SleepExplorer: A Visualization Tool to Make Sense of Correlations between Personal Sleep Data and Contextual Factors

Zilu Liang[1,2,6] **・Bernd Ploderer**[1,3] **・Wanyu Liu**[1,4,5] **・Yukiko Nagata**[6,7] **・James Bailey**[1] **・Lars Kulik**[1] **・ Yuxuan Li**[1]

[1] University of Melbourne, Parkville VIC 3010, Australia

[2] National Institute of Advanced Science and Technology, 2-3-26 Aomi, Koto-Ku, Tokyo 135-0064, Japan

[3] Queensland University of Technology, Brisbane QLD 4001, Australia

[4] Telecom ParisTech & CNRS (LTCI), Université Paris-Saclay, 75013, Paris, France

[5] Université Paris-Sud & CNRS (LRI), Inria, Université Paris-Saclay, 91405, Orsay, France

[6] The University of Tokyo, 3-8-1 Komaba, Meguro-Ku, Tokyo 153-8902, Japan

[7] Princeton University, NJ 08544, United States

Email of correspondent author: zilu.liang@aist.go.jp

Abstract Getting enough quality sleep is a key part of a healthy lifestyle. Many people are tracking their sleep through mobile and wearable technology, together with contextual information that may influence sleep quality, like exercise, diet and stress. However, there is limited support to help people make sense of this wealth of data, i.e., to explore the relationship between sleep data and contextual data. We strive to bridge this gap between sleep-tracking and sense-making through the design of *SleepExplorer*, a web-based tool that helps individuals understand sleep quality through multi-dimensional sleep structure and explore correlations between sleep data and contextual information. Based on a two-week field study with 12 participants, this paper offers a rich understanding on how technology can support sense-making on personal sleep data: SleepExplorer organizes a flux of sleep data into sleep structure, guides sleep-tracking activities, highlights connections between sleep and contributing factors, and supports individuals in taking actions. We discuss challenges and opportunities to inform the work of researchers and designers creating data-driven health and wellbeing applications.

Keywords Sleep・Health・Personal informatics・Self-tracking・Sense-making・HCI

1 Introduction

There is a growing research interest in sleep tracking in the domain of personal informatics [41]. HCI researchers have designed various systems to support the collection and analysis of personal information on sleep [38, 50], as well as factors that contribute to good sleep, like exercise [15] and emotions [30]. At the same time, a large number and variety of commercial products has become available that allow people to track their sleep in their own homes. These products include mobile apps (*e.g., SleepCycle*), wearable technology (*e.g., Fitbit*) and embedded devices (*e.g., Beddit*).

While it is now easy to collect sleep data, making sense of this data and the factors that may influence the quality of sleep remains a challenge. Firstly, we do not consciously experience our sleep. Therefore, we cannot know what exactly is going on while we are sleeping. Secondly, sleep is multi-dimensional and could be characterized by many metrics, such as the time it takes to fall asleep, length of sleep, and how fresh one feels after waking up. Hence the simplified definition 'sleep efficiency' or 'sleep quality' in most consumer sleep trackers is questionable. Thirdly, the quality of sleep is influenced by a number of contextual factors: psychological and physiological states like mood and stress, lifestyle factors like exercise, and environmental factors like room temperature and exposure to digital devices [17, 34]. Many users of commercial sleep-tracking products lack the expertise on these issues as well as tools that allow them to explore possible relationships between the quality of their sleep and contextual factors [43].

The aim of this research was to support sleep-trackers in making sense of their data, i.e., in analyzing relationships between their sleep data and contextual factors that may impact their sleep. We designed *SleepExplorer*, a web-based tool that imports various self-tracking data from commercial sleep-trackers (Fitbit) and online diaries (Google Forms), visualizes this data, and shows correlations between sleep data and contextual factors. In order to facilitate sense-making on the correlations, *SleepExplorer* also helps users understand sleep quality based on the concept of sleep structure, as well as provides guidance on how to track contextual factors. Through a two-week field study with 12 participants, we found that *SleepExplorer* helps people to better understand their sleep and discover novel relationships between sleep data and contextual factors.

In what follows, we first summarize related work in the domain of personal sleep informatics, sleep structure and sleep analysis. We then describe the implementation of *SleepExplorer* and the design of the field study. In section 4 we present our findings from the field study. Section 5 discusses challenges and opportunities for designing data-driven health and well-being tracking technology. The contributions of this study are twofold: (1) we present a novel tool *SleepExplorer* that redefines sleep quality by introducing sleep structure and emphasizes the individualism of each person's sleep, and (2) the empirical study generated considerations and opportunities that help inform the work of researchers and designers in sleep data analysis and personalized visualizations.

2 Related Work

2.1 Personal Sleep Informatics

Personal informatics refers to the tools that help people collect personal information for the purpose of self-monitoring and self-reflection [41]. By its definition, personal sleep informatics is the tools that help individual monitor and reflect on sleep. Up to date, a large number of sleep technologies have been developed in industry. These tools provide users with information about how long they sleep, how well they sleep (sleep efficiency score), the stages they sleep through, how to fall asleep and wake up with optimized freshness, and how to promote healthy sleep habits through sleep coaching tips. A detailed review of consumer sleep technologies can be found in [73].

Research Aim	Sleep Studies
Monitoring sleep and determining sleep quality	Sleepful [38]; Toss 'N' Turn [50]; iSleep [28]; Poster [57]
Measuring sleep stage and sleep duration	Best Effort Sleep Model [13]; SleepHunter [25]
Recognizing sleep patterns	Sleep pattern analysis system [48-49]
Understanding sleep environment	Lullaby [34]; Combining wearable environmental sensing [10]
Improving sleep quality and promoting sleep health	ShutEye [8]; The Wearable Lullaby [21]; EZwakeup [31]; Opportunities for computing technologies to support healthy sleep behavior [17]
Reliving sleep disorders	SNORES [27]; ZZZoo Pollows [66]; Sleep apnea detection [20]
Exploring relationships to contributing factors	SleepTight [75]
Others	Sleep quality prediction: SleepMiner [6]; Social alarm clock: Somnometer [59]; Nap supporting system [54]; Sleepwalk supporting system: Sleepstellar [33]

Table 1. A review of sleep studies in HCI.

At the same time, HCI also has devoted significant efforts in mapping personal informatics into sleep research. Table 1 summarizes recent sleep studies, indicating that various aspects of sleep are being supported, from introducing novel approach to monitor sleep [38] to supporting sleep related problems such as apnea detection [20]. However, apart from *SleepTight* [75] that is a low burden, self-monitoring application for capturing sleep and limited contributing factors, little is studied on how to make sense of sleep data through contextual information. SleepExplorer strives to collect various sleep contributing factors based on medical evidence to help individuals discover the relationships between their sleep data and these factors.

2.2 Sleep Structure

There are different methods to measure human sleep. Subjective methods include Pittsburgh Sleep Quality Index (PSQI) [11] instrument that is widely used in clinical settings for rough evaluation of sleep quality. Objective method refers to the analysis of sleep characterized by a set of metrics called sleep structure [31-32]. Some of the sleep metrics could only be measured using polysomnography in sleep clinics, while others could be monitored using commercial sleep trackers at acceptable accuracy level. Considering the practical tractability of the metrics, we used the following sleep metrics to quantify sleep quality in the study:

- Minutes asleep (MASL)
- Minutes awake (MAWK)
- Number of awakenings (NAWK)
- Minutes to fall asleep (MTFA)
- Sleep efficiency (SE)

Among these metrics, SE is determined by the ratio between MASL and total time in bed [74]. Most sleep trackers have adopted this simplified definition of sleep quality to inform users how good or bad their sleep is. However, from medical perspective, other important aspects are missing to evaluate a person's sleep [73] and likewise users find it hard to interpret this single percentage [43]. Therefore, SleepExplorer endeavors to provide sleep structure information and to help individuals discover what 'sleep quality' means to them.

2.3 Sleep Analysis

Since sleep is affected by many activities an individual does during the day and before bedtime, contextual information is of particular importance to understand one's sleep and to promote sleep health. Choe et al.'s [17] large-scale survey demonstrated that there are a number of sleep disruptors that prevent individuals from getting a good sleep. For instance, worries, temperature, sleep partner, etc.

Identifying these factors helps individuals find reasons for their bad sleep and raise awareness of healthy sleep behavior. However, the sleep research listed in Table 1 rarely takes contextual information into consideration; most of them answer the question *'what happened'* rather than '*why it happened?*' [35]. Liu et al.'s [43] sleep study illustrated that current users of commercial sleep tracking technologies encountered difficulty in interpreting data without being provided with context and they found no proper tool to integrate data from multiple sources, let alone conducting further analysis.

Previous studies on overall health and wellbeing confirmed the benefit of integrating contextual information in health data analysis [9]. Two typical studies on health self-management proposed health interweaving and provided connections among various health indicators. *Salud!* [47] aggregated data from multiple sensors and contextual streams while *Health Mashups* [9] presented statistical patterns between wellbeing data (weight, sleep, pain, and mood) and context (step count, calendar data, location, weather, and food intake) in natural language. Both research invested sufficient efforts in understanding and incorporating contextual information into self-tracking practices and investigated personal health at an overall level, yet neither of them is sleep-oriented.

Previous studies on overall health and wellbeing confirmed the benefit of integrating contextual information in health data analysis [9]. Two typical studies on health self-management proposed health interweaving and provided connections among various health indicators. *Salud!* [47] aggregated data from multiple sensors and contextual streams while *Health Mashups* [9] presented statistical patterns between wellbeing data (weight, sleep, pain, and mood) and context (step count, calendar data, location, weather, and food intake) in natural language. Both research invested sufficient efforts in understanding and incorporating contextual information into self-tracking practices and investigated personal health at an overall level, yet neither of them is sleep-oriented. *SleepTight* [75] was designed to allow users to capture contextual factors that contribute to sleep quality, like exercise and caffeine consumption. While *SleepTight* allowed users to compare these data between nights with good, neutral and poor sleep, it did not provide correlations between contextual and sleep data that confirm whether there is indeed a statistically significant relationship in the data.

In summary, SleepExplorer builds on these studies and is designed under three aims:
(1) To offer sleep-trackers insights into correlations between sleep data and contextual factors to guide behavior change;
(2) To guide individuals with medical evidence on how to track factors that may affect their sleep ;
(3) To help individuals understand sleep structure and define good sleep within their personal context.

3 Research Design

3.1 Design and Implementation of SleepExplorer

To achieve our aims, we designed SleepExplorer, a web-based tool that helps individuals understand sleep quality and to explore correlations between sleep data and contextual information. SleepExplorer was implemented on Microsoft Azure cloud based on ASP.NET Model-View-Controller (MVC) framework. SleepExplorer automatically retrieves users' data from Fitbit public API via OAuth protocol upon their authorization. The retrieved data is analyzed according to the following procedure:

- Data Cleaning: missing entries and obviously wrong data entries (e.g., sleep duration = 0 or minutes sedentary = 0) were removed.

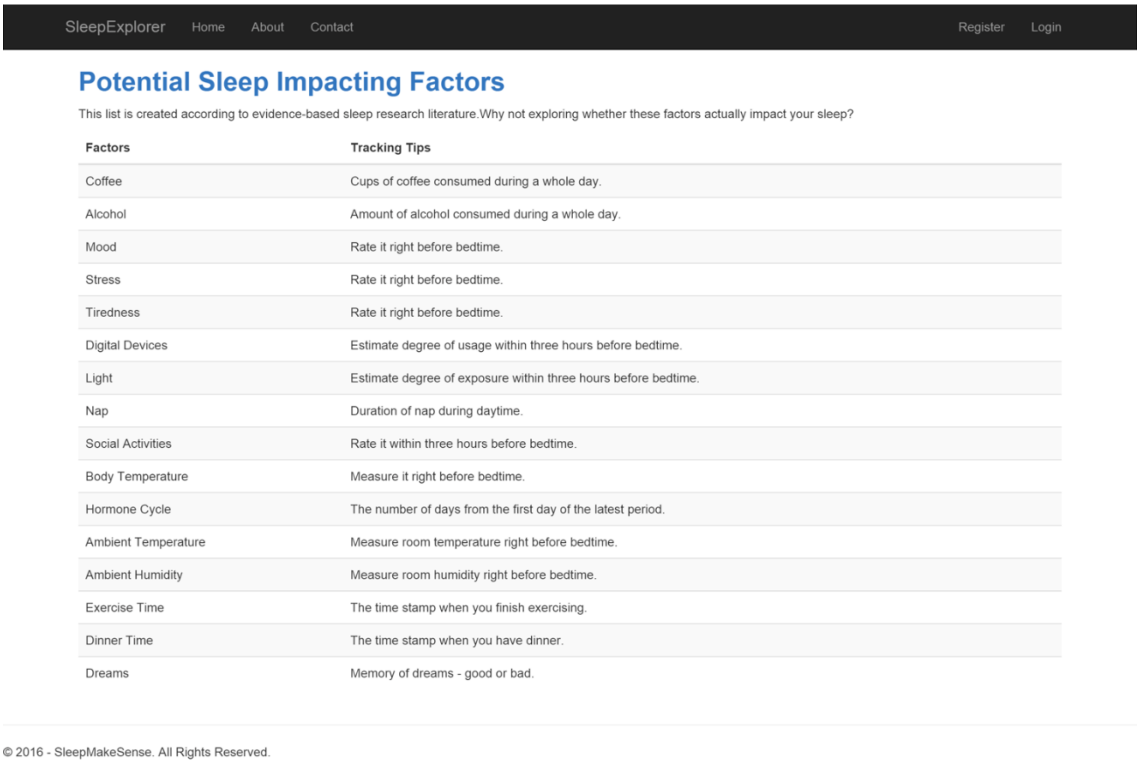
- Data Preprocessing: all time stamps (i.e., coffee time, nap time, dinner time and exercise time) were converted to numerical variables, e.g., 7:00 and 18:30 were converted to 700 and 1830 respectively.

Correlation Analysis: Spearman correlation coefficients were calculated for each pair of variables with missing values pair-wisely removed. We used Spearman correlation rather than Pearson correlation mainly for the following two reasons: (1) the relationships between sleep and potential contributing factors are not necessarily linear, e.g., previous studies suggest that the relationship between sleep and exercise is non-linear [76]; (2) variables in our study were ordinal instead of being continuous. Therefore, only Spearman correlation is suitable for the analysis.

As illustrated in Figure 2, users of SleepExplorer can see visualizations of their sleep data and correlations with contextual factors. The left column visualizes the time series plots of each of the sleep-structural metric, including MASP, MAWK, NAWK, MTFA, and SE. The right column demonstrates the identified correlated factors to each sleep metric. Only factors with Spearman correlation coefficient that is higher than 0.5 or lower than -0.5 (significance level = 0.05) were rendered for visualization. Green and red bubbles represent positively and negatively correlated factors respectively. The shade of a bubble indicates the strength of correlation. When mouse is hovered over a bubble, a tooltip will appear explaining what this bubble represents. The tooltip was added in response to formative usability tests with earlier versions of SleepExplorer to help users interpret correlations and to remove ambiguity. For example, if dinnertime was positively correlated to minutes awake, the tooltip explained that this correlation means that this user had fewer minutes awake when he or she had dinner at an earlier time.

Furthermore, SleepExplorer provides users with information about which factors may affect their sleep (see Figure 3). We created a comprehensive list of factors that are potentially related to sleep according to evidence based studies in sleep research [7,12,51-52,56], as many potential users of SleepExplorer are not likely to have sufficient domain knowledge on sleep and on sleep contextual factors [1,26,36]. The full list with the details on how each contextual factor was quantified was included in Appendix 1. Most of the contextual factors were measured based on users' subjective perception against a predefined scale. Therefore most contextual factors were measured as ordinal variables. For example, mood was quantified as a discrete variable whose possible values range between 0 (= very bad) and 6 (=very good). We were aware that there are more sophisticated way to measure these factors, and we chose the simpler approach to avoid introducing unnecessary tracking burden.

Fig.1 SleepExplorer shows time series plots of sleep-structural metrics (left side) and correlations sleep metric and contextual factors tracked by a user (right side). Green bubbles represent positively correlated factors and red bubbles show negative correlations. The shade of a bubble indicates the strength of correlation. User can access an explanation for each relationship through pop-up windows by hovering the mouse over a bubble.



SleepExplorer Home About Contact Register Login

Potential Sleep Impacting Factors

This list is created according to evidence-based sleep research literature.Why not exploring whether these factors actually impact your sleep?

Factors	Tracking Tips
Coffee	Cups of coffee consumed during a whole day.
Alcohol	Amount of alcohol consumed during a whole day.
Mood	Rate it right before bedtime.
Stress	Rate it right before bedtime.
Tiredness	Rate it right before bedtime.
Digital Devices	Estimate degree of usage within three hours before bedtime.
Light	Estimate degree of exposure within three hours before bedtime.
Nap	Duration of nap during daytime.
Social Activities	Rate it within three hours before bedtime.
Body Temperature	Measure it right before bedtime.
Hormone Cycle	The number of days from the first day of the latest period.
Ambient Temperature	Measure room temperature right before bedtime.
Ambient Humidity	Measure room humidity right before bedtime.
Exercise Time	The time stamp when you finish exercising.
Dinner Time	The time stamp when you have dinner.
Dreams	Memory of dreams - good or bad.

© 2016 - SleepMakeSense. All Rights Reserved.

Fig.2 SleepExplorer provided users with a list of contextual factors that may influence their sleep along with instructions on how to track these factors.

In our study we asked participants to select the contextual factors that they wanted to explore and we created a personalized tracking diary using Google Forms for each participant. We reminded users to fill in the form every day during the field study, and the diary data were integrated automatically into the database. (A sample Google form can be found at http://goo.gl/forms/FPmQqS96V0.)

3.2 Field Study

3.2.1 Participants

We recruited 12 participants to examine how SleepExplorer can help them understand their sleep quality and correlations with contextual factors. Participants were recruited through university mailing lists, and received a $25 book voucher per interview as an appreciation for their commitment to the study. We looked for participants who already used Fitbit to track their sleep as well as at least one contextual factor such as steps, calories and water consumed. We were seeking a diverse cohort, including people with healthy sleep as well as people with sleep problems.

Table 2 shows the participants' demographic information, device usage, and average sleep quality during the past month. This data was collected through an online survey at the start of the study and during the first interview. Two participants (Maggy and Dina) had visited a sleep clinic before but did not have severe sleep problems during the field study. One participant (Lily) was diagnosed with sleep apnea and was using a breathing machine throughout the field study.

3.2.2 Study Procedure

We commenced the study with a semi-structured interview to learn about their past sleep-tracking experiences and to explore their current data through SleepExplorer. In this interview we asked questions about participants' past self-tracking experience in general, whether and how they had analyzed personal sleep data through contextual information, and we created a baseline of their sleep quality using PSQI [11]. They were also invited to try out the SleepExplorer system to explore their current Fitbit data. At the end of the first interview, they selected extra contextual factors from the list (Appendix 1) on SleepExplorer, which they would track using the diary function of SleepExplorer in the following two-week field study.

Pseudonym	Age Range	Profession	Type of Fitbit	Months of Fitbit use	Perceived sleep quality	Wakeup freshness	Baseline PSQI
Emma	20s	Public Health Student	Charge HR	6	Good	Fairly clear-headed	1 (Good)
Lily	30s	Administrative Staff	Charge HR	2	Good	Very Drowsy	3 (Good)
Tesa	40s	Marketing Director	Charge HR	4	Good	Alert	4 (Good)
Rosa	30s	Administrative Staff	Flex	3	Good	Fairly clear-headed	4 (Good)
Maria	50s	Administrative Staff	Charge HR	5	Good	Fairly clear-headed	4 (Good)
Luke	40s	Computer Scientist	Charge HR	5	Average	Alert	4 (Good)
Luisa	50s	Administrative Staff	Flex	8	Average	Fairly clear-headed	4 (Good)
Paul	30s	Administrative Staff	Charge HR	6	Good	Drowsy	5 (Bad)
Mia	30s	Medicine Student	Charge	4	Good	Drowsy	7 (Bad)
Diana	30s	HR Consultant	Flex	3	Average	Alert	10 (Bad)
Dina	40s	Veterinary Student	Charge HR	1	Average	Fairly clear-headed	10 (Bad)
Maggy	40s	Administrative Staff	Flex	1.5	Bad	Drowsy	11 (Bad)

Table 2. Demographic profile of participants (anonymized), sorted by sleep quality (PSQI score).

The design of the field study was based on the Experience Sampling Method (ESM) approach, which is a research procedure for studying what people do, feel, and think during their daily lives [29, 37]. During the field study, all users received two daily reminders to fill in their self-tracking diary at night (to track contextual factors) and in the morning (to track wakeup freshness). Since the focus of this study was not to investigate the impact of SleepExplorer on behavior change or health improvement, the duration and scale of this study is cohort for a detailed qualitative analysis on sense-making. We will include a long-term study to work on behavior change in the future.

At the end of the field study, we conducted the second interview with participants to discuss their experience of tracking extra contextual factors during the field study, whether there was any change in sleep pattern and daily routine, and what problems they had encountered during the tracking. They were then asked to use SleepExplorer to integrate their Fitbit data with dairy data, and to analyze the correlations between sleep data and the contextual data. They were asked to think aloud how they interpreted the results, what they learnt from the analysis, and what actions they would like to take based on the analysis. All interviews were recorded and transcribed.

3.2.3 Dataset and Analysis

The study produced rich quantitative and qualitative data. We conducted quantitative analysis over participants' self-tracking data as well as qualitative analysis of all interview data, with a focus on developing an understanding of how SleepExplorer helps users make sense of their sleep data.

Self-tracking data captured by Fitbit and diary was integrated and stored in SleepExplorer database on Microsoft Azure cloud. The main purpose for analyzing these data was to investigate the difference in sleep structure between participants.

As for the qualitative data collected in the interviews, our analysis followed the process of a thematic analysis, as described by Braun and Clarke [72]. Firstly, the first, the third, and the fourth author transcribed all interviews and read through each other's transcripts to familiarize themselves with the contents. Excerpts with rich data were read and discussed with the remaining authors to discuss preliminary ideas. Secondly, initial codes were generated from the data. We did not define a coding schema beforehand but identified codes from actual data by repeatedly going through content in the interviews, guided by Weick's observations of sense-making as an ongoing retrospective activity to rationalize what people are doing [71]. Finally, we grouped our codes into four themes: organizing flux, making sense through self-tracking, making connections, and taking actions. This was done through an affinity analysis where all authors grouped the codes through post-it notes on a whiteboard and discussed suitable themes and names for each theme. The results are described in detail in the next section.

4 Results

Overall, participants liked SleepExplorer. Ten out of 12 participants said they would continue using it to analyze their data in the future. Through the following findings we attempt to show how SleepExplorer supported individuals in making sense of their sleep-tracking data as well as the correlations between sleep and contextual factors.

4.1 Organizing Flux in Sleep Data

Sense-making often starts with organizing a flux of experiences and data. SleepExplorer helped participants to understand the multiple metrics that are used in sleep research to define sleep quality and the connections within them.

4.1.1 Defining Sleep Quality

SleepExplorer helped participants understand the relevance of different metrics that define sleep quality. The Fitbit used in this study collected data on minutes asleep (MASL), minutes awake (MAWK), number of awakenings (NAWK), minutes to fall asleep (MTFA), and sleep efficiency (SE). Additionally, we collected sleep data through surveys as summarized in Table 3. These metrics captured both the psychological level (subjective) and the physiological level (objective) of sleep quality. We also explicitly mentioned the concept "sleep structure" on the user interface when rendering the visualization of the sleep metrics, and SleepExplorer calculates correlated contextual factors for each sleep metric separately. This helped some participants who did not know the concept of "sleep structure" become aware of the multi-dimensionality of sleep. As Dina mentioned: "*What surprises me is that I thought they [sleep metrics] were all the same thing, but interestingly it's not actually… I could see how time to fall asleep is different from sleep efficiency*.

Among all the metrics, PSQI (Pittsburgh Sleep Quality Index) [11] is widely used in clinical settings as a standard instrument for evaluating sleep quality. However, we found that PSQI score alone may overshadow the rich information characterized by sleep structure. As is illustrated in Figure 4, Tesa, Rosa, and Maria had the same baseline PSQI score 4, which indicates good sleep. However, when breaking down their sleep into MASL, MAWK, and MTFA, there is significant difference. Averaged over the past one month, Maria slept approximately 40 minutes shorter than Tesa and Rosa, and it took Rosa more than 10 minutes longer to fall asleep compared to Tesa and Maria. Moreover, according to the right plot in Figure 4, Rosa had more regular sleep in terms of MASL while Maria had the lowest deviation with respect to MTFA.

We also found that these sleep metrics were not equally important and that participants' definition of sleep quality varied. For example, Diana mentioned MASL as an important metric for her because if she *"thinks of having a really good sleep, it would never be under five hours"*, whereas for Maria MASL is "*not really important*" because she believed that *"There's no standard. As long as I feel good, refreshed in the morning, I don't really care how long I sleep*". We asked our participants which metric they would use to quantify sleep quality. The answers were as follows (ranked in descending order): NAWK (7 out of 12 mentioned it), WUF (7), MASL (5), MAWK (3), SE (3), and MTFA (1).

	Sleep Metrics	How to Measure
Screening survey	Subjective sleep quality	Estimated according to subjective perception over the past one month; 1 (very bad) ~ 5 (very good)
	Wake up freshness (WUF)	Estimated according to subjective perception over the past one month; 1 (very drowsy) ~ 5 (very alert)
Interviews	PSQI	A self-report questionnaire with 19 items that assess sleep quality over a 1-month time interval.
Field study	Minutes asleep (MASP)	Automatically tracked by Fitbit daily.
	Minutes awake (MAWK)	
	Number of awakenings (NAWK)	
	Minutes to fall asleep (MTFA)	
	Sleep efficiency (SE)	
	Wake up freshness (WUF)	Manually tracked using diary daily; scale: 1 (very drowsy) ~ 5 (very alert).

Table 3. Sleep metrics used in the study.

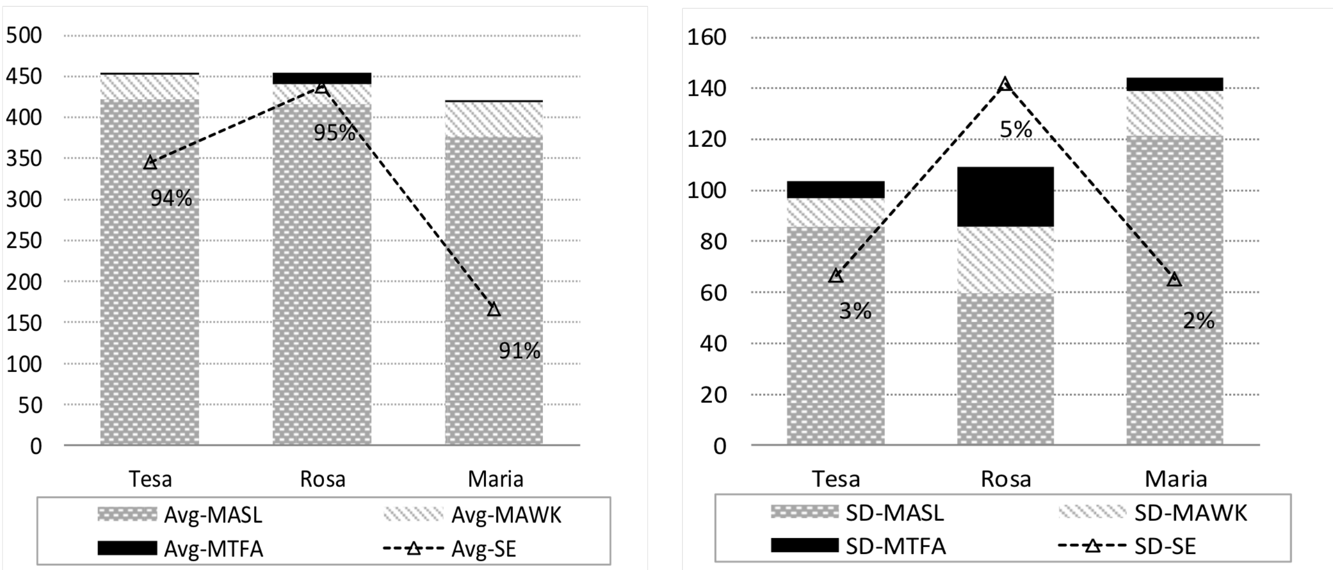


Fig.3 The same PSQI score (=4, good sleep) was mapped to different sleep structures. (left: average values over one month; right: standard deviation). Tesa, Rosa, and Maria had the same baseline PSQI score but significantly different sleep structure in terms of Minutes Asleep (MASL), Minutes Awake (MAWK), Minutes To Fall Asleep (MTFA) and Sleep Efficiency (SE).

4.1.2 Noticing Connections within Sleep Structure

The understanding of sleep structure also helps participants notice unexpected relationships among the sleep metrics. A common relationship was the connection between wake up freshness and other sleep metrics. For example, during the field study, Luke noticed that the more awakenings he had, the fresher he felt in the morning: "*It's generally quite challenging (for me to get up). But interesting enough when it's more interrupted and I don't have a big block of sleep, the morning it feels easier.*"

4.2.2 Quantifying a Dynamic Context

SleepExplorer provided information on contextual factors that are well recognized in sleep research as influencing sleep. Most of the suggested factors were highly dynamic and could change significantly throughout a day. The participants encountered several difficulties which could be classified to three categories: (1) the quantification of the factors, (2) the timing for recording the factors, and (3) the variation of the factors.

Firstly, some participants mentioned that data accuracy was the biggest problem that they encountered in the field study. The factors that were considered difficult to quantify include: caffeine consumption, alcohol consumption, electronic device usage, and light exposure before bed. For instance, Mia mentioned the difficulty on quantifying digital devices usage: "*In terms of the digital devices usage, in the questionnaire it asks many questions in one… So how to quantify the usage is not clear here. How long have you used it before going to bed? What time? What you used it for? Because you know reading books is relaxing but some people reply emails or do their homework. It's different.*"

Secondly, the timing for recording the factors may influence the logged data itself. Most participants logged contextual data before going to bed to reflect on the relevant data for the day. However, the act of logging this data negatively impacted their data on electronic device usage within the last 3 hours before sleep: *"your reminder is later than I wanted. Because I'm aware of less digital devices usage but because I was waiting for your reminder so I checked it more often to see whether it's coming." (Tesa)* Other participants filled in the diary in the morning, which did not affect their data on device usage, but meant that other data (e.g., their mood before going to bed) may have been less accurately recorded. To address this issue, one of the participants suggested tracking relatively stable factors before bed and highly dynamic ones the next morning: "*What I would rather do is the next morning I would remember how tired I was. So when you get the follow up (reminder) in the next morning, (you could ask) wake up freshness and tiredness before bed*." (Diana).

Thirdly, data variation is critical for capturing meaningful correlations. If a participant always took a cup of coffee every day at fixed time, there is no meaning tracking coffee, as it is impossible to study the impact of coffee. Several participants were aware of the importance of data variation. For instance, Luisa mentioned: "*I think caffeine keeps you awake, but I regularly have it, I don't think it affect my sleep. Maybe if I didn't have it for a while, and then have some before bed, then it may keep me awake."*

4.2.3 Reflection through Tracking

The field study also shows that sense-making could occur through tracking per se, which could have impact on behavior and mindset.

The impact of self-reflection during tracking on behavior is characterized by self-experimenting. In the field study, two participants consciously explored the possible impact of lifestyle on personal sleep through self-experimenting. Maria told us that the tracking motivated her to consume less alcohol and to reduce screen time because she hoped to "*see good records*", and she attributed her better sleep quality to the behavior change. By contrast, Luke told us that he consumed more alcohol during the field study just to increase the variation of data.

More commonly, the self-reflection during tracking impacted the mindset. Six participants mentioned that tracking the contextual factors increased their awareness of the potential impact of those factors on their sleep. Maria said: *"Tracking alcohol makes me very conscious because I need to write down if I have a glass of wine."* However, increasing awareness did not always have positive impact. In some cases, the awareness did not really have practical impact on behavior, or could even have negative impact. For instance, Mia mentioned that she "*try to (reduce caffeine consumption) like I tried beforehand. It's just I'm more aware of it"*. For Diana, tracking stress made her feel worse because she would become conscious that "*I'm so stressed, I can't sleep tonight*."

4.3 Making Sense during Retrospective Analysis

SleepExplorer helps people make sense of relationships between sleep and contextual data collected in the field study. This section presents the findings on SleepExplorer supported sense-making on personal sleep through contextual information.

4.3.1 Re-organizing Sleep Data and Contextual

Data

We found that the contextual factors of sleep are highly personal and vary for different persons. On the one hand, the factors that each participant selected to track diff ers due to varied lifestyle context. The last column in Appendix 1 shows the popularity of manually tracked contextual factors in the field study. The top five popular factors were electronic devices usage (all participants tracked it), stress (8 out of 12), tiredness (8/12), coffee and coffee time (8/12), and alcohol (8/12). On the other hand, the same contextual factor could be distinctly associated to different person's sleep, or could even be associated differently to various aspects of the same person's sleep. For example, Table 4 shows how alcohol was associated to 4 participant's sleep.

	Maria	Luke	Lily	Mia
SE	+++	/	/	++
WUF	++	+	/	/
MTFA	/	-	-	/
NAWK	--	-	/	-
MAWK	---	-	/	--
MASL	-	-	/	+

Table 4. Correlations between alcohol and each sleep metric for 4 participants (+++: very strong positive correlation; ++: strong positive correlation; +: moderate positive correlation; /: no correlation; -: moderate negative correlation; --: strong negative correlation; ---: very strong negative correlation).

The potentially distinct correlations between the same contextual factor and different sleep metrics further increased participants' awareness of the multi-dimensionality of sleep, as what Luke discovered when exploring the relationship between alcohol consumption and his sleep structure: "*The more alcohol consumed, the shorter I was awake. It's contradictory to 'the more alcohol consumed, the shorter I was asleep'. But maybe that's fine; it just means I was awake, just shorter. So not necessarily contradictory; they are different aspects of what sleep is, what certain period of sleep is.*"

4.3.2 Confirming Presumptions

Some identified relationships between sleep and contextual factors validated participants' perception or brought the relationships from back of mind to consciousness. Such validation is highly individual and may even be inconsistent with common sense. For instance, the correlation analysis indicated that Lily slept less when she was more active, and she thought it made sense because "*Most of the time that I'm using this (Fitbit), I would go to bed at the same time, but then on the days I go to gym in the morning, I had to get up earlier, so I had less sleep*."

Surprisingly, Luke also paid attention to irrelevant factors. He tracked coffee time but this factor did not show up during the analysis. Luke: "*The common sense is that don't have coffee in the afternoon if you want to be able to fall asleep quickly. That's not showing up in my data. That also corresponds to my observation as well. I can have coffee after dinner and still get a good sleep.*

4.3.3 Making Sense of Counterintuitive Relationships Using Confounding Factors

Some participants identified surprising relationships between sleep and contextual factors, and they usually attribute to other untracked confounding factors or the measurement errors of tracking devices to make sense of these surprising findings.

Participants mentioned the following common confounding factors: pets, trash collectors, partners and other factors in their daily pattern that may impact sleep. Diana's reflection on why she seemed to sleep better when she had more minutes sedentary was as follows, "*Ok, this could be because there are other factors, for example, I'm sedentary when I work, I'm quite being mentally tired with my work, and I don't drink alcohol at night… So it's not like being still makes me sleep better. It's more like, it happens to be on the days when I'm still, I'm mentally drained, and I don't eat overly at night, and I don't drink alcohol*." Identifying confounding factors requires advanced statistical and data mining techniques, which is our future work.

Some participants also showed doubt over the non-intuitive correlations and attribute these relationships to measurement errors in Fitbit data, such as Dina: "*I'm not sure how to interpret that, just because I don't trust all the Fitbit data*."

4.4 Taking Action based on Sleep Data

Reflecting on the data presented through SleepExplorer helped participants address the question "what do I do next?" While we do not have data on actual behavior change, we explored novel opportunities for action that the participants identified through the data.

Some participants showed high expectation of the system on guiding them to improve sleep. Emma told us: "*So for me, the system would show me the things I have no idea about. And then I can act on those to improve. And for people like my boyfriend who have terrible sleep, maybe it could help find out solutions."* We also found that participants' attitude to behavior change varies depending on whether the analysis validated or contradict their perception. In the following sections we present how SleepExplorer supported individuals in making informed decisions on behavioral change.

4.4.1 Following Presumptions

Several participants compared the relationships identified by SleepExplorer with their existing presumptions about which factors may influence their sleep. For participants who were able to confirm their presumption it was easy to plan actions for changing behaviors as they had already considered the consequences of their presumption prior to this study. For instance, Luke said*: "You hear people say that using a laptop or phone or something with bright light is not good for your sleep. How we do things in the moment is that you can just ignore the sleep impact. But when it's presented to me clearly, I'm prepared to change my behavior, re-arrange my day (so that I could) do more reading stuff at night and more computer stuff during the day in order to get a better sleep at night."*

4.4.2 Visual Cues for Action

Negative relationships offered more effective visual cues in triggering intensions for behavior change compared to positive relationships. Seven out of 12 participants mentioned that they would pay more attention to negatively related contextual factors (red bubbles) than positively related ones (green bubbles). The reasons varied. Someone mentioned that negative factors provided more actionable knowledge, while others thought negative ones were the things that one should fix. For example, Rosa mentioned "*because they (red bubbles) are preventing me from becoming better. They are something I should try to avoid.*"

Only Emma and Dina mentioned that they paid more attention to the green bubbles. Emma thought that it would be easier for her to strengthen the positive factors than reducing the negative ones in her lifestyle context, and Dina mentioned that she was "*already aware of the negative ones so green bubbles are more informative*". In the meanwhile, 2 participants paid more attention to darker bubbles, such as Maria: "*I probably care more about strong factors, either negative or positive.*" Paul cared about the more relevant bubbles regardless of the colors: "*It really depends on what factor that is; like the alcohol, not just because of the negative relationship. I know it could be really a problem. I need to be aware of that*".

4.4.3 Constraints for Action

Regardless of participants' intension for behavior change, many mentioned that their capacity to take real action may be bounded by uncontrollable external factors, such as daily schedule, medical conditions, and autonomous conditions (e.g., mood, stress, hormone cycle, and dream). Emma: *"Yeah I try to reduce sedentary minutes anyway. But it's just really hard to when you have an office job."* Lily: *"I'm trying to lose weight but my doctor said that I need to get more energy so I can do more exercise and get healthier."* Diana: "*Mood and stress, I think they affect my sleep but I can't really control them."*

Sometimes decisions are not just about sleep but governed by overall health, especially when participants identified counterintuitive relationships and sleep was not a major concern. Being given counterintuitive result between steps and sleep, Maria answered: "*I wouldn't reduce my steps because I think you need to be active. Walking more generally makes me feel more active and good*."

We also found that goal setting helps little in improving sleep. Goal setting has been considered as an effective way to motivate people in calorie and physical activity tracking. However, unlike calorie and physical activity tracking where actionable knowledge is straightforward, sleep tracking requires more advanced analysis in order for the users to obtain actionable knowledge. Some participants made the point that goal-setting in sleep tracking does not even motivate them to sleep longer, let alone sleep better. Emma mentioned that she did not thought the sleep goal functionality of sleep-trackers is useful, *"so most of the time I get 7 hours sleep, but like, yes of course I want to get 8 hours sleep every day, I've always had that in my mind. But how to control that?"* Participants acknowledged the importance of understanding the factors that are associated to personal sleep, but also emphasize the desire of receiving recommendations on how to change. Four out of ten participants explicitly mentioned that they would like to learn more on what they could do on the identified sleep contextual factors, such as what is the limit for coffee/alcohol consumption, how to deal with stress, and how to increase basic body temperature. Rosa: *"Maybe it could be better, for example, what would be the limit for my coffee consumption, for my alcohol. It could be great if it can provide clues or recommendations, like if I drink this amount of alcohol, it's ok. I would like to have some recommendations more than analysis.*

5 Discussions

This study has shown how visualizations of sleep data and contextual factors (like exercise, mood and coffee consumption) presented through SleepExplorer helped participants make sense of their personal data. While previous work allowed users to compare contextual factors between nights with good and poor sleep [75], (1) SleepExplorer allowed users to explore correlations between sleep quality and contextual factors. Additionally, the participants developed a better understanding of (2) the multiple aspects that define the quality of sleep as well as (3) what, when and how to track data. Based on this understanding, (4) the participants felt guided in taking actions to improve their sleep. We will now discuss these findings within the landscape of previous studies and highlight opportunities for future research.

5.1 Sense-Making through Correlations in the Data

Our main aim was to support users in exploring correlations between their sleep data and contextual factors that may influence their sleep. Previous work has demonstrated the benefit of identifying statistical patterns between wellbeing data (weight, sleep, pain, and mood) and context (step count, calendar data, location, weather, and food intake) numerically [77] and textually [9], but neither of them exclusively focuses on sleep. Sleep is complex because we do not consciously experience it, and understanding sleep structure and potential contributing factors requires background in sleep research. Building on previous work, SleepExplorer guided users to track clinically proven sleep contextual factors and visualize the detailed correlations between sleep and context, which aroused participants' awareness of the connection between sleep and various contextual factors. In the field study, users were making connections in a more systematic manner based on actual hard-cold data that represent measured fact. One participant even experimented during the tracking by intentionally adjusting his lifestyle to create more variations in data.

Whereas sense-making on intuitive correlations was straightforward, participants often attribute the non-intuitive correlations to confounding factors and measurement errors, which remain challenges for sleep-tracking technologies. In order to control the interference of confounding factors, one needs to carefully design and conduct self-tracking experiments that are similar to N-of-1 trials [78]. Although the concept of the N-of-1 trial was adopted in personalized medicine for years, it is still not widely known to the crowd. Currently self-tracking activities are usually conducted in uncontrolled settings, and many scholars question the rigidity of such experiments and the conclusions drawn from these experiments. It remains a future work to create a guideline for laymen on how to design rigid self-tracking experiments based on the principles of N-of-1 trials. As for the measurement errors of commercial sleep trackers, these errors could be reduced to certain extent through data cleaning before the data analysis. However, due to the inherent limitation of actigraphy, it is impossible to completely remove potential measurement errors. The only solution would be to develop new tracking devices based on the principle of polysomnography.

5.2 Re-Defining Sleep Quality

In order to help users make sense of the correlations between their sleep and contextual factor, SleepExplorer also helps them understand the sleep quality through the concept of sleep structure [31-32]. It is widely recognized in sleep research community that normal sleep is difficult to define because individuals vary enormously, differing in physiology, psychology, lifestyle, and living environment [19, 79]. Traditionally, sleep quality was usually characterized by a single score, such as PSQI score or sleep efficiency. However, our study suggested that PSQI may overshadow the rich information of personal sleep structure, and "sleep efficiency" is in fact one of the least considered sleep metrics when participants were asked to define sleep quality. By providing visualization of sleep metrics individually on the user interface, *SleepExplorer* improved users' awareness of sleep structure and their understanding of the multi-dimensionality of sleep.

Our study demonstrated that people's definition of sleep quality was highly individual and was closely related to how a person perceives his/her sleep. For example, if a participant suffered from many interruptions during sleep, he or she would quantify sleep quality in terms of the number of awakenings and thus fewer awakenings means better sleep for this person. On the other hand, for people who had difficulty in falling asleep, minutes to fall asleep was an important criteria when it comes to sleep quality and their definition of good sleep was fast sleep onset.

The fact that experts usually recommend 7 to 8-hour sleep to adults confused or even annoyed some of our participants, as they felt good and well-rested after 6 or less hours of sleep. In fact, recent neurology studies indicate that long sleep duration significant increase the risk of cardio vascular disease such as stroke, whereas short sleep duration does not have such impact [39]. The interpretation of "adequate sleep" may vary from person to person. Therefore, it remains to be a challenge how to establish a personalized sleep quality evaluation framework that accounts in interpersonal differences.

5.3 What, When, and How to Track

The most important issue in self-tracking is knowing what, when and how to track. In [18], Choe et al. identified "not tracking triggers and context" as one of the common pitfalls in self-tracking. This is especially true for sleep tracking, as many people do not have adequate sleep domain knowledge on what factors may be related to personal sleep [43]. Popular commercial tracking tools such as Fitbit and Jawbone could track limited contextual factors but some of these factors are not likely to be related to sleep.

SleepExplorer builds on these findings and suggested a list of potential sleep contextual factors based on sleep researches. Most of the factors that participants selected from this list turned out to be related to their sleep, whereas many of the factors tracked by Fitbit were irrelevant to personal sleep. Therefore, we argue that SleepExplorer extended the scope of users' sleep-tracking by guiding them to track more factors and at the same time keeping them focused on tracking the factors that were most likely to be related to personal sleep. Albeit previous study suggested that right before going to sleep turned out to be an opportune moment to do self-reflection [75], we found that not all data were logged before bed time.

We also identified several challenges in tracking sleep contextual factors. Firstly, it is difficult to accurately tracking some factors such as the consumption of alcohol and caffeine, the degree of digital device usage, exercise time. Second, the timing for logging the data may influence data itself; this is especially true for digital device usage, as users had to input data either on a computer or on a mobile phone. Third, the best timing for tracking some factors is hard to decide on, as those factors are highly dynamic and may significantly change in a short time. Fourth, it is difficult to collect meaningful data if a user's daily life pattern does not vary much. These issues have strong impact on the quality of the self-tracking data and thus should be addressed in future research.

5.4 Taking Actions

The ultimate purpose for identifying correlations between sleep and contextual factors is to guide behavior change for sleep improvement. However, our study suggested that the role of a factor in guiding behavior change was heavily dependent on how this factor was related to sleep: intuitively or non-intuitively, and positively or negatively.

Participants were generally willing to make behavioral changes based on intuitive correlations that confirmed participants' presumptions, but they were more cautious to take any actions based on non-intuitive correlations. Reasons include the concern of overall health, data quality, and the duration of the study. Many participants also mentioned that negatively correlations (red bubbles) were more helpful in guiding behavioral change than positively ones (green bubbles). This echoes the asymmetrical effects of positive and negative events in psychological studies [62], which states that adverse or threatening events trigger strong and rapid physiological, cognitive and emotional responses.

In [43], Liu et al. pointed out the importance of providing instructions for action. This is confirmed in our study. Our participants suggested that SleepExplorer should provide concrete recommendations on what they could do to reduce the negative factors or strengthen the positive ones. Since behavior change could be constrained by external factors that were outside a user's control, it is necessary to tailor the recommendations to the user's life context in designing and implementing future personal informatics systems. One solution is to applying context-aware computing technologies [67], which tailored the recommendations by accounting in the lifestyle context of each user.

5.5 Limitations of This Study

We have presented how SleepExplorer increase users' domain knowledge on sleep and their understanding of how sleep was associated to various contextual factors. However, the current study has its limitations in several aspects. The potential dependency among contextual factors was not considered in data analysis. The experimental set up during the field study did not control confounding factors. Contextual factors were tracked manually and were based on subjective perception rather than objective measures. The system could be further improved through implementing automated tracking and advanced data mining techniques. Our study did not focus on actual behavior change. We are aware that a longer study is necessary to investigate the effect of SleepExplorer on behavior change and sleep quality improvement, which is one of our future works.

While our study is specific to the domain of sleep, several findings have global implications for other domains in self-tracking, such as the asymmetrical effects of positive and negative contextual factors, the interference of confounding factors in sense-making, and the motivation for behavior change relating to overall health rather than a single health aspect. These findings could be applied to e.g. weight losing, diabetes self-management, smoking quitting, to name but a few.

Overall, our participants were satisfied with SleepExplorer. The system provided insights on personal sleep that people were not able to discover on their own or using existing sleep-tracking tools such as Fitbit. We also identified several opportunities for future research in developing personal informatics systems: improving self-tracking accuracy, alleviating the impact of confounding factors, and providing tailored instructions/recommendations for behavioral changes.

6 Conclusions

People now have access to a variety of sleep-tracking data from wearable technology, mobile apps and embedded technologies, but users find it difficult to make sense of this data due to lack of sleep domain knowledge and no context. This study showed how a web application (SleepExplorer) designed to visualize sleep data and to show correlations between sleep-tracking data and other self-tracking data (i.e. physiological, psychological, lifestyle and environmental data) helped users make sense of their sleep. Through a field study we found that SleepExplorer helped organize a flux of self-tracking data to personalize the definition of sleep quality in multiple dimension, guided the tracking activities, highlighted connections in the data, and led to intension for behavior change to improve personal sleep.

This research highlights three opportunities to help inform the work of researchers and designers creating data-driven health and wellbeing applications: improving the accuracy of self-tracking, alleviating the impact of confounding factors, and providing tailored instructions and recommendations for behavior change. In the next step, we plan to further develop the system to address these opportunities and to obtain evidence on how SleepExplorer impact users' sleep behavior and sleep quality in the long term.

Acknowledgment

This study was supported by Australian Government Endeavour Research Fellowship and Microsoft BizSpark. We would like to thank our study participants for contribution. We also appreciate feedback and support from Dr. Kathleen Gray, Manal Almalki and our colleagues in the Microsoft Research Centre for Social NUI at the University of Melbourne.

References

1. Mark Dodd, Christopher Findlay and David H Wilson. Risks Associated with Low Functional Health Literacy in an Australian Population. *Med J Aust* 191, 10 (2009), 530-534.
2. Manal Almalki, Kathleen Gray and Fernando Martin Sanchez. The Use of Self-Quantification Systems for Personal Health Information: Big Data

Management Activities and Prospects. *Health information science and systems* 3, Suppl 1 (2015), S1.

3. Jessica S Ancker and David Kaufman. Rethinking Health Numeracy: A Multidisciplinary Literature Review. *Journal of the American Medical Informatics Association* 14, 6 (2007), 713-721.
4. Yannick Assogba and Judith Donath. 2009. Mycrocosm: Visual Microblogging. In *Proceedings of System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. IEEE, 1-10.
5. Martin Atzmueller and Frank Puppe. Causal Subgroup Analysis for Detecting Confounding. In *Applications of Declarative Programming and Knowledge Management*, Springer, 2009, 136-148.
6. Yin Bai, Bin Xu, Yuanchao Ma, Guodong Sun and Yu Zhao. 2012. Will You Have a Good Sleep Tonight?: Sleep Quality Prediction with Mobile Phone. In *Proceedings of Proceedings of the 7th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 124-130.
7. Kate A Bartel, Michael Gradisar and Paul Williamson. Protective and Risk Factors for Adolescent Sleep: A Meta-Analytic Review. *Sleep medicine reviews* 21 (2015), 72-85.
8. Jared S Bauer, Sunny Consolvo, Benjamin Greenstein, Jonathan Schooler, Eric Wu, Nathaniel F Watson and Julie Kientz. 2012. Shuteye: Encouraging Awareness of Healthy Sleep Recommendations with a Mobile, Peripheral Display. In *Proceedings of Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1401-1410.
9. Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone and Jeff Wilson. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 5 (2013), 30.
10. Marko Borazio and Kristof Van Laerhoven. 2012. Combining Wearable and Environmental Sensing into an Unobtrusive Tool for Long-Term Sleep Studies. In *Proceedings of Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. ACM, 71-80.
11. Daniel J Buysse, Charles F Reynolds, Timothy H Monk, Susan R Berman and David J Kupfer. The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. *Psychiatry research* 28, 2 (1989), 193-213.
12. Julia KM Chan, John Trinder, Holly E Andrewes, Ian M Colrain and Christian L Nicholas. The Acute Effects of Alcohol on Sleep Architecture in Late Adolescence. *Alcoholism: Clinical and Experimental Research* 37, 10 (2013), 1720-1728.
13. Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tonmoy Choudhury and Andrew T Campbell. Year. Unobtrusive Sleep Monitoring Using Smartphones. In *Proceedings of Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*. IEEE, 145-152.

14. MD Chesson Jr, MD Coleman, MD Lee-Chiong and DDS Pancer. Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007. *Sleep* 30, 4 (2007), 519.

15. Sunny Consolvo, Katherine Everitt, Ian Smith and James A Landay. 2006. Design Requirements for Technologies That Encourage Physical Activity. In *Proceedings of Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 457-466.

16. Eun Kyoung Choe. 2014. *Designing Self-Monitoring Technology to Promote Data Capture and Reflection*. Ph.D Dissertation. Pennsylvania State University.

17. Eun Kyoung Choe, Sunny Consolvo, Nathaniel F Watson and Julie A Kientz. 2011. Opportunities for Computing Technologies to Support Healthy Sleep Behaviors. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3053-3062.

18. Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt and Julie A Kientz. 2014. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1143-1152.

19. S José Closs. Assessment of Sleep in Hospital Patients: A Review of Methods. *Journal of advanced nursing* 13, 4 (1988), 501-510.

20. Ibrahim Delibaşoğlu, Cafer Avci and Ahmet Akbaş. 2011. Ecg Based Sleep Apnea Detection Using Wavelet Analysis of Instantaneous Heart Rates. In *Proceedings of Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*. ACM, 40.

21. Elizabeth H Ehleringer and Si Jung Kim. 2013. The Wearable Lullaby: Improving Sleep Quality of Caregivers of Dementia Patients. In *Proceedings of CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 409-414.

22. Daniel A Epstein, An Ping, James Fogarty and Sean A Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 731-742.

23. Susannah Fox and Maeve Duggan. *Health online 2013*. Pew Internet & American Life Project Washington, DC. 2013.

24. Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison and James A Landay. 2009. Ubigreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1043-1052.

25. Weixi Gu, Zheng Yang, Longfei Shangguan, Wei Sun, Kun Jin and Yunhao Liu. 2014. Intelligent Sleep Stage Mining Service with Smartphones. In *Proceedings of Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 649-660.

26. Janel E Hackney, Terri E Weaver and Allan I Pack. Health Literacy and Sleep Disorders: A

Review. *Sleep Medicine Reviews* 12, 2 (2008), 143-151.

27. Jun Han, Jae Yoon Chong and Sukun Kim. Demo Abstract: Snores-Towards a Less-Intrusive Home Sleep Monitoring System Using Wireless Sensor Networks.

28. Tian Hao, Guoliang Xing and Gang Zhou. 2013. Isleep: Unobtrusive Sleep Quality Monitoring Using Smartphones. In *Proceedings of Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 4.

29. Joel M Hektner, Jennifer A Schmidt and Mihaly Csikszentmihalyi. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Sage, 2007.

30. Victoria Hollis, Artie Konrad and Steve Whittaker. 2015. Change of Heart: Emotion Tracking to Promote Behavior Change. In *Proceedings of Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2643-2652.

31. Ming-Chun Huang, Xiaoyi Zhang, Wenyao Xu, Jason Liu and Majid Sarrafzadeh. 2014. Ezwakeup: A Sleep Environment Design for Sleep Quality Improvement. In *Proceedings of CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2389-2394.

32. Christer Hublin, Markku Partinen, Markku Koskenvuo and Jaakko Kaprio. Sleep and Mortality: A Population-Based 22-Year Follow-up Study. *Sleep* 30, 10 (2007), 1245.

33. Jashanjit Kaur, Nehal Molasaria, Niyati Gupta, Shengjie Zhang and Wei Wang. 2015. Sleepstellar: A Safety Kit and Digital Storyteller for Sleepwalkers. In *Proceedings of Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 31-36.

34. Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo and Julie A Kientz. 2012. Lullaby: A Capture & Access System for Understanding the Sleep Environment. In *Proceedings of Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 226-234.

35. Stephan P Kudyba. *Healthcare Informatics: Improving Efficiency and Productivity*. CRC Press, 2010.

36. Mark Kutner, Elizabeth Greenburg, Ying Jin and Christine Paulsen. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. Nces 2006-483. *National Center for Education Statistics* (2006).

37. Reed W. Larson and Mihaly Csikszentmihalyi. 1983. "The experience sampling method". *New Directions for Methodology of Social and Behavioral Science*, 15, 41-56.

38. Shaun Lawson, Sue Jamison-Powell, Andrew Garbett, Conor Linehan, Erica Kucharczyk, Sanne Verbaan, Duncan A Rowland and Kevin Morgan. 2013. Validating a Mobile Phone Application for the Everyday, Unobtrusive, Objective Measurement of Sleep. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2497-2506.

39. Yue Leng, Francesco P Cappuccio, Nick WJ Wainwright, Paul G Surtees, Robert Luben, Carol Brayne and Kay-Tee Khaw. Sleep Duration and Risk of Fatal and Nonfatal Stroke a Prospective

Study and Meta-Analysis. *Neurology* 84, 11 (2015), 1072-1079.

40. Ian Li. 2011. *Personal informatics and context: using context to reveal factors that affect behavior*. Ph.D Dissertation. Carnegie Mellon University. Pittsburgh. Pennsylvania.

41. Ian Li, Anind Dey and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 557-566.

42. Zilu Liang and Mario Alberto Chapa-Martell. 2015. Framing self-quantification for individual-level preventive health Care. In *Proceedings of the International Conference on Health Informatics*, pages 336-343.

43. Wanyu Liu, Bernd Ploderer and Thuong Hoang. 2015. In bed with technology: challenges and opportunities for sleep tracking. *In Proceedings of the Australian Computer-Human Interaction Conference (OzCHI 2015)*.

44. Subramani Mani, Constantin F Aliferis, Alexander R Statnikov and MED NYU. 2010. Bayesian Algorithms for Causal Data Mining. In *Proceedings of NIPS Causality: Objectives and Assessment*. Citeseer, 121-136.

45. Jennifer Mankoff, Gary Hsieh, Ho Chak Hung, Sharon Lee and Elizabeth Nitao. Using Low-Cost Sensing to Support Nutritional Awareness. In *Ubicomp 2002: Ubiquitous Computing*, Springer, 2002, 371-378.

46. Jennifer Mankoff, Robin Kravets and Eli Blevis. Some Computer Science Issues in Creating a Sustainable World. *iEEE Computer* 41, 8 (2008), 102-105.

47. Yevgeniy Medynskiy and Elizabeth Mynatt. 2010. Salud!: An Open Infrastructure for Developing and Deploying Health Self-Management Applications. In *Proceedings of Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS*. IEEE, 1-8.

48. Vangelis Metsis, Georgios Galatas, Alexandros Papangelis, Dimitrios Kosmopoulos and Fillia Makedon. 2011. Recognition of Sleep Patterns Using a Bed Pressure Mat. In *Proceedings of Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 9.

49. Vangelis Metsis, Dimitrios Kosmopoulos, Vassilis Athitsos and Fillia Makedon. Non-Invasive Analysis of Sleep Patterns Via Multimodal Sensor Input. *Personal and ubiquitous computing* 18, 1 (2014), 19-26.

50. 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. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 477-486.

51. Jodi A Mindell, Lisa J Meltzer, Mary A Carskadon and Ronald D Chervin. Developmental Aspects of Sleep Hygiene: Findings from the 2004 National Sleep Foundation Sleep in America Poll. *Sleep medicine* 10, 7 (2009), 771-779.

52. M Minowa and T Tango. Impact and Correlates of Poor Sleep Quality in Japanese White-Collar Employees. *Sleep* 26, 4 (2003), 467-471.

53. Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen,

Eric Howard, Ruth West and Péter Boda. 2009. Peir, the Personal Environmental Impact Report, as a Platform for Participatory Sensing Systems Research. In *Proceedings of Proceedings of the 7th international conference on Mobile systems, applications, and services*. ACM, 55-68.

54. Daichi Nagata, Yutaka Arakawa, Takatomi Kubo and Keiich Yasumoto. 2015. Effective Napping Support System by Hypnagogic Time Estimation Based on Heart Rate Sensor. In *Proceedings of Proceedings of the 6th Augmented Human International Conference*. ACM, 201-202.

55. Adam Perer and Marc A Smith. 2006. Contrasting Portraits of Email Practices: Visual Approaches to Reflection and Analysis. In *Proceedings of Proceedings of the working conference on Advanced visual interfaces*. ACM, 389-395.

56. Peter A Poelstra. Relationship between Physical, Psychological, Social, and Environmental Variables and Subjective Sleep Quality. *Sleep: Journal of Sleep Research & Sleep Medicine* (1984).

57. Yanzhi Ren, Chen Wang, Yingying Chen and Jie Yang. 2014. Poster: Hearing Your Breathing: Fine-Grained Sleep Monitoring Using Smartphones. In *Proceedings of Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 409-412.

58. John Rooksby, Mattias Rost, Alistair Morrison and Matthew Chalmers Chalmers. 2014. Personal Tracking as Lived Informatics. In *Proceedings of Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1163-1172.

59. Alireza Sahami Shirazi, James Clawson, Yashar Hassanpour, Mohammad J Tourian, Albrecht Schmidt, Ed H Chi, Marko Borazio and Kristof Van Laerhoven. Already Up? Using Mobile Phones to Track & Share Sleep Behavior. *International Journal of Human-Computer Studies* 71, 9 (2013), 878-888.

60. Peter Spirtes, Clark N Glymour and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2000.

61. S Taheri. The Link between Short Sleep Duration and Obesity: We Should Recommend More Sleep[18] to Prevent Obesity. *Archives of disease in childhood* 91, 11 (2006), 881-884.

62. Shelley E Taylor. Asymmetrical Effects of Positive and Negative Events: The Mobilization-Minimization Hypothesis. *Psychological bulletin* 110, 1 (1991), 67.

63. Fernanda B Viégas, Scott Golder and Judith Donath. 2006. Visualizing Email Content: Portraying Relationships from Conversational Histories. In *Proceedings of Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 979-988.

64. Fernanda B Viegas, Martin Wattenberg, Frank Van Ham, Jesse Kriss and Matt McKeon. Manyeyes: A Site for Visualization at Internet Scale. *Visualization and Computer Graphics, IEEE Transactions on* 13, 6 (2007), 1121-1128.

65. Mark Whooley, Bernd Ploderer and Kathleen Gray. 2014. On the Integration of Self-Tracking Data Amongst Quantified Self Members. In *Proceedings of Proceedings of the 28th International BCS Human Computer Interaction*

Conference on HCI 2014-Sand, Sea and Sky-Holiday HCI. BCS, 151-160.

66. Shunsuke Yanaka, Motoki Ishida, Takayuki Kosaka, Motofumi Hattori and Hisashi Sato. 2013. Resolution of Sleep Deprivation Problems Using Zzzoo Pillows. In *Proceedings of Proceedings of the Virtual Reality International Conference: Laval Virtual*. ACM, 25.
67. Hengshu Zhu, Enhong Chen, Hui Xiong, Kuifei Yu, Huanhuan Cao and Jilei Tian. Mining Mobile User Preferences for Personalized Context-Aware Recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 4 (2014), 58.
68. Daytum. Retrived September 23, 2015 from http://daytum.com.
69. Grafitter. Retrived September 23, 2015 from http://grafitter.com.
70. your.flowingdata. Retrived September 23, 2015 from http://your.flowingdata.com.
71. Karl Weick, Kathleen Sutcliffe and David Obstfeld. Organizing and the Process of Sensemaking. *Organization Science* 15, 4 (2005), 409-421.
72. Virginia Braun and Victoria Clarke. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2015), 77-101.
73. PT Ko, Julie A Kientz, Eun Kyoung Choe, M Kay, CA Landis and NF Watson. Consumer Sleep Technologies: A Review of the Landscape. *J Clin Sleep Med* (2015).
74. S Ancoli-Israel, R Cole, C Alessi, M Chambers, W Moorcroft and C Pollak. The Role of Actigraphy in the Study of Sleep and Circadian Rhythms. American Academy of Sleep Medicine Review Paper. *Sleep* 26, 3 (2003), 342-392.
75. Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt and Julie A Kientz. Year. Sleeptight: Low-Burden, Self-Monitoring Technology for Capturing and Reflecting on Sleep Behaviors. In *Proceedings of Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 121-132.
76. Helen S Driver and Sheila R Taylor. Exercise and Sleep. *Sleep medicine reviews* 4, 4 (2000), 387-402.
77. Afsaneh Doryab, Mads Frost, Maria Faurholt-Jepsen, Lars V Kessing and Jakob E Bardram. Impact Factor Analysis: Combining Prediction with Parameter Ranking to Reveal the Impact of Behavior on Health Outcome. *Personal and Ubiquitous Computing* 19, 2 (2015), 355-365.
78. RL Kravitz and N Duan. Design and Implementation of N-of-1 Trials: A User's Guide. *Agency for healthcare research and quality, US Department of Health and Human Services* (2014).
79. Daniel J Buysse. Sleep Health: Can We Define It? Does It Matter. *Sleep* 37, 1 (2014), 9-17.

Appendix

Sleep Contextual Factors Collected for Users during Field Study

(Popularity = number of participants tracked this factor)

Categories	Contextual Factors	Source	Type (Unit)	Sampling	Popularity
Physiological Factors (5)	Weight	Fitbit	Continuous (Kg)	Daily	/
	Body temperature		Continuous (^{o}C)	Daily	3
	Menstrual cycles	Dairy	Days passed since last period??	Daily	3
Psychological Factors (4)	Mood	Diary	Ordinal (0 ~6, 0 = very bad, 6 = very good)	Daily	7
	Stress		Ordinal (0 ~ 6, 0 = very relaxed, 6 = very stressed)	Daily	9
	Tiredness		Ordinal (0 ~ 6, 0 = very energetic, 6 = fatigue)	Daily	8
	Dream		Ordinal (1 ~ 5, 1 = nightmare, 5 = no dream)	Daily	2
Behavior factors (18)	Steps	Fitbit	Discrete (number of cups)	Daily	/
	Minutes very active		Continuous (min)	Daily	
	Minutes fairly active		Continuous (min)	Daily	
	Minutes lightly active		Continuous (min)	Daily	
	Calories in		Continuous (kcal)	Daily	
	Calories out		Continuous (kcal)	Daily	
	Activity calories		Continuous (kcal)	Daily	
	Coffee	Diary	Discrete (count)	On demand	8
	Coffee time		Converted to continuous (18:00 → 1800)	On demand	8
	Alcohol		Continuous (ml)	On demand	8
	Electronic devices usage		Ordinal (0 ~6, 0 = none, 6 = heavy and until bed time)	Daily	12
	Evening light		Ordinal (0 ~6, 0 = none, 6 = long and until bed time)	Daily	3
	Nap time		Converted to continuous (18:00 → 1800)	On demand	2
	Nap duration		Continuous (min)	On demand	2

	Social activities		Ordinal (0 ~ 6, 0 = none, 6 = heavy and until bed time)	Daily	1
	Exercise time		Converted to continuous (18:00 → 1800)	On demand	6
	Dinner time		Converted to continuous (18:00 → 1800)	Daily	5
Environmental Factors (2)	Ambient temperature	Diary	Continuous (^{o}C)	Daily	2
	Ambient humidity		Continuous (%)	Daily	1