Making Sense of Sleep Sensors: How Sleep Sensing Technologies Support and Undermine Sleep Health

Ruth Ravichandran[1], Sang-Wha Sien[3], Shwetak N. Patel[1,2], Julie A. Kientz[3], & Laura R. Pina[2,3]
[1]Electrical Engineering, [2]Computer Science & Engineering, [3]Human Centered Design & Engineering
University of Washington, Seattle, WA USA
{vinisha, elgreco, jkientz}@uw.edu, {shwetak, lpina}@cs.washington.edu

ABSTRACT
Sleep is an important aspect of our health, but it is difficult for people to track manually because it is an unconscious activity. The ability to *sense* sleep has aimed to lower the barriers of tracking sleep. Although sleep sensors are widely available, their usefulness and potential to promote healthy sleep behaviors has not been fully realized. To understand people's perspectives on sleep sensing devices and their potential for promoting sleep health, we surveyed 87 and interviewed 12 people who currently use or have previously used sleep sensors, interviewed 5 sleep medical experts, and conducted an in-depth qualitative analysis of 6986 reviews of the most popular commercial sleep sensing technologies. We found that the feedback provided by current sleep sensing technologies affects users' perceptions of their sleep and encourages goals that are in tension with evidence-based methods for promoting good sleep health. Our research provides design recommendations for improving the feedback of sleep sensing technologies by bridging the gap between expert and user goals.

Author Keywords
Sleep tracking, health monitoring, sleep, sleep sensing, personal informatics, quantified self, behavior change.

ACM Classification Keywords
J.3. Life and medical sciences: Health.

INTRODUCTION
Adequate, restful sleep is as important to one's well-being as a healthy diet and regular physical activity. During sleep, the body and brain undergo necessary restorative activities [1], and inadequate sleep leads to reduced alertness and drowsiness [17]. In the United States, an estimated of 50 million people have poor sleep quality or have a sleep disorder such as insomnia, sleep apnea, and narcolepsy [10].

CHI 2017, May 06-11, 2017, Denver, CO, USA
© 2017 ACM. ISBN 978-1-4503-4655-9/17/05⋯$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025557

Despite the pervasiveness of sleep issues, people struggle to assess and improve their sleep. Sleep is an unconscious, passive activity and therefore—unlike diet and physical activity, which are difficult but possible to track manually [12]—accurately self-tracking sleep manually is often unattainable.

The clinical gold standard of sleep quality assessment is a polysomnographic (PSG) study. This study generally consists of a single night, clinical evaluation at a sleep clinic. The patient wears seven different physiological sensors directly on their body [3]. PSG studies are used to diagnose sleep-related disorders, such as narcolepsy (e.g., uncontrollable sleepiness) or sleep apnea. PSG studies are accurate, but expensive. They require monitoring in a highly controlled and unnatural setting, and patients find the sensors uncomfortable to wear even for a single night. These limitations make it difficult to establish a baseline of behavioral sleep patterns over time.

Commercial sleep sensing technology for use at home is a growing industry [25]. These technologies have the potential to overcome the limitations of PSG studies while providing long-term, low-cost, and accurate representations of people's daily sleep patterns in their natural and comfortable home environment. The popularity of these commercial sleep sensors is promising in that they indicate that people have an interest in understanding and obtaining good sleep health. However, literature has not examined whether commercial devices effectively sense sleep quality and provide people with meaningful feedback. Thus, we set out to answer the following research questions:

- How are people currently using commercially available sleep sensors and making sense of feedback they provide?
- What aspects of sleep sensing and feedback either facilitate or potentially undermine people's ability to understand their sleep and achieve good sleep health?
- What aspects of current sleep sensor technology designs are in line with evidence-based methods of understanding and promoting good sleep health?

To answer these questions, we collected a dataset consisting of interviews with 5 sleep experts, surveys with 87 and interviews with 12 people that have used sleep sensing devices, and 6986 consumer product reviews from the most widely used commercial sleep sensing devices. We focused on sleep sensing technologies that use physiological sensing, such as body movement, breathing rate, or heart rate to estimate sleep quality and excluded manual, self-reported sleep tracking methods such as sleep diaries. We find that:

- Self-trackers using sleep sensing technologies often develop broken mental models about what commercial sleep sensors are able to actually sense, how they work, and are frustrated with the lack of algorithmic transparency in sleep sensing technologies.
- Self-trackers find it distracting when feedback emphasizes unconscious aspects of sleep, such as time in sleep stages, over aspects of their sleep they have the ability to control and improve.
- Self-trackers can better understand and improve their overall sleep habits when feedback from sleep sensors focuses on duration, timing, and making connections to modifiable behaviors and sleep hygiene.

Our findings examine the state of sleep sensing feedback from the perspective of users' needs and sleep experts. From our results, we derive design recommendations that consider users' needs and connect them to evidence-based strategies for improving sleep quality. A set of these recommendations provides new avenues to improve sleep sensing.

RELATED WORK

In this section, we review literature on designing applications for sleep sensing and improving sleep health as well as research on helping users understand health-related data.

Designing for Sleep Tracking and Sensing

In the HCI community, there has been a recent trend focusing on novel computing-based interventions for sleep. In 2011, Choe et al. conducted a literature review and formative study to examine design opportunities for sleep from an HCI perspective [7]. The authors identified people's strong interest in lowering the barriers to track their sleep and the factors that affect their sleep. The authors also stressed the importance of supporting long-term sleep tracking to identify trends to help people create personalized sleep goals.

Building upon these opportunities, researchers have explored varying ways of capturing and providing feedback on aspects related to sleep health. SleepTight [8] is an application that lowers the barriers of manually tracking sleep and helped users make sense of behavioral factors that could be affecting their sleep. ShutEye [2] is a peripheral display on a smartphone's active wallpaper which provides timely guidance on when it is best to engage in activities that could impact sleep, such as consuming caffeine or exercising. In SleepCoacher, Daskalova et al. [13] explored the use of a personalized, automated self-experimentation system for understanding sleep health. Finally, Lullaby captures environmental factors (e.g., temperature, light, audio, and motion) in relation to sleep data captured from a Fitbit. Lullaby provides comprehensive information of users' sleep environment. This information allows users to learn about environmental factors that may affect sleep [20].

Sleep sensing can lower the burden of manual sleep tracking and improve the accuracy of sleep inference at home. Toss 'N' Turn investigated the accuracy of sleep sensing using data from seven sensors found in smartphones and found it was possible to predict aspects of sleep quality to between 81-83% accuracy [28]. DoppleSleep uses off-the shelf 24 GHz radar modules to monitor vital signs and body movements and uses that information to infer sleep stages and differentiate between sleep and wake times [30]. The contactless nature of DoppleSleep obviates the need to instrument the user's body with sensors and lowers the cost of clinical sleep studies.

Toss 'N' Turn and DoppleSleep have focused on improving sensing, but not on the feedback that would be provided to users. Liu et al. conducted an investigation of the usability and acceptability of commercial sleep sensing devices [27]. They found there are a number of issues, including discomfort, battery life, and inability for users to modify data. Our analysis confirms many of Liu et al.'s findings. We build upon this prior work by focusing on the feedback sleep sensors provide and how users interpret and take action on the feedback.

Making Sense of Health Data

Connected to the design of self-tracking technologies is the study of how people make sense of their self-tracked health data and why people abandon self-tracking when they struggle to make sense of said data [14, 22]. Previous work has examined how to represent and visualize data such that it is persuasive and offers insights that can lead to behavior-change [11,15,33]. One such way to improve data representation is to clearly convey its uncertainty [21].

Uncertainty is a point of frustration for users of physical activity inference technologies. Users of these technologies have to cope with activity inference and measurements that are prone to error. Consolvo et al. identified that users react negatively when fitness trackers incorrectly infer a particular physical activity and consequently, do not give users credit for said activity [11]. Kay et al. found that there is a disconnect between a users' perception of their weight, the precision capabilities of their scale, and clinical relevance of weight deviations [22]. Kay et al. further found that an accurate understanding of weight fluctuation is associated with greater trust in the scale itself. Work by Yang et al. [34] examined how self-trackers view the inaccuracy of sensor-driven step count inference and the process in which self-trackers engage to assess the accuracy (or lack thereof) of their fitness devices. These studies demonstrate users care about the accuracy of sensor-driven tracking, taking accuracy into account when they assess their data.

We extend such work by examining the strengths and weakness of sleep sensing feedback from the perspective of users and sleep experts. Our results indicate that sleep sensing enables or interferes with making sense of sleep quality. Our discussion provides design recommendations which connect user needs with evidence-based strategies for improving sleep quality, while still taking into account the limitations of sleep sensing technology. These design guidelines can improve sleep sensing technology and provide new avenues in sleep sensing research.

DATA COLLECTION & ANALYSIS METHODS

To understand the state of sleep research and the needs of people using sleep sensing devices to track their sleep, we 1) interviewed sleep experts and reviewed the literature on sleep research, 2) analyzed consumer product reviews of sleep sensing devices, 3) deployed an online survey, and 4) interviewed a subset of survey respondents. In this section, we describe our process and analysis methods.

Interviews with Sleep Experts

To gain an understanding of the factors contributing to sleep health, we conducted a literature review of sleep research and interviewed five experts in the field of sleep medicine (E1-E5). E1 is a Neurology professor and board certified sleep specialist. E2 is a professor in Psychiatry & Behavioral Science, co-director of a sleep research center, and editor of a major sleep research journal. E3 is a sleep researcher in a department of Family and Nursing. E4 is a professor in a department of Family and Child Nursing and focuses on pediatric sleep. Finally, E5 is a pediatric psychologist and sleep researcher.

Experts were familiar with commercial sleep sensors and the feedback they provide. These interviews helped us understand experts' perspectives on how sleep sensing technologies address sleep health needs and the practices the experts establish with patients who use sleep sensing technologies to track their sleep. During the interview, experts were asked to comment on feedback examples and discuss how they use patient-generated sleep sensing data. We analyzed the sleep expert interviews with support from the sleep literature to identify themes focused on maintaining and improving sleep.

Reviews of Sleep Sensing Products

We collected and analyzed product reviews from the most widely-used commercially available sleep sensing technologies to gather a user perspectives on sleep sensing feedback. We gathered reviews from three sources: Amazon.com, iTunes Store, and Google Play Store. Our inclusion criteria consisted of: 1) smartphone apps using phone sensors (e.g., accelerometer and/or microphone), 2) dedicated sleep sensing devices, or 3) fitness trackers which also sense sleep.

For smartphone apps, we analyzed reviews from the 4 highest-rated apps from the iTunes Store and the 5 highest-rated apps from Google Play. We selected reviews in decreasing order of word count (e.g., longest reviews first), stopping once we felt we reached data saturation. For iTunes reviews, we reached data saturation at 280 word count, analyzing 475 reviews out of a total of 2000 possible reviews. For Google Play reviews, we reached data saturation at 500 word count, analyzing 377 out of a total of 14581 possible reviews. Combining both sources, we analyzed 852 app reviews.

From Amazon.com, we collected reviews from dedicated sleep sensing devices. These are sensors that are placed under the mattress (e.g., Beddit, Withings Aura), clipped on the sleeper's pillow (e.g., Sense with Sleep Pill), or placed on the nightstand (e.g., S+). We analyzed all 683 reviews for these five dedicated sleep devices. Also from Amazon.com, we collected reviews from the top four suggested wearable fitness trackers with sleep sensing functionality: Fitbit One, Fitbit HR, Jawbone Up3, and Misfit Shine. We only included fitness tracker reviews containing the word 'sleep'. This led to 3234 Fitbit One, 4298 Fitbit HR, 893 Jawbone Up3, and 78 Misfit Shine reviews to analyze. Similar to our data saturation process for the smartphone app reviews, we read reviews in decreasing order of word count, analyzing data until we felt we reached data saturation. These reviews tended to be longer than the smartphone app reviews. The three authors coding this dataset reached saturation at different word counts for some of the devices. In total, we analyzed 2113 Fitbit One, 2452 Fitbit HR, 808 Jawbone Up3, and all 78 Misfit Shine reviews, totaling 5451 fitness tracker reviews. Combining all review datasets, we analyzed a total of 6968 reviews (Table 1).

	App Reviews (6986 reviews)
iPhone apps	Smart Alarm Clock (87), SleepBot (171), MotionX (119), Sleep Cycle (98)
Android apps	Sleep Bot (138), Sleep Cycle (84), Sleep Tracker (1), Sleep as Android (116), Sleep as Android Paid (38)
Amazon reviews	Dedicated devices (683): Sense with Sleep Pill (290), Beddit (99), Withings Aura (215), S+ (79); Fitness Trackers (5451): Fitbit One (2113), Fitbit HR (2452), Misfit Shine (78), Jawbone Up3 (808)
	Survey Demographics (87 people)
Gender	Women (50), Men (37)
Age	min 18, max 73, mean 33.6, median 31 18-23 (12), 24-29 (27), 30-39 (28), ≥40 (22)
Currently tracking?	tracking (60), discontinued (29)
Tracker type	smartphone app (3), fitness tracker (56), dedicated device (3), Other (12)
	Interview Demographics (12 people)
Gender	Women (8), Men (4)
Age	min 21, max 44, mean 31.8, median 31.5 18-23 (1), 24-29 (4), 30-39 (4), ≥40 (3)
Currently tracking?	tracking (9), discontinued (3)
Tracker type	smartphone app (18), fitness tracker (9)

Table 1. We collected data from four sources: interviews of sleep experts, app store and Amazon reviews of sleep sensing devices, a survey of people who currently use or have used sleep sensing devices, and follow-up interviews with survey respondents.

Online Survey

The themes identified from the expert interview data and the review dataset informed the list of questions to survey self-trackers using sleep tracking technologies. The 29-question survey focused on: 1) reasons why people track their sleep, 2) which sleep sensing devices people use and why those devices, 3) the type of information people wanted to collect, 4) how people make sense of the feedback from sleep sensing technology, and 5) how people connect data to their sleep quality. Questions were a mix of open-ended, Likert, and multiple choice. We recruited by posting on social networking sites, online message forums, and through a sleep blog. To incentivize participation, respondents were entered into a drawing to win one of five $20 USD Amazon gift cards. We gathered a total of 87 responses (demographics in Table 1).

Semi-Structured Interviews

Survey respondents had the option to consent to be contacted for an in-depth follow-up interview. We contacted all 46 respondents that consented. We interviewed the 12 which replied to our request (demographics in Table 1). We conducted interviews over the phone or in person. Interviews lasted between 16 to 30 minutes. With consent from participants, we recorded and transcribed interviews. Interview questions were based on respondents' survey answers and were intended to triangulate and add depth to our findings from the survey, app reviews, sleep literature, and interviews with experts. Five interviewees had been diagnosed with a sleep disorder and three had stopped tracking. We compensated interview participants with a $25 USD Amazon gift card.

Analysis

Our analysis consisted of an iterative affinity diagramming process with 6 steps to analyze our triangulated dataset [4]. In Step 1, we analyzed expert interview data and the literature. We identified 7 themes focusing on sleep hygiene, modifiable behaviors, experts' perspectives on how sleep sensing feedback can help their patients address sleep concerns, and how patients and physicians use feedback provided by sleep sensing devices. In Step 2, we analyzed the product review dataset, which generated 64 themes. In Step 3, we created our survey based on the themes generated from the two previous steps. In Step 4, we analyzed the survey data and merged it with the themes identified from the product review dataset (i.e., Step 2). In Step 5, we applied the 7 themes from the expert data to the themes generated from the survey and review dataset, but kept themes reflecting user practices and challenges. This step trimmed our themes 64 to 30. Based on these themes, we created our interview protocol to gather deeper insights. Finally, in Step 6, we integrated the interview data to identify higher level themes presented in the results. For every step of the analysis that required affinity analysis, the data was split between three authors. Each author analyzed their subset of the dataset. We then came together to merge, discuss, and iterate on themes.

BACKGROUND ON SLEEP

We now summarize the findings from our literature review on sleep and what constitutes health sleep. We explain what is sleep, how sleep quality is clinically assessed, and summarize evidence-based strategies to improve sleep. We use the terminology defined in this section in the remainder of the paper to discuss strengths and weaknesses of the current state of commercial sleep sensing feedback.

What is Sleep?

Typically, there are two main stages which we cycle through when we sleep: Rapid Eye Movement (REM) sleep and Non Rapid Eye Movement (NREM) sleep. Three-quarters of our sleep is comprised of NREM sleep, which can be further broken down into three stages: **Stage 1**, which is also known as light sleep, **Stage 2**, which is a deeper stage of sleep where one becomes disengaged from their surroundings, and **Stage 3**, which is the third and final stage is the deepest sleep. **REM** is characterized by rapid eye movement and rapid, irregular, and shallow breathing. These stages occur unconsciously and people cannot control the patterns through which they cycle through the stages or how many hours they spend in a particular sleep stage.

Sensing Sleep

In a clinical sleep assessment setting, the golden standard is a polysomnography (PSG) study. For a PSG study, patients stay overnight at a sleep lab and sleep while wearing at least seven physiological sensors. These sensors include: an electroencephalogram (EEG, used to sense brain waves), electrooculography (EOG, used to track eye movements, electromyography (EMG, used to capture electrical activity produced by skeletal muscles), pulse oximeters, and microphones [23]. The data captured from these sensors is used to classify and identify sleep stages. PSG studies are used to diagnose sleep-related disorders such as sleep apnea, restless leg syndrome, and teeth grinding. Using a standard protocol, an entire night's worth of data is manually analyzed in 30 second intervals by a trained sleep technician to identify sleep stages [3]. PSG is completed to assesses people with sleep-related disorders. It is generally not conducted with people who have poor sleep quality.

In non-clinical settings, a wearable, accelerometer based sensor, known as an Actigraph, has become a popular, clinically validated tool for continuous sleep tracking [31]. Patients wear an Actigraph to help identify wake-up, sleep times, and the amount of movement throughout the night. The data captured by Actigraphs are available to clinicians, but data or feedback is not accessible to patients for their own personal use.

Commercial sleep sensing is becoming readily available. Devices have the potential to help people learn and monitor their sleep outside of a clinical setting. These sensing technologies use accelerometers, heart rate monitors, breathing rate, and microphone sensors to infer sleep. However, these technologies have not been clinically validated, and their accuracies compared to the gold standard are not made public by the companies who sell them. For example, Montgomery-Downs et al. compared the accuracy of Fitbit and ActiWatch against a PSG study [29]. The authors found that Fitbit and ActiWatch differed significantly on recorded total sleep time, both between each other and compared to PSG. Therefore, the sleep medicine community is concerned that commercial sleep sensing technologies are providing inaccurate feedback on sleep stages and quality.

Defining Sleep Quality

In addition to sleep stages, the data captured from the PSG study is used to assess the following measures: *sleep efficiency* (the ratio of total sleep time to time spent in bed), *sleep latency* (the duration from bedtime to the onset of sleep), *arousal index* (number of awakenings after sleep onset), and other metrics to assess sleep apnea [23]. These measures help physicians diagnose sleep-related disorders. On their own, they are insufficient to assess sleep quality for people wanting to track their sleep long-term or to diagnose behavioral sleep disorders, such as insomnia.

In addition to the metrics explained, sleep quality assessment requires considering routines, as well as behaviors before going to sleep and after waking up. For example, to improve sleep quality, sleep experts and our literature review suggest users assess sleep habits and adopt *modifiable behaviors* [18]. *Modifiable behaviors* are behaviors that people have control to act on, which include examples such as: (1) keeping one's bedroom cool and dark; (2) maintaining a regular bedtime and wake time every day, even on weekends; and (3) avoiding large late-night meals. A second and related concept is *sleep hygiene,* which refers to behaviors, habits, and environmental factors that can be adjusted to promote good sleep quality [32]. Examples of sleep hygiene include avoiding caffeine later in the day, exercising regularly, and establishing a relaxing bedtime routine [18,32]. Addressing *modifiable behaviors* and *sleep hygiene* are the first two methods sleep clinicians use when patients complain about poor sleep quality.

A second component to assessing sleep quality is subjective self-assessment. SATED is a framework that uses five dimensions to measure subjective sleep quality [5]. The name of the framework is an acronym that stands for:

- **Satisfaction:** the subjective assessment of "good" or "poor" sleep
- **Alertness:** the ability to maintain attentive wakefulness
- **Timing:** the placement of sleep within the 24-hour day
- **Efficiency:** the ease of falling asleep and returning to sleep
- **Duration:** the total amount of sleep obtained per 24 hours

SLEEP SENSORS AS FACILITATORS

Sleep sensing feedback provides awareness, motivates users to prioritize sleep, helps improve sleep habits, and helps people with sleeping disorders collaborate with their physicians to better manage their condition. In this section, we discuss the strengths of sleep sensing feedback and opportunities to improve.

Promoting Awareness about Sleep Health

Inadequate sleep is the most common sleep issue in the United States [17]. Experts in our study agreed, explaining that sleep is a low priority for people: "*I don't think that people pay much attention to sleep until they have a problem"* (E3)[1]. All experts in our study said sleep sensing technologies can create awareness of the importance of sleep in typically healthy people: "*I don't think sleep hits their radar unless someone actually shows them saying look, you are not getting a lot of sleep"* (E4).

Just from the sheer availability of consumer sleep sensing technologies, people have started to utilize these emerging technologies to learn more about their sleep habits. 83.9% (73/87) of our survey respondents and a majority of online reviewers considered themselves healthy, were very interested in understanding their sleep, and discussed the benefits of having access to information about their sleep. This provided users with information they were previously unaware of: *"First useful observation was that I'll never get 7 hours of sleep if I only spend 6 hours in bed (i.e., I thought I was going to bed earlier than I really was)"* (S117). This type of feedback motivates users to prioritize sleep: "*I've drastically improved my sleep habits just by knowing that I was sleep deprived. I didn't know this was a problem but I was consistently getting less than 3 hours sleep a night. Now I'm over 6!"* (R142, Fitbit Charge)

Facilitating Adoption of Healthy Sleep Habits

In addition to increasing duration, *sleep quality* plays an equally vital role in health and well-being. To improve sleep quality, one must address *modifiable behaviors*. Sensors that capture environmental factors such as acoustical noise, room temperature, and ambient light help users identify potential environmental factors that may be impacting their sleep. R146 (Sense) says "*... After realizing my apartment was too bright, I added curtains to darken the room.*" With respect to *sleep hygiene,* feedback representing sleep duration and sleep interruptions over time helps people better understand the impact of irregular sleep schedules. I11 says, *"... being able to look at the last couple of days and be like, ... It's more important to stay on a regular schedule so that you're not throwing yourself out of whack every couple of days by staying up until three one day and then trying to go to bed at 10 the next. The Fitbit gave me kind of a quantified view into my sleep schedule."*

Compared to other health conditions, sleep quality is highly subjective. The number of hours, modifiable behaviors, and changes in sleep hygiene to improve sleep quality may vary for every person. From using sleep sensing technology, S131 discovered that: *"I tend to operate on about half the amount of sleep as other people."* The feedback provided by sleep sensing technologies helped users understand how many hours they need to feel rested, thus providing a more objective measure: *"I sleep well (barely any restless sleep) but only sleep for 5 hours. I'm still rested, though"* (S116). Although people believe it is important to prioritize sleep and

[1] We refer to 'users' as product reviewers and survey and interview participants. We use RXX to refer to a quote from the review dataset, SXX for the survey responses, and IXX from the interview responses. The expert data will be referred by E [1-5].

taking proactive steps to address sleep quality, they also have their own beliefs and metrics on sleep. Personal beliefs can be carefully examined by incorporating a self-assessment framework. The SATED framework can be used to identify the quality of a person's sleep and personalize what adequate sleep means to a specific user.

Managing Sleep Disorders or Chronic Conditions that Affect Sleep

Feedback from sleep sensing also has the potential to manage sleep disorders. This data can help experts work with patients to identify and manage patients' sleep conditions. Sleep sensors can improve assessment and screening. Currently, to assess sleep, patients self-track their sleep to report on the five dimensions of the SATED scale: satisfaction, alertness, timing, duration, and efficiency (i.e., ease of falling asleep and returning to sleep). E1 says, "*I think whenever we have somebody comes in with a sleepiness complaint, we always want to get their sleep schedule to get a rough idea as to whether or not they're sleep deprived.*" The data provided by sleep sensors have the opportunity to provide longitudinal data for the five dimensions for the SATED scale and assess if the sleep issue is a chronic condition or just poor sleep habits: *"If it's accurate, I think it (a sleep sensor) can help identify if a sleep problem is present and really help parse it out between, is the problem not enough sleep? or is there a problem with the sleep quality itself?"* (E1).

Because PSG studies take place in a clinic, they do not represent a patient's natural sleep environment. They are therefore not well-suited to study non-physiological disruptors of sleep. Experts discussed the importance of understanding a patient's environment, how it impacts people's sleep, and how home sleep sensing can provide this type of information: "*... the advantage of these devices is that they're more ecologically valid because they're measuring sleep in the patient's typical sleep environment, and they're measuring it over a multiple of nights. Those are two huge advantages of this over sonography in the lab ... You're in a strange sleep environment, you're hooked up to all of these equipment…"* (E1).

Sleep sensing feedback can help patients determine the effectiveness of a treatment for a particular sleep disorder, such as using a CPAP (Continuous Positive Airways Pressure) machine for sleep apnea. Experts stated that patients struggle to adhere without longitudinal data on the effects of the treatment: "*Is the treatment working? Something beyond their subjective sense of whether or not they're better, but some objective data to show that their sleep quality is better" (E1).* Connecting sleep with treatment effectiveness is crucial, especially since over time, patients' motivation to adhere to treatment decreases: "*… they forget what it was like before, they get uncertain as to whether or not they're better. The device might be able to increase that certainty to motivate ongoing compliance with treatment"* (E1). For patients using a CPAP machine to prevent sleep apnea, sleep sensing feedback helps them determine if they are frequently moving throughout the night and if they need to adjust their CPAP machine: "*... My Fitbit One allows me to monitor how well I sleep and how often I wake up, so if my apnea ever worsens and I need an adjustment to the settings on my CPAP, I will know right away" (R1946, Fitbit One).* Furthermore, people with chronic conditions wanted to share the data from sleep sensors to better understand their prognosis: *"I think just really being able to go back to, say, my endocrinologist and my sleep doctor and say, 'Look it. Here's a pattern of restless legs. In this particular phase of my menstrual cycle, my restless legs is more intense or less intense, so maybe we need to adjust the pharma, or we need to adjust other aspects of treating my legs and treating my sleep disorder based on where I am in the month'" (I4).*

Related to I4's comment above, people track sleep to find correlations between sleep and their other health conditions, not only track to manage sleep disorders: "*I had prostate cancer and have a frequent urination problem. I wanted to know how many times that I got up as well as deep and light sleep*" (S56). In cases where the treatment for a condition involved medication, users wanted to monitor the effects of medication on sleep: "*…it's especially helpful because it helps me track a side effect of a medication I'm tracking (insomnia) to help inform decisions on whether or how to change the medication*." (S75). Later in the paper, we present opportunities to support scientific self-experimentation.

SLEEP SENSORS AS BARRIERS

Although sleep sensing devices provide useful and objective feedback that is beneficial to users, our analysis identified areas of improvements and opportunities incorporate evidence-based strategies to sleep sensing feedback. Feedback from the sleep sensors tends emphasize estimating the number of hours users spent in various sleep stages and assessing sleep quality using computed, single-point measures such as Sleep Efficiency or Sleep Score. *Sleep efficiency* is the ratio of total sleep time to time spent in bed [3]. However, variations in hardware sensitivities result in small movements classified as "restlessness". This results in variations in computed sleep efficiency scored across devices. Furthermore, the algorithms used to compute these values are proprietary and not made available to the public.

As discussed in our background section, people cannot voluntarily control the number of hours spent in a particular sleep stage. Current feedback tends to focus on these sleep stages. or neurotypical people, a breakdown of the time spent in the different sleep stages is not helpful feedback for improving sleep quality. The focus on sleep stages leads to users developing inaccurate mental models of how current sleep sensors work and what it means to get good quality

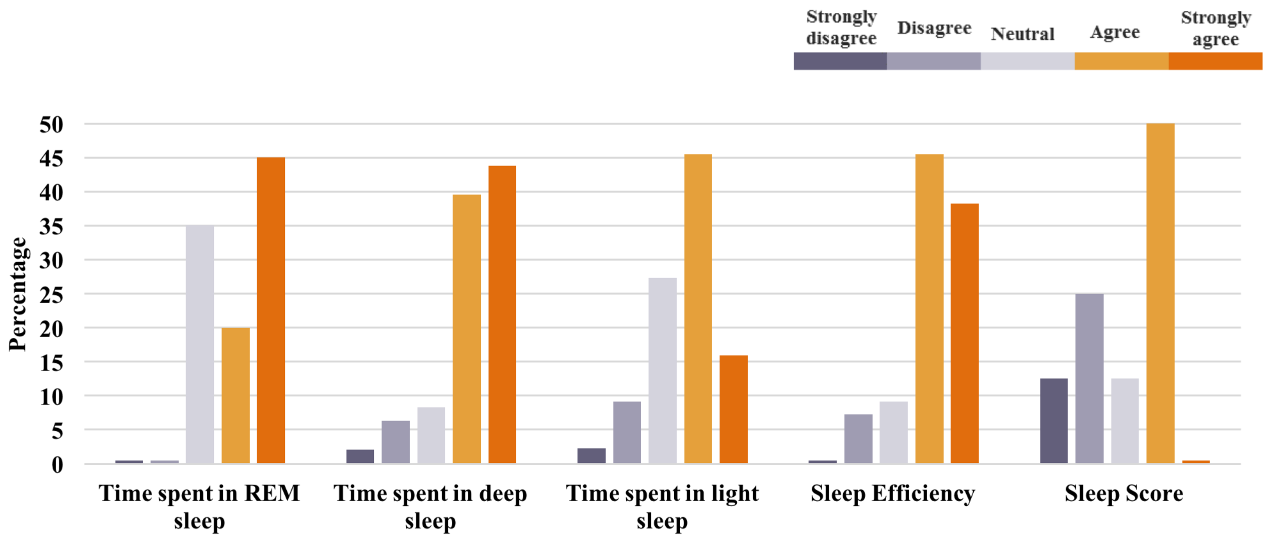

Figure 1. Percentage of survey respondents who believed their sleep quality was related to various sleep metrics

sleep. Feedback on sleep stages distracts users from focusing on adapting *modifiable behaviors* to improve *sleep hygiene,* which really could have a positive impact on their health. SATED's dimensions combine subjective perspectives of sleep quality such as *Satisfaction* along with objective measures such as *Sleep Efficiency*. These dimensions provide a more holistic representation of sleep quality [5].

Inconsistency in Sleep Quality Inference

Emphasizing the SATED dimensions of sleep, E5 said: *"I think for a healthy night's sleep, the person is getting adequate sleep for their developmental needs and has a low number of interference with that sleep during the night. They wake feeling fairly well rested and satisfied with their night's sleep. They are able to maintain good alertness during the day and not feel fatigued."* Experts noted that there is a trait variability to sleep and some measures are very specific to each person, "*Some people can … have poor quality sleep and not really feel many ill effects from it. Other people can have just minor decrements in those sleep factors that have a pretty big impact on sleep quality (E1)*." Satisfaction, one of the five SATED dimensions, is highly subjective and specific to each person. This measure can be gathered from getting users' own perception of their sleep quality and how rested they feel on waking up.

To provide feedback to users about their sleep, commercial sleep sensors often focus on determining objective measures such as *sleep efficiency*, *sleep latency*, and the different stages of sleep. Commercial sensors tend to focus less on subjective measures. However, the focus on objective measures led many users to have a broken mental model of what sensors can infer and what information is useful to address sleep concerns. R2066 (Fitbit One) says, "*The lack of explanation as to the formula/algorithm that lead to the results are very maddening. What does the 94% effective sleep rating actually mean?"*

Focusing solely on a *single* objective value such as sleep score distracts users from adopting *modifiable behaviors* proven to help improve sleep, such as maintaining a regular bedtime and wake time every day or identifying environmental factors such as ambient light or sound that may be disrupting their sleep. Users were, instead, focused on trying determine what sleep score to target. "*I would also love some benchmark information or links to what is a good target for sleep efficiency?"* (R565, smartphone app).

Mismatch between score and user perception

The confusion on the feedback is further compounded when the sleep scores do not correlate with users' own perceptions of sleep satisfaction. R7 (Beddit) reports, "*After using it for a week during which I slept as poorly as I normally do, I earned not one, but two perfect 100 sleep scores as well as a 99 last night, where I tossed and turned and woke twice from nightmares."* R7 felt they had slept badly for three nights, but the feedback reflected a high sleep score and therefore a good night's sleep, which was opposite of what was expected. Users in this situation struggled to determine what the sleep score and percentages represented.

In our survey, 50% of our respondents agreed or strongly agreed that their sleep score or sleep efficiency was related to their sleep quality (Figure 1). However, these scores may not necessarily provide users an accurate picture of their sleep quality because sleep efficiency scores vary on hardware specification and sensing sensitivities. "*The problem is that it is nowhere near sensitive enough on normal and way too sensitive on sensitive setting. I had a restless night the night I had it on normal, waking multiple times, and it recorded 15 minutes of restless sleep and no wake times. The next night, on sensitive, I had much better sleep, and it recorded only 3 1/4 hrs. of sleep and the rest waking or activity!"* (R10, Fitbit One).

One benefit of sleep scores and sleep efficiency numbers is the potential to provide longitudinal feedback over several weeks. E1 suggested, "*I would focus more on the trends than I would on just the night to night scores. They [consumers] need to approach it with a dose of skepticism and then follow the trends more than the night time score.*" Another suggestion from E5 was, "*There needs to be something tracked alongside that, whether it's their subjective rating of sleep quality or their sleep habits. Something that actually is an actionable item. I don't think just giving someone a number, like you slept 440 minutes last night, is enough of the right kind of feedback to lead to changes.*"

Placing Undue Emphasis on Sleep Stages

Reviewers of sensing devices place high value in sleep sensing devices that can infer sleep quality based on sleep stages. One review said: "*The only thing it doesn't really do that my mom's Jawbone does, is that it doesn't tell me about my sleep cycles (stage 1/stage 2/REM etc.) it just tells me when I'm asleep, awake, and 'restless'*" (R784, Fitbit Charge). Survey respondents reflected the same perspective, considering sleep stages to be representative of sleep quality (see Figure 1). More than 60% agreed or strongly agreed that time spent in specific stages such as REM, deep was related to sleep quality. I8 said, "*The Jawbone ... gave me light sleep and the deep sleep separation and, of course, the awake states. Since I didn't have many awake states during the night, which was good, I only pretty much had the amount of hours that I was having light sleep and deep sleep ... What I eventually understood was that I was having not enough deep sleep*".

The sleep experts we interviewed believe people generally misunderstand the relationship between sleep quality and sleep stages. "*I think that for a layperson, that word deep implies more restful sleep and so I would imagine that's the only word that they would think about…" (E5).* Although there has been evidence that time spent in specific sleep stages, such as REM, helps with memory consolidation [16], the focus on sleep stages has two concerns:

First, to accurately determine sleep stages such as REM, a device would need to capture brain waves and eye movement, as is done in a clinical PSG study. Experts are skeptical of the accuracy of sleep stages inferred from a combination of movement, breathing, and heart rate data as a proxy to EEG and EOG data, which commercially available products currently provide. Like sleep scores, every sleep sensing device has its own proprietary algorithm to determining sleep stages. Sleep users began to notice the inconsistencies: "*I tracked a few days of sleep with both devices and the results are in the Fuse vs UP3 Sleep comparison… It is very apparent that they have completely different ideas about what light and deep sleep means*" R195 (Jawbone Up3). Figure 2 shows R195's comparison of the sleep stage inference provided by two different sensors on the same night. MioFuse inferred R195 had spent 64% of sleep time in deep sleep, while Jawbone Up3 inferred 16.5% of sleep time in deep sleep. Similarly, participants that had previously completed a PSG sleep study noticed the difference between what the clinicians reported and what their devices were reporting. I8 said: "*The band gave me too much deep sleep when compared with the actual exam …*" P8.

Second, there is limited research on how a person can take actionable steps towards affecting the number of hours spent in a particular sleep stage. "*I'm not aware of anything necessarily that can increase REM sleep. Many medications, particularly psychiatric medications, can affect sleep architecture sum. What the effect that a medication would have on any given individual is there's probably some variability to that" (E1).* The feedback on sleep stages provided by commercial sensors promotes incorrect mental models on what these sensors can infer and how these stages actually impact sleep quality. "*It's … a lot of useless feedback… what I would like to see more clearly is really that the lay public understands there is no scientific basis for these numbers" (E5)*. Experts instead want feedback to focus on issues people actually have control over, such as *sleep hygiene* and *modifiable behaviors*. *"That feedback might be if you're not getting a lot of deep sleep, they might interpret that as a poor night of sleep and think that their getting bad sleep. Again, it doesn't really lend itself to being actionable; so what are they supposed to do about that necessarily?" (E5).* Experts expressed a desire to help users understand what these sensors can actually infer about sleep: "*I think it [feedback from the device] needs to be scaled back into what we can expect them [users] to realistically understand and do something about" (E5)*.

Making unscientific correlations based on sleep stages

In line with previous work [34], some users conducted a self-experiments on their sleep and make correlations from their findings. However, the focus on sleep stages led users to make unscientific correlations between daily behaviors and specific sleep stages. For example, R8 correlated deep sleep with a late meal: "*Last night, I had a late meal and that wasn't the best for me… I wake up this morning and see that… I got very little [deep] sleep compared to my average*

Figure 2. R195's comparison of one nights' sleep stage feedback from two separate devices. (Left) Mio Fuse infers 64% of sleep spent in deep sleep. (Right) Jawbone Up3 infers 16.5% deep sleep (shown as 1hr 22mins out of 8hrs and 14mins).

…. I always knew it was bad to have a big meal before bed, but now I actually have the data to back it up" (R8, Jawbone Up3). For good sleep hygiene, avoiding large late night meals is recommended, but correlating with deep sleep stage might scientifically incorrect. Users may not have understood is that a big meal might cause restless sleep, and therefore cutting down on large meals might have allowed them to have less restlessness, leading to better sleep quality overall.

In some cases, such inferences can sometimes lead to actions that can be potentially detrimental to health. For instance, users like R168 (Zeo) experimented with medication in an attempt to increase the duration of REM: "*I can also see a day-to-day trend of how the iodine supplement I have just re-started is helping my sleep. It has increased my deep and REM and I feel better, even though my overall sleep time is not that much more."*

Like Yang et al., our data reflects that users were making unscientific correlations and lacked support to conduct self-experiments that can identify causal inferences [34]. Users want the means to self-experiment: *"I basically want a sleep tracker that has three or four variable knobs ... 'You had three really good nights' sleep, and here are the variables that you played with'"* (P4). Many people who track sleep additionally wanted to add notes to add context to their sleep to help assess what is affecting their sleep. "*I had some theories about what was causing me to sleep well or not and I had to track those in a different app. I would have been nice to track them in the same [sleep app] to help me see trends*" (S86). These types of information not only provide context to people's sleep quality, but can also help address the subjective aspect of sleep quality: *"[I would like to track] Number of sleep hours and quality [to] cross-check against what I feel during the day"* (S177).

Supporting self-experimentation as a scientific process will help users better identify personal triggers affecting their sleep [13]. Our data reflects that many people who track sleep want to test a hypothesis they have about their sleep quality. Users want to test a variety of factors that could be affecting their sleep and related to *sleep hygiene*. These include medication, stress, diet, aspects of their environments, and other aspects of their health including time within a menstrual cycle.

DISCUSSION AND DESIGN RECOMMENDATIONS

Our findings show that sleep sensors increase awareness in prioritizing sleep and help users address *modifiable behaviors* and their *sleep hygiene*. On the other hand, current feedback focuses on sleep metrics people do not have control to directly change (e.g., time in sleep stages) and this distracts users from focusing on aspects they have control over that improve sleep health. We now provide design recommendations for on the feedback sleep sensing technology can provide to users. Our guidelines draw from our results, and connect to evidence-based strategies that focus on *sleep hygiene, modifiable behaviors,* and the SATED framework for good sleep quality. Our design recommendations aim to mitigate the tension between user-driven goals, expert recommendations, and the sensing limitations of current commercial sleep sensing technologies.

Include Subjective Sleep Quality Assessment

Sleep quality is inherently subjective. A poor night's sleep for one person can be satisfactory and rested sleep for another person. Furthermore, the effects of a poor night's sleep vary from person to person. Sleep quality self-assessments is often used by clinicians to assess the severity of sleep-related issues [6]. We recommend that subjective self-assessments be incorporated as part of the analysis that sleep sensing technologies execute to calculate people's sleep quality for a given night. To assess subjective sleep quality, we recommend incorporating the five dimensions of the SATED framework, such as Satisfaction and Alertness with more objective measures such as Efficiency, Timing, and Duration. Incorporating users' subjective assessment should also be integrated into algorithms that personalize feedback or calculate a sleep score. Furthermore, self-assessments should be used to learn and assess which types of *modifiable behaviors* worked best in helping a user improve their sleep over time.

Contextualize Sleep Quality with Journaling

The current state of feedback does not support long-term perspectives on sleep trends. We recommend sleep technologies support long-term visualizations of bed time, wake time, and sleep duration. Long-term visualizations can provide a richer and more holistic view on variability compared to daily feedback focused on sleep stage. Viewing long-term trends will help users address aspects of sleep hygiene related to maintaining a consistent bedtime and wake time.

We also recommend allowing users to log major life events such as job changes, the birth of a newborn, or the start or end of college semesters. These logs will help users identify events in their daily life that might be impacting their sleep. Integrating long-term trends and life logs will contextualize lifestyle changes and help users to assess and focus on aspects that positively or negatively affect their sleep. Contextualizing sleep data and supporting sleep self-assessments will also help physicians diagnose what is affecting a patient's sleep. Physicians and sleep clinicians could use longitudinal sleep data, self-assessments of sleep, and information on sleep routines to gather a more holistic view of a patient's health.

Focus on Actionable Feedback

We find that feedback helps users connect their daytime behaviors, pre-bedtime behaviors, and environmental conditions of their bedrooms to their sleep quality, which in turn helps them act accordingly. This confirms previous research [9], [20]. To help people draw meaningful conclusions from sleep data, designs need to develop ways of presenting feedback to users beyond correlational graphs. Moreover, support for more systematic tests such as through self-experiments [13,19], can make this process less frustrating than simple trial and error. Systems can allow people to test behaviors such as the timing of caffeine consumption or installing noise and light blocking curtains. Reviewing these experiences will help people identify the impact of that change on their sleep duration, timing, or satisfaction.

Finally, technologies can promote good sleep health by delivering timely behavior change suggestions or actions, such as turning off electronic devices close to bedtime or automatically dimming lights at night. To provide these suggestions, experts recommended a two-week period of data collection before offering personalized suggestions. This two-week period would serve as a baseline to understand daily behaviors.

Give Feedback in Ranges, not Single Point Values

Currently, sleep scores and sleep efficiency are presented as a precise, single point value, such as 92%. Clinically, these metrics are calculated using brain waves to identify the onset of sleep and arousal. However, current sleep sensing technologies infer these same metrics based on physiological signals such as body movement, breathing, and heart rate. Physiological signals cannot currently accurately differentiate between awake in bed and asleep in bed. This substitution in sensing modality introduces a certain level of inaccuracy.

We recommend reporting sleep scores in ranges rather than single values to avoid false precision. Similar to users' perceptions of changes in weight by several pounds [22], daily fluctuations in sleep scores or sleep efficiency does not imply drastic changes in sleep quality and only causes users to be unnecessarily concerned. To improve quantitative metrics, systems need to incorporate self-assessments on sleep satisfaction using clinically valid frameworks like SATED. Systems can provide sleep score ranges instead of a single-point value, based on sleep sensing data and self-assessments. These ranges will focus on overall sleep quality, move away from making sleep quality a single value. Doing so will embrace the inherent sensing inaccuracies without compromising on the metrics.

Increase Transparency in Formulae and Algorithms

Our results indicate that users have broken mental models about how sleep sensing technologies work. Publicly documenting the algorithms and formulae used to calculate sleep score in a straightforward manner can bridge the gap between sensing capabilities and users' expectations. Work by Lim et al. [26] suggests explanations can help improve the intelligibility of context-aware intelligent systems. This can be applied to sleep sensing technologies to equip users to better interpret sleep feedback results in a meaningful way.

Sleep quality feedback may appear to have an understated role in affecting people's health-related decisions. In a culture where people are encouraged to do more with less sleep, presenting users feedback that misleads them to make unscientific correlations could lead to practices which are potentially detrimental to health. Therefore, we believe that tool makers have an ethical and social obligation to avoid accidentally promoting false precision and to avoid non-actionable feedback that steers users' focus away from making healthy choices. Users should to be able to act on their own health and use sleep sensing technologies to experiment and determine what makes them healthy. We hope tool makers will continue to innovate on new metrics for sleep health beyond what is currently possible in the clinic. We also want to emphasize that new, experimental features, such as new sleep measurements, should be clearly labeled as experimental. Tool makers should ensure tools are designed primarily with the shared goal of improved sleep health, rather than marketing new features which may not be scientifically validated.

CONCLUSION & FUTURE WORK

Sleep sensing technology provides people with rich information about their sleep. These technologies help people learn about their sleep habits and how to improve sleep health by providing feedback on their sleep. However, certain types of feedback lead users to develop broken mental models about what sleep sensors have the ability to sense and distract users from habits and behaviors that are actually affecting their sleep. Across different commercial sensors, the metrics used to give sleep quality feedback vary and sometimes conflict with clinical standards, potentially undermining people's ability to improve their sleep. The focus on sleep stages, which are difficult to infer from the set of sensors sleep technologies use, leads users to focus on aspects of their sleep difficult to control, such as REM sleep. This focus derails users from focusing on modifiable behaviors and sleep hygiene.

Our findings provide a review of the state of current sleep sensing technology from the perspective of users and sleep experts. We suggest future tools display data in ranges rather than single point values. Tools should focus on actionable feedback that integrates modifiable behaviors. Designers should make the algorithms behind sleep sensing devices transparent. Sleep self-assessments can help personalize and contextualize sleep sensing feedback. We hope this work leads to new designs which better align sleep sensing technologies with user's needs and integrate evidence-based frameworks and strategies created by the sleep research community.

ACKNOWLEDGEMENTS

We would like to thank Katie Siek and our reviewers for providing feedback to improve this paper. This research was reviewed and granted exemption by the University of Washington's Institutional Review Board. The work was supported by National Science Foundation #1344613, the University of Washington Research Innovation Award, and the Intel Science & Technology Center on Pervasive Computing.

REFERENCES

1. Kirstine Adam. 1980. Sleep as a Restorative Process and a Theory to Explain Why. *Progress in Brain Research* 53: 289–305. http://doi.org/10.1016/S0079-6123(08)60070-9

2. Jared Bauer, Sunny Consolvo, Benjamin Greenstein, et al. 2012. ShutEye. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, ACM Press, 1401. http://doi.org/10.1145/2207676.2208600

3. Richard B Berry, Rita Brooks, Charlene E Gamaldo, et al. The AASM Manual for the Scoring of Sleep and Associated Events.

4. Hugh Beyer and Karen Holtzblatt. 1997. *Contextual design: defining customer-centered systems*. Elsevier. Retrieved September 21, 2016 from https://books.google.com/books?hl=en&lr=&id=JxQaQgOONGIC&oi=fnd&pg=PP2&dq=beyer+and+holtzblatt+affinity+diagramming&ots=-S-_qGc-v3&sig=RZ0DfmsXqCjrsTN2qAr634BlWkk

5. Daniel J Buysse. 2014. Sleep health: can we define it? Does it matter? *Sleep* 37, 1: 9–17. http://doi.org/10.5665/sleep.3298

6. Daniel J. Buysse, Charles F. Reynolds, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. 1989. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research* 28, 2: 193–213. http://doi.org/10.1016/0165-1781(89)90047-4

7. Eun Kyoung Choe, Sunny Consolvo, Nathaniel F. Watson, and Julie A. Kientz. 2011. Opportunities for computing technologies to support healthy sleep behaviors. *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, ACM Press, 3053. http://doi.org/10.1145/1978942.1979395

8. Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: Low-burden, Self-monitoring Technology for Capturing and Reflecting on Sleep Behaviors. *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*: 121–132. http://doi.org/10.1145/2750858.2804266

9. Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, ACM Press, 121–132. http://doi.org/10.1145/2750858.2804266

10. Harvey R Colten, Bruce M Altevogt, and Institute of Medicine (US) Committee on Sleep Medicine and Research. 2006. *Sleep Disorders and Sleep Deprivation*. National Academies Press (US). http://doi.org/10.17226/11617

11. Sunny Consolvo, Ryan Libby, Ian Smith, et al. 2008. Activity sensing in the wild. *Proc. CHI '08, ACM Press*, August 2015: 1797. http://doi.org/10.1145/1357054.1357335

12. Felicia Cordeiro, Daniel A. Epstein, Edison Thomaz, et al. 2015. Barriers and Negative Nudges. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, 1159–1162. http://doi.org/10.1145/2702123.2702155

13. Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, et al. 2016. SleepCoacher. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*, ACM Press, 347–358. http://doi.org/10.1145/2984511.2984534

14. Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A lived informatics model of personal informatics. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, ACM Press, 731–742. http://doi.org/10.1145/2750858.2804250

15. Jon Froehlich, Shwetak Patel, James A. Landay, et al. 2012. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, ACM Press, 2367. http://doi.org/10.1145/2207676.2208397

16. Lisa Genzel, Marijn C.W. Kroes, Martin Dresler, and Francesco P. Battaglia. 2014. Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes? *Trends in Neurosciences* 37, 1: 10–19. http://doi.org/10.1016/j.tins.2013.10.002

17. Lauren Hale. 2014. Inadequate Sleep Duration as a Public Health and Social Justice Problem: Can We Truly Trade Off Our Daily Activities for More Sleep? *SLEEP*. http://doi.org/10.5665/sleep.4228

18. Leah A Irish, Christopher E Kline, Heather E Gunn, Daniel J Buysse, and Martica H Hall. 2015. The role of sleep hygiene in promoting public health: A review of empirical evidence. *Sleep medicine reviews* 22: 23–36. http://doi.org/10.1016/j.smrv.2014.10.001

19. Ravi Karkar, Jasmine Zia, Roger Vilardaga, et al. 2015. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association*.

20. Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, et al. 2012. Lullaby. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 226. http://doi.org/10.1145/2370216.2370253

21. Matthew Kay, Tara Kola, Jessica R Hullman, and

Sean A Munson. 2016. When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI '16)*. http://doi.org/10.1145/2858036.2858558

22. Matthew Kay, Dan Morris, Mc Schraefel, and Julie a. Kientz. 2013. There's No Such Thing as Gaining a Pound: Reconsidering the Bathroom Scale User Interface. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*: 401–410. http://doi.org/10.1145/2493432.2493456

23. Sharon Keenan and Max Hirshkowitz. 2010. Monitoring and Staging Human Sleep. *Principles and Practice of Sleep Medicine: Fifth Edition*: 1602–1609. http://doi.org/10.1016/B978-1-4160-6645-3.00141-9

24. Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, ACM Press, 557. http://doi.org/10.1145/1753326.1753409

25. Life Sciences & Transformational Health. *Global Market Study on Sleep Aids: Sleep Apnea to Witness Highest Growth by 2020*. Retrieved from http://www.persistencemarketresearch.com/market-research/sleep-aids-market.asp

26. Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. *Why and why not* explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, ACM Press, 2119. http://doi.org/10.1145/1518701.1519023

27. Wanyu Liu, Bernd Ploderer, and Thuong Hoang. 2015. In Bed with Technology. *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction on - OzCHI '15*, ACM Press, 142–151. http://doi.org/10.1145/2838739.2838742

28. Jun-ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I Hong. 2014. Toss " N " Turn : Smartphone as Sleep and Sleep Quality Detector. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*: 477–486. http://doi.org/10.1145/2556288.2557220

29. Hawley E. Montgomery-Downs, Salvatore P. Insana, and Jonathan A. Bond. 2012. Movement toward a novel activity monitoring device. *Sleep and Breathing* 16, 3: 913–917. http://doi.org/10.1007/s11325-011-0585-y

30. Tauhidur Rahman, Alexander T. Adams, Ruth Vinisha Ravichandran, et al. 2015. DoppleSleep. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, ACM Press, 39–50. http://doi.org/10.1145/2750858.2804280

31. Avi Sadeh and Christine Acebo. 2002. The role of actigraphy in sleep medicine. *Sleep Medicine Reviews* 6, 2: 113–124. http://doi.org/10.1053

32. Edward J Stepanski and James K Wyatt. 2003. Use of sleep hygiene in the treatment of insomnia. *Sleep medicine reviews* 7, 3: 215–25. http://doi.org/10.1053/SMRV.2001.0246

33. Rayoung Yang and Mark W. Newman. 2013. Learning from a learning thermostat. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*, ACM Press, 93. http://doi.org/10.1145/2493432.2493489

34. Rayoung Yang, Eunice Shin, Mark W. Newman, and Mark S. Ackerman. 2015. When fitness trackers don't "fit." *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, ACM Press, 623–634. http://doi.org/10.1145/2750858.2804269