Bachelor's Programme in Science and Technology

Evaluation of Sleep Quality From Passive Data

Fathima Afrooz Abdul Mahir

Aalto University

BACHELOR'S THESIS

Aalto University
BACHELOR'S THESIS 2024

Evaluation of Sleep Quality From Passive Data

Fathima Afrooz Abdul Mahir

Thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Technology.
Otaniemi, 4 Sep 2024

Supervisor: Maarit Korpi-Lagg
Advisors: Talayeh Aledavood, Nguyen Luong

Aalto University
School of Science
Bachelor's Programme in Science and Technology

Author
Fathima Afrooz Abdul Mahir

Title
Evaluation of Sleep Quality From Passive Data

School School of Science

Degree programme Bachelor's Programme in Science and Technology

Major Data Science **Code** SCI3095

Supervisor Maarit Korpi-Lagg

Advisors Talayeh Aledavood, Nguyen Luong

Level Bachelor's thesis **Date** 4 Sep 2024 **Pages** 22 **Language** English

Abstract

Sleep quality is a complex phenomenon, made up of various sleep metrics. Studies in the past have used multiple methods to detect these sleep metrics. This study utilizes longitudinal data from the Mobile Monitoring of Mood (MoMo-Mood) study, to collect data from four groups: healthy users, patients with with bipolar disorder, patients with borderline personality disorder, and patients with major depressive disorder. This thesis aims to investigate how sleep patterns differ across healthy users and patients with varying mood disorders. Additionally, sleep patterns from actigraphy and bed sensor data collection methods were analysed to ensure consistent measurements for a reliable study in sleep quality. Moreover, the thesis used subjective data from surveys to test the hypothesis of whether factors like pet ownership and co-sleeping affected the accuracy of sleep measurement results.

Results revealed a noticeable dispersion in the data as the sleep duration measured by the two methods differed significantly for some users. Nevertheless, the control group consistently had shorter nocturnal sleep durations in comparison to the mood disorder groups. The Kruskal-Wallis H-statistic tests, conducted at significance levels of 0.05 for both devices, revealed statistically significant differences between the nocturnal sleep durations among the four user groups (actigraphy: H = 22.101, p-value = 6.214e-05, bed sensor: H = 11.105,p-value = 0.011). Actigraphy data exhibited a stronger significance compared to the bed sensor data, indicating more pronounced variations among the groups. Additionally, independent t-tests at a significance level of 0.05 were conducted to examine the influence of pets and co-sleeping on sleep measurement devices. No significant difference in sleep duration was detected due to the ownership of pets for both devices (Actigraphy: t = -0.518, p-value = 0.604, bed sensor: t = 0.999,p-value = 0.318). However, for co-sleeping, results revealed a significant difference with the bed sensor (t = 1.989, p-value = 0.047), but no significant difference was found with actigraphy (t = 1.632, p-value = 0.103). When measured by a bed sensor, the median sleep duration was slightly higher and the box plot had a broader range for those who sleep alone. Furthermore, the results revealed that on average, the sleep duration measured by the bed sensor is 0.78 hours less than the sleep duration measured by the actigraph for those who sleep alone.

The discrepancy in sleep duration from both devices suggests that bed sensors may be more sensitive to subtle disturbances triggered by bed-sharing. This highlights the significance of utilising multiple measurement methods for a comprehensive evaluation of sleep quality. Additionally, the results indicated that the nocturnal sleep distribution for the control group is around 8 hours from

both measurement methods, which aligns with the NSF's suggested sleep time duration of 7-9 hours for adults. Moreover, the extended sleep durations within mood disorder groups aligns with findings that suggest an association between hypersomnia and depression, specifically amongst individuals with bipolar disorder. However, there are several limitation to this study. It should be considered that while the Kruskal Wallis test displays a significant overall difference between the groups, it does not account for within-subject variability. Furthermore, the study's focus on nocturnal sleep limits understanding of broader sleep quality factors affecting participants. The reliance solely on nocturnal duration neglects important metrics such as sleep architecture and wakefulness.

Keywords actigraph, bed sensor, mood disorder, sleep quality, sleep duration

urn https://aaltodoc.aalto.fi

Contents

1. Introduction

Sleep and health are interconnected and can affect both body and mind. Low quantity and quality of sleep are significant components that increase the probability of type 2 diabetes [1]. Additionally, a study [2] revealed that prolonged sleep disruption predicts an increase in the risk of diseases, such as hypertension. Another study has linked shorter sleep duration to decreased heart rate and blood pressure recovery after exposure to stressful stimuli [3]. On the other hand, poor sleep can double the probability of experiencing depression and anxiety, thereby increasing the risk of mental disorders [2].

Depressive disorders intricately manifest themselves as uneasiness, addiction to drugs, and behavioural disturbances [4]. In particular, the most widespread of these is major depressive disorder (MDD) [5]. Another long-term disorder is bipolar disorder (BD), which influences above 1% of the population [6]. Similarly, borderline personality disorder (BPD) is characterised by psychological issues, such as enhanced unpredictability, anger, and annoyance, as recognised by both MDD and BD [7]. Patients diagnosed with such mood disorders are susceptible to insomnia or are likely to sleep for prolonged hours [5]. For instance, a study [8] states that 65% of individuals with MDD experience sleep disturbances and report a decrease in nocturnal sleep hours throughout manic episodes. Likewise, patients diagnosed with BD were reported to encounter either sleeplessness or excessive daytime sleepiness, known as hypersomnia, during varying periods.

On this note, sleep quality is a complex phenomenon with various dimensions [9]. Previous studies on sleep quality have recognised that a standardised definition of sleep quality is lacking, which poses a challenge in sleep research [10]. To address this issue, studies have individually evaluated the effect of objective measures, such as polysomnography and

actigraphy, as well as subjective measures, such as questionnaires, to assess sleep quality [10], [11]. Nevertheless, the findings indicate a weak correlation between questionnaires like the Pittsburgh Sleep Quality Index (PSQI) score and polysomnographic measures of sleep quality in the elderly [11]. According to Krystal et al. [10], this inconsistency is likely due to the nature of questionnaires that assess typical sleep routines rather than the quality of specific nights.

A recent study by [9] introduced composite measures as a means to define sleep quality. Focused solely on objective sleep measures, they defined sleep quality as a combination of sleep efficiency, sleep latency, the number of awakenings longer than 5 minutes, and wake after sleep onset. However, this may overlook the subjective experience of sleep quality. Hence, it is necessary to analyse the connection between subjective sleep quality and objective sleep variables. Additionally, further research is required to assess composite measures for evaluating sleep quality [9].

Therefore, this paper aims to validate previous research findings by identifying key indicators of sleep quality. This is achieved by using longitudinal data from the Mobile Monitoring of Mood (MoMo-Mood) study [12]. Additionally, the paper aims to evaluate the nocturnal sleep duration metric of sleep quality across different mood disorder patients and control groups using two frequently applied measurement methods—actigraphy and bed sensors. Recognising the differences in the captured metrics could assist in evaluating the reliability of actigraphy and bed sensors, thus leading to an accurate definition of sleep quality. Furthermore, compatible results between actigraphy and bed sensors would strengthen the credibility of the provided figures. Such consistencies are critical for validating the measurement of sleep quality in future studies. Moreover, the thesis seeks to understand whether subjective factors, such as survey responses, could influence these measurements and help recognise the strengths and constraints of these methods in recording accurate sleep data. Overall, this thesis seeks to yield insights for both clinical and personal health applications and may contribute to interventions that could lead to improved health outcomes.

The remainder of the thesis is divided into five sections. Section 2 provides the methods and indicators to assess sleep quality. Section 3 presents the results that are analysed in Section 4, along with the limitations of the study. Lastly, Section 5 summarises the findings of this paper and compares it to existing research.

2. Literature review

The rise in wearable technology has led to the creation of various advanced and affordable consumer devices. These devices gather data from sensors to provide insights into users' behaviours and sleep patterns by measuring bio-signals [13]. The MoMo-Mood Pilot study [12] demonstrates the integration of these technologies in research. For the purposes of this study, Section 2.1 focuses on actigraphy, bed sensors, questionnaires, and experience sampling (ESM) as methods to assess sleep quality. Moreover, Section 2.2 provides a summary of the sleep-related results from this study.

Furthermore, it is necessary to identify key indicators across different stages of life to quantify sleep quality. The National Sleep Foundation (NSF), a U.S.-based organisation, supported by a panel of experts, conducted a comprehensive framework for assessing sleep quality [9]. In Section 2.3, NSF's findings are consolidated with other studies to highlight primary indicators and provide a broad overview of the metrics used to assess sleep quality.

2.1 Methods to Assess Sleep Quality

Sleep research increasingly utilises mobile and wearable technologies due to the nature of data that they produce, both for the general population [14] as well as for patients with mental disorders [15]. Continuous data collection allows researchers to investigate the relationship between sleep patterns and daytime activities [13]. This allows for the accumulation of extensive passive data without interfering with users' daily routines.

2.1.1 Objective Measures

Actigraphy

According to multiple studies, actigraphy is a suitable tool to document and examine sleep-related physical activity, providing a convenient and economical method to track sleep-wake patterns in natural settings over time [16], [17], [18]. Actigraphy employs compact, computerised devices attached to the body and is highly beneficial for observing group variations, sleep patterns, and different interventions impacts, such as lifestyle changes [16].

A study by [17] revealed that actigraphy measurements were comparable to the gold standard, polysymnography, achieving 86.3% accuracy in sleep classification. They argued that wrist-worn actigraphy along with established algorithms effectively measured the duration of sleep and periods of wakefulness after falling asleep. The study also suggested that actigraphy provided higher accuracy for sleep-wake pattern estimates than self-reports in longitudinal research. Conversely, another study by [18] highlighted the challenge of low precision in observing wake after sleep onset with actigraphy. Therefore, to validate results, researchers used sleep diaries to identify disturbances [18], or alternatively, they recommend maintaining detailed daily logs and extended recording periods [16].

Ballistocardiography

Ballistocardiography (BCG) bed sensors are non-intrusive devices that monitor cardiovascular activities by measuring body movement caused by heartbeats and blood flow [19]. According to a study by [20], bed sensors that track movement detected voluntary motion with a mean error rate of 3.22% (±0.54). They concluded that the performance of the detector remained consistent regardless of aspects such as weight or bed properties, rendering this method both simple and robust.

Another study by [21] explores the applications of BCG bed sensors in sleep staging by combining heart-beat interval (HBI) and motion data. The authors examined the movement features to identify and assess episodes of wakefulness for sleep classification. This cost-effective method demonstrated reliable performance in conducting sleep staging with a high degree of accuracy.

2.1.2 Subjective Measures

Patient Health Questionnaire-9 (PHQ-9)
The Patient Health Questionnaire-9 (PHQ-9) is a framework for depression that consists of nine items with criteria for diagnosing major depression. The PHQ-9 scores each item from zero to three, deriving a total score from zero to 27 [22].

One specific item of the PHQ-9, item 3, which assesses sleep problems, has been successful in recognising sleep disturbances in primary care settings. Research on PHQ-9 indicates that a score of one or higher on item 3 effectively identifies patients with sleep issues, demonstrating high sensitivity (82.5%) and precision (84.5%) [23]. However, while the sleep item from PHQ-9 is useful, studies suggest that it is not sufficiently comprehensive for a thorough sleep assessment [24].

Experience Sampling Method
The Experience Sampling Method (ESM) is a tool that demands participants document their thoughts, feelings, and activities during or just after an experience [25]. The use of ESM is applicable for sleep studies, as it enables researchers to observe real-time interaction between sleep and wakefulness in participants' natural environments. By encouraging participants to report their experiences multiple times a day, ESM reduces recall bias and offers consistent data across different participants during the same real-life periods [26].

In the MoMo-Mood Pilot study [12], participants responded to ESM questionnaires on their phones, providing morning assessments on their sleep from the previous night and evening evaluations on their daily activities. This approach enhanced the accuracy and significance of the findings from their study.

2.2 Findings from MoMo-mood Pilot Study

The MoMo-Mood Pilot study [12], [27] explored the feasibility of using mobile applications and wearable devices to collect data from participants with major depressive disorder (MDD). The research involved 14 patients and 22 controls. The sleep-related findings of the study revealed that the median rest duration for patients diagnosed with MDD was higher compared to healthy controls for both devices. Moreover, there were some

outliers in the data which indicated variability among individuals within each group. Additionally, the actigraphy measurements of the sleep fragments from the total rest duration revealed that the control group mostly had a fraction of sleep close to 1, suggesting that when they were resting, they were likely sleeping. On the other hand, the patients displayed a wider range of fractions, indicating that they had longer resting times without necessarily sleeping. Moreover, the results also exhibited a significant difference in rest duration measurements between the actigraph and bed sensor devices for both participant types. According to [12], the devices may have variations in their rest-detection algorithms, leading to a dissimilarity in the results.

2.3 Indicators of Sleep Quality

2.3.1 National Sleep Foundation's Sleep Quality Recommendations: First Report

The panel identified 12 practical measurements of sleep quality, which were categorised into three groups in the NSF report [9].

1. Sleep continuity variables including awakenings greater than 5 minutes, sleep efficiency, sleep latency, and wake after sleep onset.

2. Sleep architecture variables including arousal, N1 sleep, N2 sleep, N3 sleep, and rapid eye movement (REM) sleep.

3. Nap variables covering days per week with at least one nap, nap duration, and number of naps per 24 hours.

According to the report [9], REM activity and number of awakenings were considered suitable measures of sleep quality for newborns. Additionally, N3 sleep was unresolved, and the remainder of measures were regarded as inappropriate and were excluded. For infants, N1 sleep and N2 sleep were deemed unsuitable, while nap duration and frequency remained uncertain but were included among other suitable measures. In all other age categories, all identified indicators were considered relevant. Table 2.1 summarizes the indicators considered by the panel, sourced directly from the NSF report [9].

2.3.2 Analysis of Sleep Quality Metrics and Indicators

The subsection focuses on analysing continuity measures which consistently reflect sleep quality across all age groups, as identified by the NSF [9]. Conversely, inconsistent data and limited consensus among experts in that study hinder the analysis of sleep architecture metrics and nap indicators. For example, although it was concluded that raised levels of REM sleep in adults does not suggest positive sleep quality, the authors of this study stated that further research is required to recognise the affect of nap duration on sleep quality across life stages.

Sleep Latency

Sleep latency is the period between when the subject switches off the lights and falls asleep. As specified by [28], sleep latency indicates whether the corresponding bedtime aligns with the patient's usual bedtime schedule. They examined that if bedtime is set too early or too late, latency will appear longer or shorter, respectively. Therefore, it is crucial to match bedtime with the patient's regular schedule for accurate results. According to the NSF [9], a sleep latency ≤ 15 minutes generally implies good sleep quality across every age bracket, whereas a latency exceeding 60 minutes suggests low sleep quality.

Awakenings > 5 Minutes

Recurrent episodes of awakenings exceeding 5 minutes can impact sleep quality. Across every age group, having one or fewer awakenings per night is regarded a sign of good sleep quality, whereas having 4 or more awakenings per night is recognised to be an indicator of poor sleep quality. Specifically for teenagers, having greater than 3 awakenings per night evidences to poor sleep quality [9].

Wake After Sleep Onset (WASO)

WASO measures periods of wakefulness after transitioning into sleep [28]. For all age brackets, a WASO of ≤ 20 minutes signifies good sleep quality, whereas ≥ 51 minutes suggests low sleep quality. Similarly, a WASO ≥ 41 minutes is also generally not representative of good sleep quality for all ages [9].

Sleep Efficiency

Sleep efficiency is the fraction of time a person sleeps while occupying the bed, which offers a measure of sleep quality that does not differentiate be-

tween recurring or short awakenings [28]. Higher sleep efficiency implies better sleep quality, with ≥ 85% considered suitable for sufficient sleep quality for all age ranges. In contrast, a sleep efficiency of ≤ 74% suggests poor sleep quality for most age brackets, while ≤ 64% demonstrates low sleep quality specifically for young adults [9].

Term	Measurement	Definition
Sleep efficiency	Percent (%)	Ratio of total sleep time to time in bed
Sleep latency	Minutes	Length of time, in minutes, it takes to transition from wake to sleep
REM sleep[a]	Percent (%)	Ratio of time spent in REM sleep to total sleep time
N1 sleep[a]	Percent (%)	Ratio of time spent in N1 sleep to total sleep time
N2 sleep[a]	Percent (%)	Ratio of time spent in N2 sleep to total sleep time
N3 sleep[a]	Percent (%)	Ratio of time spent in N3 sleep to total sleep time
Naps	Number (#)	Number of naps per 24-h period
Nap duration	Minutes	Average length of each nap, in minutes
Nap frequency	Days	Number of days, in the past 7, that a nap occurred
Arousals	Number per hour	An abrupt change from “deeper” stage of NREM sleep to a “lighter” stage, or from REM sleep toward wakefulness, with the possibility of awakening at the final outcome. May be accompanied by increased tonic electromyographic activity and heart rate, as well as by an increased number of body movements
Awakenings (>5min)	Number per night	Number of episodes, per night, in which an individual is awake for greater than 5 min
Wake after sleep onset	Minutes	Amount of time, in minutes, spent awake after sleep has been initiated and before final awakening

[a] REM, N1, N2, and N3 sleep are acronyms commonly defined in sleep medicine.

Table 2.1. Indicators considered by the panel. Adapted from [9]

3. Methodology

This section focuses on the methodological aspects of this research. Section 3.1 describes the data collection processes of the study by [27]. Additionally, section 3.2 discusses the details of the recruitment of participants and the stages of the longitudinal study. Finally, section 3.3 and 3.4 provides a description of the dataset, and outlines the steps for data processing.

3.1 Data Collection

The Mobile Monitoring of Mood (MoMo-Mood) study, was a joint effort by Aalto University, Helsinki University Central Hospital, and University of Helsinki. It focused on the potential of wearable devices to track behaviours and emotional states in individuals with psychiatric conditions [12], [29], [27]. The study unfolded in two distinct phases [12]. The active phase spanned two weeks in which participants had to report daily mood data through both subjective and objective measures. On the other hand, the passive phase lasted a year, with the option for participants to discontinue at any moment. This phase involved collecting data from smartphone sensors and conducting monthly PHQ-9 assessments [29].

For comprehensive data collection, several devices were utilised in addition to questionnaires and ESM (experience sampling) [30]:

- Philips Actiwatch 2: These actigraphy devices were worn on the wrist to monitor activity levels and sleep patterns over two weeks without the need to recharge.

- Murata SCA11H: These ballistocardiography based sensors measured physiological data, with data transmitted via a pre-arranged WiFi router.

- AWARE framework [31]: Smartphones were equipped with the specialized AWARE application, which collected extensive smartphone usage data.

Finally, Aalto University's Niima platform was utilised to compile the dataset [27].

3.2 Participant Recruitment and Study Procedure

In the MoMo-Mood study, healthy controls and patient cohorts were recruited [12]. First, healthy controls with no psychiatric conditions were enlisted using email lists from universities and student health services. Second, participants diagnosed with psychiatric disorders were recruited from the outpatient clinics at Helsinki University Hospital. To be eligible for the study, individuals were required to have a diagnosis of major depressive disorder and a PHQ-9 score that is greater than or equal to 10. Furthermore, participants needed to have an Android smartphone to allow data collection through sleep sensors and WiFi.

The initial procedures involved providing participants with detailed briefing and obtaining their informed consent. Researchers then conducted interviews and had the participants complete questionnaires [29]. The active monitoring phase began with the installation of a specialised application on the participants' smartphones. After two weeks, participants transitioned into the passive monitoring phase.

3.3 Dataset Descriptions

The dataset utilised for this study involved 164 participants in total, with each categorised into four groups: 31 participants for the control group, 21 patients diagnosed with bipolar disorder (BD), 27 patients having borderline personality disorder (BPD), and 85 patients diagnosed with major depressive disorder (MDD) [32].

3.4 Data Processing

Data preprocessing steps such as the removal of missing information were attempted to acquire sleep variables from the actigraph and bed sensor

datasets. All timestamps before 3 PM were adjusted to the previous day to ensure that all nighttime activities were consistently captured on the same date. Thereafter, the data was chronologically sorted by the timestamps.

3.4.1 Sleep Periods

To identify sleep periods, the actigraphy data was filtered to recognise continuous periods of low activity using the ‘REST-S’ status (see Appendix A). For the bed sensors, sleep periods were recognised based on the occupancy status (status 1) that indicated sleep (see Appendix B). The algorithms included a five-minute threshold for continuity so that sleep periods within five minutes of each other were considered to be the same sleep period. The longest identified sleep period were designated as nocturnal sleep, while all other sleep periods were recognised as naps. Additionally, nocturnal sleep periods were filtered to be between 3 to 15 hours long.

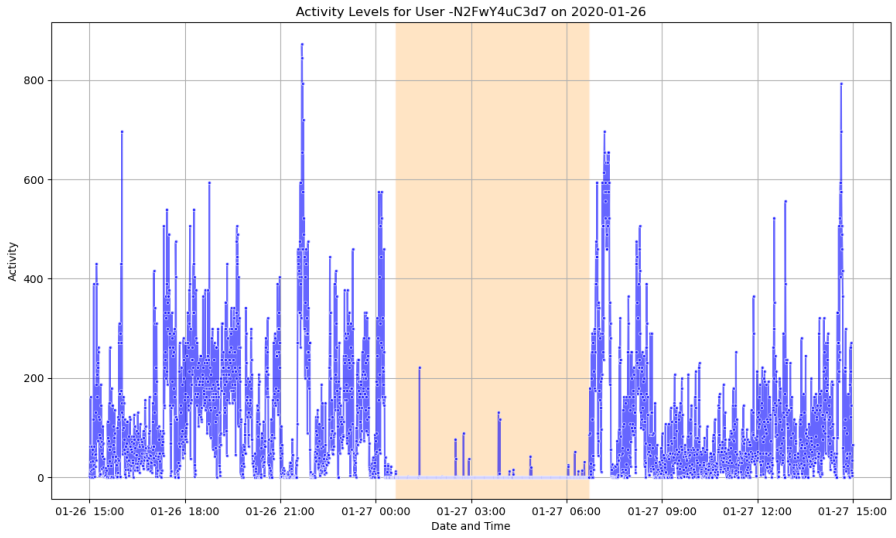


Figure 3.1. Daily activity levels for a sample user from actigraph. The cream-coloured band indicates a likely sleep period from approximately 01:00 on January 26, 2020, to 07:00 on January 27, 2020. During this period, activity levels are significantly lower, indicating minimal movement, which is consistent with sleep.

Figure 3.1 illustrates a user’s daily activity pattern from actigraphy data, showcasing a distinct sleep period that lasts approximately 6 hours. Moreover, the highlighted region indicates the time frame where the status is consecutively at ‘REST-S’, indicating low activity. Therefore, status is a reliable measure of nocturnal sleep periods.

On the other hand, figure 3.2 displays the change in status monitored by the bed sensor for the same user, on the same date. In accordance with the findings of Triana et al. [12] as discussed in section 2.2, there is a notable

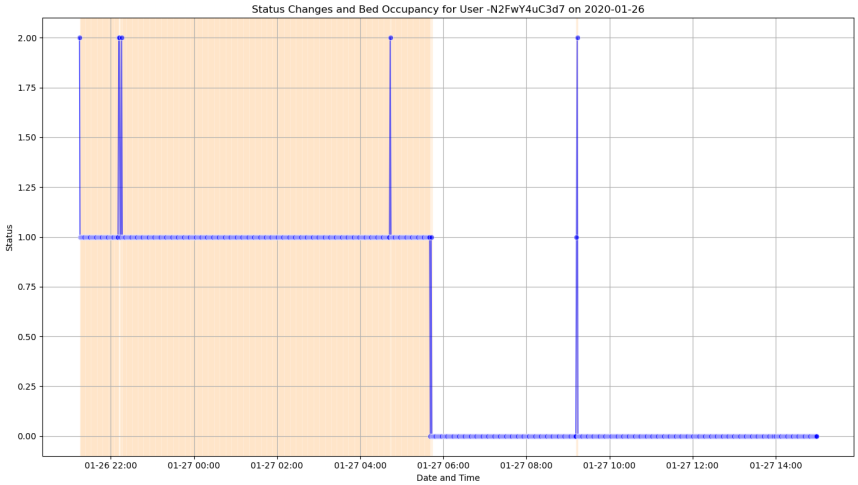


Figure 3.2. Status changes and bed occupancy for a sample user from bed sensor. The cream-coloured band indicates a likely sleep period from approximately 21:00 on January 26th, 2020, to 05:45 on January 27, 2020. During this period, the bed sensor recorded the user as being in a low activity state (status 1), which is consistent with sleep.

difference between the sleep duration acquired from both the devices. The nocturnal sleep period identified by the bed sensor is generally longer than the actigraph. It is hypothesised that this variation could be caused by the algorithms' inconsistencies in detecting when pets or people other than the participant lays on the same bed.

3.4.2 Sleep Variables

Awakenings during the nocturnal sleep period were identified by tracking transitions between sleep and wake states. The algorithm recognised periods where the status changed from asleep to a non-sleep state and back to asleep. If the duration of the awakening was more than 5 minutes, it was counted as an awakening. Similarly, WASO was calculated by summing up the duration of wake periods within the identified nocturnal sleep period.

For the actigraph data, sleep latency was calculated by detecting the transition from lights off to the start of the nocturnal sleep period. According to [32], a light source is considered to be switched on if the luminance is over 40 lux. Therefore, the duration between which the luminance value was lower than 40 and the start of the nocturnal sleep was observed as the sleep latency. Conversely, sleep latency was calculated as the time period from which the bed was engaged when the user was awake (state 2) and the start of the sleep period.

Other sleep metrics such as sleep effectiveness were computed as the difference between the duration of nocturnal sleep and WASO. Sleep efficiency was then computed as the ratio of effective sleep to nocturnal

sleep. After calculating these metrics, the results were merged into a single data-frame for each participant group. Finally, rows with zero hours of nocturnal sleep were filtered out.

3.4.3 Survey Data

The survey questions and responses, originally in Finnish, were mapped to their English equivalents. Additionally, each user in the survey data was categorised into their respective groups (control, MDD, BD, BPD) to allow for group-specific analysis. Finally, the data was filtered based on relevant survey responses to analyse the relationships between different variables and methods.

3.5 Statistical Tests and Models

The Kruskal Wallis H statistic test [33] is typically used to test whether independent samples originate from identical distributions. Additionally, the non-parametric nature of this test prevents any normality assumptions of the data. In this thesis, the Kruskal Wallis tests are conducted at a significance level of 0.05 to determine if there are statistically significant differences between the nocturnal sleep duration among the four groups: control, MDD, BPD, and BD. The null hypothesis states that all independent groups have the same central tendency and therefore come from the same population, revealing no significant differences in their rank sums.

Furthermore, the independent t-test, at a significance level of 0.05, is conducted to compare the differences in nocturnal sleep hours due to pet ownership and co-sleeping. The test differentiates the means of the two groups when accompanied by or in absence of the specific living arrangements of the participants. For this test, the null hypothesis states that the means of both groups are the same, establishing no significant difference between the sleep duration. Moreover, a linear regression model is designed to investigate whether there is a difference in sleep duration as measured by a bed sensor compared to the actigraph, depending on whether individuals sleep alone or not.

4. Results

This section presents the results of the study. Section 4.1 focuses on the analysis of sleep patterns obtained from the both the actigraph and bed sensor datasets. Section 4.2 presents the statistical tests used to investigate the significance of sleep variations among participants with mood disorders. Finally, section 4.3 provides statistical tests to identify the effect of pets and co-sleeping on the measurements captured by the actigraphs and bed sensors.

4.1 Evaluating Sleep Patterns: Actigraphy vs. Bed Sensor Data

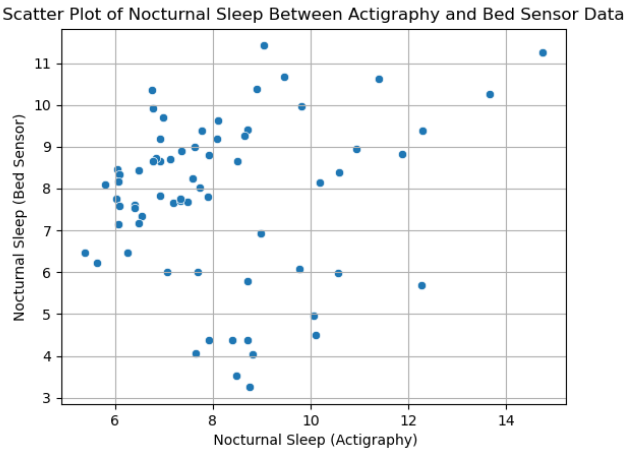
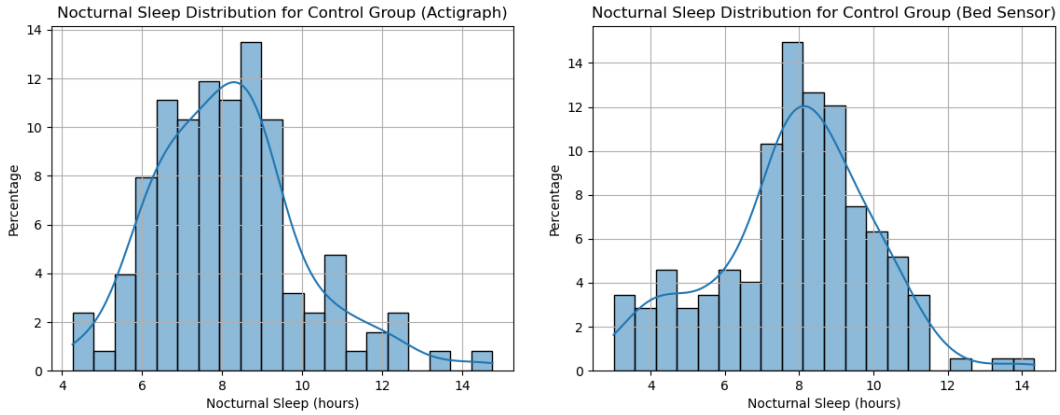


Figure 4.1. Comparison of nocturnal sleep hours measured by actigraph and bed sensors for the control group.

Figure 4.1 displays a low positive correlation (r = 0.14) between the nocturnal sleep hours measured by actigraphy and bed sensors. A notable cluster of points is around 7-9 hours on both axes, indicating a consistent sleep duration range for all groups measured by both devices. However, there is a noticeable dispersion in the data as the sleep duration measured by the two methods differs significantly for some users. Moreover, the

points are spread more widely along the x-axis (actigraphy) compared to the y-axis (bed sensor), suggesting that actigraphy measurements might show more variability than bed sensor measurements.

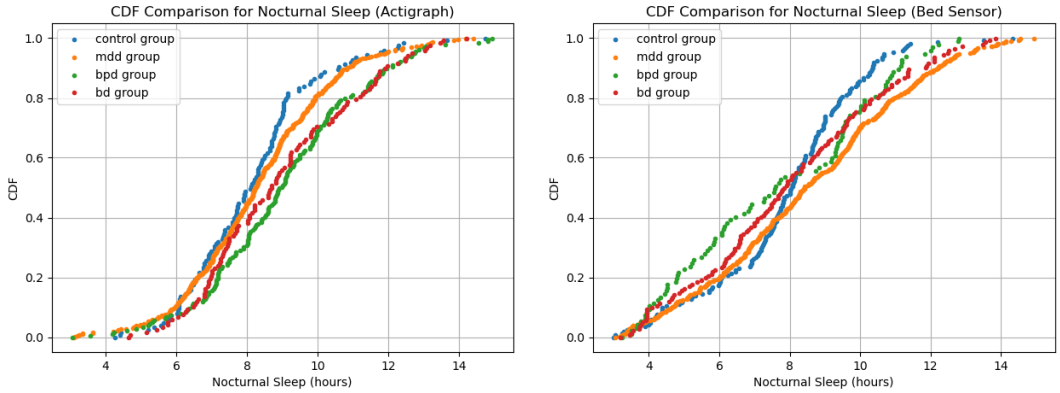


(a) Measured by actigraph **(b)** Measured by bed sensor

Figure 4.2. Distribution of nocturnal sleep hours for the control group as measured by actigraph and bed sensor

Similarly, in figure 4.2 both methods show the mean sleep duration for the control group is around 8 hours, indicating consistency between the two measurement techniques in capturing average sleep duration. Both distributions are right-skewed. However, the actigraphy data shows a more pronounced tail towards the right, indicating a slightly higher variability in capturing longer sleep duration.

4.2 Nocturnal Sleep Differences Across Various Mood Disorders



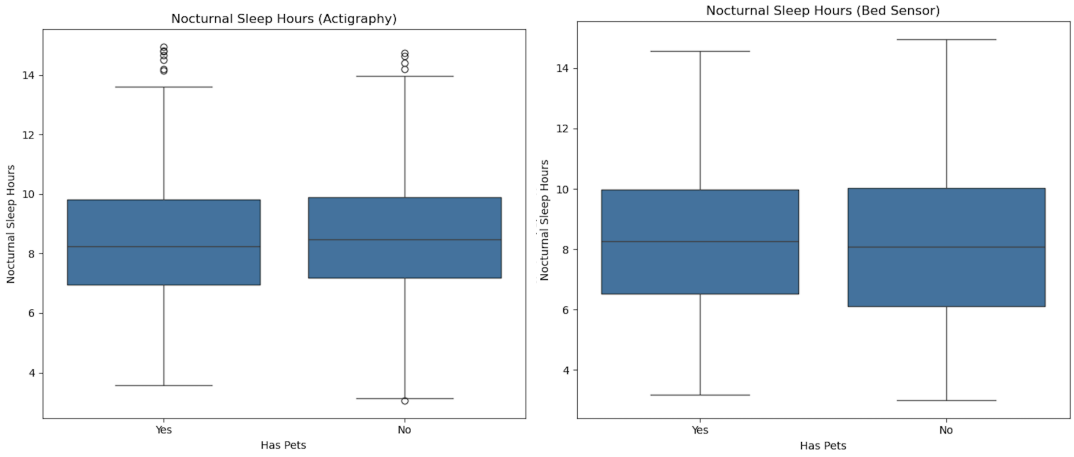
(a) Measured by actigraph **(b)** Measured by bed sensor

Figure 4.3. Distribution of nocturnal sleep hours for the control group and patients with different depressive disorders as measured by actigraph and bed sensor

Figure 4.3 indicates that the control group consistently has shorter nocturnal sleep durations in comparison to the patient groups. Additionally,

the actigraphy data shows a slightly steeper rise in the cumulative distribution function (CDF) for the control group compared to the bed sensor data. There are also variations in the positioning of the CDF curves between the two methods, indicating some dissimilarity in sleep measurement between actigraph and bed sensors. Moreover, the statistical tests suggest that the distinction in sleep duration among the groups are statistically significant for the actigraph (H-statistic = 22.101, p-value < 0.001) and the bed sensor (H-statistic = 11.105, p-value = 0.01). However, the actigraph data exhibits a stronger significance compared to the bed sensor data, indicating more pronounced variations among the groups.

4.3 Effect of Pets and Co-Sleeping on Sleep Metrics

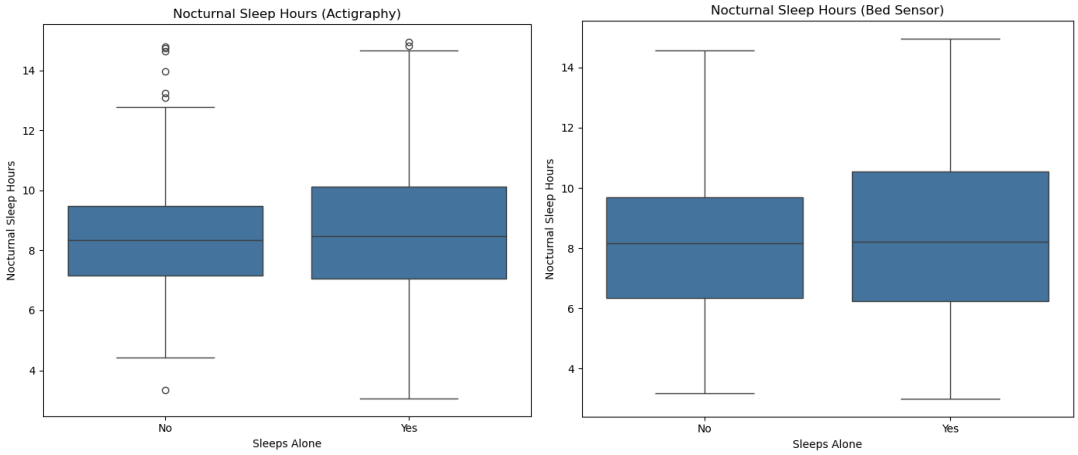


(a) Measured by actigraph **(b)** Measured by bed sensor

Figure 4.4. Comparison of nocturnal sleep hours for individuals who have pets versus those who do not, as measured by actigraph and bed sensor

The box plots in figure 4.4 display similar median sleep duration for participants with and without pets, when measured by both actigraphs and bed sensors. This indicates that there is minimal difference in sleep duration based on pet ownership. The results indicated no significant difference in sleep duration with the actigraph (t-statistic = -0.518, p-value = 0.60) and the bed sensor (t-statistic = 0.999, p-value of 0.32).

Likewise, the box plots in figure 4.5 display similar median sleep duration for participants who sleep alone and those who do not, when measured by actigraphs and bed sensors. In particular, when measured by a bed sensor, the median sleep duration is slightly higher and the plot has a broader range for those who sleep alone. The statistical test revealed no significant difference in sleep duration for the actigraph (t-statistic = 1.632, p-value



(a) Measured by actigraph **(b)** Measured by bed sensor

Figure 4.5. Box plot comparing nocturnal sleep hours for individuals who sleep alone versus those who do not, as measured by actigraph and bed sensor

= 0.10). However, it indicated a statistically significant difference with the bed sensor (t-statistic = 1.989, p-value of 0.05). This indicates that individuals who sleep alone have slightly longer sleep duration with bed sensors.

Linear Model of Sleep Duration Difference Based on Sleeping Alone

	Sleep Duration Difference (Bed Sensor - Actigraph)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.08	-0.48 – 0.33	0.712
sleep_alone	-0.78	-1.30 – -0.25	**0.004**
Observations	450		
R^2 / R^2 adjusted	0.019 / 0.017		

Figure 4.6. Linear model analysis of sleep duration difference based on sleeping alone

Furthermore, the model in figure 4.6 reveals that individuals who sleep alone have a significantly lower difference in sleep duration between the two devices. Specifically, the estimate for sleeping alone is -0.78 with a 95% confidence interval of -1.30 to -0.25 (p-value = 0.004).

5. Discussion

This section delves into the significance and relevance of the results.

In accordance with the findings from the MoMo-Mood Pilot study [12], the results reveal a notable deviation between the nocturnal sleep duration of the actigraph and the bed sensor. In addition to a low correlation, the scatter plot reveals that the metrics from the actigraph are generally higher than the duration observed by the bed sensor. This finding aligns with previous studies [34], [35] that have revealed the tendency of actigraphs to overestimate sleep duration. Additionally, the results indicate that the nocturnal sleep distribution of the control group is approximately 8 hours for both measurement devices. This finding is in line with the NSF's recommended sleep time duration of 7-9 hours for adults [36]. Hence, the variability and skewness from the model are diversions that indicate probable concerns in sleep quality. Furthermore, healthy participants exhibited shorter sleep duration in contrast to participants with mood disorders, which is in line with the findings by [37]. These results indicate an association between hypersomnia and depression, specifically amongst individuals with BD.

Moreover, the hypothesis that pet ownership affects sleep measurements was not supported by the results. Both devices were able to reliably differentiate between pets and the users, leading to reliable readings. On the other hand, the hypothesis concerning the effect of co-sleeping on sleep duration had interesting results. Participants who slept alone had marginally extended sleep duration with the bed sensors, but no deviations were found with the actigraphs. Additionally, the model suggests that individuals who sleep alone tend to have a lower difference in sleep duration between the bed sensor and actigraph compared to those who do not sleep alone. This indicates that co-sleeping has a significant effect on the discrepancies between the two measurement methods. This suggests that

bed sensors may be more sensitive to subtle disturbances triggered by bed-sharing, highlighting the significance of utilising multiple measurement methods for a comprehensive assessment of sleep quality.

6. Conclusion

This thesis aimed to explore sleep patterns across healthy users and patients with mood disorders, as well as to compare the reliability of actigraphy and bed sensors in the measurement of sleep quality. Results indicated that the healthy users had consistently shorter sleep durations than the patients, with notable variation in sleep patterns. When metrics from the two devices were compared, actigraphy presented more variability and less sensitivity to transitions in sleep compared to bed sensors. Furthermore, the hypothesis that pet ownership or co-sleeping could affect the measurements was partly supported as bed sensors displayed some sensitivity to co-sleeping, but not to pet ownership. Generally, bed sensors were more responsive and displayed more accuracy in detecting movements in bed. Nonetheless, all recorded sleep disturbances were not directly attributed to changes in sleep quality.

An unexpected finding was the discrepancy in nocturnal sleep duration between the actigraph and bed sensor, which could be due to variations in the algorithms or responsiveness of the devices. The study also has various limitations. The scope was limited since the thesis was centred on the nocturnal sleep duration metric of sleep quality. While sleep duration is the primary metric that is used to determine other variables, like sleep efficiency, it is also crucial to examine further sleep variables for a comprehensive analysis. Additionally, the Kruskal-Wallis H-statistic test and the independent T-statistic test were conducted with a 0.05 significance level, which is arbitrary to some degree. So, it is important to note that different thresholds could yield distinct interpretations of the data. Moreover, the accuracy of the Kruskal-Wallis test is limited as it may not account for within-subject variability.

Future research could use a more extensive approach by including more sleep metrics for a better analysis of overall sleep quality. Additionally, the

results could be compared with other existing sleep measurement devices, such as polysomnography, to validate the findings. Finally, the study could have a more diverse and greater sample size to improve the generalisability of the results.

Bibliography

[1] F. P. Cappuccio, L. D'Elia, P. Strazzullo, and M. A. Miller, "Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis," *Diabetes care*, vol. 33, no. 2, pp. 414–420, 2010.

[2] Y. S. Bin, "Is sleep quality more important than sleep duration for public health?" *Sleep*, vol. 39, no. 9, pp. 1629–1630, 2016.

[3] E. J. Mezick, K. A. Matthews, M. H. Hall, J. Richard Jennings, and T. W. Kamarck, "Sleep duration and cardiovascular responses to stress in undergraduate men," *Psychophysiology*, vol. 51, no. 1, pp. 88–96, 2014.

[4] R. E. Jorge, "Mood disorders," *Handbook of clinical neurology*, vol. 128, pp. 613–631, 2015.

[5] R. M. Benca, M. Okawa, M. Uchiyama, S. Ozaki, T. Nakajima, K. Shibui, and W. H. Obermeyer, "Sleep and mood disorders," *Sleep medicine reviews*, vol. 1, no. 1, pp. 45–56, 1997.

[6] I. Grande, M. Berk, B. Birmaher, and E. Vieta, "Bipolar disorder," *The Lancet*, vol. 387, no. 10027, pp. 1561–1572, 2016.

[7] M. Mneimne, W. Fleeson, E. M. Arnold, and R. M. Furr, "Differentiating the everyday emotion dynamics of borderline personality disorder from major depressive disorder and bipolar disorder." *Personality Disorders: Theory, Research, and Treatment*, vol. 9, no. 2, p. 192, 2018.

[8] M. J. Peterson and R. M. Benca, "Sleep in mood disorders," *Psychiatric Clinics*, vol. 29, no. 4, pp. 1009–1032, 2006.

[9] M. Ohayon, E. M. Wickwire, M. Hirshkowitz, S. M. Albert, A. Avidan, F. J. Daly, Y. Dauvilliers, R. Ferri, C. Fung, D. Gozal *et al.*, "National sleep foundation's sleep quality recommendations: first report," *Sleep health*, vol. 3, no. 1, pp. 6–19, 2017.

[10] A. D. Krystal and J. D. Edinger, "Measuring sleep quality," *Sleep medicine*, vol. 9, pp. S10–S17, 2008.

[11] D. J. Buysse, C. F. Reynolds III, T. H. Monk, C. C. Hoch, A. L. Yeager, and D. J. Kupfer, "Quantification of subjective sleep quality in healthy elderly men and women using the pittsburgh sleep quality index (psqi)," *Sleep*, vol. 14, no. 4, pp. 331–338, 1991.

[12] A. M. Triana, A. Martikkala, I. Baryshnikov, R. Heikkilä, T. Alakörkkö, R. K. Darst, J. Ekelund, E. Isometsä, and T. Aledavood, "Mobile monitoring of mood (momo-mood) pilot: A longitudinal, multi-sensor digital phenotyping study of patients with major depressive disorder and healthy controls," *medRxiv*, pp. 2020–11, 2020.

[13] M. De Zambotti, N. Cellini, A. Goldstone, I. M. Colrain, and F. C. Baker, "Wearable sleep technology in clinical and research settings," *Medicine and science in sports and exercise*, vol. 51, no. 7, p. 1538, 2019.

[14] N. Luong, G. Mark, J. Kulshrestha, and T. Aledavood, "Sleep during covid-19 pandemic: Longitudinal observational study combining multisensor data with questionnaires," 2023. [Online]. Available: https://arxiv.org/abs/2310.01652

[15] T. Aledavood, J. Torous, A. M. Triana Hoyos, J. A. Naslund, J.-P. Onnela, and M. Keshavan, "Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders," *Current psychiatry reports*, vol. 21, pp. 1–9, 2019.

[16] C. Acebo and M. K. LeBourgeois, "Actigraphy." *Respiratory care clinics of North America*, vol. 12, no. 1, pp. 23–30, 2006.

[17] M. Marino, Y. Li, M. N. Rueschman, J. W. Winkelman, J. M. Ellenbogen, J. M. Solet, H. Dulin, L. F. Berkman, and O. M. Buxton, "Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography," *Sleep*, vol. 36, no. 11, pp. 1747–1755, 2013.

[18] L. J. Meltzer, H. E. Montgomery-Downs, S. P. Insana, and C. M. Walsh, "Use of actigraphy for assessment in pediatric sleep research," *Sleep medicine reviews*, vol. 16, no. 5, pp. 463–475, 2012.

[19] L. Giovangrandi, O. T. Inan, R. M. Wiard, M. Etemadi, and G. T. Kovacs, "Ballistocardiography—a method worth revisiting," in *2011 annual international conference of the IEEE engineering in medicine and biology society*. IEEE, 2011, pp. 4279–4282.

[20] A. M. Adami, M. Pavel, T. L. Hayes, and C. M. Singer, "Detection of movement in bed using unobtrusive load cell sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 481–490, 2009.

[21] J. M. Kortelainen, M. O. Mendez, A. M. Bianchi, M. Matteucci, and S. Cerutti, "Sleep staging based on signals acquired through bed sensor," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 776–785, 2010.

[22] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[23] K. L. MacGregor, J. S. Funderburk, W. Pigeon, and S. A. Maisto, "Evaluation of the phq-9 item 3 as a screen for sleep disturbance in primary care," *Journal of General Internal Medicine*, vol. 27, pp. 339–344, 2012.

[24] T. Schulte, D. Hofmeister, A. Mehnert-Theuerkauf, T. Hartung, and A. Hinz, "Assessment of sleep problems with the insomnia severity index (isi) and the sleep item of the patient health questionnaire (phq-9) in cancer patients," *Supportive Care in Cancer*, vol. 29, no. 12, pp. 7377–7384, 2021.

[25] N. Van Berkel, D. Ferreira, and V. Kostakos, "The experience sampling method on mobile devices," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–40, 2017.

[26] L. Mulligan, G. Haddock, D. Littlewood, and S. D. Kyle, "Experience sampling in the study of sleep and wakefulness," in *Experience Sampling in Mental Health Research*. Routledge, 2019, pp. 67–80.

[27] T. Aledavood, N. Luong, I. Baryshnikov, R. Darst, R. Heikkila, J. Holmen, A. Ikaheimonen, A. Martikkala, K. Riihimaki, O. Saleva *et al.*, "Mobile monitoring of mood (momo-mood): a multimodal digital phenotyping study with major depressive patients and healthy controls," *medRxiv*, pp. 2024–06, 2024.

[28] D. Shrivastava, "How to interpret the results of a clinical polysomnogram," *Indian Journal of Sleep Medicine (IJSM)*, vol. 7, no. 4, 2012.

[29] A. Ikäheimonen, N. Luong, I. Baryshnikov, R. Darst, R. Heikkilä, J. Holmen, A. Martikkala, K. Riihimäki, O. Saleva, E. Isometsä *et al.*, "Predicting and monitoring symptoms in diagnosed depression using mobile phone data: An observational study," *medRxiv*, pp. 2024–06, 2024.

[30] I. Baryshnikov, T. Aledavood, T. Rosenström, R. Heikkilä, R. Darst, K. Riihimäki, O. Saleva, J. Ekelund, and E. Isometsä, "Relationship between daily rated depression symptom severity and the retrospective self-report on phq-9: A prospective ecological momentary assessment study on 80 psychiatric outpatients," *Journal of Affective Disorders*, vol. 324, pp. 170–174, 2023.

[31] D. Ferreira, V. Kostakos, and A. K. Dey, "Aware: mobile context instrumentation framework," *Frontiers in ICT*, vol. 2, p. 6, 2015.

[32] A. Gunay *et al.*, "Modelling of sleep behaviors of patients with mood disorders," Master's thesis, 2023.

[33] E. Ostertagova, O. Ostertag, and J. Kováč, "Methodology and application of the kruskal-wallis test," *Applied mechanics and materials*, vol. 611, pp. 115–120, 2014.

[34] L. de Souza, A. A. Benedito-Silva, M. L. N. Pires, D. Poyares, S. Tufik, and H. M. Calil, "Further validation of actigraphy for sleep studies," *Sleep*, vol. 26, no. 1, pp. 81–85, 2003.

[35] L. J. Meltzer, L. S. Hiruma, K. Avis, H. Montgomery-Downs, and J. Valentin, "Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents," *Sleep*, vol. 38, no. 8, pp. 1323–1330, 2015.

[36] M. Hirshkowitz, K. Whiton, S. M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E. S. Katz, L. Kheirandish-Gozal *et al.*, "National sleep foundation's sleep time duration recommendations: methodology and results summary," *Sleep health*, vol. 1, no. 1, pp. 40–43, 2015.

[37] A. Murru, G. Guiso, M. Barbuti, G. Anmella, N. Verdolini, L. Samalin, J.-M. Azorin, J. J. Angst, C. L. Bowden, S. Mosolov *et al.*, "The implications of hypersomnia in the context of major depression: results from a large, international, observational study," *European Neuropsychopharmacology*, vol. 29, no. 4, pp. 471–481, 2019.

[38] R. Darst and contributors, “Murata bsn configuration information,” https://github.com/digitraceslab/koota-server/wiki/MurataBSN, 2020, accessed: 2024-07-18.

A. Dataset Description for Philips Actiwatch

Attribute	Type	Description
user	object	Anonymous user id
device	object	Device identifier
time	float64	Timestamp
time_str	object	Human-readable timestamp
line	int64	Line number in the dataset
activity	float64	Count of activity by means of built-in accelerometer
marker	int64	Marker for the data point
white light	float64	Illuminance measure (lux/x^2)
sleepwake	float64	Measure related to sleep/wake status
intervalstatus	object	Active/Rest status based on an algorithm: **ACTIVE**: High activity, user is awake **REST**: Low activity, user is resting **REST-S**: Low activity, user is likely to be sleeping

Table 1.1. Adapted from [32]

B. Dataset Description for Murata SCA11H

Attribute	Type	Description
time	float64	The time reported by the sensor (updated every six hours).
hr	float64	Heart rate measured in beats per minute.
rr	float64	Respiration rate measured in breaths per minute.
sv	float64	Heart stroke volume, a dimensionless unit with no absolute scale.
hrv	float64	Heart rate variance in milliseconds.
ss	int64	Signal strength
status	int64	Bed occupancy test status: **0**: Low signal, the user is not in bed **1**: OK signal, the bed is occupied and the user is likely to be sleeping **2**: High signal, the bed is occupied but the user is likely to be awake **3**: Signal overload, the measured heart rate is near maximum (120 bpm)
btt0, btt1, btt2	float64	Beat-to-beat time in milliseconds between the last two heartbeats.
time2	float64	An alternate time calculation based on data upload time; generally equivalent to 'time'.

Table 2.1. Adapted from [38]

C. Dataset Description for Survey Data

Attribute	Type	Description
user	object	Anonymous user id
device	object	Identifier of the device used to capture the data
id	object	Unique identifier for the record
access_time	int64	Access time in seconds since epoch
time	int64	Timestamp in seconds since epoch
question	object	Question asked in the survey
answer	object	Answer provided by the user (optional)
order	int64	Order of the question in the survey
choice_text	object	Text of the choice selected by the user

Table 3.1. Dataset description for survey data