January 2024

Exploring The Factors Of Sleep Quality: An Integrated Analysis Of Data From Questionnaires And Environmental Sensor Data

Jiaqian Xu
jiaqian.kx@gmail.com

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Exploring the Factors of Sleep Quality: An Integrated Analysis of Data from Questionnaires and Environmental Sensor Data

Kristine Jiaqian Xu

Abstract

Sleep quality plays a crucial role in human health, but its determinants remain incompletely understood. This integrated analysis investigates the relationships of demographic characteristics, living habits, as well as psychological factors on sleep quality. It also ranks the importance of sleep discomfort and examines the effect of bedroom air quality on sleep quality. Results showed stress had a significant negative correlation (r = -0.66) with sleep quality. Furthermore, nocturnal bathroom usage was the most influential predictor, contributing approximately 17.5% to sleep quality ratings using random forest classification. Air quality measures, including carbon dioxide (CO₂) levels, temperature, relative humidity, and particulate matter (PM2.5), demonstrated no significant impact on sleep quality. This research builds a stepstone for more precise sleep intervention by looking at the interaction of impacting factors and creates methods to generate new features from real-time air quality sensor data.

1. Introduction:

Sleep plays a crucial role in human health, with insufficient sleep being linked to numerous chronic diseases. A comprehensive review of 41 studies demonstrated the detrimental effects of irregular sleep patterns on health. This review, which involved over 92,000 participants across 14
countries, summarised that inconsistent sleep habits are associated with adverse health outcomes like mortality and cardiovascular disease (Chaput et al., 2020). Furthermore, another systematic review and meta-analysis Lee et al. found that poor sleep quality correlates with higher hemoglobin A1c levels among type 2 diabetes indicating suboptimal blood sugar control based on an analysis encompassing 20 articles (Lee et al., 2017). A prospective study of 60,586 individuals found that poor sleep quality, characterized by restless sleep and difficulty in falling asleep, significantly heightens the risk of coronary heart disease (Lao et al., 2018). Inadequate sleep quality was also linked to an elevated risk of coronary heart disease based on a meta-analysis reviewing 74 studies (Kwok et al., 2018). Moreover, data from 26,553 subjects across various studies suggest poor sleep quality, independent of sleep duration, notably increases the likelihood of obesity and overweight in younger demographics (Fatima et al., 2016).

Additionally, poor sleep is tied to mental health challenges, such as depression. A systematic review identified various sleep disorders and disturbances in the body's circadian rhythm as potential precursors to developing depression (Zhang et al., 2022). Despite its clear importance, sleep and the factors influencing it are not completely understood.

Significant research has been done examining factors affecting sleep including demographic information, living habits, psychological condition and sleep air quality environment. However, these have all been studied in isolation. Mong & Cusmano showed that females have a higher percentage of insomnia compared to males for 13 years and older (Mong & Cusmano, 2016). Gu et al. demonstrated age is the most important demographic factor impacted by the self-reported sleep quality and duration among older adults (Gu et al., 2010). Some living habits have also been shown to impact sleep. For instance, Yazdannik et al. showed that there is a positive effect
of earplugs and eye mask use on patients’ perceived sleep quality in intensive care units (Yazdannik et al., 2014). According to Fan et al., window opening habits during the heating season significantly improved sleep quality in Danish homes (Fan et al., 2022). Psychological factors also impact sleep quality. In a study using self-reported questionnaires, higher stress levels are significantly associated with poor sleep quality (Alotaibi et al., 2020). Previous studies have also shown an impact on air quality and sleep. Xu et al. found an increase in indoor CO₂ levels is associated with a significant decline in sleep quality, evidenced by difficulty falling asleep quickly (long sleep onset latencies) and shorter durations of slow-wave sleep (Xu et al., 2020). The analysis of another study found that in a group of 62 participants tracked for two weeks with activity monitors and sleep logs, higher bedroom levels of air pollution (particulate matter <2.5 μm [PM2.5]), carbon dioxide, noise, and temperature were all linked independently to lower sleep efficiency. Through this work, we have gained insight into how some factors impact sleep individually, but their interplay remains to be investigated.

Previous studies used different ways to measure air quality in sleep studies which may lead to inconsistent results. Most studies used outside air quality data and limited studies used real-time air quality sensors. Even for studies using real-time sensors, most studies did not analyze the data beyond using the mean (Liu et al., 2020). For example, Shen et al. used outdoor ambient exposure monitoring stations to collect air quality data to study sleep-disordered breathing in urban Taiwan (Shen et al., 2018). Some studies only used self-reported questionnaire data on air quality or factors that affect sleep quality. Wei et al. used the self-report questionnaire to collect air quality data related to cooking oil fumes and studied its impact on sleep (Wei et al., 2017). These self-reported data may introduce significant bias into the data set (Althubaiti, 2016). A few
studies used indoor air quality monitoring. Lapparat et al. used personal air sampling and placed them in the bedroom for three consecutive days in each season. They then used the mean of air quality to study the severity of obstructive sleep apnea (Lapparat et al., 2018).

Factors that impact sleep are complex and diverse (Ying et al., 2023). Few studies have been able to assess the interplay between the factors that impact sleep quality. In order to assess the interplay between factors affecting sleep, in this study, we comprehensively collected information on participant demographics, living habits, psychological conditions and sleep quality from a cohort of 36 in Switzerland. We use a network analysis to examine the relationship between all these factors and sleep quality. Most previous studies used the outdoor air quality monitor to study sleep. However, in this cohort, people sleep indoors in their bedrooms. Therefore, we deployed a real-time sensor in each participant’s bedroom to measure the air quality including CO₂, relative humidity, temperature and PM2.5 to collect air quality measures. Beyond using the mean of air quality measures, we used statistical methods to measure the air changes per hour and the correlation of change between different air quality measures to study their impact on sleep. We also collect subjective reasons why participants have a difficult time falling asleep to further examine their impact on sleep quality and use Random Forest models to solve the non-linear relationship between sleep discomfort and sleep quality challenges. Our findings could help with building better sleeping health intervention strategies to empower individuals to optimize their sleep environment and habits for improved well-being.
2. Methods:

In this study, we reutilize the dataset from González-Serrano et al. and build upon the studies by Gonzalez et al and Vanfleteren to further investigate the dynamics of indoor air quality and sleep.

In Gonzalez et al. personal exposure monitors and stationery real environmental monitors tracked CO₂ levels, air temperature, relative humidity and PM 2.5 in 36 households. They collected three consecutive days of real-time sensor data and personal exposure monitor data in participants' bedrooms. Real-time sensor data was collected every minute (González-Serrano et al., 2023). Participants in each household also completed survey questions including demographics, behaviour, and sleep patterns (Vanfleteren, 2022).

Table 1 reveals the demographic information of this cohort of 36 participants. We can see that sex is relatively evenly distributed with 16 of female and 20 male participants. The majority of the participants have a master’s degree or higher and are 45 years old or younger.

<table>
<thead>
<tr>
<th>Category</th>
<th>Demographic</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Female</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>20</td>
</tr>
<tr>
<td>Educational Level</td>
<td>High School</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Bachelor’s</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Master’s</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>PhD or higher</td>
<td>11</td>
</tr>
<tr>
<td>Age Group</td>
<td>16-25 years old</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>26-35 years old</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>36-45 years old</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>46-55 years old</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>56-65 years old</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Demographic Information of 36 participants
All categorical variables with ordinal patterns have been coded with numeric values. For example, the demographic variable, “high school” is coded with 0 and “PhD or higher” is coded with 3 in Education Level. For the living habits and psychological factors, responses are frequencies. For example, for the Feel stress variable, participants are asked to select the frequency of feeling stressed. In this variable “do not feel stressed” is coded as 0, “occasionally feel stressed” is coded as 1 and “regularly feel stressed” is coded as 2. A visualization of the frequency distribution of living habits can be found in Appendix Figure App1. For the sleep discomfort variable, participants were asked to provide frequencies on different sleep discomfort for the past month which is also an ordinary variable that was replaced with numeric values. One exception is the demographic variable Sex which is not an ordinary variable was coded with “Male” as 0 and “Female” as 1 to understand its correlation with other variables such as demographics, living habits, psychological condition, and sleep quality rating.

2.1 Correlational Analysis of Demographic, Living Habits, and Psychological Factor on Sleep Quality

The interested survey variables are grouped into four categories: 1) Demographic: sex, age and education; 2) Living Habits (measured by frequencies): sleep with the door open, sleep with the window open, use essential oil or herbal tea, use ear plug while asleep, use eye mask while asleep, alcohol consumption, tea consumption, and exercise; 3) Psychological condition (measured by frequencies): feeling stressed; 4) Outcome Variable: sleep quality rating. We calculated their linear correlation and associated p values. We utilized a map of significant
correlation (p value < 0.05) given these four categories of variables to understand the pathway that impacts sleep quality.

2.2 Role of Sleep Discomfort Indicators on Sleep Quality Rating

To quantify the impact of sleep disturbance factors on sleep quality, we employed a Random Forest regression model. The model incorporated seven key features identified as potential sleep quality disruptors: disturbed by noise, feeling hot, experiencing pain, having bad dreams, disturbed by stuffy air, disturbed by using the bathroom and having difficulties staying awake. Random Forest models are ideal because (1) the Random Forest model is a better method to study the non-linear relationship between sleep discomfort indicators and sleep quality rating (Auret & Aldrich, 2012); (2) the Random Forest model can help with identifying which feature is the most important in the sleep quality prediction, then help us understand which factors are most important in determining the sleep quality.

2.3 Real-Time Sensor Data Feature Extraction

The first feature included the minimum, maximum, and mean values of temperature and relative humidity, as well as concentrations of PM2.5. These features were chosen to represent the variability and average conditions of the indoor environment both at night and over the entire day. In this study, nighttime is defined as 11 pm to 8 am to capture most participants' sleep duration based on the questionnaires.
The second feature is Spearman’s rank correlation between changes in CO$_2$, PM2.5, temperature, and relative humidity. The correlation was calculated based on the log of change of the four environmental quality indicators including CO$_2$, PM2.5, temperature, and relative humidity. We computed these Spearman's rank correlation coefficients using three temporal resolutions: 1 minute, 15 minutes, and 1 hour. We want to pick a stable time range of correlation and preserve as many data points as possible to study its association with sleep quality rating.

Lastly, we extracted the weighted average of air changes per hour (ACH) or air change rate to model the effectiveness of ventilation in the bedroom using CO$_2$ concentrations. We applied a rolling average using a time window of five minutes to smooth out short-term CO$_2$ level fluctuations. When the average CO$_2$ levels dropped, this indicated that fresh air was entering the room, diluting the CO$_2$ during “decay periods”. We assumed a constant level of outdoor CO$_2$ (400ppm) and used the formula below to calculate ACH:

$$\ln \left( \frac{c - c_{out}}{c_0 - c_{out}} \right) = -\frac{Q}{V} \cdot t$$

Here, $c$ is the indoor CO$_2$ concentration at the time $t$, $c_{out}$ is the outdoor CO$_2$ concentration, $c_0$ is the initial indoor CO$_2$ concentration, $Q$ is the ventilation rate, and $V$ is the volume of the room. The slope of the line $\frac{Q}{V}$ formed by plotting the natural logarithm of the adjusted CO$_2$ concentration over time gives us the rate at which CO$_2$ is removed from the room. Multiplying this slope by 3600 converts it into the air change rate per hour (ACH). We then used time-
weighted average ACH at night (11 pm to 8 am) to model the impact on sleep quality from environmental sensor measurement.

2.4 Statistical Analysis

In our analysis, logistic regression was employed to assess the influence of environmental features extracted from the real-time sensors on sleep outcomes. In the survey design, there are four subjective sleep quality ratings including sleep “Very bad”, “Fairly Bad”, “Fairly Good” and “Good” for the past month. The logistic model combines these measures into two categories “Sleep Bad” and “Sleep Good” as the values of outcome variable sleep quality rating. Predictor variables are features extracted in section 2.2.

3. Results

3.1 Pathways from Demographics, Behaviors and Psychological Condition to Sleep Quality
**Figure 1**: Network visualization of correlations among demographics, habits, psychological factors, and sleep quality rating. The nodes represent variables categorized by demographics (orange), habits and psychological factors (pink), and the central node for sleep quality rating (green). Edges indicate significant correlations, with positive correlations shown in red and negative correlations in blue. The correlation coefficients are labeled on the edge.

To comprehensively understand factors influencing sleep quality, we analyzed a network of interrelated demographic characteristics, lifestyle habits, psychological conditions, and their collective impact on sleep quality rating. From Figure 1, there are two clusters of correlation. The right cluster shows that sleeping with the window open and sex are negatively correlated, indicating that males are more likely to sleep with the window open. However, neither sleeping with the window open nor sex significantly correlates with sleep quality ratings. On the left side
of the figure, two demographic factors and four living habits and psychological factors are directly or indirectly connected. Most significantly, people with higher tea consumption also show a higher frequency of use of essential oils or herbal tea. This indirectly correlates with exercise and feeling stressed. As shown in Figure 1, feeling of stress is the only factor significantly correlated with sleep quality rating: it shows a strong negative correlation ($r = -0.66$) suggesting that higher stress levels are associated with poorer sleep quality.

### 3.2 Nighttime Discomforts and Sleep Quality

![Figure 2: Comparative Visualization of Feature Importance in Sleep Quality Prediction Using Random Forest](image)

Figure 2 shows the rank of importance of sleep discomforts in predicting sleep quality. The leading predictor, frequently using the bathroom at night for the past month, has the highest
importance score (approximately 0.175) in the prediction model. This means this feature contributes to 17.5% of the overall importance in determining the output variable, sleep quality rating. The next important factor was experiencing pain, with importance scores slightly lower than the primary factor. In this model, feeling hot is the least important predictor, contributing to less than 10% of overall importance in determining the sleep quality rating. However, these factors all similarly impact sleep quality predictions.

### 3.3 Air Quality and Sleep

![Figure 2: Spearman's rank correlation coefficient distributions for CO₂, PM2.5, temperature, and relative humidity across 1-minute, 15-minute, and 1-hour intervals, depicting the stability and direction of their relationships.](image-url)
The Figure displays Spearman's rank correlation coefficients for environmental parameters across three temporal resolutions. For the 1-minute interval correlations, the majority of correlations for each participant’s bedroom data are densely centered around zero except correlation of CO$_2$ and relative humidity. In the 15-minute and 1-hour intervals, the correlation plots showed more consistent patterns with each other.

For the 15-minute and 1-hour resolution, CO$_2$, temperature, and relative humidity demonstrated a positive association with each other. On the other hand, PM2.5 did not demonstrate a clear pattern of association with CO$_2$, temperature, and relative humidity. To study the impact of individualized correlation impacts on the sleep quality rating, we chose the 15-minute interval as it demonstrated a more stable pattern compared to the 1-minute interval. When compared with 1-hour intervals of correlation patterns, 15-minute intervals also did not show a significant difference.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Odds Ratio</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min Temperature (night)</td>
<td>0.42</td>
<td>0.24</td>
<td>1.53</td>
<td>3.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Max Temperature (night)</td>
<td>0.05</td>
<td>0.23</td>
<td>1.05</td>
<td>0.22</td>
<td>0.83</td>
</tr>
<tr>
<td>Mean Temperature (night)</td>
<td>0.25</td>
<td>0.24</td>
<td>1.28</td>
<td>1.05</td>
<td>0.30</td>
</tr>
<tr>
<td>Min Relative Humidity (night)</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.99</td>
<td>-0.23</td>
<td>0.82</td>
</tr>
<tr>
<td>Max Relative Humidity (night)</td>
<td>0.01</td>
<td>0.03</td>
<td>1.01</td>
<td>0.17</td>
<td>0.86</td>
</tr>
<tr>
<td>Mean Relative Humidity (night)</td>
<td>0.01</td>
<td>0.03</td>
<td>1.01</td>
<td>0.28</td>
<td>0.78</td>
</tr>
<tr>
<td>Min PM2.5 (night)</td>
<td>0.25</td>
<td>0.39</td>
<td>1.29</td>
<td>0.64</td>
<td>0.52</td>
</tr>
<tr>
<td>Max PM2.5 (night)</td>
<td>0.01</td>
<td>0.01</td>
<td>1.01</td>
<td>1.07</td>
<td>0.58</td>
</tr>
<tr>
<td>Mean PM2.5 (night)</td>
<td>0.00</td>
<td>0.08</td>
<td>1.00</td>
<td>0.05</td>
<td>0.96</td>
</tr>
<tr>
<td>Mean Temperature (day and night)</td>
<td>0.26</td>
<td>0.25</td>
<td>1.29</td>
<td>1.04</td>
<td>0.30</td>
</tr>
<tr>
<td>Mean Relative Humidity (day and night)</td>
<td>0.00</td>
<td>0.03</td>
<td>1.00</td>
<td>0.06</td>
<td>0.95</td>
</tr>
<tr>
<td>Mean PM2.5 (day and night)</td>
<td>-0.01</td>
<td>0.04</td>
<td>0.99</td>
<td>-0.35</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Table 1: Air quality metrics and sleep quality logistic regression results: The min, max and mean values during all three consecutive days and nights (11 pm to 8 am) were generated using the 15-minute interval measurement. Each of these was correlated to sleep quality using logistic regression.

The results of the logistic regression analysis, incorporating air quality metrics as predictors of sleep quality, are presented in Table 1. Minimum nighttime temperatures were associated with sleep quality, with an odds ratio (OR) of 1.53, although this association did not reach statistical significance (p = 0.08). Other temperature measures at night also did not show a significant relationship with sleep quality. Similarly, minimum, maximum, and mean values in relative humidity and particulate matter (PM2.5) levels during the whole day and night only did not demonstrate a statistically significant impact on sleep quality.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Statistic</th>
<th>Odds Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp. &amp; Rel. Humidity Cor.</td>
<td>-2.58</td>
<td>2.05</td>
<td>-1.26</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>PM2.5 &amp; Temp. Cor.</td>
<td>-1.35</td>
<td>2.34</td>
<td>-0.58</td>
<td>0.26</td>
<td>0.56</td>
</tr>
<tr>
<td>CO2 &amp; Temp. Cor.</td>
<td>-0.55</td>
<td>2.23</td>
<td>-0.25</td>
<td>0.58</td>
<td>0.81</td>
</tr>
<tr>
<td>PM2.5 &amp; Rel. Humidity Cor.</td>
<td>0.77</td>
<td>1.74</td>
<td>0.44</td>
<td>2.16</td>
<td>0.66</td>
</tr>
<tr>
<td>CO2 &amp; Rel. Humidity Cor.</td>
<td>-0.75</td>
<td>2.38</td>
<td>-0.32</td>
<td>0.47</td>
<td>0.75</td>
</tr>
<tr>
<td>CO2 &amp; PM2.5 Cor.</td>
<td>1.56</td>
<td>1.99</td>
<td>0.78</td>
<td>4.74</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 2: Spearman's Rank Correlation Coefficients Between Environmental Variables and Sleep Quality Using Logistic Regression: this table shows the results of a Spearman's rank correlation analysis, which examines the log-transformed variations in sleep quality using environmental
parameters, including temperature, relative humidity, PM2.5, and CO₂ concentrations in a 15-minutes range.

The results, summarized in Table 2, reveal no significant correlations between these environmental factors and sleep quality. Specifically, the strongest correlation was observed between CO₂ and PM2.5, with an estimate of 1.56 and an odds ratio of 4.74; however, the correlation did not reach statistical significance (p = 0.42). The other correlations varied in magnitude, but none presented compelling evidence to suggest a reliable association within the context of this analysis. These findings suggest that, within the measured intervals, the environmental factors under study do not have a discernible impact on sleep quality according to the data captured.

**Figure 3:** Air change per hour and CO₂ levels over time (ID 123)

We picked a single participant and used Figure 3 to demonstrate the CO₂ level and air change per hour. For this participant, we noticed that the bedroom CO₂ level started to rise around 9 pm each day and had a steady drop in the morning around 6 am consistently for three consecutive
days. Based on the CO₂ level, air change per hour was modelled using the methods described in section 2.3. Air changes per hour is the highest at around noon on the third day, following the second highest at around 9 pm on the first day.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Statistic</th>
<th>Odds Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Weighted Average Air Change/Hour</td>
<td>-3.5738</td>
<td>9.0198</td>
<td>-0.396</td>
<td>0.028</td>
<td>0.69</td>
</tr>
</tbody>
</table>

**Table 3:** Time-weighted average air change per hour and sleep quality using logistic regression

Based on Table 3, the time-weighted average air change per hour (11 pm – 8 am) did not demonstrate a significant impact on sleep quality.

### 4. Discussion

In this study, we aim to comprehensively understand the factors impacting sleep quality. We used a correlation network analysis to examine the connected relationship between demographics, living habits, psychological condition and their impacts on sleep quality. We found that only the psychological condition of feeling stressed is significantly and negatively associated with sleep quality. Moreover, we found using the bathroom is the most important feature based on the random forest model, contributing to around 17.5% of importance in determining the sleep quality rating and feeling hot is the least important. However, all seven discomfort factors are very similar in terms of their importance. Lastly, we extracted features from environmental air quality monitors. Results show that minimum, maximum and average values of bedroom temperature, PM2.5 and relative humidity have no impact on sleep quality.
rating for both day and nighttime. Additionally, the correlation of the change in bedroom temperature, PM2.5, relative humidity as well as air change per hour also have no impact on the sleep quality rating.

From the network analysis results in Figure 1, sleep quality involves complicated interaction from multiple factors. Sleep with the window open does not correlate with sleep quality. This is in contrast to Fan et al. which demonstrated sleep with window open improved sleep quality in Danish homes. However, this study was conducted in Switzerland and data was collected during the winter season. Previous studies showed that seasonality can significantly impact sleep quality which might be one reason why these two cohorts have different findings (Mattingly et al., 2021). In this study, demographics and living habits did not demonstrate a significant correlation with sleep quality, however, they indirectly correlated with sleep quality. Understanding the network of factors impacting sleep is essential to provide better interventions to improve sleep.

This study shows that using the bathroom at night during sleep is the most important sleep discomfort factor in predicting sleep quality. One hypothesis is frequent urination at night is because of the psychological condition of feeling stressed. Nocturia is the need to get up at night to urinate (Leslie et al., 2019). A systematic review of results suggested that there is a strong association between anxiety and nocturia (Breyer et al., 2013). However, the direction of impact remains unclear. We are unsure whether frequent nighttime bathroom usage causes stress or if stress leads to frequent nighttime bathroom usage.
In this study, we did not find significant impacts of air quality measures on sleep quality which is not consistent with other works in the field. There might be several reasons that lead to these results. First, the indoor air quality in most participants' bedrooms might not have exceeded certain thresholds to affect sleep quality. For example, the mean of averaged PM2.5 for all participants is 6 µg/m³ (see Figure App2 in Appendix), which is well within WHO-recommended PM2.5 values of less than 15 µg/m³ for a 24-hour average (Particulate Matter (PM 2.5 and PM 10, n.d.). Similarly, relative humidity has a mean of around 45% which is within the range of the recommended indoor humidity level according to the United States Environmental Protection Agency (US EPA, 2014). Another potential reason is people have different thermal perceptions (Schweiker et al., 2018). Factors like age and gender can impact the perception of cold or warmth (Harju, 2002) and this may impact sleep. We did not include the variation of subjective thermal perception into consideration in this study which may lead to inaccurate results.

There are several limitations to this study. Firstly, there might be bias due to subjective measures from questionnaire data. For example, the questionnaire collected the reported frequency of tea consumption as an indicator of living habits. However, tea consumption frequency may fluctuate over time, and thus the data might not accurately reflect participants' overall lifestyle patterns. Similarly, when correlating sleep discomfort data, participants are asked to rate the frequency of sleep discomfort in the past month, which may also introduce recall bias into the study. Moreover, we only collect three consecutive days of air quality data. This study relies on the assumption that these three days' air quality data are representative of the overall air quality in participants' households, which might not be accurate. All these factors may introduce bias into
our study. The second limitation is that we have a relatively small sample size, which may reduce statistical power. We may fail to detect significant factors that impact sleep quality. Additionally, we cannot conclude any causal relationships or control for confounding variables in this study; therefore, we do not know what causes poor sleep quality.

Despite these limitations, this study lays the groundwork for future research. Through this study, we found how living habits vary across demographic factors such as age, gender, and education level. Future investigations could study the sub-population with different demographics like different age and gender groups to better draw insights and provide sleep interventions using more personalized approaches. Future studies could conduct randomized controlled trials to explore causal relationships based on the significant association between stress and sleep quality. To enhance the reliability of findings, future studies should recruit more participants, and incorporate more objective measures alongside questionnaires, such as employing smart watches for tracking participants' habits and sleep patterns, thereby reducing bias.

In summary, this study demonstrates the complex interplay between demographic variables, behavioural patterns, and psychological and sleep quality thereby providing evidence for future research aimed at optimizing sleep health interventions. We also use an innovative approach to draw insights from time series data by extracting more complicated features including correlation of change between different air quality measurements and the ventilation effectiveness to study sleep quality. The future integration of objective metrics alongside traditional survey methods will help with mitigating biases.
Acknowledgement: I am deeply grateful to my thesis and academic advisors Prof. Krystal Pollitt and Prof. Joshua Warren for their support, guidance, encouragement, and invaluable mentorship throughout my MPH study at Yale. I would also like to extend my appreciation to my colleagues and collaborators who made great efforts to collect the data as well as provide insightful perspectives to my research. Furthermore, I am thankful to all my professors and mentors at Yale for their teachings, inspiration, and encouragement along the way. Finally, I want to extend my deepest gratitude to my family and friends for their unconditional love. Their encouragement and belief in me have been the driving force behind my achievements, and I am forever grateful for their presence in my life.

References:


24. Particulate matter (PM 2.5 and PM 10. (n.d.).
https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdf


Appendix:

Figure App1: Descriptive statistics for selected living habits and psychological factors
Figure App2: Box plots for mean of average air quality concentrations for each participant