# Daily Activity Analysis with Fitbit Tracker Data

Riku Tikkanen

December 2021

# Contents

# 1 Introduction

Technology is already part of most people's day-to-day life, and new tools emerge continuously to improve and facilitate our well-being. For example, accelerometer-based activity monitors have already been used over a decade ago in free-living settings to monitor the behavior of older people (Davis & Fox, 2007). Similar, although much more sophisticated and technologically advanced activity tracking devices have been available to the general public for a while now, and they may be useful for monitoring and promoting physical activity.

Wearable fitness trackers, such as the Fitbit tracker, enable their wearer to easily monitor personal physical activity and progress. Small, lightweight, unobtrusive and a relatively low price are some of the features which make modern fitness trackers like the Fitbit widely available and popular. Fitbit devices capture body motion using microelectronic triaxial accelerometers, and analyze the motion data using special algorithms. Users of the Fitbit tracker can upload data to an interactive website which provides them with information about calory expenditure, steps taken and distance traveled (Sasaki et al., 2015). Furthermore, users can track their progress and share personal data with others in a user-friendly manner.

Wearable fitness trackers are becoming quite popular in general. According to Pew Research, an American nonpartisan research company, 21% of Americans said that they were using some sort of smart watches or fitness trackers in 2019 (Vogels, 2020). A German counterpart, Bitkom research, also mentions in their 2019 report that German's interest in fitness trackers has increased steadily for the last 5 years (e.V., 2019). According to their survey, approximately 29% of Germans are using some sort of fitness tracker in 2019. The global popularity of fitness trackers is further emphasized in a recent study by CCS insight, a market intelligence and advisory firm. Their survey shows that the total amount of shipments of wearables has increased by 20% in 2021 including 142 million units of smartwatches and 90 million units of simpler fitness trackers (Karanis, 2021).

The recent growth in fitness tracker sales is somewhat boosted by the ongoing global Covid-19 pandemic which has made people more aware of their health, fitness and well-being (Karanis, 2021). The regular lockdowns imposed by governments worlwide also close many forms of pass time activities making sports and the outdoors increasingly popular. Even without the pandemic there has been an increasing interest towards self-improvement and health monitoring which is good considering that many people do not meet the daily activity recommendations. In fact more than one in four adults and more than three in four adolescents do not meet the recommendations for aerobic excercise (Guthold et al., 2018, 2020).

Fitbit wristbands have been studied in the past to compare their accuracy with other brands and to validate their output. For example, Adam et al. found in their 2013 research that the Fitbit accelerometers underestimate energy expenditure compared to indirect calorimetry for inclined activities (Adam Noah et al., 2013). Twenty-three participants with mean age of 26.65 ± 7.55 years took part in the study which main purpose was to determine the reliability and validity of the Fitbit and Fitbit Ultra fitness trackers. The study concluded that both, the Fitbit and the Fitbit Ultra are reliable and valid for monitoring over-ground energy expenditure. Another study conducted by Paul et al. aimed to determine the accuracy of the Fitbit in older people aged 60 and over (Paul et al., 2015). Thirty-two participants took part in the study and the results show that the Fitbit is an excellent device for tracking step counts in older people as well.

With the amount of Fitbit users available and with proof that the Fitbit devices produce valid and reliable data, there is a new opportunity to analyze Fitbit users' personal data and create a larger picture of their health and behavior in general. This study analyzes the behavior of 24 participants using data generated by the Fitbit fitness tracker. The activity of the participants is compared with global health recommendations to produce insights about their health and lifestyles. The results show that most of the participants live an unhealthy lifestyle and are at risk of developing chronic diseases.

# 2 Problem formulation

The goal of this study is to find out more about the health and behavior of Fitbit users by analyzing their daily activity. The objective is to find correlations in users' activity to find if there is common behavior, and to compare their activity levels with recommendations set by authorities and non-profit organizations such as the World Health Organization (WHO).

# 3 The data

The data used for this project is available at Kaggle (Möbius, 2021) and the original source is available at Zenodo (Furberg et al., 2016). The data describes the daily activities of 30 volunteers and has been has been collected using the Fitbit wristband fitness tracker. All the generated data has been collected to a distributed survey via Amazon Mechanical Turk between 12.04.2016-12.05.2016. The personal tracker data has been anonymized and it includes minute-level output for physical activity, heart rate and sleep monitoring. The data is time series and consists of eighteen csv files with numerical and boolean features.

## 3.1 Preprocessing and feature selection

This study is interested in the daily activity and behavior of Fitbit users and therefore the analysis is based on the sections of data that represent information on a daily level.

After inspecting the data, three eligible datasets were first selected: dailyActivity_merged.csv, sleepDay_merged.csv and weightLogInfo_merged.csv. However, the number of eligible volunteers differed for the selected datasets. The dailyActivity_merged.csv dataset had 33 unique user Id's, the sleepDay_methed.csv had 24 unique user Id's and the weightLogInfo_merged.csv had only 8 unique user Id's. Given that there only existed weight information for 8 unique users, it would be impossible to make conclusions or recommendations about the weight for all the volunteers. Therefore, after careful consideration the weightLogInfo_merged.csv dataset was not included in the analysis because the number of unique users was simply too low.

To make future analysis easier the datasets dailyActivity_merged.csv and sleepDay_merged.csv were merged into one dataset. First the features of both datasets representing the date were changed to datetime format and renamed to 'Date'. Next the datasets were merged by features 'Id' and 'Date'. The format of the merged dataset is shown in Table 1.

| | number of features | number of unique users |
|---|---|---|
| merged_data | 18 | 24 |

Table 1: Basic information about the merged dataset.

# 4 Methods

In order to analyze the data and solve the presented problem, the Python programming language is used. A range of Python libraries are used to help analyze the data and visualize behavior and activities of users. The libraries used in this study are Numpy (Harris et al., 2020), Pandas (pandas development team, 2020), Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021). Preprocessing is accomplished using Pandas which is a helpful library for representing data in table format. During preprocessing the data is cleaned, missing records are removed and the data is grouped by user id and date. More information about preprocessing is given in section 5.1. The Numpy library is used mainly for manipulating the data, and the Matplotlib and Seaborn libraries are used to present the results visually.

# 5 Results

## 5.1 Physical activity analysis

### 5.1.1 Importance of regular physical activity

Any bodily movement that is produced by skeletal muscles and which results in more energy expenditure than resting expenditure can be defined as activity

(Pate et al., 1995). The association between health and physical activity has been recognized for a long time and there exists a large amount of studies which have proven the correlation many times.

An increased physical activity is known to reduce the risk of chronic diseases such as obesity, coronary heart disease, type 2 diabetes and psychological disorders (Melzer et al., 2004). In the United States alone approximately 70% of such chronic diseases, including cardiovascular disease and stroke could be prevented by healthy nutrition and an active lifestyle, yet, 78% of Americans are at elevated health risk for not exercising enough (Aldana et al., 2005).

It is also known that the association between chronic diseases and inactivity is evident regardless of body weight (Kokkinos, 2012). Furthermore, it has been proven that physical activity can reduce mortality rates by as much as 30% (Melzer et al., 2004).

As mentioned, physical activity can also prevent psychological disorders and elevate the mood states of persons and increase psychological well-being (Tomporowski, 2003). Proof shows that people living an inactive lifestyle also have an increased risk of becoming depressed and developing Alzheimer's disease (Camacho et al., 1991; Friedland et al., 2001).

The treatment of diseases causes rising costs for health systems and therefore the prevention of diseases is highly important for societies. There have been impressive pharmaceutical and technological advances, yet, maintaining an active and healthy lifestyle is the best improvement for public health (Aldana et al., 2005). Furthermore, the health benefits of physical health also highly outweigh the risks (Melzer et al., 2004).

The WHO's most recent guidelines from the year 2020 (Bull et al., 2020) recommend a weekly average of at least 150-300 minutes of moderate intensity or 75-150 minutes of vigorous intensity aerobic physical activity for adults. An equivalent combination of moderate and vigorous intensity activity is also acceptable. The WHO does not specify a quantitative threshold for sedentary behavior but does recommend minimizing it. It also highly recommends increasing the amounts of physical activity for increased health benefits.

### 5.1.2 Physical activity of participants

To analyze the physical activity behavior of the participants the different activity levels were visualized using a histogram as shown in Figure 1.

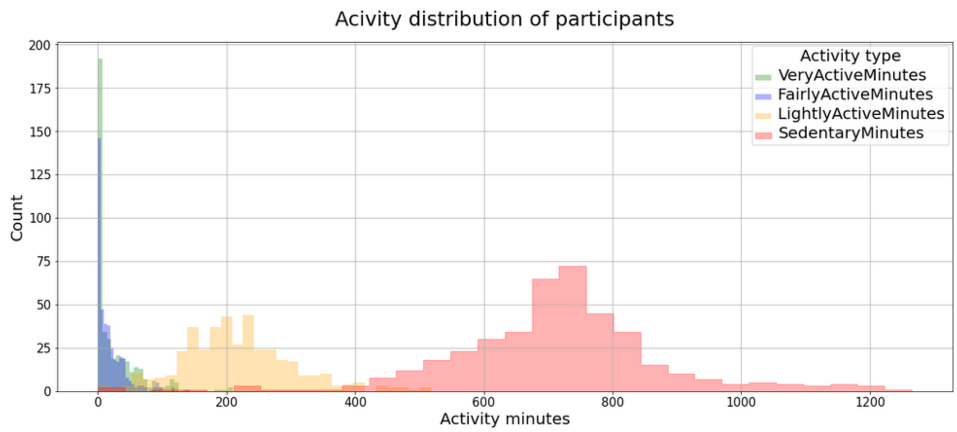


Figure 1: Physical activity distribution of the participants.

Clearly the records showed that participants spent most of their time being sedentary and a very short time being very active or moderately active. On average participants spent 712 minutes or just below 12 hours being sedentary. On a daily level that would translate to participants spending nearly half of their day sitting and being inactive. Another 216 minutes or 3.6 hours were spent being lightly active on average. In total participants spent only 42 minutes being very or moderately active on average.

Surprisingly, the activity distribution of the participants did not change much during the week. The only notable difference was a larger lower quartile for 'SedentaryMinutes' during Saturdays and Sundays. In other words, participants spent a little less time being sedentary during the weekends.

More interesting results were achieved by separating the activity records by into two groups, namely 'active' and 'inactive'. As mentioned in the previous section, the WHO recommends 150-300 minutes of moderate or 75-150 minutes of vigorous intensity exercise per week. A combined recommendation would therefore be approximately 113 minutes of moderate and vigorous activity per week. On a daily basis that translates to only 16 minutes of activity which was considered too low as a threshold for this comparison. For the comparison between active and inactive lifestyles the participants were labeled as 'active' if the sum of their 'VeryActiveMinutes' and 'FairlyActiveMinutes' was at least 60 minutes, and as 'inactive' if it was less than 60 minutes. A total of 19 participants were classified as inactive and only 5 participants as active.

The results indicated an obvious correlation between active users having higher calorie consumption and step count than their inactive counterparts as shown in Figure 2. Active users had an average calorie consumption of 2800 kcals per day and inactive users 2200 kcals. On average, active users spent over 100 minutes being very or moderately active during the day while inactive participants spent only under 20 minutes daily for said activities. Surprisingly, active users spent more time being sedentary on average than inactive users.

Figure 2: Comparison between active and inactive participants

Interestingly, as the violin plot shows, the calorie consumption for both user types had two peaks. Both, inactive and active users had a lower peak at a level of 2000 kcals. The upper peak for inactive users showed a calory consumption of 3000 kcals while active users had an upper peak of 3700 kcals. The levels of the peaks suggested that they were both dependent on the sex of the participants. The lower peaks were indications of female calorie consumption and the upper peaks male calorie consumption.

Some of the Participants had alarmingly low average activity levels with a combined value of under 20 minutes of activity per day. Such low activity levels might already indicate an increased risk of developing chronic diseases.

## 5.2 Sleep analysis

### 5.2.1 Importance of sleep

Life on earth follows a circadian rhythm and for us humans it means following an approximately 24-hour day-night cycle. This rhythm reflects on us humans at both a psychological and a physiological level and it can have some effect on a persons activities and health (Aledavood et al., 2015). Sleep is a basic human need and it is essential for health, well-being and performance during the day.

A reduction in sleep health, including circadian disorders, insufficient sleep duration, irregular timing and poor sleep quality are not uncommon in modern society and they are often associated with multiple disease risks (Laposky et al., 2016).

Poor sleep health causes physiological, behavioral and psychological problems. Common physiological problems are cardiovascular disease, metabolic disease, obesity and diabetes (Laposky et al., 2016). Some behavioral problems caused by insufficient sleep include the increased habit of smoking and drinking (Vail-Smith et al., 2009). All psychological disorders are highly related with sleep and often sleep disturbances are a sign of mood disorders and a risk factor for developing anxiety and depression (Aledavood et al., 2019).

The World Health Organization (WHO) indicates that sleep health is affected by environmental and psychological stressors, and also the lifestyle of a person (Organization et al., 2004). The WHO also separates sleep related health risks into short, medium and long term sections. The short term risks of disturbed sleep include the increased risk of accidents, high blood pressure and stress hormones. Medium term risks are said to include cardiovascular disease and cognitive performance issues. Long term effects consist of mental issues, immune system related problems and also cardiovascular disease.

According to a consensus study conducted by the American Academy of Sleep Medicine (AASM) and Sleep Research Society (SRS) adults should sleep 7 hours or more per night to promote optimal health (Panel et al., 2015). Less than 7 hours of sleep is associated with risks mentioned before and sleeping over 9 hours is considered oversleeping. However, it is uncertain if oversleeping is associated with health risks (Panel et al., 2015).

### 5.2.2 Sleep behavior at a community level

The healthy amount of sleep depends on the age of a person, and since the data used for this study was anonymized, it was assumed that the participants were all adults. The recommended amounts of sleep were based on the suggestions explained in the previous section.

To analyze the overall sleep behavior of participants, the sleep records were first separated into three categories based on the minutes asleep using the 'TotalMinutesAsleep' feature. Records indicating less than seven hours of sleep were labeled as 'Inadequate', records with seven to nine hours were labeled as 'Adequate' and records with over nine hours of sleep were labeled as 'Oversleep'. A histogram plot was constructed to visualize the results as shown in Figure 3

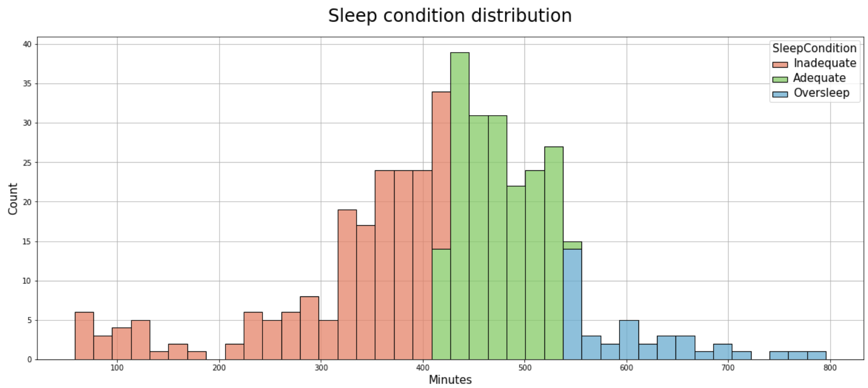


Figure 3: Distribution of sleep records

Based on Figure 3 the sleep duration of the participants seemed to be mainly adequate. However, there were also many records implying that there were unhealthy levels of sleep, suggesting that the duration of sleep was irregular at least for some participants. There were not many reported cases of oversleeping.

Based on the data the average total amount of sleep per day for the participants was 419 minutes which is an almost even seven hours of sleep. The lowest recorded total sleep duration was an alarmingly low 58 minutes for a single day. The longest recorded total duration was 796 minutes which adds up to surprisingly long total sleep duration of 13 hours. The median for the total amount of sleep was 432 minutes and therefore at least half of the participants had an averagely healthy total sleep duration for each day.

To learn more about the distribution of sleep during different weekdays, a box plot was constructed as shown in Figure 4.

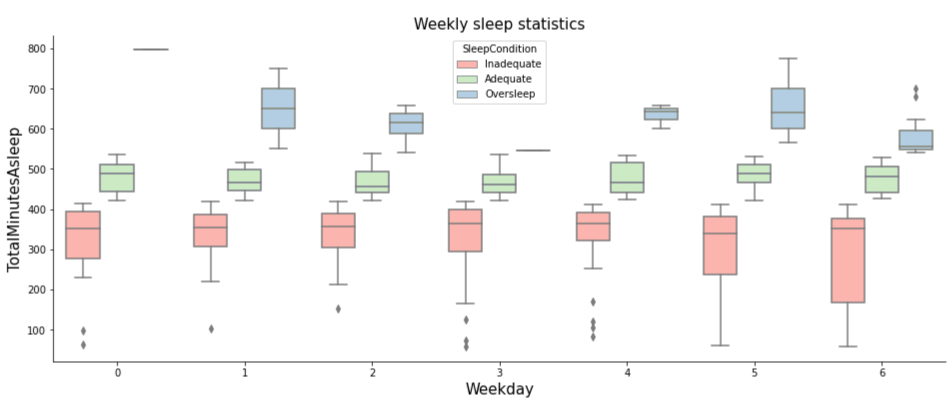


Figure 4: Weekly sleep statistics

The weekly sleep distribution shows that there is little to no shift in the median for adequate amounts of sleep, indicating routine like consistency. The median for inadequate sleep is also consistently at a level of about 350 minutes. The median for oversleeping records is shifting a lot during the week suggesting that oversleeping happens irregularly and more spontaneously. Unsurprisingly, most of the inadequate sleep records are in the lower quartile on weekends, suggesting that participants are staying up later on weekends.

### 5.2.3 Participants with inadequate sleep levels

The previous section analyzed sleep behavior on a community level. In this section some participants with low levels of sleep are analyzed. To find extreme levels of sleep in participants the average levels of sleep for each user was calculated.

The average sleep records indicated dangerously low levels of sleep for three participants during the test duration. Only one participant appeared to regularly oversleep. On closer inspection, however, it became clear that most of the inadequate sleepers had only few sleep records and therefore were considered unreliable. After removing individuals with less than ten recordings of sleep there were still five participants with unhealthy levels of sleep. The sleep patterns for these five participants are presented in Figure 5.

Three of those five subjects (3977337714, 4702921684, 1503960366) showed consistent sleep behavior during weekdays and a minor increase in total sleep duration during the weekends. The two other participants (4388161847, 4445114986) had quite irregular sleep patterns during the test period.

Of the participants 3977337714 had the lowest average duration of sleep of only 293 minutes during the whole test period. That is a 30% decrease from the recommended amount of sleep which already might indicate increased medium to long-term risks such as cognitive performance issues and cardiovascular disease.

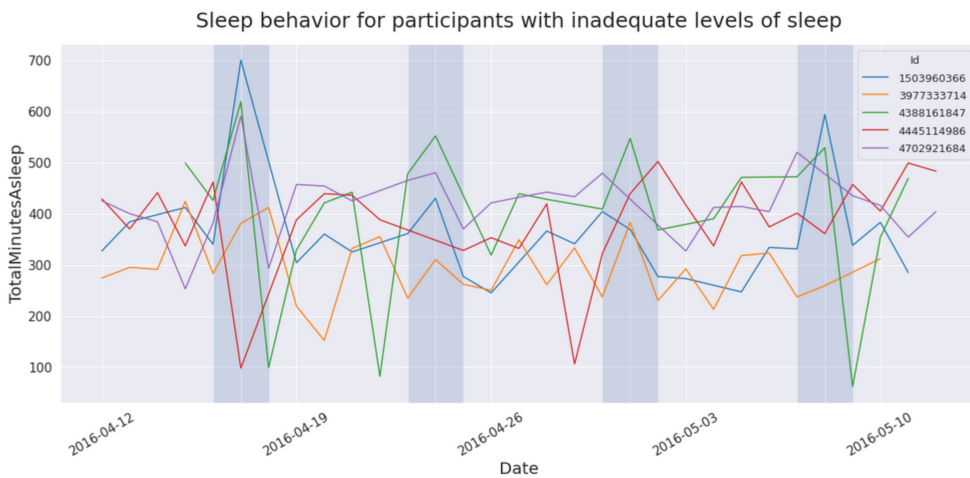


Figure 5: Sleep behavior of participants with low levels of sleep (highlighted weekends).

Interestingly, all except one of the participants with inadequate levels of sleep and a minimum of 10 daily records during the test period were also the most inactive participants. The participants 1503960366, 4388161847, 4445114986 and 4702921684 all showed very low levels of sleep and very low levels of moderate to vigorous levels of activity during the day. All of these participants are at elevated risk of developing chronic diseases and long-term issues such as cardiovascular disease, diabetes, obesity and cognitive performance issues.

# 6 Conclusion & Discussion

The goal of this study was to learn more about the health and behavior of Fitbit users by analyzing their daily activity. The objective of this study was to find correlations in users' activity to find common behavior and to compare their activity levels with the recommendations set by authorities and non-profit organizations such as the World Health Organization (WHO).

The results are in agreement with the presented studies and recommendations about activity and sleep. Most of the participants showed unhealthy levels of activity and sleep during the test period, and some participants showed very unhealthy lifestyles. The participants with least healthy lifestyles were both inactive and sleeping too little which confirms the importance of regular sleep and activity for a healthy life.

While the results look promising, only 24 participants were analyzed in the end which limits the validity of the results. In the future more participants with regular activity logs are required to make better conclusions. With the number of fitness trackers rising worldwide, it seems plausible to do so.

# References


Davis, M. G., & Fox, K. R. (2007). Physical activity patterns assessed by accelerometry in older people. *European journal of applied physiology*, *100*(5), 581–589.

Sasaki, J. E., Hickey, A., Mavilia, M., Tedesco, J., John, D., Keadle, S. K., & Freedson, P. S. (2015). Validation of the fitbit wireless activity tracker for prediction of energy expenditure. *Journal of Physical Activity and Health*, *12*(2), 149–154.

Vogels, E. A. (2020). *About one-in-five americans use a smart watch or fitness tracker* [Accessed: 2021-12-24].

e.V., B. (2019). *Zukunft der consumer technology - 2019* [Accessed: 2021-12-24].

Karanis, A. (2021). *Good times for the smart wearables market* [Accessed: 2021-12-24].

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1· 9 million participants. *The lancet global health*, *6*(10), e1077–e1086.

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2020). Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1· 6 million participants. *The Lancet Child & Adolescent Health*, *4*(1), 23–35.

Adam Noah, J, Spierer, D. K., Gu, J., & Bronner, S. (2013). Comparison of steps and energy expenditure assessment in adults of fitbit tracker and ultra to the actical and indirect calorimetry. *Journal of medical engineering & technology*, *37*(7), 456–462.

Paul, S. S., Tiedemann, A., Hassett, L. M., Ramsay, E., Kirkham, C., Chagpar, S., & Sherrington, C. (2015). Validity of the fitbit activity tracker for measuring steps in community-dwelling older adults. *BMJ open sport & exercise medicine*, *1*(1), e000013.

Möbius. (2021). *Fitbit fitness tracker data* [Accessed: 2021-12-24].

Furberg, R., Brinton, J., Keating, M., & Ortiz, A. (2016, May). Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016. https://doi.org/10.5281/zenodo.53894

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

pandas development team, T. (2020, February). *Pandas-dev/pandas: Pandas* (Version latest). Zenodo. https://doi.org/10.5281/zenodo.3509134

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., King, A. C., et al. (1995). Physical activity and public health: A recommendation from the centers for disease control and prevention and the american college of sports medicine. *Jama*, *273*(5), 402–407.

Melzer, K., Kayser, B., & Pichard, C. (2004). Physical activity: The health benefits outweigh the risks. *Current Opinion in Clinical Nutrition & Metabolic Care*, *7*(6), 641–647.

Aldana, S. G., Greenlaw, R. L., Diehl, H. A., Salberg, A., Merrill, R. M., Ohmine, S., & Thomas, C. (2005). Effects of an intensive diet and physical activity modification program on the health risks of adults. *Journal of the American Dietetic Association*, *105*(3), 371–381.

Kokkinos, P. (2012). Physical activity, health benefits, and mortality risk. *International Scholarly Research Notices*, *2012*.

Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta psychologica*, *112*(3), 297–324.

Camacho, T. C., Roberts, R. E., Lazarus, N. B., Kaplan, G. A., & Cohen, R. D. (1991). Physical activity and depression: Evidence from the alameda county study. *American journal of epidemiology*, *134*(2), 220–231.

Friedland, R. P., Fritsch, T., Smyth, K. A., Koss, E., Lerner, A. J., Chen, C. H., Petot, G. J., & Debanne, S. M. (2001). Patients with alzheimer's disease have reduced activities in midlife compared with healthy control-group members. *Proceedings of the National Academy of Sciences*, *98*(6), 3440–3445.

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P., Chastin, S., Chou, R., et al. (2020). World health organization 2020 guidelines on physical activity and sedentary behaviour. *British journal of sports medicine*, *54*(24), 1451–1462.

Aledavood, T., Lehmann, S., & Saramäki, J. (2015). Digital daily cycles of individuals. *Frontiers in Physics*, *3*, 73. https://doi.org/10.3389/fphy.2015.00073

Laposky, A. D., Van Cauter, E., & Diez-Roux, A. V. (2016). Reducing health disparities: The role of sleep deficiency and sleep disorders [NHLBI Workshop on Reducing Health Disparities: The Role of Sleep Deficiency and Sleep Disorders]. *Sleep Medicine*, *18*, 3–6. https://doi.org/https://doi.org/10.1016/j.sleep.2015.01.007

Vail-Smith, K., Felts, W. M., & Becker, C. (2009). Relationship between sleep quality and health risk behaviors in undergraduate college students. *College Student Journal*, *43*(3), 924–930.

Aledavood, T., Torous, J., Hoyos, A. M. T., Naslund, J. A., Onnela, J.-P., & Keshavan, M. (2019). Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. *Current psychiatry reports*, *21*(7), 1–9.

Organization, W. H. et al. (2004). *Who technical meeting on sleep and health: Bonn germany, 22–24 january 2004* (tech. rep.). World Health Organization. Regional Office for Europe.

Panel, C. C., Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., et al. (2015). Recommended amount of sleep for a healthy adult: A joint consensus statement of the american academy of sleep medicine and sleep research society. *Journal of Clinical Sleep Medicine*, *11*(6), 591–592.