# Title here

Author 1 [*]

Department of YYY, University of XXX

and

Author 2

Department of ZZZ, University of WWW

August 15, 2022

**Abstract**

The text of your abstract. 200 or fewer words.

*Keywords:* 3 to 6 keywords, that do not appear in the title

# 1 Introduction

This template demonstrates some of the basic latex you'll need to know to create a ASA article.

# 2 Background

Physical activity is associated with human health conditions in many aspects. (Doherty et al. 2017) Previous studies based on self-reported frequency and duration of participation in activity were not inefficient to draw the relationship between physical activity and clinical trails and health recommendations, as it is difficult to quantify total physical activity across multiple levels of intensity. To bridge the gap, the development of objective methods for measuring physical activity worked as a complement to the self-reported assessment. Accelerometry using wearable devices has been a widely accepted method for objective measurement of physical activity. Accelerometers are sensors which measure the acceleration of objects in motion along reference axes. The application of accelerometer with wearable devices can be roughly divided into 2 main categories: controlled and uncontrolled experiments. Controlled experiment require users perform certain actions including walking, sitting, falling etc. (Nho et al. 2020, Casale et al. (2011)) Uncontrolled ones are usually carried out with wrist-worn devices. Users wearing them will act normally in daily life. The obtained accelerometry data can be analyzed with different purposes. One is mechanical modelling which estimates some certain features of the motion of human beings. An example was estimating step length from wearable accelerometer devices. (Alvarez et al. 2006). Another usage is for activity recognition and detection. Physical activity information can be extracted from accelerometry with certain signal processing techniques. Such information can be primarily be used for activity characterization. (Willetts et al. 2018, Ravi et al. (2005), Yang & Hsu (2010)) Activity information can also be related to health outcomes. Previous studies validated the relationships between accelerometry-derived activity or sleep and some specific disease: obesity, cardiovascular disease, cardio-metabolic disease, breast cancer, psychiatric disorder and Parkinson's disease.(Guo et al. 2019, Guo et al. (2020), Ramakrishnan et al. (2021), Barker et al. (2019), Cassidy et al. (2018), Dennison et al.

(2021), Nikbakhtian et al. (2021)) However, in this paper we will create bio-markers for general health conditions. Machine learning methods has been very efficient for detection of diseases. (Kubota et al. 2016) In this paper, we will combine statistical analysis and machine learning models to relate accelerometry-derived activity with health conditions.

# 3   Data

## 3.1   Dataset Type

In this paper, three different types of datasets were involved: Summary Data, Raw Accelerometry and ECG Data.

Summary Data consisted of both users' demographic information and actigraphy. The actigraphy was obtained by summarizing Raw Accelerometry for each individual user. Raw Accelerometry recorded the signals from wearable device sensors to track human movement. (ECG data)

## 3.2   Compliance

Compliance measures whether a user was in accordance with established guidelines for using the wearable devices. Data from users with compliance are more reliable to support further analysis.

In Summary Data, there were three categorical variables ($data\_quality\_good\_wear\_time$, $data\_quality\_good\_calibration$, and $data\_quality\_calibrated\_on\_own\_data$) already provided by UK Biobank to determine data quality. In addition, actigraphy data were necessary for the analysis on relationship between activity and health conditions. So a user was considered to be compliant if 2 conditions were met:

1. Data quality variables equal to *Yes*;
2. Actigraphy data were present.

## 3.3  Summary Data Analysis

### 3.3.1  3-peak Distribution of Accelerometry

In Summary Data, there were 67 fields recorded the fractions of acceleration less than or equal to 1, 2, . . . , 2000 milli-gravities respectively for each user. Subtracting each fraction by its previous one generated the concrete fraction for each acceleration equal to 1, 2, . . . , 1900 milli-gravities. Figure 1 showed a sample histogram of acceleration fractions. All the distributions of compliant users' accelerometry followed the same pattern as Figure 1. There were 3 peaks existing which divided the acceleration into 3 ranges: 1 - 19 milli-gravities, 20 - 95 milli-gravities and 100 - 1900 milli-gravities.
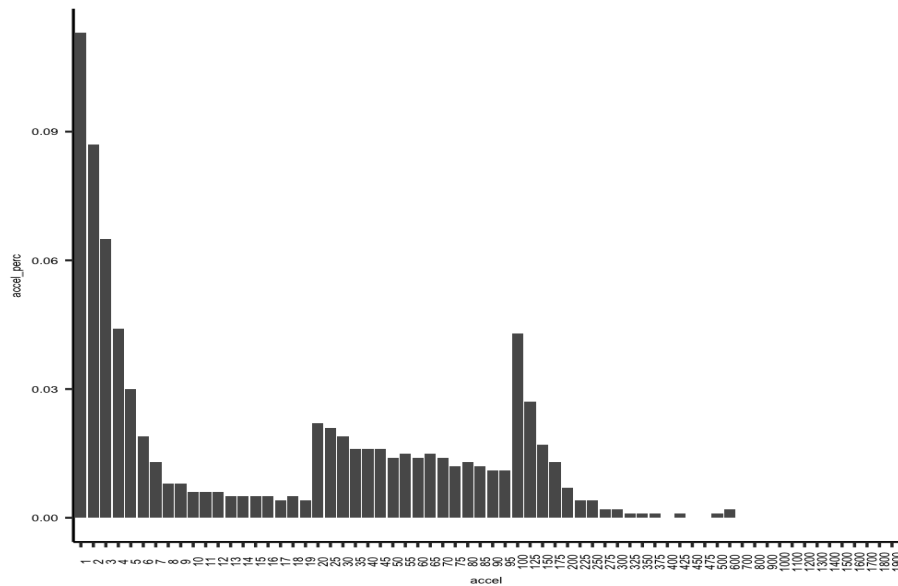


Figure 1: Sample Accelerometry Distribution

### 3.3.2  Correlation Between Accelerometry and Activity

As the intensity of activity is positively correlated to acceleration, these 3 ranges of acceleration can be related to low, moderate and high intensity activities respectively. By summing up all the fractions in each range, the total fraction depicted the percentage of acceleration triggered by the different kinds of activities. For example, the total fraction of acceleration from 1 to 19 milli-gravities is proportional to the amount of low-intensity

4

Table 1: Map of Acceleration and Activity Group

| acceleration | activity |
| --- | --- |
| 1 - 19 milli-gravities | low-intensity activity |
| 20 - 95 milli-gravities | mid-intensity activity |
| 100 - 1900 milli-gravities | high-intensity activity |
| 1 - 1900 milli-gravities | total activity |

activity. In following analysis, sum of acceleration fractions would be taken as the value of certain activity amount directly. Table 1 shows the map between acceleration and activity groups.

### 3.3.3 Bivariate Analysis: Activity Amount vs. Medical Conditions

The Summary Data contained multiple variables indicating users' sleep, psychological and medical conditions. For example, *overall_health_rating* taking values of *Excellent*, *Good*, *Fair*, *Poor*, *Do not know* and *Prefer not to answer* indicated the overall health condition of a user. Figure 2 is the box plot of activity amount grouped by *overall_health_rating*.

In the subplot of high-intensity activity, the distributions of activity amount for *Excellent*, *Good*, *Fair* and *Poor* users had significant differences. The health ratings were positively correlated to the amount of high-intensity activity. Similar patterns can also be found in other medical variables.

### 3.3.4 Significance Validation

According to Central Limit Theorem, mean of activity amount for users with a certain medical condition approximates a normal distribution. Thus the differences between the mean of activity amount among users with multiple medical conditions should also approximate a normal distribution. Taking the minimal p-value for difference greater or less than 0 as the final statistics, we were able to validate whether activity amount can distinguish medical condition effectively.
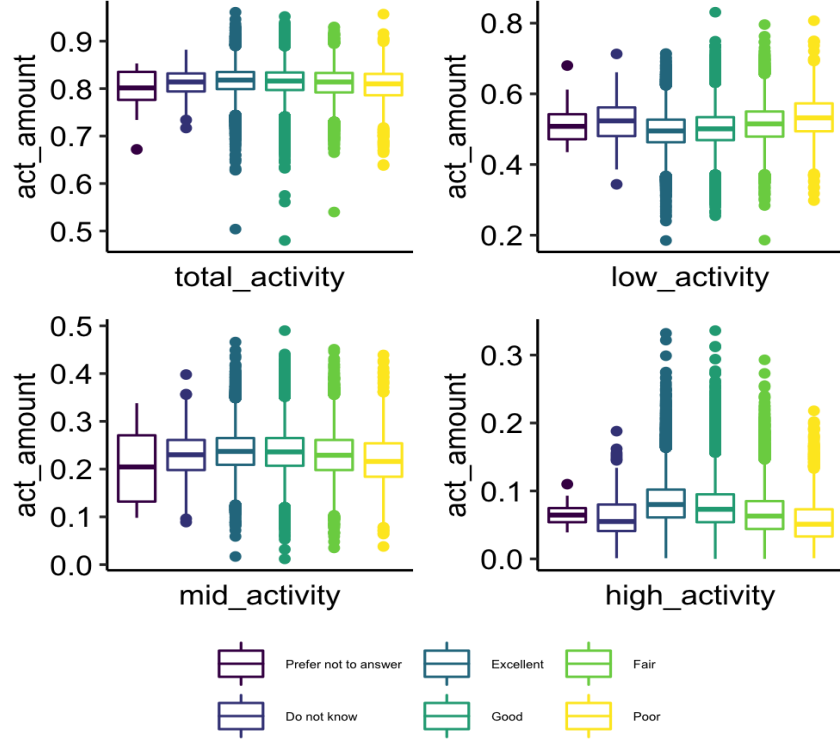
Figure 2: Box Plot of Activity Amount Group by Overall Health Rating

$$p.value = min\{p.value(mean \ \ difference > 0), \ \ p.value(mean \ \ difference < 0)\}$$

Figure 3 showed the approximated normal distribution density plot for the mean difference of high activity amount between users with *overall_health_rating* as *Excellent* and *Poor*. The final p-value was 0, which meant the high activity amount was significant to distinguish between users with "best" and "worst" health ratings.

Taking a series of medical variables and conducting the significance validation generated Table 2. It was obvious to see activity amount worked effectively to distinguish the differences for a considerable number of combinations in medical conditions.

Table 2: Significance Validation of Activity Amount Differences Between Medical Conditions

| activity | medical.variable | Condition.1 | Condition.2 | p.value |
| --- | --- | --- | --- | --- |

6

| | | | | |
|---|---|---|---|---|
| high | nap during day | Never/rarely | Usually | 0.0000 |
| mid | nap during day | Never/rarely | Usually | 0.0000 |
| low | nap during day | Never/rarely | Usually | 0.0000 |
| high | neuroticism score | 0 | 11 | 0.0256 |
| mid | neuroticism score | 0 | 11 | 0.1130 |
| low | neuroticism score | 0 | 11 | 0.0111 |
| high | overall health rating | Excellent | Poor | 0.0000 |
| mid | overall health rating | Excellent | Poor | 0.0000 |
| low | overall health rating | Excellent | Poor | 0.0000 |
| high | sleep duration | 18 | -3 | 0.0256 |
| mid | sleep duration | 18 | -3 | 0.3628 |
| low | sleep duration | 18 | -3 | 0.0073 |
| high | sleeping too much | No | Yes | 0.0000 |
| mid | sleeping too much | No | Yes | 0.0000 |
| low | sleeping too much | No | Yes | 0.0000 |
| high | sleeplessness insomnia | Never/rarely | Usually | 0.0000 |
| mid | sleeplessness insomnia | Never/rarely | Usually | 0.0000 |
| low | sleeplessness insomnia | Never/rarely | Usually | 0.0000 |
| high | snoring | No | Yes | 0.0000 |
| mid | snoring | No | Yes | 0.0000 |
| low | snoring | No | Yes | 0.0000 |
| high | trouble falling asleep | No | Yes | 0.0272 |
| mid | trouble falling asleep | No | Yes | 0.0116 |
| low | trouble falling asleep | No | Yes | 0.0008 |

# References

Alvarez, D., González, R. C., López, A. & Alvarez, J. C. (2006), Comparison of step length estimators from weareable accelerometer devices, *in* '2006 International Conference of
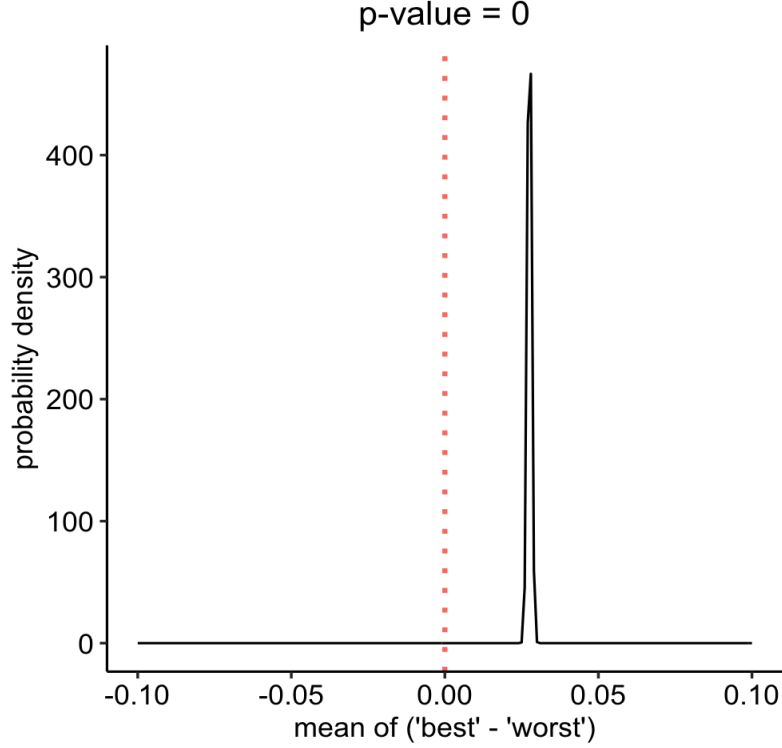
Figure 3: Normal Distribution Approximation for Mean of High Activity Amount Difference Between Excellent and Poor

the IEEE Engineering in Medicine and Biology Society', IEEE, pp. 5964–5967.

Barker, J., Smith Byrne, K., Doherty, A., Foster, C., Rahimi, K., Ramakrishnan, R., Woodward, M. & Dwyer, T. (2019), 'Physical activity of uk adults with chronic disease: cross-sectional analysis of accelerometer-measured physical activity in 96 706 uk biobank participants', *International Journal of Epidemiology* **48**(4), 1167–1174.

Casale, P., Pujol, O. & Radeva, P. (2011), Human activity recognition from accelerometer data using a wearable device, *in* 'Iberian conference on pattern recognition and image analysis', Springer, pp. 289–296.

Cassidy, S., Fuller, H., Chau, J., Catt, M., Bauman, A. & Trenell, M. I. (2018), 'Accelerometer-derived physical activity in those with cardio-metabolic disease compared to healthy adults: a uk biobank study of 52,556 participants', *Acta diabetologica* **55**(9), 975–979.

Dennison, C. A., Legge, S. E., Bracher-Smith, M., Menzies, G., Escott-Price, V., Smith, D. J., Doherty, A. R., Owen, M. J., O'Donovan, M. C. & Walters, J. T. (2021), 'Association of genetic liability for psychiatric disorders with accelerometer-assessed physical activity in the uk biobank', *PloS one* **16**(3), e0249189.

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., White, T., Van Hees, V. T., Trenell, M. I., Owen, C. G. et al. (2017), 'Large scale population assessment of physical activity using wrist worn accelerometers: the uk biobank study', *PloS one* **12**(2), e0169649.

Guo, W., Fensom, G. K., Reeves, G. K. & Key, T. J. (2020), 'Physical activity and breast cancer risk: results from the uk biobank prospective cohort', *British journal of cancer* **122**(5), 726–732.

Guo, W., Key, T. J. & Reeves, G. K. (2019), 'Accelerometer compared with questionnaire measures of physical activity in relation to body size and composition: a large cross-sectional analysis of uk biobank', *BMJ open* **9**(1), e024206.

Kubota, K. J., Chen, J. A. & Little, M. A. (2016), 'Machine learning for large-scale wearable sensor data in parkinson's disease: Concepts, promises, pitfalls, and futures', *Movement disorders* **31**(9), 1314–1326.

Nho, Y.-H., Lim, J. G. & Kwon, D.-S. (2020), 'Cluster-analysis-based user-adaptive fall detection using fusion of heart rate sensor and accelerometer in a wearable device', *IEEE Access* **8**, 40389–40401.

Nikbakhtian, S., Reed, A. B., Obika, B. D., Morelli, D., Cunningham, A. C., Aral, M. & Plans, D. (2021), 'Accelerometer-derived sleep onset timing and cardiovascular disease incidence: a uk biobank cohort study', *European Heart Journal-Digital Health* **2**(4), 658–666.

Ramakrishnan, R., Doherty, A., Smith-Byrne, K., Rahimi, K., Bennett, D., Woodward, M., Walmsley, R. & Dwyer, T. (2021), 'Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the uk biobank cohort study', *PLoS medicine* **18**(1), e1003487.

Ravi, N., Dandekar, N., Mysore, P. & Littman, M. L. (2005), Activity recognition from accelerometer data, *in* 'Aaai', Vol. 5, Pittsburgh, PA, pp. 1541–1546.

Willetts, M., Hollowell, S., Aslett, L., Holmes, C. & Doherty, A. (2018), 'Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 uk biobank participants', *Scientific reports* **8**(1), 1–10.

Yang, C.-C. & Hsu, Y.-L. (2010), 'A review of accelerometry-based wearable motion detectors for physical activity monitoring', *Sensors* **10**(8), 7772–7788.