparison of the estimates obtained through the iLM with those of a MLM demonstrates that individual-level estimates get systematically attenuated in a group-based approach when there is between-person heterogeneity in the effects (**Figure 7**). When there was a strong effect and small between-person variability in the tested associations, as with the association between Sleep Duration and Concurrent Sleep Quality, the individual linear models provided nearly identical slope estimates to those estimated through a multilevel model (**Figure 7**, top right). However, for associations that showed a weaker effect and greater degree of between-person variability, group-level approaches systematically flatten individually-tailored effect sizes, or even reverse the sign of the association. We observed this in the lead-lag associations between Stress and Sleep Duration (as highlighted by the two asterisks on Figure 7), and to some degree in the lagged association between Sleep Quality and Sleep Duration.

Figure 7. Within-individual estimates of the association between Stress and Sleep Duration show systematic attenuation in the group-based linear model. Estimates of the intercepts and slopes from the individual-level linear models (iLM; x-axis) for each participant compared to the random effects estimated from a group-based multilevel linear model (MLM; y-axis), presented separately for associations between Sleep Duration and Sleep Quality (top row), and associations between Stress and Sleep Duration (bottom row). Symbol shading signifies statistically significant slopes in the iLMs. Dashed diagonal lines represent identical estimates between the individual- and group-based approaches. Asterisks highlight attenuation of individual estimates in the MLM relative

to the iLMs.



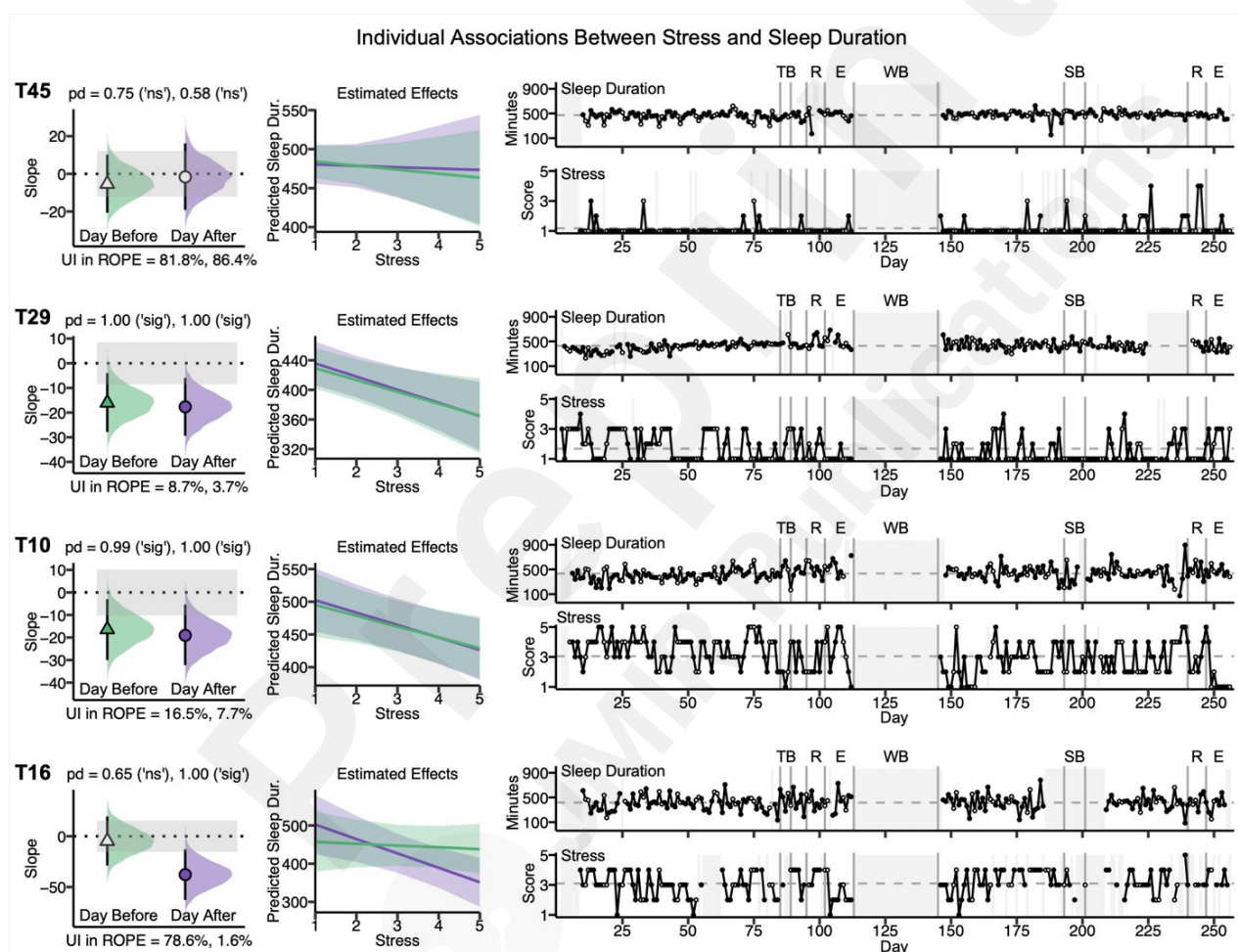
Examining Raw Within-individual Data Informs the Individual Linear Model Results

A closer look at individual participants' data reveals important considerations for the application and interpretation of the iLMs (**Figure 8**). As with other statistical models, the iLM is unlikely to detect stable associations for participants with too little variability in their data. For example, participant T45 had consistently low Stress levels over the course of the year, with occasional, small increases in Stress tied to periods with increased academic demands (e.g., ahead of midterm and final exams periods). This participant's individual linear models showed null results for both directions of stress-sleep associations ($pds < 0.975$). Meanwhile, participant T29 also had low baseline Stress levels, but presented more frequent and substantial rises in Stress throughout the year. Participant T29's iLMs showed bidirectional negative associations between Stress and Sleep Duration ($pds > 0.975$).

While variability in daily observations over time is necessary to detect linear associations within the iLM, variability alone is not sufficient. Participant T10 showed substantial fluctuations in Stress and Sleep Duration throughout the year, and their iLMs found bidirectional negative associations between Stress and Sleep Duration ($pds > 0.975$). Meanwhile, participant T16 also presented substantial variability in Stress and Sleep Duration, and while their iLM detected a significant association between Sleep Duration and Stress the Day After ($pd = 1.00$), no association was detected in the opposite direction ($pd = 0.65$).

Figure 8. Intensive within-individual longitudinal data reveal differences between individuals. Results for individual-level models testing the association between Sleep Duration and Stress the

Day Before (green) and between Sleep Duration and Stress the Day After (purple). The first column shows the models' estimated slopes (computed as the median of the posterior distribution) and uncertainty metrics. Symbol shading signifies statistically significant slopes. Error bars show 95% uncertainty intervals (UI). Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPE); pd = probability of direction. The second panel shows the models' predicted Sleep Duration as a function of Stress; shading around the lines indicates 95% UI. The time series on the right show daily observations of actigraphy-derived Sleep Duration and survey-based perceived Stress. Gray dashed lines show the participant's mean value across the study period. Vertical lines indicate landmark events in the academic calendar, labeled at the top. Gray shading indicates missing data during the school terms or Winter Break, which were excluded from the individual-level linear models. TB = Thanksgiving Break; R = Reading period; E = Exams period; WB = Winter Break; SB = Spring Break.



Discussion

Stress levels and sleep duration interact in an individual's daily life, but research has yielded mixed findings regarding the temporal directionality of their associations, with changes in stress preceding changes in sleep, changes in sleep preceding changes in stress, or both. Here, we leveraged a novel individual-level linear regression modeling framework (iLM) to obtain precision estimates of day-to-

day associations between self-reported stress levels and actigraphy-derived sleep duration, in a sample of first-year college students studied continuously for a full academic year. While most participants showed a negative association between daily stress levels and sleep duration, the temporal direction of the association varied, with all types of lead-lag association previously reported at the group-level present within distinct individuals in our sample.

In agreement with prior literature, our results within individuals confirm that stress levels and sleep duration are closely and inversely related in daily life [10-19]. Nearly four out of ten participants—each considered an independent test of the association—showed a statistically significant effect (in either temporal direction), revealing that day-to-day changes in stress levels or sleep duration can reliably predict one another in the real world in a substantial portion of the sample. The slope estimate for the association between stress and sleep duration was negative in 86% of the fitted models. Within the estimated slopes that reached statistical significance, all but one were negative. This suggests that, for most individuals, increased stress levels are associated with shorter rather than longer sleep duration in the surrounding days, consistent with findings showing that periods of heightened stress levels coincide with periods of reduced sleep [11–14]. Critically, we provide a framework to obtain individually tailored estimates of these day-to-day associations. The individual-level slope estimates ranged from 10 to 38 minutes in shorter sleep duration per 1-unit increase in a 5-point perceived stress scale, suggesting that a change in stress levels can be associated with substantial changes in sleep duration, especially when daily stress levels increased by several units. It should be noted that while negative associations between sleep duration and stress levels predominated in our sample, it is possible that some individuals in the wider population show positive associations but were not captured in our study.

Our precision approach further revealed that the temporal directionality of the association between stress and sleep duration varied from person to person, representing all patterns of results reported by prior, group-level studies. For some participants, heightened stress during the day associated with shorter sleep later that night, but not vice versa, in agreement with group results reported by [15] and [16]; for others, shorter sleep associated with heightened stress the next day, but not vice versa, in agreement with group results in [17]; and yet others showed both directions of association, in agreement with group results in [18] and [19]. Daily psychological and behavioral experiences such as perceived stress and sleep duration thus appear to interact in person-specific ways, rather than being uniform across the population.

For individuals showing the stress-then-reduced sleep phenotype, experiences of heightened stress (e.g., due to an impending final exam or social conflict) might elicit hyperarousal and

rumination [13, 51–53], as well as behaviors aimed at mitigating the source of stress (e.g., studying or socializing with friends late into the night), all of which can delay sleep and reduce its overall duration. Moreover, for those showing the reduced sleep-then-stress phenotype, shortened sleep durations might enhance their sensitivity to, and/or undermine their ability to cope with, academic, interpersonal, or other stressors and thus make them more likely to experience heightened stress levels [10, 54, 55]. For some individuals, both patterns of effect might occur, with changes in stress levels and sleep duration reinforcing one another and resulting in chains of days with heightened stress and nights of short sleep that succeed one another.

Critically, our results suggest that group-level studies might report inconsistent findings, at least partly, because the dynamic interaction between daily stress levels and sleep duration varies from person to person. When data are aggregated at the group level, individual phenotypes might be obscured, and the resulting group-level estimates are suggestive of generalized effects when in fact they might only apply to a fraction of the sample. Even when hierarchical group models allow for fitting individual estimates, these are biased by the group and tend to attenuate (or “shrink”) the estimation of individual effects [26]. Comparison of individual slope estimates of stress-sleep associations derived from our iLMs to those derived from a multilevel linear model (MLM) with random intercepts and slopes starkly demonstrated this group bias: the group MLM estimated individual slopes that were systematically attenuated, or even sign-reversed, compared to those estimated by our individual-level models.

The present results demonstrate the utility of individual-level modeling for characterizing real-world behavioral and psychological dynamics [27–29, 34]. Our approach leverages mobile and wearable technology and provides a fit-for-purpose methodology that can turn these devices’ large-scale longitudinal measurements into meaningful insights. The iLM’s model specification is parsimonious by design and easy to interpret, including a single linear term for the main predictor of interest and a day of week covariate to account for the weekly structure in the outcome variable. Diagnostic checks confirmed that this simple model specification was powerful enough to capture real-world, stable linear associations between psychological phenomena and objective behaviors, while accounting for the time-related dependencies in the daily observations. Individual tailoring is achieved by fitting only one person’s data within a model, but the model specification remains identical across individuals, allowing for direct between-person comparisons of results, including estimating the relative prevalence of different phenotypes in the group.

The iLM framework is readily applied to a single individual’s data, making it useful for multiple real-world purposes beyond fundamental research that are increasingly gaining interest in

the fields of consumer health informatics and digital health [21, 31, 32]. The iLM can be applied to data collected through personal devices for self-monitoring as well as for precision approaches in health care settings, where clinicians might use a patient's data to triage intervention plans. Taking the present results as an example, stress management interventions might be first prioritized among individuals for whom heightened stress precedes shortened sleep, while sleep interventions might be prioritized among individuals for whom shortened sleep precedes heightened stress. Moreover, the present work might inform a growing body of research and products combining multiple data streams from wearable and smartphones along with machine learning techniques to "predict" experiences of stress [56, 57]. Although prediction was not the focus of our work, understanding the person-specific association between stress and sleep duration, as well as the weekly behavioral patterns revealed by our actigraphy and survey data and individualized approach, could potentially contribute to the identification of periods when individuals are more likely to experience stress so that timely interventions can be offered.

The iLM offers a simple yet powerful precision framework for the estimation of real-world psychological and behavioral associations within the individual, but results should be interpreted carefully in light of its assumptions and limitations. First, the iLM's assumptions of linear, stable associations between the predictor and outcome variables are a deliberate attempt to simplify real-world behavioral and psychological dynamics that are highly complex. In the context of the present study, these features allowed us to estimate college students' day-to-day stress-sleep associations that are stable across students' fluctuating demands within the school semesters, as well as straightforward to interpret. However, it is possible that the association between stress and sleep duration is context-dependent, varying as a function of the specific source of stress experienced by the individual (e.g., academic versus interpersonal) and/or the broader seasonal demands (e.g., whether school is in session). In fact, given the possibility of the latter, we decided a priori to exclude data collected during winter break from our models. Future research could examine how these contextual demands influence stress-sleep associations, as well as explore non-linear associations or cumulative effects over time.

Our analyses leveraged leading and lagging patterns in each individual's time series of stress and sleep duration measures to ascertain the temporal directionality of their association (e.g., stress during the day as a predictor of subsequent sleep duration later that night), but we did not implement a controlled experimental manipulation and cannot establish a causal relationship. While our use of intensive longitudinal data collected in the real world grants our results ecological validity, it also exposes them to confounders. For example, it is possible that a significant association between short

sleep and higher stress the next day could be explained by the anticipated demands of the next day, such as an exam. Rather than short sleep duration *causing* higher stress the next day, the exam might be the primary cause behind both the short sleep (staying up late to prepare) and the stress reported the next day (heightened stress during test-taking).

While our current intensive longitudinal dataset and individual-level modeling framework passed sanity checks and diagnostics that confirmed data quality and adequate model specification, future application of our framework to other datasets and research questions should be careful to scrutinize the appropriateness of the data and the model before interpreting results. Mobile and wearable technology enable the collection of large behavioral datasets over time, but quantity does not guarantee quality, and long study periods require extra vigilance to ensure participants remain compliant over time. Quality checks should confirm that the data collected are capturing expected real-world behavior, e.g., as suggested by the intuitive changes we observed in students' behavior between school terms and breaks and between weekdays and weekends. Even if the available data are substantial and participant compliance is high, small variability in the metrics under study could still impede obtaining meaningful estimates of their associations, as we demonstrated in some of our participants with minimal fluctuations in stress levels over the course of the academic year. Moreover, investigators should be careful to identify appropriate uses, sanity checks, and interpretations of the iLM for their population of study. For example, part of our iLM validation process included testing the expected positive association between objective sleep duration (measured via the actigraphy wristband) and the participant's subjective rating for the same sleep event (reported via a daily smartphone survey). While we expect that generally healthy participants will tend to rate nights of shorter-than-usual sleep duration as lower quality, assuming a simple linear association between objective sleep duration and subjective sleep quality might not always be advisable, especially when modeling data from patients with sleep and/or psychiatric disorders.

Conclusions

Our novel iLM framework leveraged intensive longitudinal data from mobile and wearable devices to obtain individually tailored estimates of day-to-day associations between subjective stress levels and objective sleep duration. While stress and sleep duration were inversely related in most participants, the iLM revealed that the temporal direction of these associations is person-specific, identifying a variety of individual phenotypes that may account for the diverse group-level findings reported in prior literature. Our results demonstrate the utility of individual-level modeling approaches for the assessment of behavioral and psychological associations in the real world. An individualized approach offers a foundation for the characterization of life dynamics at both the

individual and the group levels, as well as for the development of precision health and wellbeing interventions tailored to the individual.

Acknowledgments

We thank Laura Farfel, Marisa Marotta, Erin Phlegar, Lauren DiNicola, Arpi Youssoufian for their help collecting data. Timothy O'Keefe, Harris Hoke, Lily Jeong provided valuable assistance in neuroinformatics support. Kenzie W. Carlson provided support with Beiwe. Katherine Miclau, Amira Song, Emily Iannazzi and Hannah Becker helped with the actigraphy quality control procedure. This work was supported by a generous gift from Kent and Liz Dauten, NIMH grants U01MH116925 and DP2MH103909, NIH grant T90DA022759, and the Sackler Scholar Program in Psychobiology.

Data Availability

The data sets generated during this study are not publicly available due to concerns related to participant identifiability but may be available from the corresponding author on reasonable request.

Authors' Contributions

G.C., J.P.O., J.T.B., R.L.B. designed the research. G.C., J.P.O., R.L.B. performed the research. C.M.V., G.C., H.R., R.L.B. analyzed the data. P.M. provided statistical support. C.M.V. and R.L.B. wrote the paper, and all other authors reviewed the final manuscript.

Competing Interests

J.P.O is a cofounder and board member of Phebe, a commercial entity that operates in digital phenotyping. J.T.B has received consulting fees from Verily Life Sciences as well as consulting fees and equity from Mindstrong Health Inc. for work unrelated to the present work. R.L.B. has received consulting fees from Pfizer, Roche, Alkermes, and Cognito for work unrelated to the present work. All other authors declare no competing interests.

Abbreviations

DPSleep: Deep phenotyping of sleep actigraphy data processing pipeline

ESS: Effective sample size

iLM: Individual-level linear model

MCMC: Monte Carlo Markov Chain

Pd: Probability of direction

ROPE: Region of practical equivalence

UI: Uncertainty interval

References

1. Dyson R, Renk K. Freshmen adaptation to university life: Depressive symptoms, stress, and coping. *J Clin Psychol*. 2006;62(10). doi:10.1002/jclp.20295
2. Ross S, Niebling B, Heckert T. Sources of Stress among College Students. *Coll Stud J*. 1999;33(2).
3. Taylor ZE, Doane LD, Eisenberg N. Transitioning from high school to college: Relations of social support, ego-resiliency, and maladjustment during emerging adulthood. *Emerg Adulthood*. 2014;2:105-115. doi:10.1177/2167696813506885
4. Bose M, Oliván B, Laferrère B. Stress and obesity: the role of the hypothalamic–pituitary–adrenal axis in metabolic disease. *Curr Opin Endocrinol Diabetes Obes*. 2009;16(5):340. doi:10.1097/MED.0b013e32832fa137
5. Kendler KS, Hettema JM, Butera F, Gardner CO, Prescott CA. Life Event Dimensions of Loss, Humiliation, Entrapment, and Danger in the Prediction of Onsets of Major Depression and Generalized Anxiety. *Arch Gen Psychiatry*. 2003;60(8):789. doi:10.1001/archpsyc.60.8.789
6. Luyster FS, Strollo PJ Jr, Zee PC, Walsh JK. Sleep: A Health Imperative. *Sleep*. 2012;35(6):727-734. doi:10.5665/sleep.1846
7. Pêgo JM, Sousa JC, Almeida O, Sousa N. Stress and the Neuroendocrinology of Anxiety Disorders. In: Stein MB, Steckler T, eds. *Behavioral Neurobiology of Anxiety and Its Treatment*. Current Topics in Behavioral Neurosciences. Springer; 2010:97-118. doi:10.1007/7854_2009_13
8. Rosengren A, Hawken S, Ôunpuu S, et al. Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *The Lancet*. 2004;364(9438):953-962. doi:10.1016/S0140-6736(04)17019-0
9. Zhang J, Paksarian D, Lamers F, Hickie IB, He J, Merikangas KR. Sleep Patterns and Mental Health Correlates in US Adolescents. *J Pediatr*. 2017;182:137-143. doi:10.1016/J.JPEDI.2016.11.007
10. Goldstein AN, Walker MP. The Role of Sleep in Emotional Brain Function. *Annu Rev Clin Psychol*. 2014;10(1):679-708. doi:10.1146/annurev-clinpsy-032813-153716
11. Kim EJ, Dimsdale JE. The Effect of Psychosocial Stress on Sleep: A Review of Polysomnographic Evidence. *Behav Sleep Med*. 2007;5(4):256-278. doi:10.1080/15402000701557383
12. Lund HG, Reider BD, Whiting AB, Prichard JR. Sleep patterns and predictors of disturbed sleep in a large population of college students. *J Adolesc Health Off Publ Soc Adolesc Med*. 2010;46(2):124-132. doi:10.1016/j.jadohealth.2009.06.016
13. Sadeh A, Gruber R. Stress and Sleep in Adolescence: A Clinical-Developmental Perspective. In: Carskadon MA, ed. *Adolescent Sleep Patterns: Biological, Social, and Psychological Influences*. Cambridge University Press; 2002:236-253. doi:10.1017/CBO9780511499999.017
14. Van Reeth O, Weibel L, Spiegel K, Leproult R, Dugovic C, Maccari S. Interactions between stress and sleep: from basic research to clinical situations. *Sleep Med Rev*. 2000;4(2):201-219. doi:10.1053/smr.1999.0097
15. Marcusson-Clavertz D, Sliwinski MJ, Buxton OM, Kim J, Almeida DM, Smyth JM. Relationships between daily stress responses in everyday life and nightly sleep. *J Behav Med*. 2022;45(4):518-532. doi:10.1007/s10865-021-00281-3

16. Slavish DC, Asbee J, Veeramachaneni K, et al. The Cycle of Daily Stress and Sleep: Sleep Measurement Matters. *Ann Behav Med*. 2021;55(5):413-423. doi:10.1093/abm/kaaa053
17. Sin NL, Almeida DM, Crain TL, Kossek EE, Berkman LF, Buxton OM. Bidirectional, Temporal Associations of Sleep with Positive Events, Affect, and Stressors in Daily Life Across a Week. *Ann Behav Med*. 2017;51(3):402-415. doi:10.1007/s12160-016-9864-y
18. Doane LD, Thurston EC. Associations among sleep, daily experiences, and loneliness in adolescence: Evidence of moderating and bidirectional pathways. *J Adolesc*. 2014;37(2):145-154. doi:10.1016/j.adolescence.2013.11.009
19. Yap Y, Slavish DC, Taylor DJ, Bei B, Wiley JF. Bi-directional relations between stress and self-reported and actigraphy-assessed sleep: a daily intensive longitudinal study. *Sleep*. 2020;43(3):zsz250. doi:10.1093/sleep/zsz250
20. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspect Psychol Sci*. 2016;11(6). doi:10.1177/1745691616650285
21. Onnela JP, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*. 2016;41(7):1691-1696. doi:10.1038/npp.2016.7
22. Torous J, Kiang MV, Lorme J, Onnela JP. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment Health*. 2016;3(2):e16. doi:10.2196/mental.5165
23. Sano A, Taylor S, McHill AW, et al. Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study. *J Med Internet Res*. 2018;20(6):e210. doi:10.2196/jmir.9410
24. Vidal Bustamante CM, Coombs G, Rahimi-Eichi H, et al. Fluctuations in behavior and affect in college students measured using deep phenotyping. *Sci Rep*. 2022;12(1):1932. doi:10.1038/s41598-022-05331-7
25. Wang R, Chen F, Chen Z, et al. Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In: *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM Press; 2014:3-14. doi:10.1145/2632048.2632054
26. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications; 2002. Accessed August 20, 2019. <http://us.sagepub.com/en-us/nam/hierarchical-linear-models/book9230>
27. Barlow DH, Nock M, Hersen M. *Single Case Experimental Designs: Strategies for Studying Behavior for Change*. Pearson/Allyn and Bacon; 2009.
28. Haynes SN, Mumma GH, Pinson C. Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clin Psychol Rev*. 2009;29(2):179-191. doi:10.1016/j.cpr.2008.12.003
29. Molenaar PCM. A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Meas Interdiscip Res Perspect*. 2004;2(4):201-218. doi:10.1207/s15366359mea0204_1
30. Piccirillo ML, Rodebaugh TL. Foundations of idiographic methods in psychology and applications for psychotherapy. *Clin Psychol Rev*. 2019;71:90-100.

doi:10.1016/j.cpr.2019.01.002

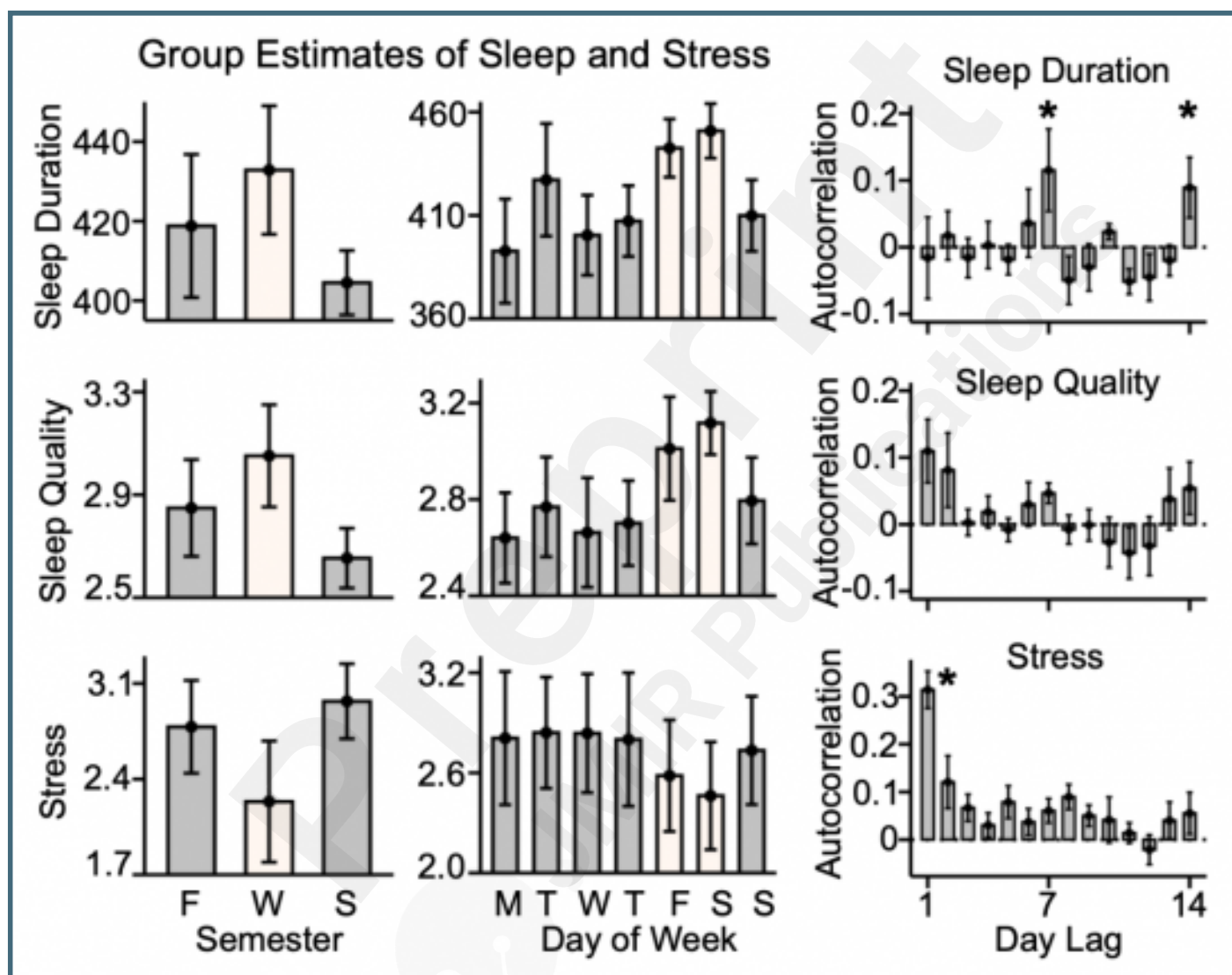
31. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Pers Med*. 2018;15(5):429-448. doi:10.2217/pme-2018-0044
32. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M. The New Field of 'Precision Psychiatry.' *BMC Med*. 2017;15(1):80. doi:10.1186/s12916-017-0849-x
33. Scala JJ, Ganz AB, Snyder MP. Precision Medicine Approaches to Mental Health Care. *Physiology*. 2023;38(2):82-98. doi:10.1152/physiol.00013.2022
34. Wright AGC, Woods WC. Personalized Models of Psychopathology. *Annu Rev Clin Psychol*. 2020;16(1):49-74. doi:10.1146/annurev-clinpsy-102419-125032
35. Onnela JP, Dixon C, Griffin K, et al. Beiwe: A data collection platform for high-throughput digital phenotyping. *J Open Source Softw*. 2021;6(68):3417. doi:10.21105/joss.03417
36. Rahimi-Eichi H, Coombs 3rd G, Vidal Bustamante CM, Onnela JP, Baker JT, Buckner RL. Open-source Longitudinal Sleep Analysis From Accelerometer Data (DPSleep): Algorithm Development and Validation. *JMIR MHealth UHealth*. 2021;9(10). doi:https://doi.org/10.2196/29849
37. Pickup M. *Introduction to Time Series Analysis*. SAGE Publications, Inc.; 2015. doi:10.4135/9781483390857
38. Muth C, Oravecz Z, Gabry J. User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *Quant Methods Psychol*. 2018;14(2):99-119. doi:10.20982/tqmp.14.2.p099
39. McElreath R. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. CRC Press; 2020.
40. Kruschke JK, Liddell TM. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev*. 2018;25(1):178-206. doi:10.3758/s13423-016-1221-4
41. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2023. <https://www.R-project.org/>
42. Stan Development Team. The Stan Core Library. Published online 2018. <http://mc-stan.org>
43. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan. Published online 2023. <https://mc-stan.org/rstanarm/>
44. Kay M. tidybayes: Tidy Data and Geoms for Bayesian Models. Published online 2021. <http://mjskay.github.io/tidybayes/>
45. Makowski D, Ben-Shachar MS, Lüdtke D. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *J Open Source Softw*. 2019;4(40):1541. doi:10.21105/joss.01541
46. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457-472. doi:10.1214/ss/1177011136
47. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis*. 3rd ed. Chapman and Hall/CRC; 2013. doi:10.1201/b16018
48. Makowski D, Ben-Shachar MS, Chen SHA, Lüdtke D. Indices of Effect Existence and Significance in the Bayesian Framework. *Front Psychol*. 2019;10. Accessed July 11, 2023. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02767>
49. Kruschke JK. Rejecting or Accepting Parameter Values in Bayesian Estimation. *Adv Methods*

- Pract Psychol Sci.* 2018;1(2):270-280. doi:10.1177/2515245918771304
50. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Second. Erlbaum; 1988.
51. Drake CL, Pillai V, Roth T. Stress and Sleep Reactivity: A Prospective Investigation of the Stress-Diathesis Model of Insomnia. *Sleep*. 2014;37(8):1295-1304. doi:10.5665/sleep.3916
52. Harvey AG, Tang NKY, Browning L. Cognitive approaches to insomnia. *Clin Psychol Rev.* 2005;25(5):593-611. doi:10.1016/j.cpr.2005.04.005
53. Kalmbach DA, Cuamatzi-Castelan AS, Tonnu CV, et al. Hyperarousal and sleep reactivity in insomnia: current insights. *Nat Sci Sleep*. 2018;10:193-201. doi:10.2147/NSS.S138823
54. Minkel JD, Banks S, Htaik O, et al. Sleep deprivation and stressors: Evidence for elevated negative affect in response to mild stressors when sleep deprived. *Emotion*. 2012;12(5):1015-1020. doi:10.1037/a0026871
55. Zohar D, Tzischinsky O, Epstein R, Lavie P. The Effects of Sleep Loss on Medical Residents' Emotional Reactions to Work Events: a Cognitive-Energy Model. *Sleep*. 2005;28(1):47-54. doi:10.1093/sleep/28.1.47
56. Vos G, Trinh K, Sarnyai Z, Rahimi Azghadi M. Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review. *Int J Med Inf.* 2023;173:105026. doi:10.1016/j.ijmedinf.2023.105026
57. Razavi M, Ziyadidegan S, Jahromi R, et al. Machine Learning, Deep Learning and Data Preprocessing Techniques for Detection, Prediction, and Monitoring of Stress and Stress-related Mental Disorders: A Scoping Review. *arXiv*. Published online August 8, 2023. doi:10.48550/arXiv.2308.04616

Supplementary Files

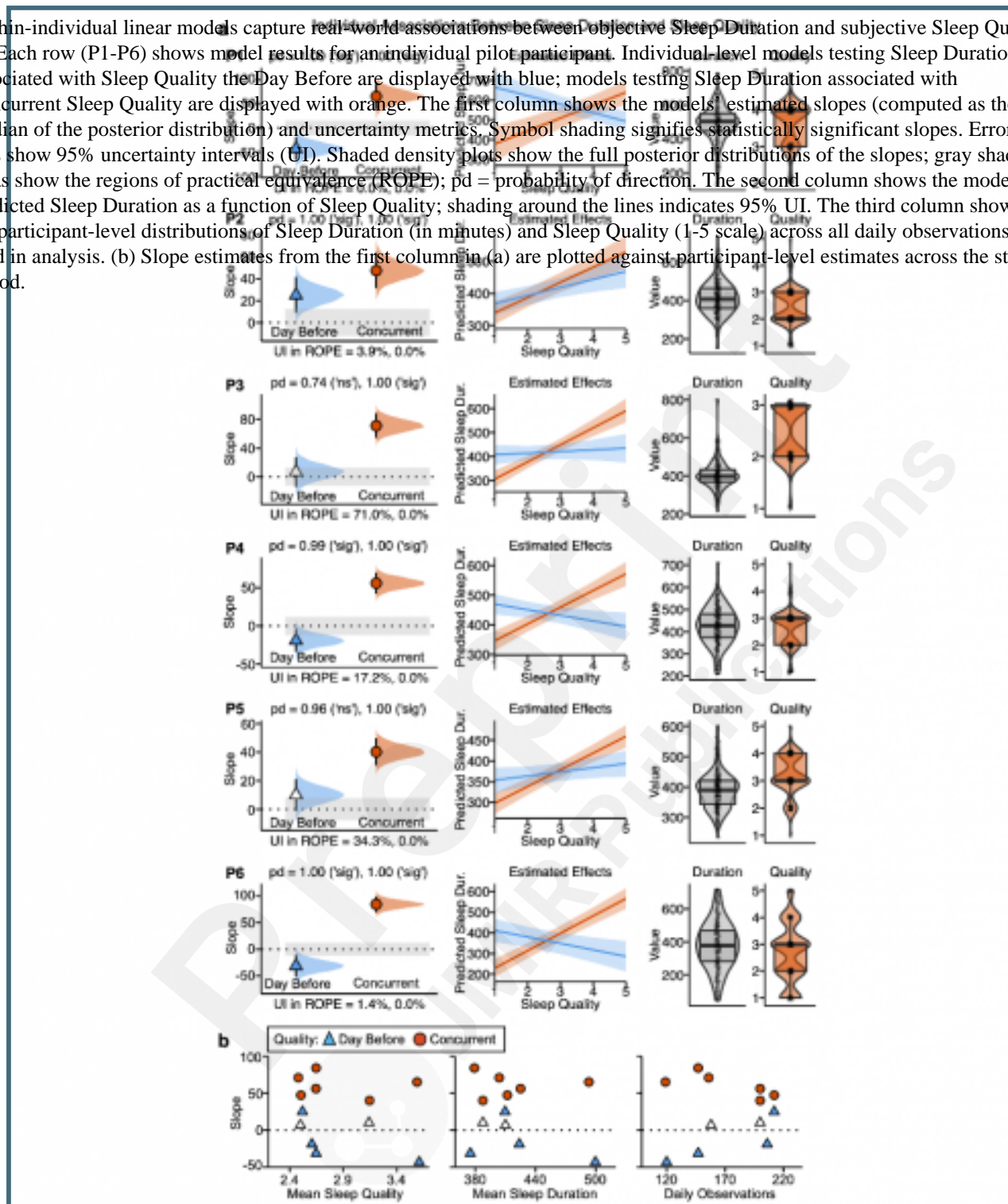
Figures

Group Sleep and Stress metrics reveal structure related to the academic calendar. Pilot participants' group (between-person) means of Sleep Duration (in minutes, top row), Sleep Quality (5-point Likert scale, middle row), and perceived Stress (5-point Likert scale, bottom row), aggregated by school semester (Fall and Spring) and Winter Break (first column) and by day of the week (second column). The third column shows between-person means of autocorrelation estimates over a 14-day window. Asterisks highlight that autocorrelations were strongest at a 7-day lag (and again at 14-day lag) for Sleep Duration and at 1-day and 2-day lags for perceived Stress. Error bars show standard error of the mean. F = Fall semester, W = Winter break, S = Spring semester; M = Monday, T = Tuesday, W = Wednesday, T = Thursday, F = Friday, S = Saturday, S = Sunday.



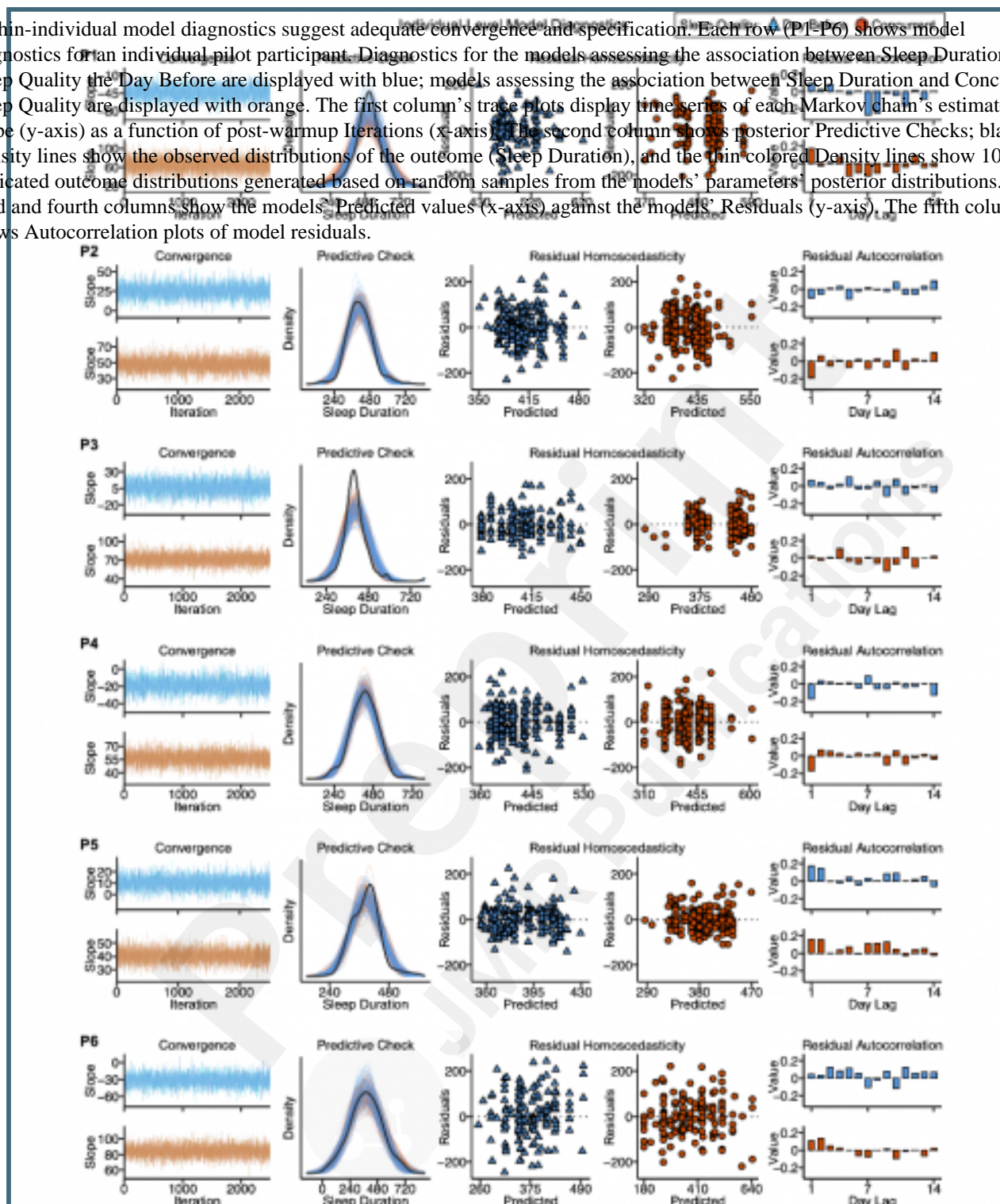


Within-individual linear models capture real-world associations between Objective Sleep Duration and subjective Sleep Quality. (a) Each row (P1-P6) shows model results for an individual pilot participant. Individual-level models testing Sleep Duration associated with Sleep Quality the Day Before are displayed with blue; models testing Sleep Duration associated with Concurrent Sleep Quality are displayed with orange. The first column shows the models' estimated slopes (computed as the median of the posterior distribution) and uncertainty metrics. Symbol shading signifies statistically significant slopes. Error bars show 95% uncertainty intervals (UI). Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPE); pd = probability of direction. The second column shows the models' predicted Sleep Duration as a function of Sleep Quality; shading around the lines indicates 95% UI. The third column shows the participant-level distributions of Sleep Duration (in minutes) and Sleep Quality (1-5 scale) across all daily observations used in analysis. (b) Slope estimates from the first column in (a) are plotted against participant-level estimates across the study period.

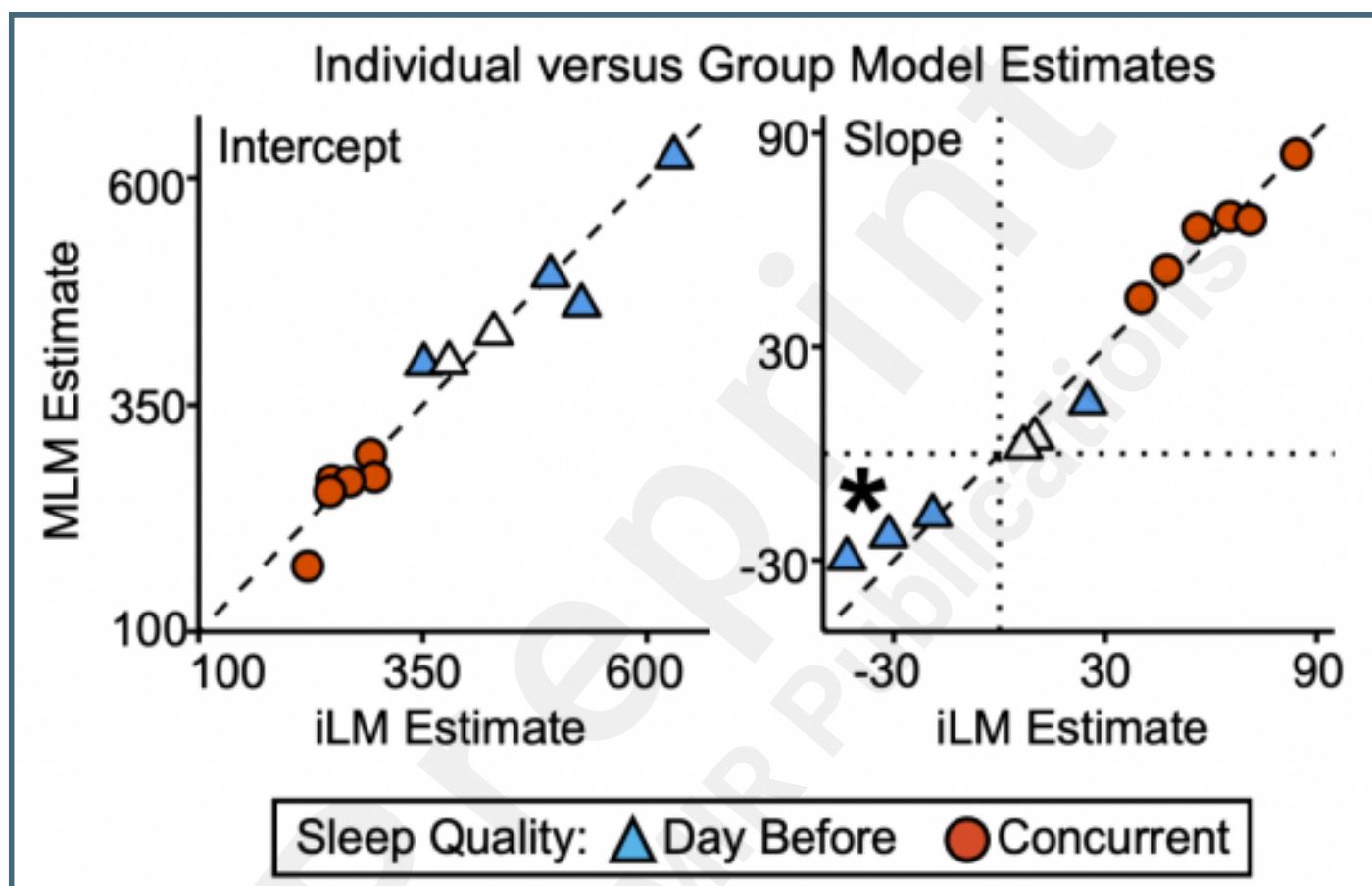




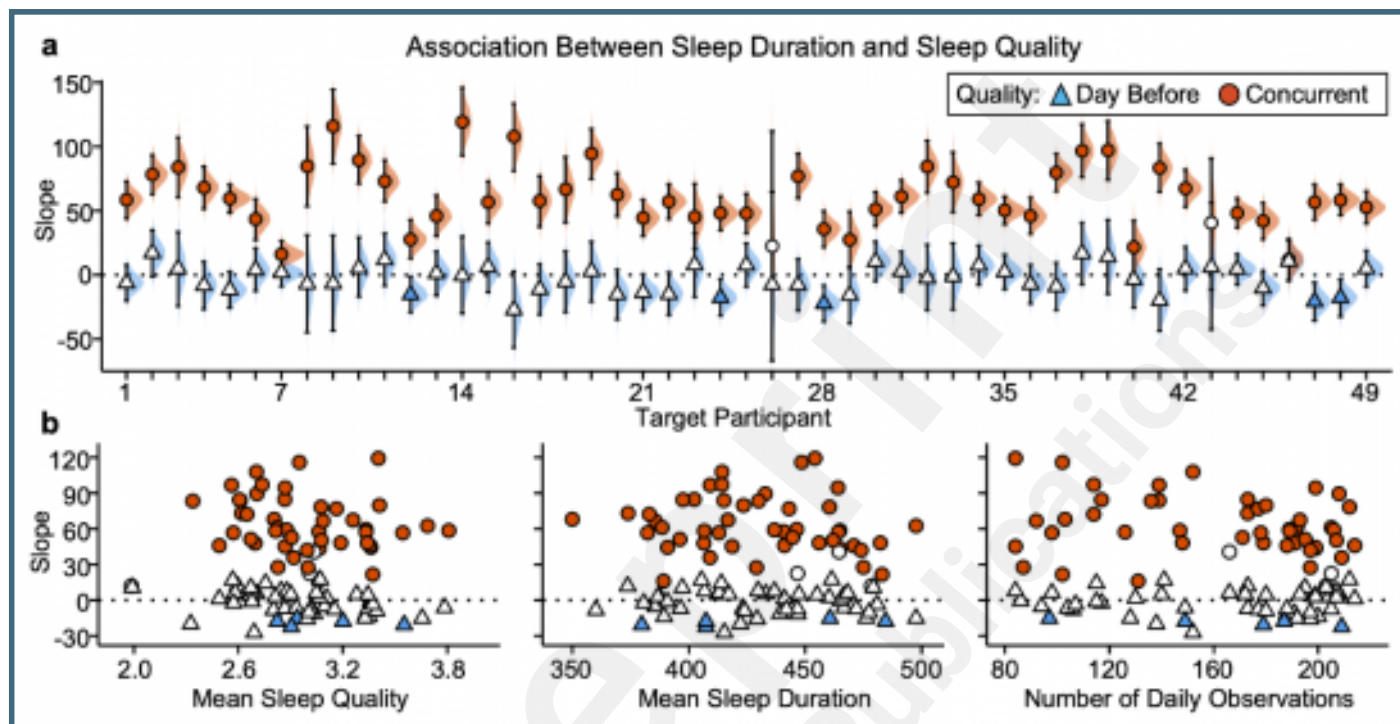
Within-individual model diagnostics suggest adequate convergence and specification. Each row (P1-P6) shows model diagnostics for an individual pilot participant. Diagnostics for the models assessing the association between Sleep Duration and Sleep Quality the Day Before are displayed with blue; models assessing the association between Sleep Duration and Concurrent Sleep Quality are displayed with orange. The first column's trace plots display time series of each Markov chain's estimated Slope (y-axis) as a function of post-warmup Iterations (x-axis). The second column shows posterior Predictive Checks; black Density lines show the observed distributions of the outcome (Sleep Duration), and the thin colored Density lines show 100 replicated distributions generated based on random samples from the models' parameters' posterior distributions. The third and fourth columns show the models' Predicted values (x-axis) against the models' Residuals (y-axis). The fifth column shows Autocorrelation plots of model residuals.



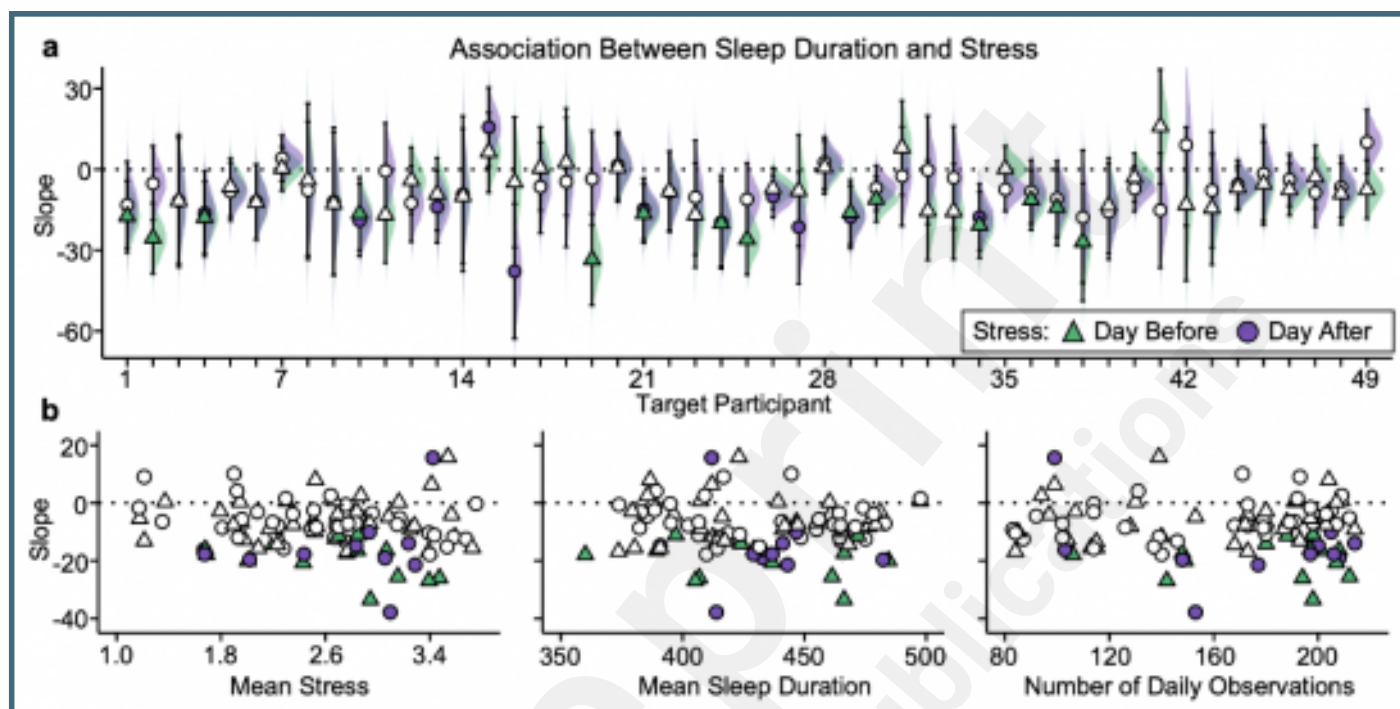
Within-individual linear models of the association between Sleep Duration and Sleep Quality yield similar, but not identical, estimates to a group-based linear model. Estimates of the intercepts and slopes from the individual-level linear models (iLM; x-axis) for each participant compared to the random effects estimated from a group-based multilevel linear model (MLM; y-axis). Triangles show estimates for the models examining the association between Sleep Duration and Sleep Quality the Day Before; circles show estimates for the models examining the association between Sleep Duration and Concurrent Sleep Quality. Symbol shading signifies statistically significant slopes in the iLMs. Dashed diagonal lines represent identical estimates between the individual- and group-based approaches. Asterisk highlights attenuation of individual estimates in the MLM relative to the iLMs.



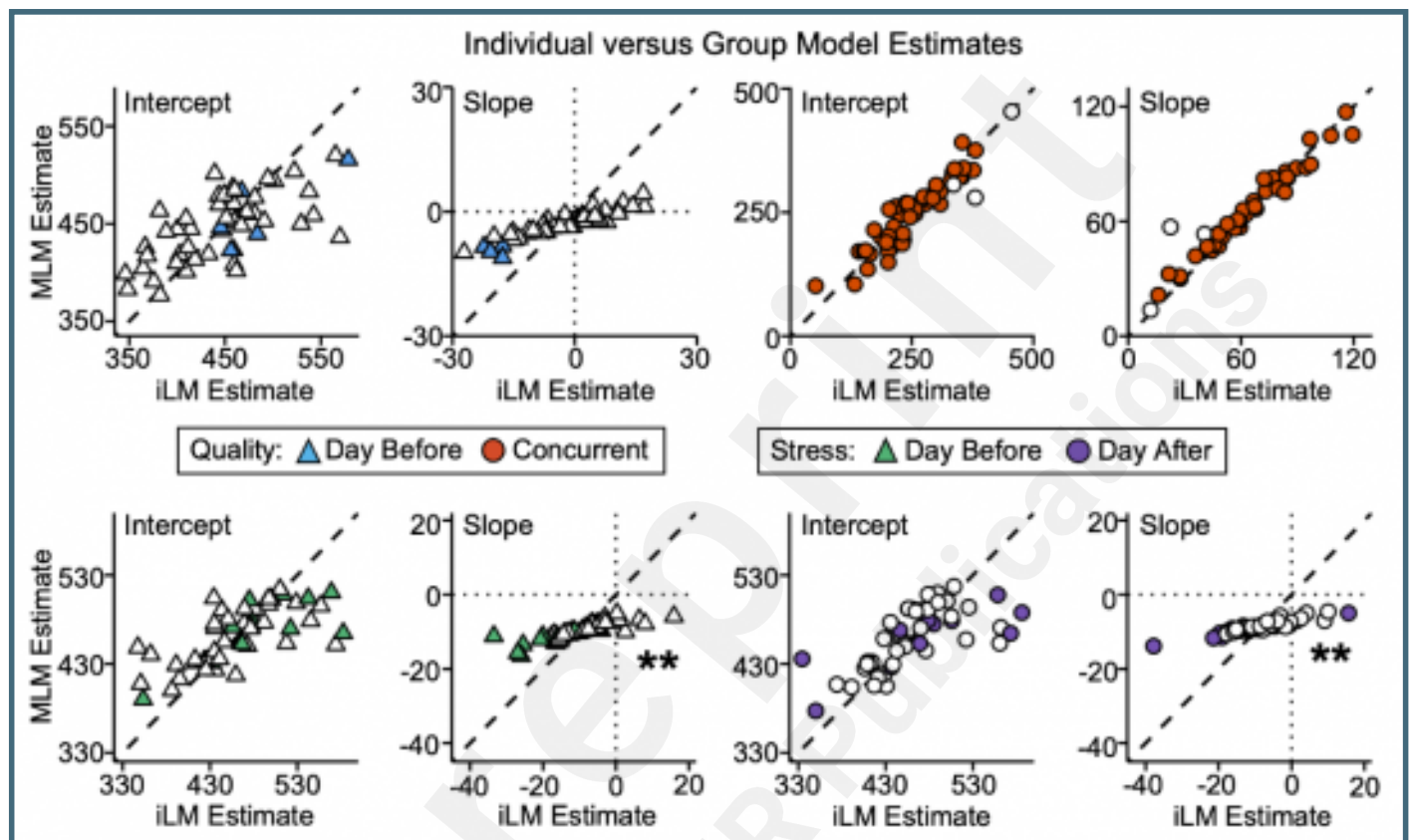
Longer objective Sleep Duration associates with Higher subjective Sleep Quality in most individuals. (a) Estimated Slopes (corresponding to the median of the posterior distribution) from the individual-level models assessing Sleep Duration associated with Sleep Quality the Day Before (triangles) and with Concurrent Sleep Quality (circles) are plotted along the y-axis ordered by participant (x-axis). Statistically significant slope estimates are shaded blue (Day Before) or orange (Concurrent). Error bars show 95% uncertainty intervals and shaded density plots show the full posterior distributions of the slopes. (b) Slope estimates from panel (a) are plotted against participant-level estimates across the study period.



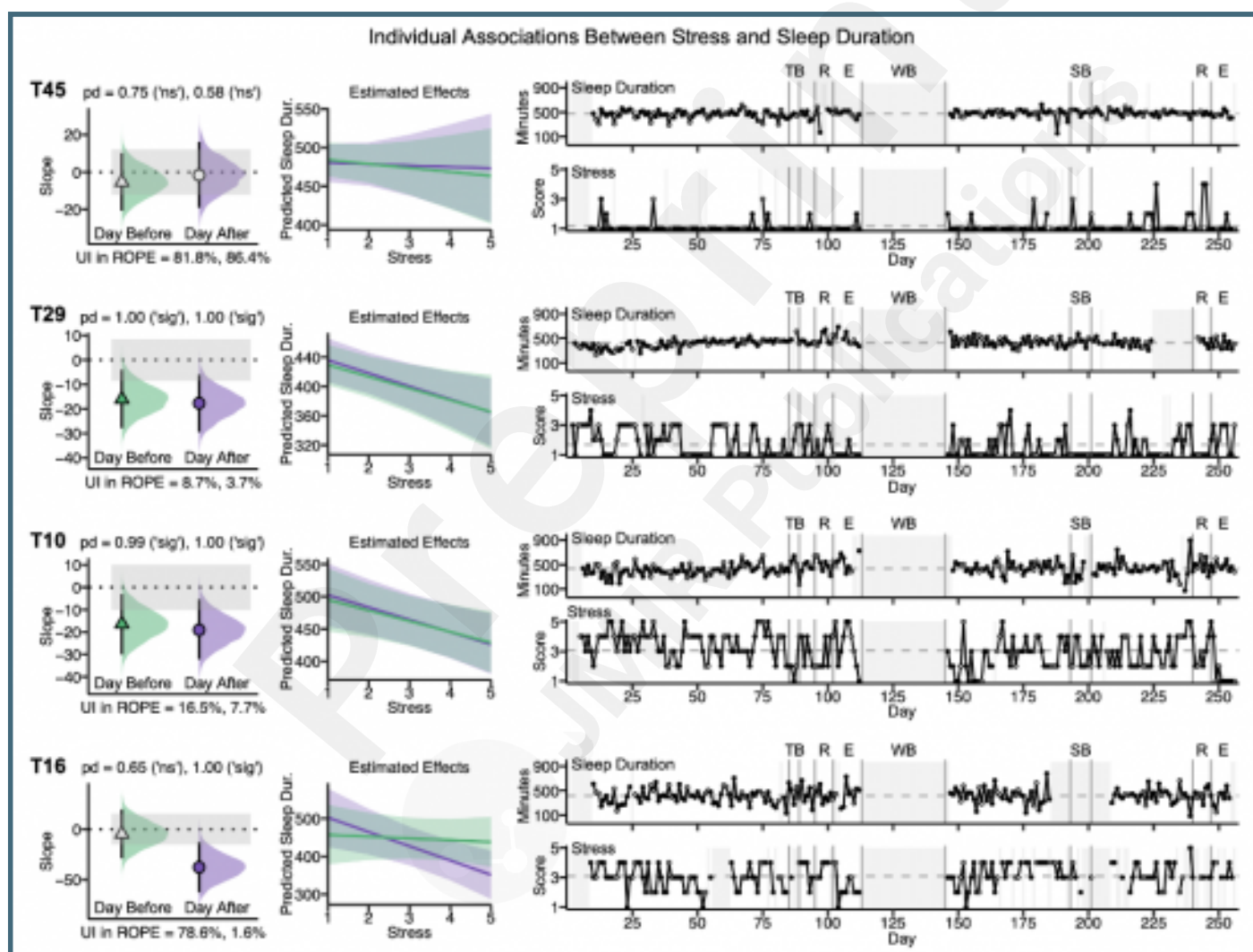
Higher subjective Stress associates with shorter objective Sleep Duration in most individuals. (a) Estimated Slopes (corresponding to the median of the posterior distribution) from the individual-level models assessing Sleep Duration associated with Stress the Day Before (triangles) and with Stress the Day After (circles) are plotted along the y-axis ordered by participant (x-axis). Statistically significant slope estimates are shaded green (Day Before) or purple (Concurrent). Error bars show 95% uncertainty intervals and shaded density plots show the full posterior distributions of the slopes. (b) Slope estimates from panel (a) are plotted against each participant-level estimates across the study period.



Within-individual estimates of the association between Stress and Sleep Duration show systematic attenuation in the group-based linear model. Estimates of the intercepts and slopes from the individual-level linear models (iLM; x-axis) for each participant compared to the random effects estimated from a group-based multilevel linear model (MLM; y-axis), presented separately for associations between Sleep Duration and Sleep Quality (top row), and associations between Stress and Sleep Duration (bottom row). Symbol shading signifies statistically significant slopes in the iLMs. Dashed diagonal lines represent identical estimates between the individual- and group-based approaches. Asterisks highlight attenuation of individual estimates in the MLM relative to the iLMs.



Intensive within-individual longitudinal data reveal differences between individuals. Results for individual-level models testing the association between Sleep Duration and Stress the Day Before (green) and between Sleep Duration and Stress the Day After (purple). The first column shows the models' estimated slopes (computed as the median of the posterior distribution) and uncertainty metrics. Symbol shading signifies statistically significant slopes. Error bars show 95% uncertainty intervals (UI). Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPE); pd = probability of direction. The second panel shows the models' predicted Sleep Duration as a function of Stress; shading around the lines indicates 95% UI. The time series on the right show daily observations of actigraphy-derived Sleep Duration and survey-based perceived Stress. Gray dashed lines show the participant's mean value across the study period. Vertical lines indicate landmark events in the academic calendar, labeled at the top. Gray shading indicates missing data during the school terms or Winter Break, which were excluded from the individual-level linear models. TB = Thanksgiving Break; R = Reading period; E = Exams period; WB = Winter Break; SB = Spring Break.



Multimedia Appendixes

Supplementary Figures.

URL: <http://asset.jmir.pub/assets/c2c70ed36c362334f858f0194da4c44f.docx>



Related publication(s) - for reviewers eyes onlies

Pre-existing data were utilized to develop and test the viability of our current manuscript's model approach (from this 2021 paper published on JMIR mHealth and uHealth) before applying them to the present target cohort. We cited this publication in the current manuscript.

URL: <http://asset.jmir.pub/assets/410ae77aaf9f87733bba06e212760fdd.pdf>

This 2022 paper published in Scientific Reports used the same target dataset as in the current study but for completely different analyses focused on between-subject and group effects, not within-individual modelling as explored here. We cited this paper in the current manuscript.

URL: <http://asset.jmir.pub/assets/5d9e8ea660664c17d487fcedc9605a0c.pdf>