Computers in Industry 92 (2017) 32–42

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Using consumer-grade wearables and novel measures of sleep and activity to analyze changes in behavioral health during an 8-month simulated Mars mission

Jocelyn Dunn, Ph.D.[a,*], Erich Huebner[b], Siyu Liu[a], Steve Landry, Ph.D.[a], Kim Binsted, Ph.D.[c]

[a] School of Industrial Engineering, Purdue University, 315 N Grant St., West Lafayette, IN 47907, United States
[b] Information Technology at Purdue University, 155 S Grant St., West Lafayette, IN 47907, United States
[c] Information and Computer Sciences, University of Hawaii, 2500 Campus Rd., Honolulu, HI 96822, United States

ARTICLE INFO

Article history:
Received 2 December 2016
Received in revised form 25 March 2017
Accepted 9 June 2017
Available online xxx

Keywords:
Wearables
Actigraphy
Sleep
Activity
Behavioral health
Data analytics
Health monitoring
Stress
Preventative care
Astronauts
Mars
Space exploration

ABSTRACT

For analyzing changes in an individual's health over time, this research has developed objective measures for comparing behavioral patterns, including sleep quality and activity scores. These novel measures of behavioral health have provided insight about how sleep debt accumulates after long, extended days, how sleep disruption and recovery from wakefulness occur during the night, and when cross-correlations exist between measures. This data-driven approach to quantifying behavioral patterns is informed by minute-by-minute data from consumer-grade, wrist-worn wearables. In this 8-month longitudinal study, Jawbone UP wristbands and the Jawbone UP API were utilized to collect minute-by-minute data about the behavior of crewmembers participating in a simulated Mars mission. To study the challenges of living and working on the planet Mars, for eight months, these crewmembers were confined to a Mars-like habitat, living in close quarters, isolated from the rest of humanity at a high elevation on Mauna Loa volcano in Hawaii, wearing mock spacesuits while exploring the volcanic terrain, consuming shelf-stable foods, restricted in water usage, relying on solar energy, and delayed in communications with 20-min lag-times for delivering messages to and from the crew. Analyzing the behavior of these astronaut-like individuals has led to the development of objective measures for quantifying sleep patterns, that have potential for contributing to the development of next-generation, smart wearables.

1. Introduction

Toward the goal of understanding behavioral health and performance challenges that astronauts will face on the first human mission to Mars, researchers have designed Mars analog environments for immersing crews in Mars-like conditions to measure biological, psychological, and social changes over time. During an 8-month simulated Mars mission conducted by Hawaii Space Exploration Analog and Simulation (HI-SEAS), six crew-members were isolated from the rest of humanity, at an elevation of 8200-feet on the red, rocky slopes of Mauna Loa volcano in Hawaii (Appendix A). Crewmembers were confined to a living space of 1000 square-feet in a geodesic dome habitat, except when going outside the habitat to explore the volcanic terrain while wearing mock spacesuits. HI-SEAS crewmembers had to cope with group-living in close quarters, only consumed shelf-stable foods, were restricted in water usage, relied on solar energy, and had 20-min lag-times on e-mail systems to simulate Earth to Mars communication.

Conducting wearables research in a Mars analog research environment is an intermediate step between laboratory work and real-world data collection. In contrast to monitoring directed behaviors for short periods of time in a controlled laboratory setting, crewmembers are behaving autonomously without any direction from researchers on how to schedule their sleep and activity periods, which results in complex data that is similar to real world behavior. This research environment has limited modes of entertainment, nutrition, social interaction, and work projects for the crew, as compared to the innumerable factors on health and behavior in the real world. During the 8-month mission, wrist-worn wearables were implemented for recording sleep, activity, and heart rate data on a minute-by-minute basis., These data have

* Corresponding author.
E-mail address: dunn11@purdue.edu (J. Dunn).

http://dx.doi.org/10.1016/j.compind.2017.06.003
0166-3615/

potential for improving self-awareness and crew performance by promoting behavioral health changes, such as increased activity and more consistent sleep schedules.

Minute-by-minute wearable device data collection about the behavior of these “astronaut-like” crewmembers is a first step in the development of unobtrusive behavioral health and performance monitoring systems for future astronauts. As currently planned, a mission to Mars will be a 3-year journey for a small crew of 4-6 astronauts who will need to be resilient to the stressors of isolated and confined living conditions in a highly complex work environment [1]. Remote health monitoring has useful applications for astronauts on a mission to Mars, but also for many other health and performance applications, including at-home patient care, athletic training, and military deployments. Wearables, mobile applications, and telehealth systems enable examinations of health to extend beyond intermittent visits with coaches and medical professionals. Wearables data can be transferred to healthcare providers or other leaders and coaches as an automated, reliable method for gaining information about behavioral patterns that reflect lifestyle choices and health status, and wearable users can receive interpretations of these data as well as reminders and incentives for continuing treatment plans or practice regimens to promote behavioral health changes and improved performance. Other uses of behavioral monitoring with wearables may include health education, parenting and care-taking in families, as well as personal challenges to incentivize sleep hygiene and consistent exercise, including competitions among coworkers, friends, and family.

By recording sleep and activity data, wearables are collecting volumes of information about users. However, in the research community, there is uncertainty about the accuracy of consumer-grade, wrist-worn wearables for activity, sleep, and heart rate measurement. Consumer-grade devices have not reached a level of acceptability to replace gold-standard research methods, such as polysomnography; however, collecting data about a large population offers “big data” value, as it is estimated that over 60 million people in the United States are using consumer-grade wearables, such as Jawbone and Fitbit wristbands [2]. Although consumer-grade, wrist-worn wearables data are less accurate than clinical research tools for sleep analysis, this research is leveraging the high volume of data collected from these devices to analyze changes over time [3].

Here, concepts of sleep debt, sleep quality, activity patterns, and relationships between these measures are defined and quantified. These measures have shown promising potential for answering both practical questions of wearable device users, such as how much sleep is needed for a given individual, when is sleep debt accumulating over time, how do activity levels impact sleep quality, and open research questions about how to define what behavioral guidelines are most applicable for a given individual, how to measure when lifestyle choices are impacting biological health, and how does allostatic load accumulate over time [4,5].

2. Background

For comparing sleep performance, both the quantity and quality of sleep should be objectively measured. Astronauts on the ISS sleep on average about 6 hours per night which is less than is typical on Earth, and astronauts have been shown to have circadian misalignment while in space that is correlated to lower perceived sleep quality [6,7]. Sleep quality reporting has been shown to be more related to sleepiness than sleep duration [8]. Rather than relying on subjective assessments of perceived sleep quality, this research aims to develop objective sleep quality measures from wearables data for analyzing patterns of sleep and wakefulness during the night.

With Jawbone devices, sleep phases are defined as: deep sleep (D), light sleep (L), and awakening (A). In addition, Jawbone has a proprietary algorithm for computing a quality score for each sleep event. With the proprietary Jawbone quality score, longer duration sleeps tend to receive high quality scores, and short duration sleeps, even if they are entirely deep sleep, still receive low quality scores. Intuitively, it seems that a sleep session that has primarily deep sleep and is without awakenings should be scored as high quality; furthermore, research has shown that perceived sleep quality is not correlated with sleep duration [8]. However, as shown in Fig. 1, the Jawbone “quality” score is 93% correlated with sleep “quantity,” or sleep duration. There is a need for sleep quality measures that differentiate various patterns of sleep and are not based solely on quantity.

The sleep quality score presented here is a weighted sum of sleep phase durations with a novel method of assigning penalties for awakenings during the night (Table 1 and Fig. 2). Sleep measures in the literature include sleep onset, wake after sleep

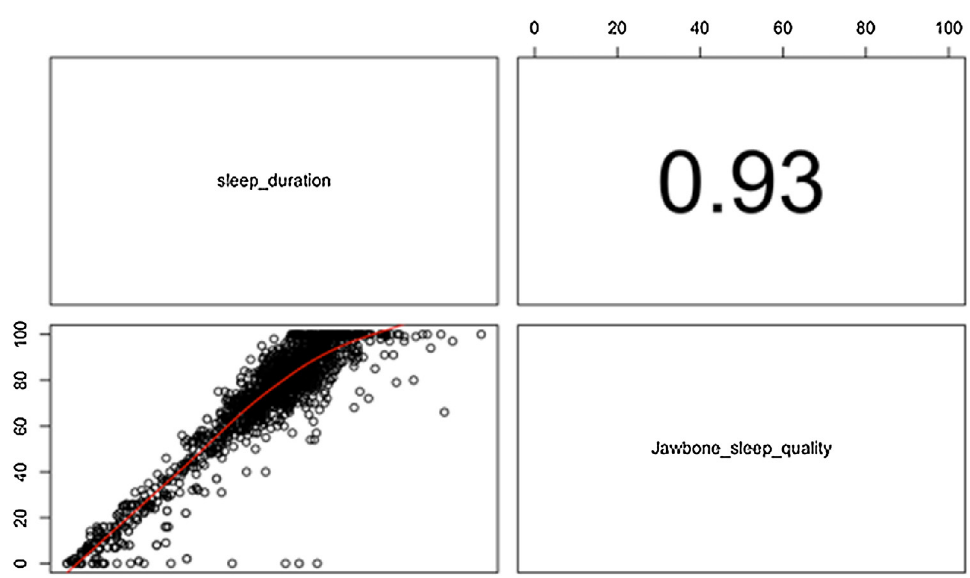

Fig. 1. Jawbone sleep quality score is highly correlated with sleep duration.

Table 1
Weights W_p for sleep phase types $P \in [D, L, A]$..

Sleep Phase Types (P):	D for Deep sleep	L for Light sleep	A for Awakening
Weights (W_P):	$W_D = 2$	$W_L = 1$	$W_A \in [-1,0]$

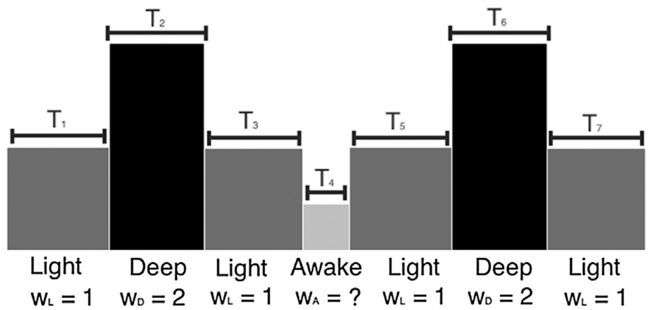

Fig. 2. Phases of sleep are weighted differently to reward for deep sleep and light sleep and to penalize for awakenings. Deep sleep phases are assigned a weight of 2, light sleep phases are assigned a weight of 1, and awakenings are assigned from [-1,0] by a penalty function.

onset, fragmentation or movement index, and sleep efficiency [9]. In brief, these measures quantify the amount of awake time before falling asleep and the amount of awake time versus the amount of sleep time during the sleep event. For example, sleep efficiency is the percentage of the time spent in bed in which the individual was actually sleeping. This quantitative measure does not consider how awakenings are impacting subsequent phases of sleep. For example, being awake for six 10-min intervals during the night will be scored the same sleep efficiency as being awake for one 60-min interval; these are two qualitatively different sleep experiences that the novel measure presented here is able to differentiate.

Similarly, the movement index is a count on the number of minutes in which one or more movements were recorded (i.e. the duration, or quantity of awakenings, not the pattern of awakenings or their impact on sleep quality). The fragmentation index (FI) is computed by counting the number of very short time periods of sleep, such as periods lasting less than 1-min, and dividing this count by the total number of sleep periods of any duration. FI computes the proportion of sleep that was disturbed by brief awakenings, but it does not measure if various awakenings are more or less disruptive.

3. Methods

The participants in this study were living and working together in an 8-month simulated Mars mission called HI-SEAS (Hawaii Space Exploration Analog and Simulation). HI-SEAS is a University of Hawaii NASA-funded study characterizing the behavioral health and performance (BHP) risks for long-duration human spaceflight missions. This research is being conducted at HI-SEAS in accordance with University of Hawaii and NASA Institutional Review Board approved protocols for human subject research. This 8-month HI-SEAS mission III began on October 15, 2014 and ended June 12, 2015. In brief, HI-SEAS immerses 6-person “astronaut-like” crews in a Mars-like habitat on Mauna Loa volcano, isolated from the rest of humanity for a long duration, confined to a 1000-square-foot living space, except for spacewalks wearing mock spacesuits, while living on limited energy, water, and food resources, and working with a 20-min lag-time on all communications to and from the habitat. For statistical analysis of these data, we used mixed effects modeling and analysis of variance (ANOVA) where participants are a “random” factor and day of week or month of mission are fixed factors. Tukey post-hoc tests for comparing means were conducted with Bonferroni adjustment for multiple comparisons.

4. Theory/calculations

To compute a sleep quality score that differentiates the variable impact of sleep disturbances, or awakenings, this research has developed a novel approach of applying awakening penalties. Consider that the simplest and best case for a night of sleep would be if a person could have an entire night of deep sleep with no awakenings, and the worst case would be if a person remained awake for the entire night. In between these extreme cases are sleeps that have intermittent disturbances. In order to analyze relationships between the proposed sleep quality score, other behavioral measures, such as length of day, sleep debt, sleep onset, and activity score have been defined and computed as described in sub-sections 4.1–4.7.

4.1. Sleep quality score

The sleep quality score presented here is a weighted sum of sleep phase durations with a novel method of assigning penalties for awakenings during the night. Wearable devices usually record various types of sleep phases. With Jawbone devices, these are defined as: deep sleep (D), light sleep (L), and awakening (A). More deep sleep during the night and fewer awakenings is considered as high sleep quality [10]. Therefore, the quality score (Q_S) for sleep s is defined as a weighted sum with deep sleep phase durations having the highest weight of 2, light sleep having a weight of 1, and awakenings are assigned negative weights or penalties ranging from -1 to 0 (Eq. (1)). The sleep score is normalized by the ideal case of deep sleep for the total sleep duration: $T_{total}W_D$

$$Q_s = \frac{\sum_{i=1}^{n} T_i W_P}{T_{total} W_D} \text{ for } i = [1, n] \text{ phases of duration } T_i \text{ and phase type P } \varepsilon[D, L, A] \quad (1)$$

4.2. Awakening penalty

The awakening penalty function for quantifying the impact of a given awakening is defined in this research as the ratio of two factors: recovery and disruption. Recovery ($T_{Recovery}$) is defined by the duration of the first deep sleep phase that occurs after the awakening. Disruption ($T_{Disruption}$) is defined as the time elapsed from the start of the awakening until deep sleep resumes. These factors are depicted in Fig. 3. The awakening penalty or weight W_A of a given awakening is computed by Eq. (2).

$$W_A = (T_{Recovery}/T_{Disruption}) - 1 \quad (2)$$

The penalty is constrained to range from $[-1, 0]$. Consider the extreme cases of $(T_{Recovery}/T_{Disruption}) > 1$ which is a best case and $T_{Recovery} = 0$ which is a worst case. The highest penalty case, or worst case of sleep disturbance, is when there is not a recovery into deep sleep after an awakening. In this case $T_{Recovery} = 0$ and therefore, W_A is computed as -1. The lowest penalty case would be if disruption time is shorter than recovery. One can imagine that a quick bathroom break in the middle of the night may lead to some relief and allow for a quick recovery to a deep sleep phase. In this case the awakening led to a long deep sleep phase $(T_{Recovery}/T_{Disruption}) > 1$, and the awakening penalty is computed as a positive value. However, positive values are re-assigned to 0 in this study. For cases in between, as recovery time approaches the disruption time,

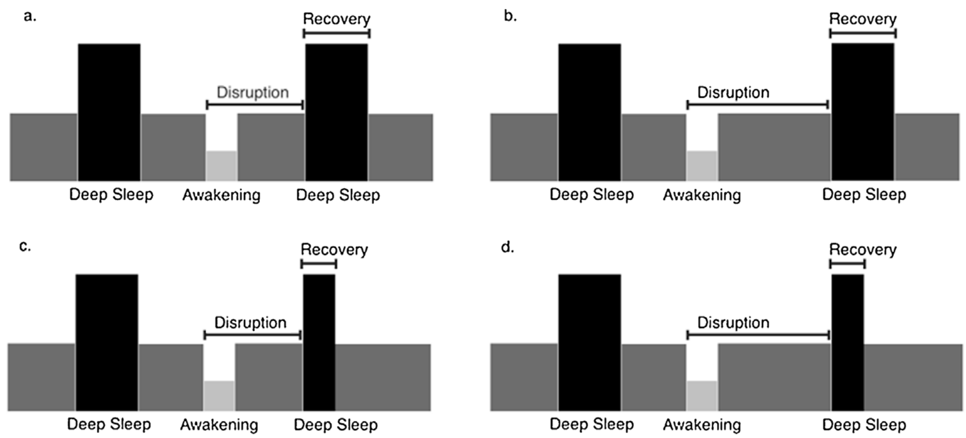

Fig. 3. The example in 3a shows nearly equal time periods of recovery and disruption, 3b shows increasing disruption, 3c shows decreasing recovery, and 3d shows both increased disruption and decreased recovery time compared to 3a.

the penalty becomes less severe, closer to 0. For example, depicted in Figs. 4 and 5 and computed in Table 2 are four cases of various awakening penalties

4.3. Length of day

To analyze our proposed sleep quality score and to determine relationships between length of day and other sleep measurements, instead of assuming that individuals are by default on a 24 h-cycle (i.e. ideal daily pattern of 8 h of sleep and 16 h of activity), here the actual length of day is computed. We quantified how long each wearable device user has stayed awake in between sleeps to consider how length of day before sleeping may impact sleep quality, as defined in Eq. (3).

$$length\_of\_day_n = asleep\_time(n+1) - awake\_time(n) \text{ for day } n \quad (3)$$

4.4. Sleep debt

The duration of the first deep sleep phase in each sleep is recorded for this measure called "sleep debt." If a given sleep did not have any deep sleep phases, then sleep debt is recorded as 0. This measure has been defined to test the hypothesis that when an individual is sleep-deprived, he or she will have a longer first deep phase duration than when he or she is not sleep-deprived. Consider the feeling of "knocking out" or the analogy of "slept like a rock" that is used to describe the perception of having had a deep sleep without any movement when a person is very tired. Wearables also use movement to categorize deep sleep, so it was hypothesized that a long stationary, deep sleep period is one possible measure of sleep debt.

$$sleep\_debt_m = first\_deep\_phase\_duration(m) \text{ for sleep } m \quad (4)$$

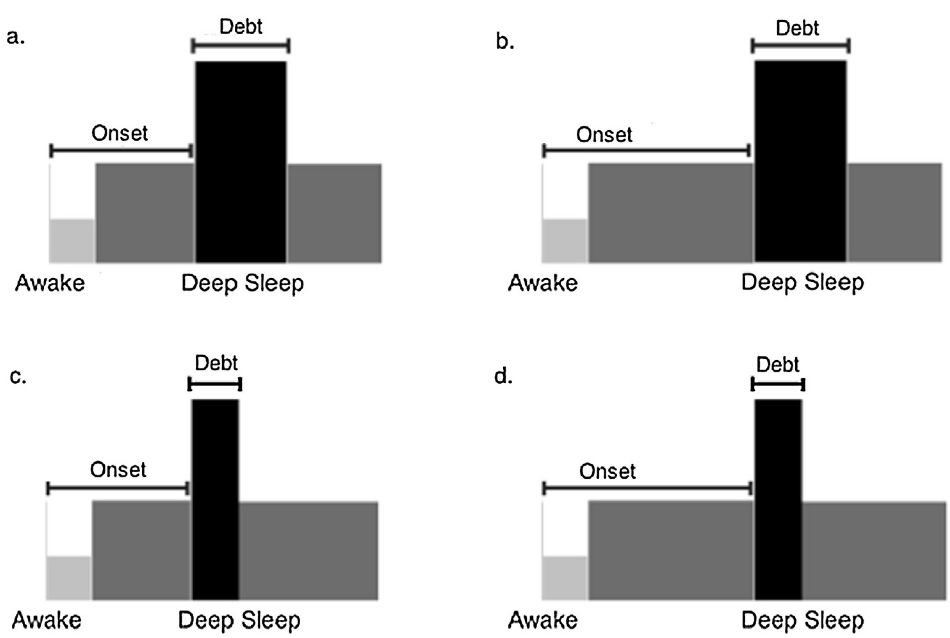

Fig. 4. Examples of various sleep readiness patterns, 4a shows time elapsed until deep sleep is nearly equal to the duration of the deep sleep phase, 4b shows the case of the onset time being longer than the deep sleep duration, 4c shows a deep sleep phase duration that is shorter than the onset time, and 4d shows the case of a longer onset and a shorter deep sleep.

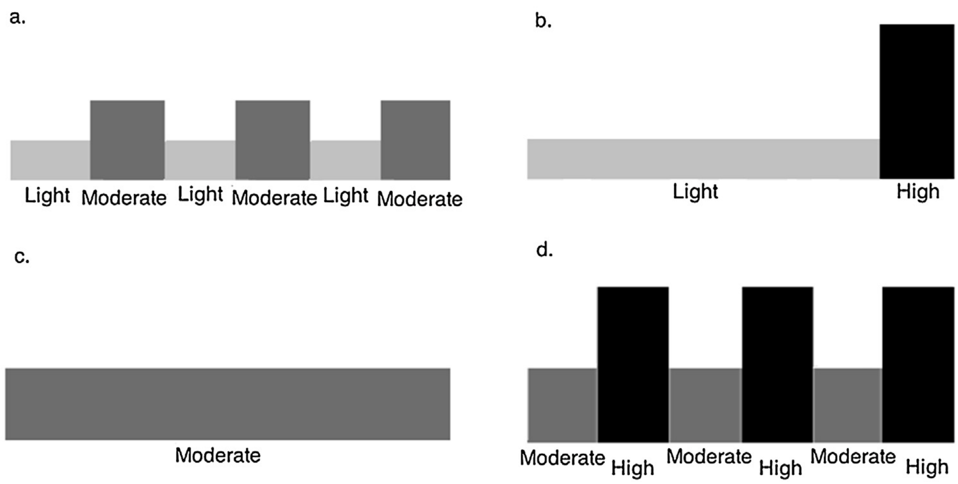

Fig. 5. Examples of activity patterns, 5a shows a balance of light and moderate activity, 5b shows light activity except for one hour of high activity, 5c shows constant moderate activity, and 5d shows a balance of moderate and high activity. Activity scores are computed in Table 4.

Table 2
Computing awakening penalties for the examples cases of Fig. 3 by assuming values of $T_{Recovery}=(2||4)$ minutes and $T_{Disruption}=(6||8)$ minutes.

Case:	$T_{Recovery}$	$T_{Disruption}$	$T_{Recovery}/T_{Disruption}$	W_A
Fig. 3a	4 min	6 min	2/3 = 0.67	−0.33
Fig. 3b	4 min	8 min	1/2 = 0.50	−0.50
Fig. 3c	2 min	6 min	1/3 = 0.33	−0.67
Fig. 3d	2 min	8 min	1/4 = 0.25	−0.75

4.5. Deep sleep onset

The time elapsed from the beginning of the sleep event until the start of the first deep sleep phase is computed and stored as "deep sleep onset." If there was not a deep sleep phase in a given sleep, then deep sleep onset is set equal to total sleep duration. This measure is related to the sleep debt measure, as sleep debt is measuring how long deep sleep lasted, and deep sleep onset is measuring how long it took to enter that deep sleep.

$$sleep\_onset_m = deep\_phase\_start(m) - sleep\_start\_time(m)\ \text{for sleep}\ m \tag{5}$$

4.6. Sleep readiness

The relationship between "deep sleep onset" and "sleep debt" is similar to the previously discussed concepts of "disruption" and "recovery" for computing the penalty of a mid-sleep awakening. Here, instead of a penalty for being awake though, we are computing the ease of falling into a deep sleep. Therefore, similar to Eq. (2), here in Eq. (6), we take the sleep debt and divide by deep sleep onset. This measure is called the "sleep readiness" and is considered to be a positive, rather than a negative, so we do not subtract 1 from this factor. In Fig. 4 and Table 3, there are examples of calculating sleep readiness for various values of sleep onset and sleep debt.

$$W_S = T_{Debt}/T_{Onset} \tag{6}$$

Table 3
Computing sleep readiness for the examples cases of Fig. 4 by assuming values of $T_{Debt}=(2||4)$ minutes and $T_{Onset}=(6||8)$ minutes.

Case:	T_{Debt}	T_{Onset}	$W_S = T_{Debt}/T_{Onset}$
Fig. 4a	4 min	6 min	2/3 = 0.67
Fig. 4b	4 min	8 min	1/2 = 0.50
Fig. 4c	2 min	6 min	1/3 = 0.33
Fig. 4d	2 min	8 min	1/4 = 0.25

Table 4
Computing activity score for the examples cases of Fig. 5.

Case:	Light	Moderate	High	Activity Score
Fig. 5a	3 h	3 h	0 h	3 + 6 = 9
Fig. 5b	5 h	0 h	1 h	5 + 3 = 8
Fig. 5c	0 h	6 h	0 h	12
Fig. 5d	0 h	3 h	3 h	6 + 9 = 15

4.7. Activity score

Jawbone devices count the number of steps taken on a minute-by-minute basis as well as hourly and daily timescales. The activity score defined here is a weighted count on the number of hours per day spent in three different activity levels. The three different categories defined here are: light activity (L), moderate activity (M), and high activity (H) based on the number of steps taken hourly. Light activity ranges from 50 to 400 steps per hour and weighs 1 point. Moderate activity ranges from 400 to 750 steps taken per hour and weighs 2 points. High activity includes more than 750 steps taken per hour and weighs 3 points. The scoring system is rewarding more points if the user has consistent activity throughout the day or some high activity hours, and it rewards less points if the user is sedentary. The activity score is calculated by adding the appropriate weight for the activity level in each hour of the day.

$$Q_A = \sum_{j=1}^{m} W_C(\mathrm{i})\ \text{for}\ hour\ j = [1, m]\ hours \tag{7}$$

$$\text{where}\ W_C(\mathrm{j}) = \begin{cases} 1, if(steps(j) \geq 50 \cap steps(j) < 400) \\ 2, if(steps(j) \geq 400 \cap steps(j) < 750) \\ 3,\ if(steps(j) \geq 750) \end{cases} \tag{8}$$

5. Results

In contrast with the proprietary Jawbone quality score, which was found to be 93% correlated with sleep duration (Fig. 1), the novel sleep quality score presented here is only 13% correlated

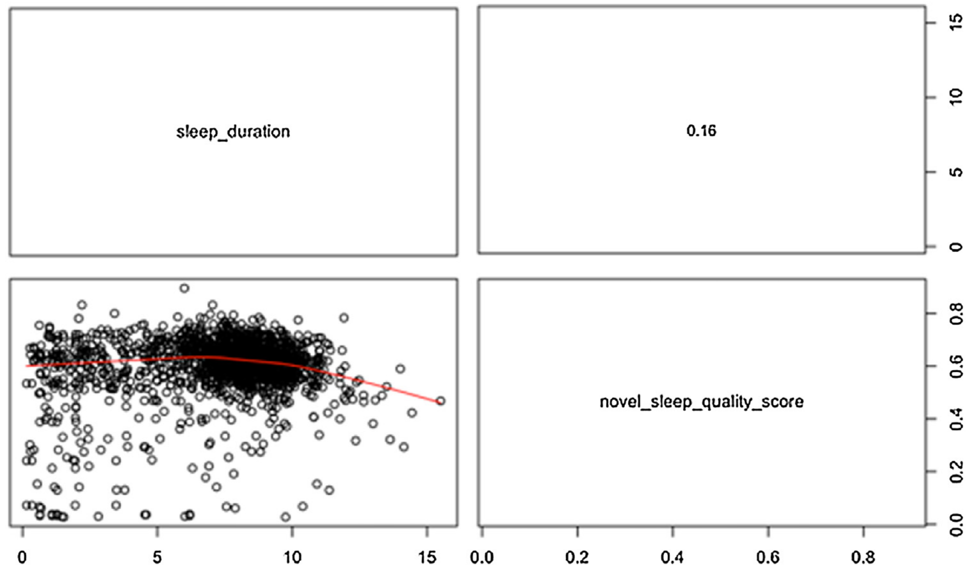

Fig. 6. The novel sleep quality score presented here is not correlated with sleep duration, which is contrasted with the Jawbone quality score in Fig. 1 which due to the high correlation with sleep duration is more of a quantity measure than a quality measure.

with sleep duration as shown here in Fig. 6. The "quality" of sleep reported by Jawbone being highly correlated to the duration of sleep indicates it is not a sensitive measure of sleep quality. The lack of correlation of our novel measure with duration does not prove that we have a sensitive measure of sleep quality, but this does show promise in that our sleep quality score is meeting our first criteria of not being dominated by sleep duration.

Sleep duration for each of the six crewmembers for each month of the mission are shown in Fig. 7. Sleep duration ranged from 6 to 10 h per night on average depending on the individual. Some crewmembers had more stable sleep durations during the mission, whereas crewmembers A and F had divergent patterns in February, March, and April. Sleep quality score for each of the six crewmembers for each month of the mission are shown in Fig. 8. Crewmember D had the most pronounced change in sleep quality from January to February.

Day of week trends during the long time period of observation over the 8-month mission are shown in Figs. 9 and 10. Crewmembers had significantly longer sleep durations on Saturdays compared to Fridays ($p = 0.01$). Crewmembers did not have significant difference in other sleep measure by day of week as a group, but there are individual differences. For example, these measures elucidated that crewmember D had higher sleep readiness on Fridays and Saturdays (Fig. 11). Both the average sleep duration and hte length of day were the longest on Saturdays, indicating that crewmembers slept-in on Saturday mornings and then stayed up late on Saturday nights (Figs. 9 and 12).

To investigate correlations among these measures, Bayesian "sum of trees" based models were fit in an attempt to understand the underlying unknown relations between these novel measures. To draw a strong inference an ensemble learning technique was used to identify what factors are contributing to sleep quality. Here

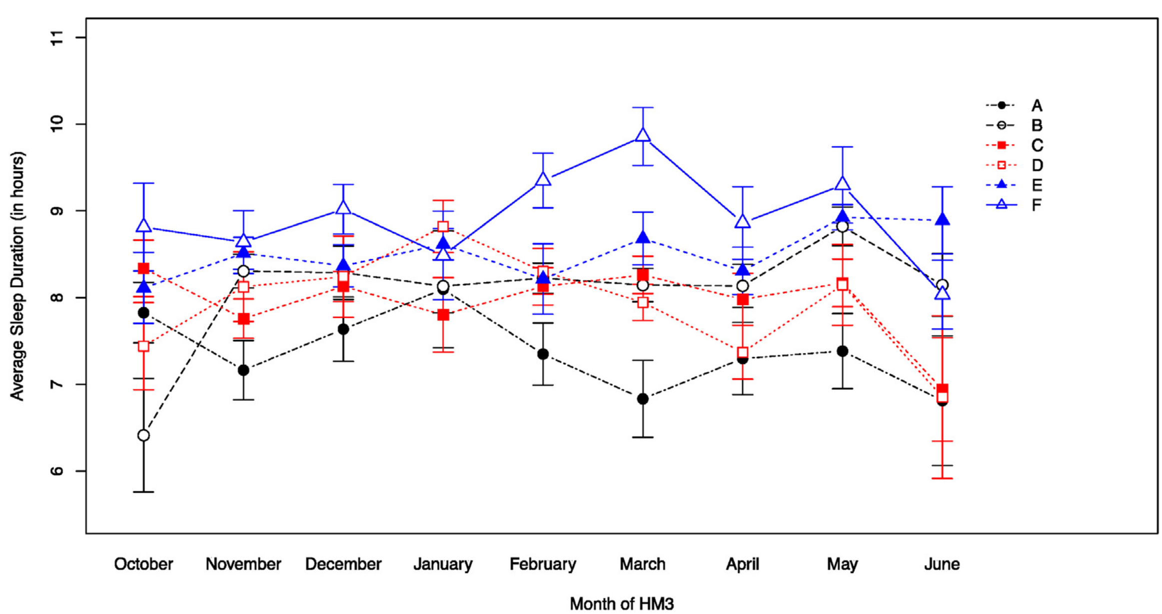

Fig. 7. Average sleep duration in hours for each crewmember during each month.

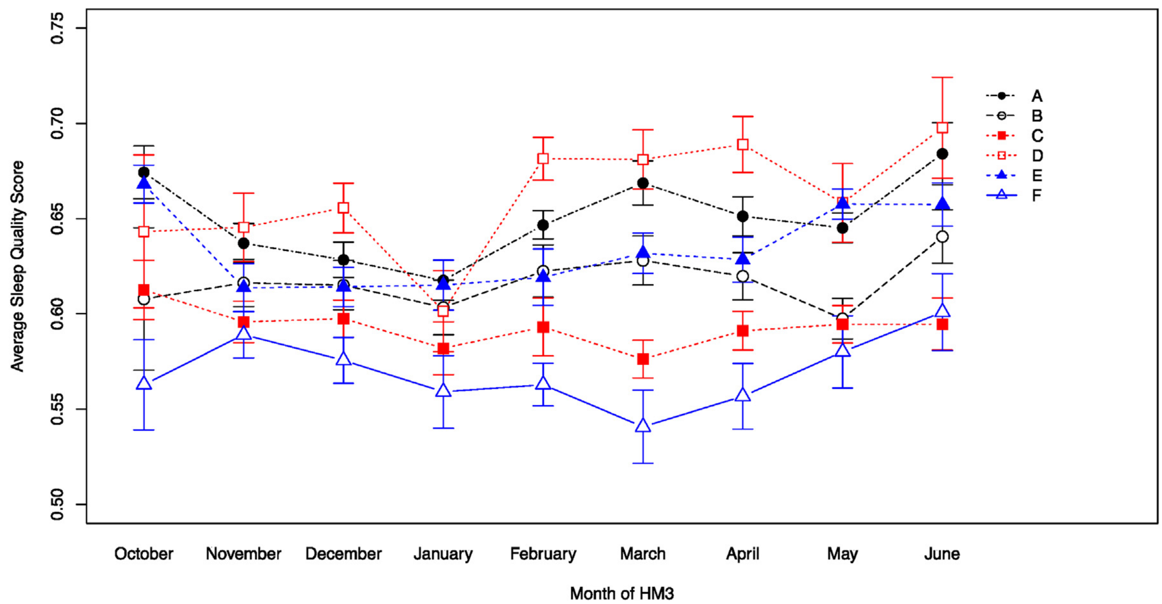

Fig. 8. Average sleep quality score for each crewmember during each month.

all data are considered in aggregate. The Bayesian approach partitions the data in an ad-hoc manner. The model adapts to the "basis" of the conditional data to generate a mixture model from decision trees by using Bayesian probabilistic models at the nodes/ leaves of the decision tree.

Results shown in Figs. 13 and 14, show the partial dependence of sleep quality on some of the other measures presented here, including sleep readiness, activity score, steps, and length of day. Sleep quality has the highest partial dependence on sleep readiness, as sleep readiness also is computed from the patterns of sleep phases, namely the first phases in the onset of sleep. Sleep quality score also has some partial dependence on activity measures including activity score, steps, and length of day. However, computing correlations directly between these measures does not show significant correlation. Therefore, the sleep quality score is likely to be a sensitive measure that provides additional, non-redundant information about user behavior.

6. Discussion

In sleep research, commonly-used measures that are computed from actigraph data include the movement index, fragmentation index, and sleep efficiency. These factors are more indicative of duration, or quantity, than quality. For example, sleep efficiency is the percentage of time spent in bed actually sleeping; this is a simple aggregate of sleep versus wakefulness which does not consider how patterns of sleep and wakefulness during the night are impacting sleep quality. Similarly, the movement index is a count on the number of minutes in which one or more movements were recorded (i.e. the duration, or quantity, of awakenings), and

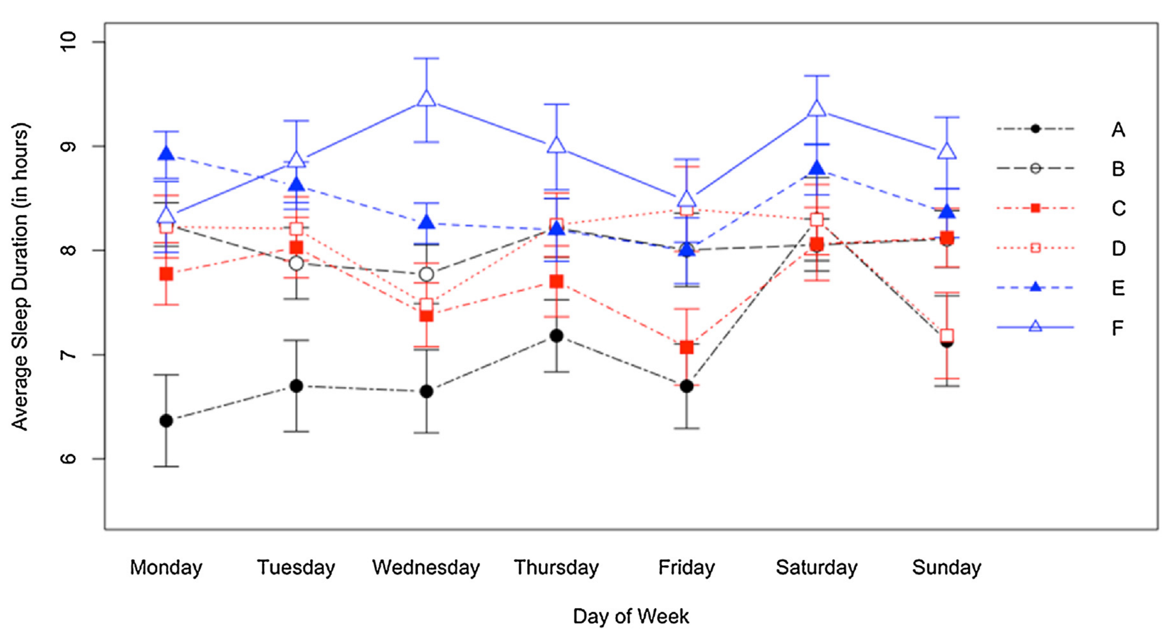

Fig. 9. Average sleep duration for each crewmember for each day of the week.

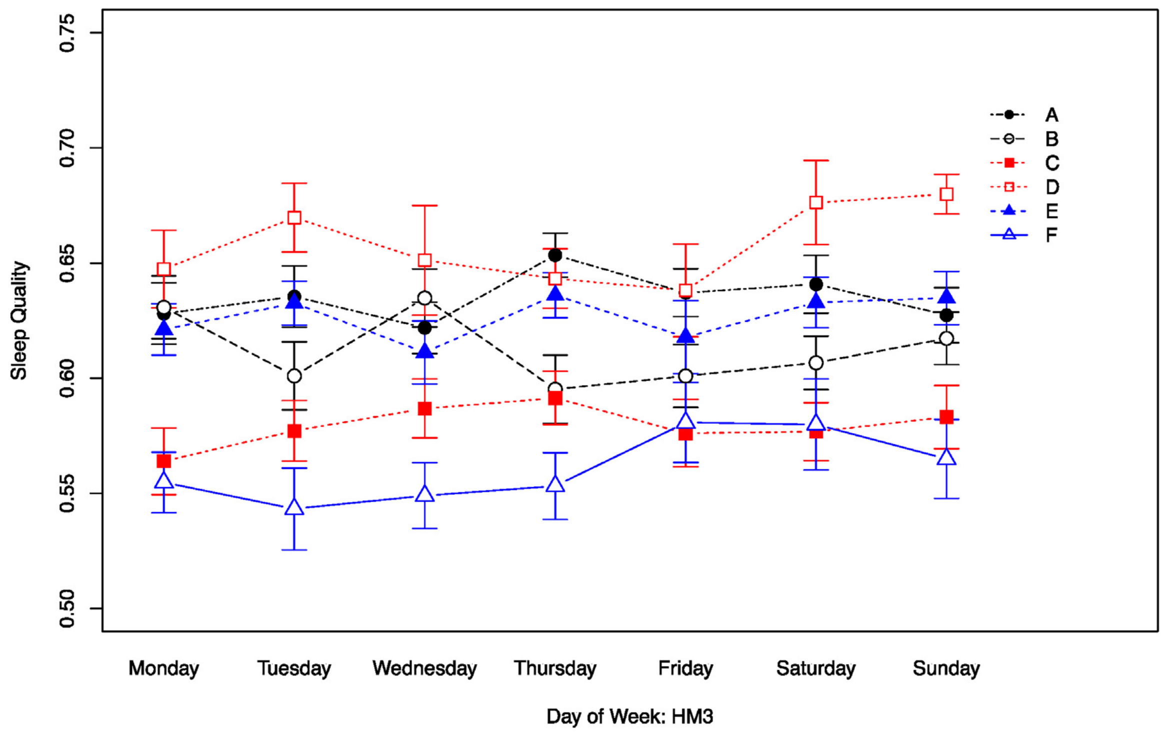

Fig. 10. Average sleep quality score for each crewmember for each day of the week.

the fragmentation index computes the proportion of sleep that was disturbed by brief awakenings. Both of these are also aggregating the wakefulness data and not considering sleep as a process over time that can be disturbed by awakenings.

These measures do not analyze how subsequent sleep phases are impacted by awakenings. For example, multiple factors may affect sleep quality, such as the negative impact of an awakening in the middle of sleep followed by a non-restorative sleep [11]. Here, we have hypothesized that sleep quality is significantly impacted by the longitudinal patterns of sleep phases and awakenings. The "quality" scores provided by the Jawbone UP API are highly correlated with sleep duration 0.93, whereas the sleep quality score presented here is not correlated with duration 0.13. This research developed novel measures of sleep quality considering each awakening along with its impact on subsequent sleep activities. Prior longitudinal research on the time course of behavioral health patterns includes results showing that the primary source of inefficiency within sleep periods was an

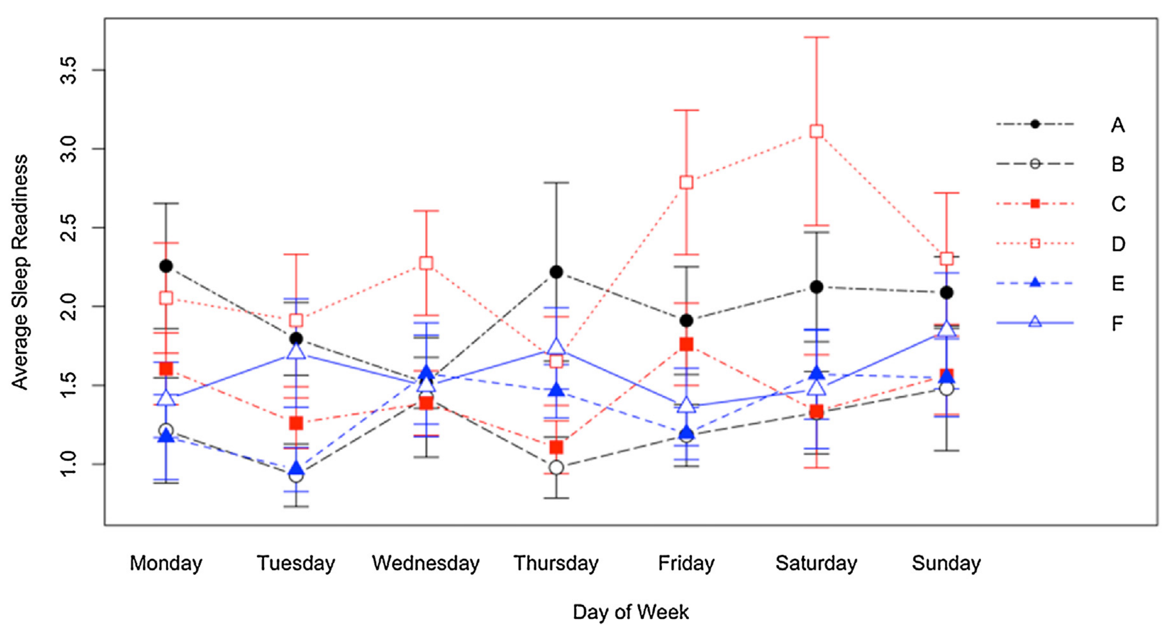

Fig. 11. Measure of Sleep Readiness.

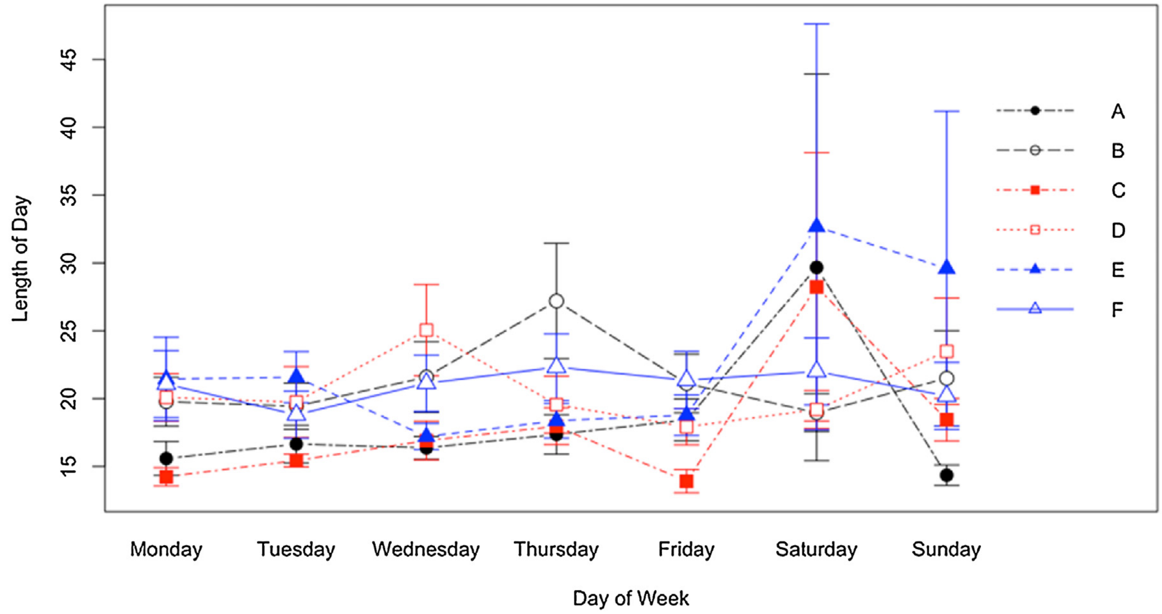

Fig. 12. Measure of Length of Day.

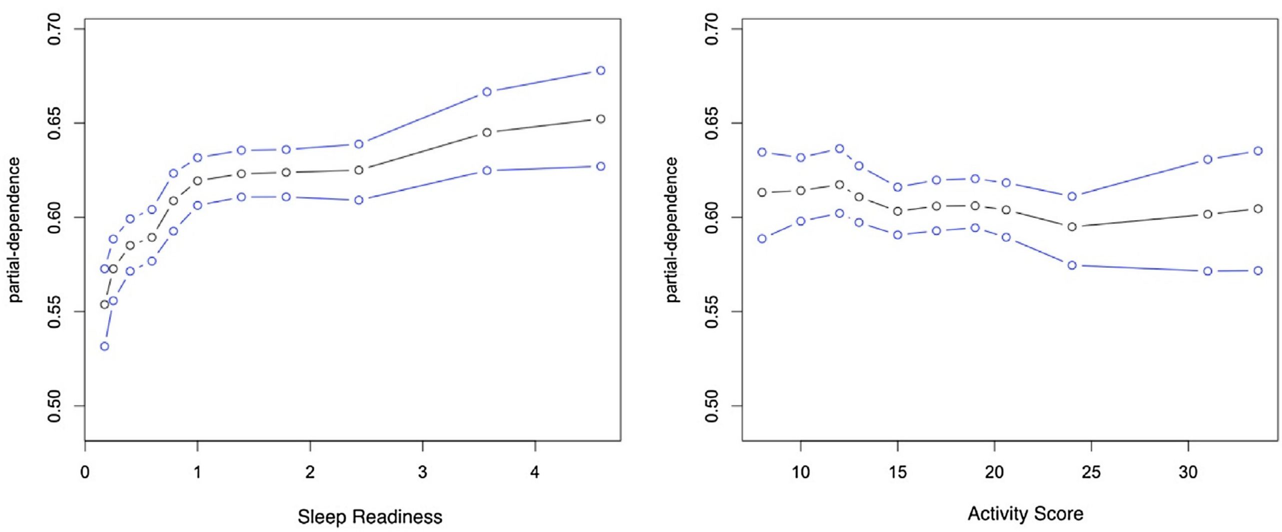

Fig. 13. The partial dependence of sleep quality score on sleep readiness and activity score.

increasing in wakefulness after sleep onset. The wakefulness after sleep onset was the distinguisher between good sleeper and poor sleeper [12,13].

Physical exercise, such as participation in exercise training or self-perceived exercise before bedtime time, has been shown in prior research to have positive correlations with high sleep quality and sleep efficiency [10,14]. Sleep and exercise affect each other. Physical activity is known to be beneficial to improving sleep quality, and conversely, sleep deficiencies or disturbances could affect a person's ability to perform physical activity and increase the risk of injury [15]. Wearable actigraphy recording devices have the advantage of being easy to use but have the challenge of not meeting the high accuracy of all PSG measurements and reliability on estimating wake-sleep stages using activity information. Although significant correlation has been found between subjective and objective sleep quality, consumer-grade wearables devices have limitations regarding identification of times awake during sleep and sleep efficiency percentage measurement [11,16–18]. Assuming that wearables are providing reliable data as shown in validation studies [3,9], then advanced analytics can be designed to accurately suggest behavioral changes for improving health. For example, increasing daily exercise is the most common way to improve health. However, for those who sleep less than the recommended eight hours of sleep they will benefit more from increasing the number of hours sleeping than increasing the number of hours exercising [19].

This approach can be used to objectively assess impacts on behavioral health, including sleep routines, activity patterns, sleep timing, and awakenings. By measuring behavioral changes over time, this research can enable hypothesis testing of many practical questions, such as "Does sleep quality decrease when not following a consistent sleep schedule?" or "Does staying awake for an extra hour for three nights in a row lead to accumulating sleep debt?" Future studies with these measures are recommended to evaluate patterns of behavioral stress that are most detrimental. Developing behavioral predictors of biological or psychological impacts will

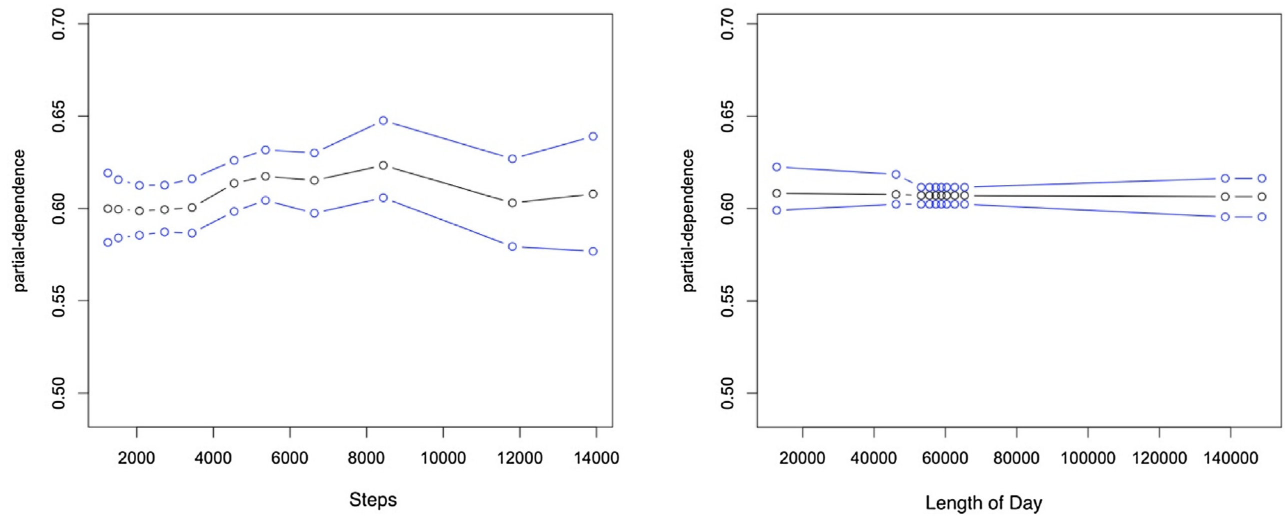

Fig. 14. The partial dependence of sleep quality score on number of steps and length of day.

provide data-driven information for analyzing and assessing personal health management strategies, such as adequate sleep, regular exercise, and proper nutrition [20]. Strategies for promoting behavioral health should be practiced, recorded, and assessed in a longitudinal manner to evaluate efficacy and develop predictive models of impact over time. Data-driven information from smart wearables will lead to improved self-awareness, self-report to health providers, and telemedicine support systems.

7. Conclusions

Wearable device users are already finding benefit in basic data summaries and simple data visualizations provided in mobile applications. However, the attrition rate of wearable users is notably high, as most users only use the devices for about six months [21,22]. There is both a short-term need and long-term opportunity to provide users with added benefits by developing more advanced, but still understandable, measures for analyzing sleep and activity patterns over time. To gain accuracy, develop further capabilities, and promote sustained usage, deeper analysis of existing data is required. Even though many wearable devices are in "beta testing" or require more development and validation, this approach is using the wealth of existing wearables data to analyze behavioral health changes over time. Advanced analytics is a necessary step toward the next-generation of smart wearables for reliably predicting the likelihood and impact of behavioral changes. Preventing the development of unhealthy states, or mitigating the negative impacts of behavioral health challenges, requires methods for quantifying changes over time and analyzing relationships between various activity and sleep patterns. This research has developed measures for objectively assessing how sleep debt accumulates after extended days, how sleep disruption and recovery from wakefulness occur during the night, and when cross-correlations exist between patterns of activity levels and sleep quality. This data-driven information can be applied to prevent health and performance deficits in astronaut crews and Earth-bound citizens alike.

Acknowledgments

To Martha Lenio, Sophie Milam, Allen Mirkadyrov, Neil Scheibelhut, and Zak Wilson, thank you for your invaluable contributions to this research. Dr. Eric Nauman, Dr. Barrett Caldwell, and Dr. Sara McComb for providing helpful feedback and for broadening my perspective during PhD advisory committee meetings. To Jordy Turner and Prithvi Sulaya for contributions to this work though undergraduate research projects. This research would not have been possible without NASA Human Research Program support, as this HI-SEAS mission was funded by NASA grant NNX13AM78G, "Key Contributors to the Maintenance and Regulation of Team Function and Performance on Long Duration Exploration Missions."

Appendix A.

References

[1] M. Basner, D.F. Dinges, D.J. Mollicone, I. Savelev, A.J. Ecker, A. Di Antonio, J.P. Sutton, Psychological and behavioral changes during confinement in a 520-day simulated interplanetary mission to mars, PLoS One 9 (3) (2014) e93298, doi:http://dx.doi.org/10.1371/journal.pone.0093298.

[2] L. Berkowitz, Health Wearables: Early Days, PwC's Health Research Institute, 2014, pp. 12. Retrieved from http://www.pwc.com/us/en/industry/

entertainment-media/publications/consumer-intelligence-series/assets/pwc-hri-wearable-devices.pdf.

[3] J. Mantua, N. Gravel, R.M.C. Spencer, Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography, Sensors (Switz.) 16 (5) (2016), doi:http://dx.doi.org/10.3390/s16050646.

[4] S. Cohen, E. Frank, W.J. Doyle, D.P. Skoner, B.S. Rabin, J.M. Gwaltney Jr., Types of stressors that increase susceptibility to the common cold in healthy adults, Health Psychol. 17 (3) (1998) 214–223.

[5] S. Priyadarshini, P. Aich, Effects of psychological stress on innate immunity and metabolism in humans: a systematic analysis, PLoS One 7 (9) (2012), doi:http://dx.doi.org/10.1371/journal.pone.0043232.

[6] Laura K. Barger, Erin E. Flynn-Evans, Alan Kubey, Lorcan Walsh, Joseph M. Ronda, Wei Wang, Kenneth P. Wright, Charles A. Czeisler, Prevalence of sleep deficiency and use of hypnotic drugs in astronauts before, during, and after spaceflight: an observational study, Lancet Neurol. 13 (9) (2014) 904–912.

[7] Erin E. Flynn-Evans, Laura K. Barger, Alan A. Kubey, Jason P. Sullivan, Charles A. Czeisler, Circadian misalignment affects sleep and medication use before and during spaceflight, npj Microgravity 2 (2016).

[8] June J. Pilcher, Douglas R. Ginter, Brigitte Sadowsky, Sleep quality ver- sus sleep quantity: relationships between sleep and measures of health, well- being and sleepiness in college students, J. Psychosom. Res. 42 (6) (1997) 583–596.

[9] Carol A. Landis, Christine A. Frey, Martha J. Lentz, James Rothermel, Dedra Buchwald, Joan L.F. Shaver, Self-reported sleep quality and fatigue correlates with actigraphy in midlife women with fibromyalgia, Nurs. Res. 52 (3) (2003) 140–147.

[10] S. Brand, N. Kalak, M. Gerber, R. Kirov, U. Pühse, E. Holsboer-Trachsler, High self-perceived exercise exertion before bedtime is associated with greater objectively assessed sleep efficiency, Sleep Med. 15 (9) (2014) 1031–1036, doi:http://dx.doi.org/10.1016/j.sleep.2014.05.016.

[11] C.-E. Kuo, Y.-C. Liu, D.-W. Chang, C.-P. Young, F.-Z. Shaw, S.-F. Liang, Development and evaluation of a wearable device for sleep quality assessment, IEEE Trans. Biomed. Eng. PP (99) (2016), doi:http://dx.doi.org/10.1109/tbme.2016.2612938 1-1.

[12] W.B. Webb, H.W. Agnew, Sleep efficiency for sleep-wake cycles of varied length, Psychophysiology 12 (6) (1975) 637–641, doi:http://dx.doi.org/10.1111/j.1469-8986.1975.tb00063.x.

[13] C.M. Morin, S.E. Gramling, Sleep patterns and aging: comparison of older adults with and without insomnia complaints, Psychol. Aging 4 (3) (1989) 290–294. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2803622.

[14] P.-Y. Yang, K.-H. Ho, H.-C. Chen, M.-Y. Chien, Exercise training improves sleep quality in middle-aged and older adults with sleep problems: a systematic review, J. Physiother. 58 (3) (2012) 157–163, doi:http://dx.doi.org/10.1016/s1836-9553(12)70106-6.

[15] M. Chennaoui, P.J. Arnal, F. Sauvet, D. Léger, Sleep and exercise: a reciprocal issue? Sleep Med. Rev. 20 (2015) 59–72, doi:http://dx.doi.org/10.1016/j.smrv.2014.06.008.

[16] C.V. Visovsky, K.E. Kip, J.L. Rice, M. Hardwick, P. Hall, Choosing instruments for research: an evaluation of two activity monitors in healthy women, J. Novel Physiother. 03 (05) (2013), doi:http://dx.doi.org/10.4172/2165-7025.1000171.

[17] L. de Souza, A. Benedito-Silva, M. Pires, et al., Further validation of actigraphy for sleep studies, Sleep 1 (2003) 81–85.

[18] J. Merilahti, A. Saarinen, J. Parkka, K. Antila, E. Mattila, I. Korhonen, Long-term subjective and objective sleep analysis of total sleep time and sleep quality in real life settings, 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2007), doi:http://dx.doi.org/10.1109/iembs.2007.4353514.

[19] M.E. Rosenberger, M.P. Buman, W.L. Haskell, M.V. McConnell, L.L. Carstensen, Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices, Med. Sci. Sports Exerc. 48 (3) (2016) 457–465, doi:http://dx.doi.org/10.1249/MSS.0000000000000778.

[20] L. Varvogli, C. Darviri, Stress management techniques: evidence-based procedures that reduce stress and promote health, Health Sci. J. 2 (2011) 74–89.

[21] P. Andrew, D. Maeve, Americans' internet access: 2000–2015, Pew Res. Center (2015) 1–13.

[22] L. Piwek, D.A. Ellis, S. Andrews, A. Joinson, The rise of consumer health wearables: promises and barriers, PLoS Med. (2016) 1–9, doi:http://dx.doi.org/10.1371/journal.pmed.1001953.