A quality control framework for sleep metrics and its impact on association analysis with demographic and health outcomes in a large cohort study.

Andy Liu, PhD [1], Matt Patterson, PhD [1], Ali Neishabouri, PhD [1], Christine Guo, PhD [1]
[1]ActiGraph LLC, Pensacola, FL 32502, USA

ActiGraph™
Pioneering the Digital Transformation of Clinical Research

SLEEP 2024
HOUSTON, TX | JUNE 1-5

1 Overview

- Sleep is an essential aspect of our health and can be affected by numerous diseases and health disorders. Wrist-worn accelerometer, or actigraphy, provides an effective, lightweight and non-invasive tool that can measure sleep parameters such as total sleep time, sleep efficiency, and number of awakenings.
- Automated sleep assessments consists of two steps: 1) the detection of the total sleep opportunity window, and 2) the classification of stages such as sleep and wake. While the classification of sleep status has been the focus of many wearable sleep research studies and have been shown to be highly concordant with PSG (gold standard), the detection of total sleep opportunity remains challenging.
- To define the total sleep opportunity window, algorithms typically rely on a priori assumptions. A common sleep period algorithm (Van Hees) relies on: 1) the individual having a regular circadian rhythm, 2) the device being worn for greater than 16 hours per day, and 3) exactly 1 sleep period window per day, with a maximum gap of 1 hour between sleep segments. Deviations from these assumptions could lead to misspecified sleep opportunity windows, which in turn will result in inaccurate estimation of sleep parameters.
- While the sleep opportunity window can be manually corrected, this can be a subjective and cumbersome process, which may not be achievable in a large dataset with many individuals.
- In our work, we developed an automated outlier detection method for detecting potential abnormalities in sleep opportunity window detection by searching for discrepancies between minute-level sleep and wear status.
- We apply our outlier detection method to the National Health and Nutrition Examination Survey (NHANES) 2013-2014 cohort, deriving quality metrics for each individual and identifying days with anomalous sleep windows. We demonstrate that excluding days based on quality metrics can impact statistical association results, and potentially uncover associations that may otherwise go undetected.
- Our work highlights the importance of proper quality control in wrist-worn sensor-based sleep detection.

2 Methods

Data

- Wrist acceleration data (GT3X, ActiGraph) was collected from 793 subjects aged 50-59 in the National Health and Nutrition Examination Survey (NHANES) 2013-2014 for up to 7 complete days.
- Data from the tri-axial accelerometer was sampled at 80hz and was processed using the sleep analysis pipeline from ActiGraph.

Sleep processing pipeline

- The sleep processing pipeline first consisted of wear detection, followed by total sleep opportunity detection and then finally sleep / wake classification within the total sleep opportunity window. The total sleep opportunity algorithm detects one sleep opportunity per 24-hour cycle (in noon-to-noon windows) and allows for up to 60 minutes of continuous sleep disruption within the sleep opportunity.
- Numerous sleep metrics were calculated in each total sleep opportunity using the DACNN sleep / wake classification algorithm, including: total sleep time (TST), wake after sleep onset (WASO) and sleep efficiency (SE).

Outlier Detection

- Our outlier detection method searches for discrepancies between three components of accelerometer-generated data and derives an outlier score:

Step 1: Derive list of candidate outlier metrics that could predict inaccurate sleep opportunity windows based on training set data.

Training Set: Wear/sleep status, Sleep window → Candidate Outlier Metrics: extra.sleep, extra.sleep.weighted, sleep.nonwear.overlap

Step 2: Visualize training set data and annotate each daily sleep opportunity window as “good”, "truncated", "interrupted", "nonwear" or "misaligned".

Training Set: Wear/sleep status, Sleep window → Good / Truncation, Interruption, Nonwear

Step 3: Based on the annotations, learn which candidate outlier metrics are most relevant to detecting sleep opportunity window anomalies. Establish filtering criteria based on these metrics.

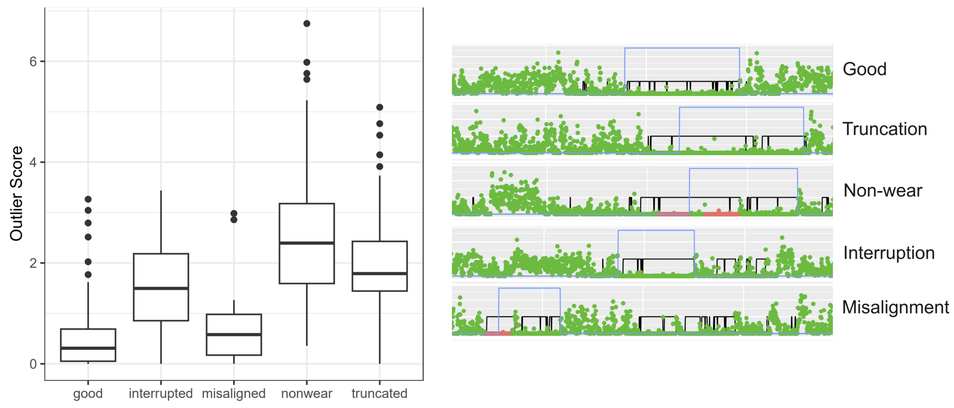
Candidate Outlier Metrics ↔ Annotations → Filtering Criteria

Step 4: Apply filtering criteria to validation dataset, compare reliability/association results for filtered vs unfiltered data.

Validation Set: Wear/sleep status, Sleep window ↔ Filtering Criteria → Reliability, Sensitivity, Association Power

3 Results: Outlier Detection

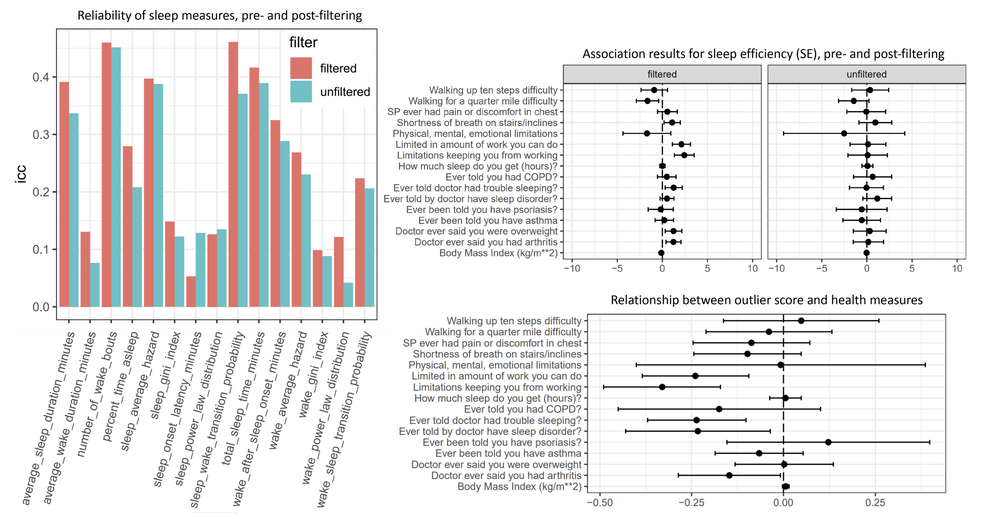
- We randomly sampled 200 individuals (100 training and 100 test cases) from the NHANES age 50-59 cohort and annotated each day (7 days per individual).
- In our test set, most common problematic annotation were truncations (23.2%), followed by interruptions (19.1%), non-wear (18.6%) and finally misalignments (2.7%). Examples of each annotation are shown below.
- For each day, we calculated an outlier score, which was a linear combination of the most relevant candidate outlier metrics.
- The outlier score metric provides strong differentiation between interruptions, nonwear and truncations from good days, while misalignments remain challenging.
- Using the training set, we determined the optimal cutoff for the outlier score that maximizes sensitivity while minimizing false positive rates.



Annotation	N	Sensitivity	False-positive Rate
Good	254 (36.4%)	0.870	0.080
Truncation	162 (23.2%)	0.857	0.192
Interruption	133 (19.1%)	0.820	0.232
Non-wear	130 (18.6%)	0.962	0.209
Misalignment	19 (2.7%)	0.474	0.786

4 Results: Application to NHANES Cohort

- Overall, reliability of sleep measures increased after removing outlier days
- Removing days identified as outliers results in improved association power in NHANES cohort for several measures, (sleep efficiency shown)
- In addition, individuals with higher outlier scores were more likely to be identified as having a sleep disorder, or trouble sleeping.



5 Discussion

- Accurate definition of total sleep opportunity window is an essential component to the accurate estimation of sleep outcome measures.
- While sleep / wake classification at the epoch-level has been validated by PSG, the validation of sleep period window is not well established – a major challenge is the lack of robust, objective ground truth.
- In our work, we propose an automated approach for detection of outlier sleep period windows – in particular, our method can detect many anomalies with high accuracy.
- Detection of truncations are important, as truncated sleep period windows can result in underestimated total sleep time, number of awakenings and sleep onset latency, while overestimating sleep efficiency.
- We also demonstrate that our outlier scores could potentially be used as a sleep outcome measure itself, as higher values of our truncation detection metrics are associated with sleep disorders in the NHANES dataset
- Further research is needed to determine the optimal way of using our outlier measures. In the case of quality control, removing the days with high outlier scores could result in bias, since higher scores are correlated with sleep disorders
- Manual input is still required in correcting these sleep period windows

References

- van Hees, V.T., Sabia, S., Jones, S.E. et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 8, 12975 (2018). https://doi.org/10.1038/s41598-018-31266-z
- Nunes AS, Patterson MR, Gerstel D, et al. Domain Adversarial Convolutional Neural Network Improves the Accuracy and Generalizability of Wearable-based Sleep Assessment Technology. Research Square; 2022. DOI: 10.21203/rs.3.rs-2354368/v1.