# Non-wear or sleep? Evaluation of five non-wear detection algorithms for raw accelerometer data

Matthew N. Ahmadi, MS [1] (ORCID: 0000-0002-3115-338X)

Nicole Nathan, PhD[2-5]

Rachel Sutherland, PhD[2-5]

Luke Wolfenden, PhD[2-5]

Stewart G. Trost, PhD[1] (ORCID: 0000-0001-9587-3944)

1. Institute of Health and Biomedical Innovation at QLD Centre for Children's Health Research, School of Exercise and Nutrition Sciences, Queensland University of Technology, Brisbane, Australia
2. Hunter New England Population Health, Hunter New England Area Health Service, Newcastle; Australia
3. School of Medicine and Public Health, The University of Newcastle, Newcastle NSW Australia
4. Priority Research Centre for Health Behaviour, The University of Newcastle, Newcastle NSW, Australia
5. Hunter Medical Research Institute, New Lambton Heights NSW, Australia

**Corresponding Author:**

Matthew Ahmadi, MS
Institute of Health and Biomedical Innovation at QLD Centre for Children's Health Research
Level 6, 62 Graham Street
South Brisbane, QLD 4101
Australia

Phone: +61 7 3069 7342
Email: matthew.ahmadi@qut.edu.au

**Abstract**

Detection of non-wear periods is an important step in accelerometer data processing. This study evaluated five non-wear detection algorithms for wrist accelerometer data and two rules for non-wear detection when non-wear and sleep algorithms are implemented in parallel. Non-wear algorithms were based on the standard deviation (SD), the high-pass filtered acceleration, or tilt angle. Rules for differentiating sleep from non-wear consisted of an override rule in which any overlap between non-wear and sleep was deemed non-wear; and a 75% rule in which non-wear periods were deemed sleep if the duration was < 75% of the sleep period. Non-wear algorithms were evaluated in 47 children who wore an ActiGraph GT3X+ accelerometer during school hours for 5 days. Rules for differentiating sleep from non-wear were evaluated in 15 adults who wore a GeneActiv Original accelerometer continuously for 24 hours. Classification accuracy for the non-wear algorithms ranged between 0.86 – 0.95, with the SD of the vector magnitude providing the best performance. The override rule misclassified 37.1 minutes of sleep as non-wear, while the 75% rule resulted in no misclassification. Non-wear algorithms based on the SD of the acceleration signal can effectively detect non-wear periods, while application of the 75% rule can effectively differentiate sleep from non-wear when examined concurrently.

KEYWORDS: accelerometry, physical activity, measurement, wearable sensor, sleep

## Introduction

Due to their unobtrusive size, robustness, and low cost, accelerometer-based motion sensors have become the method of choice for measuring physical activity and sedentary behaviour in free-living samples [1, 2]. Detection of non-wear periods is an important step in accelerometer data processing. The determination of valid monitoring days, time spent in sedentary behaviour, and adjusting activity estimates for differences in daily wear time during data analysis, are all dependent on the accurate detection of non-wear periods. Inaccuracies in non-wear detection can lead to substantial variations in physical activity estimates which can impact the validity of associations in observational studies, and increase the risk of bias in intervention trials.

Current methods use strings of consecutive zero counts ranging between 20-90 minutes in duration as an indicator of non-wear time [3-9]. However, these algorithms have been developed for monitoring protocols in which hip or waist mounted accelerometers are worn during the waking hours and are based on proprietary activity counts which cannot be generalised across different accelerometers. As such, they cannot be applied to more contemporary activity monitoring protocols in which raw tri-axial accelerometer signal is collected at the wrist over a full 24-h period [10-12].

To date, non-wear detection algorithms for raw acceleration signal have received little research attention. Van Hees [13, 14] developed a non-wear detection algorithm based on the standard deviation and magnitude of acceleration in each axis, calculated for 60-minute windows with a 45-minute overlap. Additionally, a series of rules is applied to test the plausibility of short wear periods located in-between long periods of non-wear. The algorithm, is based on data from a single lab experiment in which accelerometers were left motionless on a flat surface for 30 minutes and visual inspection of accelerometer data from a single observational study [13]. However, to our knowledge, the validity of the algorithm has not been independently tested and reported on in the scientific literature.

Another important methodological issue related to automatic non-wear detection is the misclassification of sleep as non-wear. Sleep is essential for health and well-being in both children and adults; and there is consistent evidence linking short sleep duration to weight gain and obesity through reductions in physical activity, increases in sedentary behaviours, and alterations in dietary intake [15]. On the weight of this evidence, public health organisations have issued 24-hour movement guidelines recommending a healthy combination of physical activity, screen time, and sleep across the day [16-18]. In studies employing 24-h monitoring protocols to objectively measure physical activity, sedentary time, and sleep and assess compliance with these guidelines, it is likely that non-wear detection algorithms will misclassify sleep as non-wear periods. Therefore, it is necessary to derive rules that can be applied when sleep and non-wear algorithms are implemented in parallel. One approach has been to apply, a simple override rule which allows the researcher to decide if overlapping non-wear periods should replace sleep or vice versa [19, 20]. However, like the aforementioned non-wear detection algorithm, the validity of the override rule has not been formally evaluated.

With this in mind, the purpose of the current study was to evaluate the validity of an algorithm for detection of non-wear periods in raw acceleration data. In addition, to expand the current knowledge in this area, the validity of four new non-wear detection algorithms was evaluated and compared. To address the challenges of implementing non-wear and sleep detection algorithms in parallel, the performance of two rule based heuristics for the simultaneous detection of sleep and non-wear periods was examined.

## Methods

Two completely de-identified wrist accelerometer datasets were used in the study. Dataset 1 comprised accelerometer data collected as part of the physical activity intervention trial involving 12 primary schools in New South Wales, Australia.  Because the study's evaluation plan required participants to only wear the accelerometer during school hours, the dataset provided distinct periods of known non-wear time, which allowed for the evaluation of the non-wear algorithms. Dataset 2 comprised accelerometer data from a study in Queensland, Australia evaluating the performance of a wrist-worn accelerometer activity classification model in free living adults [21]. This dataset was selected because it provided periods of sleep without non-wear periods, which allowed for the evaluation of the decision rules for differentiating sleep from non-wear. The comparative performance of the five non-wear detection algorithms was tested using Dataset and  the best performing non-wear detection algorithm was implemented in parallel with a sleep detection algorithm in Dataset 2. Descriptive characteristics for each sample and information about the monitoring protocols are reported below.

### *Dataset 1*

Forty-seven children (46.7% female; mean age = 8.7 ± 0.4 y) wore an ActiGraph GT3X+ accelerometer (ActiGraph Corporation, Pensacola, FL) on their non-dominant wrist during school attendance on five consecutive days, with the sampling frequency set to 30 Hz. Teachers distributed the monitors individually to students to wear as they arrived to school and collected the monitors from the students as they left school each day. Teachers recorded the time students put on and took off the monitors each day in a wear-time log, which is consistent with previous studies evaluating non-wear algorithms [7, 22-24]. This protocol provided ground truth non-wear periods from midnight to before school and the end of school to midnight, with no sleep periods. Additionally, on one day, a subsample of children removed the accelerometer for class pictures between 11:50 am and 1:20 pm, thus providing an additional non-wear period for comparison. A total of 175 monitoring days were available for analysis.

*Dataset 2*

A total of 15 adults (38% female; mean age = 27.6 ± 6.2 y) wore a GENEActiv Original accelerometer (Activinsights Ltd., Cambridgeshire, UK) on their non-dominant wrist for a 24-h period, with the sampling frequency set to 30 Hz. Participants were instructed to wear the monitor for the full 24-hour period and report times when they may have removed the monitor. Data was considered valid only if the monitor was worn for the full time period. Thus, similar to the approach implemented by Choi and colleagues [3, 4] any non-wear detected by an algorithm was known to be an error (false positive). The day after the 24-h monitoring period, participants completed the Multimedia Activity Recall for Children and Adults (MARCA) [25], a computer-based time use survey that requires participants to recall the activities they completed on the previous day in time intervals of 5 minutes or more, including sleep. Data from the MARCA provided ground-truth for sleep duration over the 24-h monitoring period. The use of self-reported time use data for identifying sleep periods is consistent with the approach used by Van Hees et al. [26, 27] in the development and evaluation of the sleep algorithm implemented in the current study; and has been used in other studies evaluating the validity of sleep-wake period detection algorithms [27-29].

*Non-wear algorithms*

Five non-wear detection algorithms were evaluated. Sample code and data for the non-wear detection algorithms can be found in the following link: https://github.com/MA-QUT/Nonwear.

The first algorithm (VH_30/80), developed by Van Hees et al. [13,14], identified non-wear periods using the standard deviation (SD) and range of the acceleration signal along the x, y, and z axes recorded for 60-minute windows, with a 45-minute overlap. Windows were classified as non-wear if the SD was < 13 m*g* (*g* = gravity-based acceleration units, where 1 *g* = 9.81 $m/s^2$; m*g* = milli *g* where 1000 m*g* = 1*g*) and the range was < 150 m*g* for at least two of the axes. Additionally, all wear periods of < 6 hours and < 30% of the combined duration of their bordering non-wear periods were classified as non-wear. Wear periods of < 3 hours and 80% the duration of their bordering non-wear periods were also classified as non-wear.

The second algorithm (SD_XYZ) was based solely on the SD of the acceleration signal recorded in the x, y, and z axes. Windows were classified as non-wear if the SD of all 3 axes was < 13 mg's for 30 consecutive minutes. The third algorithm (SD_VM) was based on the SD of the vector magnitude (VM) of the tri-axial acceleration signal. Windows were classified as non-wear if the SD of the VM was < 13 mg for 30 consecutive minutes. The fourth algorithm (SUM_HPF), applied a high-pass filter ($4^{th}$ order Butterworth filter with $\omega_0 = 0.25$ Hz) to remove the static gravitational component and residual noise from the acceleration signal leaving only acceleration due to movement [30, 31]. The absolute value of the resultant acceleration signal was then summed for each axis over the entire 30-min window. Windows were classified as non-wear if the sum of the acceleration signal for all three axes was 0 mg for 30 consecutive minutes.

The fifth algorithm (TILT) was based on a lack of change in monitor tilt angle in all three axes relative to the horizontal. Tilt for each axis was determined using the equation:

$$[\mathrm{atan}([\text{axis of interest}]/\sqrt{(\text{1st orthogonal axis}^2 + \text{2nd orthogonal axis}^2)}) * (180/\mathrm{pi})]$$

Windows were classified as non-wear if the change in tilt was < 1.0° in all three axes for 30 consecutive minutes. The 1.0° threshold was chosen because when not worn tilt angle will not change by more than 1.0° for all three axes and a change of more than 1.0° will only occur from a 17.5mg displacement [32].

For the SD_XYZ, SD_VM, SUM_HPF, and TILT algorithms, wear-time periods were classified as non-wear if the duration was < 30 min and < 30% of a bordering non-wear period. This rule was implemented to account for brief wear periods located in-between long periods of non-wear. The 13mg threshold for the VH_30/80, SD_XYZ and SD_VM algorithms was dictated by the residual noise associated with the electrical components, battery voltage, and sampling frequency in the ADXL accelerometers used in Geneactiv, and Kionix accelerometers used in ActiGraph monitors [33, 34].

--Insert Table 1 near here--

*Non-wear and sleep rules*

When the best performing non-wear detection algorithm was implemented in parallel with the Van Hees sleep detection algorithm [27] in Dataset 2, two rule-based heuristics were examined. The first approach implemented a simple override rule. Any overlap between windows of predicted non-wear and predicted sleep time were classified as non-wear. The second approach employed a 75% overlap rule. If the duration of an overlapping non-wear period was < 75% of the predicted sleep period, the non-wear period was classified as sleep. Conversely, if the overlapping non-wear period was > 75% of the predicted sleep period, the sleep period was classified as non-wear. Sample code and data for the decision rules can be found at: https://github.com/MA-QUT/Nonwear.

*Algorithm Evaluation*

For the evaluation of the non-wear algorithms in Dataset 1, accuracy at the instance level (i.e., each 1 second epoch) was evaluated by calculating percent agreement, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve [35]. When the monitoring day served as the unit of analysis, non-wear time estimates were evaluated by calculating the mean bias and 95% limits of agreement (LoA), root mean square error

(RMSE), and mean absolute percent agreement (MAPE). To evaluate the performance of the rules to distinguish sleep from non-wear periods in Dataset 2, estimates of sleep duration were compared to those reported in the MARCA, and mean bias and 95% LoA were calculated.

## Results

Percent agreement, sensitivity, specificity, and area under the ROC curve for the five non-wear algorithms are reported in Table 2. Percent agreement for all five algorithms was excellent. The TILT (96.1%) algorithm had the highest agreement, followed by the SD_VM (95.9%) algorithm. The VH_30/80 (90.0%) algorithm had the lowest percent agreement, but was still excellent. In addition, all five algorithms exhibited excellent sensitivity and good to excellent specificity. The area under the ROC curve provides a measure of classification accuracy that jointly considers sensitivity and specificity [34]. Applying the rubric of Metz [36] the SD_XYZ, SD_VM, SUM_HPF, and TILT algorithms demonstrated excellent classification accuracy, while the VH_30/80 algorithm demonstrated good classification accuracy.

--Insert Table 2 near here--

The results of the day level analysis are presented in Table 3. Based on the daily wear time logs completed by the classroom teacher, average non-wear time was 1089.8 min/day. In comparison, estimated daily non-wear time ranged between 1064.0 mins for the VH_30/80 algorithm to 1091.1 mins per day for the TILT algorithm. Non-wear time was marginally underestimated by four out of the five non-wear algorithms, with SD_VM (0.8 min/day) and TILT (-1.3 min/day) displaying the lowest mean bias. LoA's were comparable in magnitude for SD_XYZ, SD_VM, SUM_HPF, and TILT, whilst VH_30/80 had an LoA that was more than 2-fold larger. RMSE's for the five non-wear algorithms ranged from 63.0 min/day for

TILT to 179.6 min/day for VH_30/80. MAPE for the SD_XYZ, SD_VM, SUM_HPF, and TILT ranged from 4.7% to 5.2%, whilst the VH_30/80 algorithm exhibited a MAPE of 12.9%.

--Insert Table 3 near here--

The performances of the two rules for differentiating sleep from non-wear are reported in Table 4. Relative to the MARCA, the Van Hees sleep detection algorithm overestimated sleep duration by an average of 11.6 min/day (95% LoA = -25.1 min/day to 48.5 min/day). When the best performing non-wear algorithm (SD_VM) was implemented in parallel with this sleep detection algorithm, the override rule resulted in an average of 37.1 min of sleep misclassified as non-wear (95% LoA = 15.1 - 57.9 min). When implementing the 75% overlap rule, there was no misclassification of sleep as non-wear.

--Insert Table 4 near here--

## Discussion

The current study evaluated the validity of five non-wear detection algorithms and two rule-based heuristics to differentiate sleep from non-wear. All five non-wear detection algorithms provided accuracies greater than 0.86, with the SD_VM and TILT algorithms providing the best overall performance. Due to its ease of implementation, the SD_VM is recommended. Additionally, implementation of the 75% rule eliminated all misclassification of sleep as non-wear when the two algorithms were implemented in parallel.

The results suggest that, regardless of the method used to determine non-wear, the length of the non-wear time window was the most influential factor in determining algorithm performance. The SD_XYZ, SD_VM, SUM_HPF, and TILT required 30 consecutive minutes below their representative threshold for a period to be considered non-wear and these algorithms displayed greater than 95% classification accuracy. In comparison, the VH 30/80 required 60 minutes below the respective threshold for a period to be classified as non-wear, which resulted in lower classification accuracy and an RMSE twice that of the other algorithms. The lower performance of the VH 30/80 relative to the other four may thus be explained, at least in part, to its longer window length and different plausibility rules for brief wear periods. For example, in Dataset 1, removing the monitors for 90 minutes for class pictures during the school day resulted in two separate wear periods that were less than 6 hours and less than 30% of bordering non-wear periods. Implementation of the VH 30/80 algorithm thus resulted in the entire school day being classified as non-wear. Additionally, if the monitor was removed a few minutes before the end of the school day, a wear period of less than 6 hours was created, and the entire school day was misclassified as non-wear. Notably, when the non-wear time window for VH 30/80 was reduced to 30 minutes, with a 75% overlap, and the same plausibility rules as the other four algorithms were applied, the performance of the VH 30/80 improved substantially, with ROC-AUC increasing to 0.92, and RMSE and MAE decreasing to 86.3 min and 6.1%, respectively.

Although the VH 30/80, SD_XYZ, and SD_VM non-wear algorithms are all based on the SD of the acceleration signal, each algorithm has its advantages and disadvantages. The VH 30/80 is very sensitive to movement but the least robust to internal mechanical noise, because only two out of the three accelerometer axes have to be above the 13mg threshold. The SD_XYZ algorithm is slightly less sensitive to movement but more robust to noise because all three axes have to be above the threshold. Conversely, the SD_VM is the most sensitive to movement while also being the most robust to noise. When stationary, the VM value will be dominated by acceleration due to gravity on the vertical axis (1000 mg) and thus the SD of the VM will be minimally affected by internal noise that may occur on the two orthogonal axes. When movement occurs, the value of the VM will be dominated by the direction of movement occurring along any of the three orthogonal axis and therefore is more sensitive to movement than the VH 30/80 and SD_XYZ algorithms. The TILT algorithm, in contrast, works on the principle that tilt angle will not change during extended periods of non-wear. The algorithm thus has the advantage of not being reliant on the variability of the acceleration signal. Under static conditions when the monitor is not worn, the orientation will not change by more than 1.0° along any of the axes. However, under dynamic conditions the angle measurements become unstable which may affect the reliability of the derived angles without additional processing. The SUM_HPF algorithm uses a high-pass filter to eliminate any acceleration due to internal noise and gravity, however as there is no universal band-pass filter in the literature, it is also the most imprecise algorithm. Despite the subtle differences between each algorithm, they all are based on lack of monitor movement, which is why they all performed similarly with the most influential factor being the length of period needed to be considered non-wear.

When the best performing non-wear detection algorithm was applied in parallel with a previously validated sleep detection algorithm, brief episodes within the sleep period were identified as non-wear; reflective of the lack of arm movement during portions of nightly sleep. The simple override rule resulted in an average of 37.1 minutes of non-wear classification during predicted sleep periods. We chose the option of replacing sleep with non-wear because: 1) false positives for sleep during prolonged periods of non-wear would lead to inaccurate sleep duration estimates and bias sleep parameter results; and 2) data imputation methods can be applied to instances of non-wear, but not predicted sleep periods. Implementation of the 75% overlap rule completely eliminated the misclassification of sleep as non-wear.  This is because the 75% overlap rule considers the entirety of the sleep duration and not just the period of overlap. In addition, by providing a percentage limit rather than an absolute time limit, the 75% overlap rule affords flexibility for changes in nightly sleep duration.

The current study had several strengths. First, the non-wear algorithms were applied to free-living accelerometer datasets using two accelerometer brands that are widely used by physical activity researchers. Prior research suggests that different accelerometer brands provide different raw acceleration values [37-39]. Therefore, it was important to develop non-wear algorithms that are dependent on change in acceleration (i.e., standard deviation or tilt) and not absolute acceleration values. In addition, the use of time-invariant features allows for the algorithm to be applied over different window lengths. Second, our study evaluated the performance of non-wear detection algorithms in both child and adult samples, thus increasing the generalisability of the results. Third, the study introduced simple rules that can be easily applied when a non-wear algorithm is implemented in parallel with a sleep algorithm.

Opposing these strengths were several limitations. First, although Dataset 1 provided known periods of non-wear before and after the school, the exact start and end of non-wear periods were based on daily wear-time logs completed by classroom teachers, However, logs have been used to evaluate non-wear algorithms in prior studies [7, 22-24]. To determine if inaccuracies in the log were contributing to biased results, a sensitivity analysis was conducted in which the start and end of each day was adjusted by 15 minutes. Notably, the performance of the five algorithms did not change. Second, the algorithms were tested in students who wore an accelerometer during a six-hour school day. As such, each monitoring day did not include short periods of non-wear related to removal of the monitor for bathing/showering or other self-care activities. . Future studies should examine the performance of our algorithms in samples wearing the accelerometer over longer periods, such as studies that monitor activity levels during the waking hours. Third, because Dataset 2 did not contain any periods of non-wear, the rules for differentiating sleep from non-wear only evaluated instances of non-wear misclassification during sleep. Future studies should assess the performance of the rules when differentiating instances of sleep misclassification that occur during monitor non-wear.

In summary, algorithms based on the standard deviation of the VM and tilt angle can be used for automatic detection of non-wear from raw accelerometer data collected at the wrist. The use of a 30-minute rather than 60-minute non-wear windows allows for the identification of legitimate brief non-wear periods that would otherwise be undetected. Application of algorithms that are not dependent on the signal magnitude allows for a consistent approach to non-wear detection across different monitor brands. Additionally, the 75% overlap rule provides a simple method to differentiate sleep periods from non-wear periods when non-wear and sleep detection algorithms are implemented in parallel.

References

1. Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. J Appl Physiol (1985). 2008;105(3):977-87. Epub 2008/07/19. doi: 10.1152/japplphysiol.00094.2008. PubMed PMID: 18635884.

2. Trost SG. State of the art reviews: measurement of physical activity in children and adolescents. J American Journal of lifestyle medicine. 2007;1(4):299-314.

3. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc. 2011;43(2):357-64. Epub 2010/06/29. doi: 10.1249/MSS.0b013e3181ed61a3. PubMed PMID: 20581716; PubMed Central PMCID: PMCPMC3184184.

4. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. Med Sci Sports Exerc. 2012;44(10):2009-16. Epub 2012/04/25. doi: 10.1249/MSS.0b013e318258cb36. PubMed PMID: 22525772; PubMed Central PMCID: PMCPMC3443532.

5. Esliger DW, Copeland JL, Barnes JD, Tremblay MS. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. J Journal of Physical Activity and Health. 2005;2(3):366-83.

6. Evenson KR, Terry Jr JWJRqfe, sport. Assessment of differing definitions of accelerometer nonwear time. 2009;80(2):355-62.

7. Jaeschke L, Luzak A, Steinbrecher A, Jeran S, Ferland M, Linkohr B, et al. 24 h-accelerometry in epidemiological studies: automated detection of non-wear time in comparison to diary information. Sci Rep. 2017;7(1):2227. Epub 2017/05/24. doi: 10.1038/s41598-017-01092-w. PubMed PMID: 28533553; PubMed Central PMCID: PMCPMC5440390.

8. Masse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. Med Sci Sports Exerc. 2005;37(11 Suppl):S544-54. Epub 2005/11/19. PubMed PMID: 16294117.

9. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008;40(1):181-8. Epub 2007/12/20. doi: 10.1249/mss.0b013e31815a51b3. PubMed PMID: 18091006.

10. Doherty A, Jackson D, Hammerla N, Plotz T, Olivier P, Granat MH, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. PLoS One. 2017;12(2):e0169649. Epub 2017/02/02. doi: 10.1371/journal.pone.0169649. PubMed PMID: 28146576; PubMed Central PMCID: PMCPMC5287488 study. PO has previously been a director of Axivity Ltd. NH has previously consulted for Axivity Ltd. The partners of DJ and PO own shares in Axivity. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

11. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. Med Sci Sports Exerc. 2016;48(2):245-53. Epub 2015/09/17. doi: 10.1249/MSS.0000000000000771. PubMed PMID: 26375253.

12. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014;48(13):1019-23. Epub 2014/05/02. doi:

10.1136/bjsports-2014-093546. PubMed PMID: 24782483; PubMed Central PMCID: PMCPMC4141534.

13. van Hees VT, Gorzelniak L, Dean Leon EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One. 2013;8(4):e61691. Epub 2013/04/30. doi: 10.1371/journal.pone.0061691. PubMed PMID: 23626718; PubMed Central PMCID: PMCPMC3634007.

14. van Hees VT, Renstrom F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS One. 2011;6(7):e22922. Epub 2011/08/11. doi: 10.1371/journal.pone.0022922. PubMed PMID: 21829556; PubMed Central PMCID: PMCPMC3146494.

15. Atkinson G, Davenne D. Relationships between sleep, physical activity and human health. *Physiol Behav*. 2007;90(2–3):229–35

16. Tremblay MS, Carson V, Chaput J-P, et al. Canadian 24-hour movement guidelines for children and youth: An integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metab*. 2016;41(6 (Suppl. 3)):S311–27.

17. Okely AD, Ghersi D, Hesketh KD, et al. A collaborative approach to adopting/adapting guidelines - The Australian 24-Hour Movement Guidelines for the early years (Birth to 5 years): An integration of physical activity, sedentary behavior, and sleep. *BMC Public Health*. 2017;17(Suppl 5)

18. World Health Organization. *Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age*. 2019.

19. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. *Sleep Heal*. 2015;1(4):275–84

20. van Hees VT, Fang Z, Zhao JH, Heywood J, Mirkes E, Sabia S. Package 'GGIR' 2019.

21. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. J Sci Med Sport. 2017;20(1):75-80. Epub 2016/07/04. doi: 10.1016/j.jsams.2016.06.003. PubMed PMID: 27372275.

22. Winkler EA, Bodicoat DH, Healy GN, Bakrania K, Yates T, Owen N, et al. Identifying adults' valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol. Physiol Meas. 2016;37(10):1653-68. Epub 2016/09/23. doi: 10.1088/0967-3334/37/10/1653. PubMed PMID: 27652827.

23. Arguello D, Andersen K, Morton A, Freedson PS, Intille SS, John D. Validity of proximity sensor-based wear-time detection using the ActiGraph GT9X. J Sports Sci. 2018;36(13):1502-7. Epub 2017/11/04. doi: 10.1080/02640414.2017.1398891. PubMed PMID: 29099649; PubMed Central PMCID: PMCPMC6086128.

24. Zhou SM, Hill RA, Morgan K, Stratton G, Gravenor MB, Bijlsma G, et al. Classification of accelerometer wear and non-wear events in seconds for monitoring free-living physical activity. BMJ Open. 2015;5(5):e007447. Epub 2015/05/15. doi: 10.1136/bmjopen-2014-007447. PubMed PMID: 25968000; PubMed Central PMCID: PMCPMC4431141.

25. Gomersall SR, Olds TS, Ridley K. Development and evaluation of an adult use-of-time instrument with an energy expenditure focus. J Sci Med Sport. 2011;14(2):143-8. Epub 2010/10/12. doi: 10.1016/j.jsams.2010.08.006. PubMed PMID: 20932797.

26. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. PLoS One. 2015;10(11):e0142533. Epub 2015/11/17. doi: 10.1371/journal.pone.0142533. PubMed PMID: 26569414; PubMed Central PMCID: PMCPMC4646630.

27. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimaki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep. 2018;8(1):12975. Epub 2018/08/30. doi: 10.1038/s41598-018-31266-z. PubMed PMID: 30154500; PubMed Central PMCID: PMCPMC6113241.

28. Barreira TV, Schuna JM, Jr., Mire EF, Katzmarzyk PT, Chaput JP, Leduc G, et al. Identifying children's nocturnal sleep using 24-h waist accelerometry. Med Sci Sports Exerc. 2015;47(5):937-43. Epub 2014/09/10. doi: 10.1249/MSS.0000000000000486. PubMed PMID: 25202840.

29. Kinder JR, Lee KA, Thompson H, Hicks K, Topp K, Madsen KA. Validation of a hip-worn accelerometer in measuring sleep time in children. J Pediatr Nurs. 2012;27(2):127-33. Epub 2012/02/22. doi: 10.1016/j.pedn.2010.11.004. PubMed PMID: 22341191; PubMed Central PMCID: PMCPMC3285433.

30. Kangas M, Konttila A, Lindgren P, Winblad I, Jamsa T. Comparison of low-complexity fall detection algorithms for body attached accelerometers. Gait Posture. 2008;28(2):285-91. Epub 2008/02/26. doi: 10.1016/j.gaitpost.2008.01.003. PubMed PMID: 18294851.

31. Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans Inf Technol Biomed. 2006;10(1):156-67. Epub 2006/02/01. PubMed PMID: 16445260.

32. Pedley M. Tilt sensing using a three-axis accelerometer. J Freescale semiconductor application note. 2013;1:2012-3.

33. Analog Devices. Accelerometers. Available from: https://www.analog.com/en/parametricsearch.

34. Kionix. Accelerometers. Available from: https://www.kionix.com/parametric/Accelerometers.

35. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561-77. Epub 1993/04/01. PubMed PMID: 8472349.

36. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8(4):283-98. Epub 1978/10/01. PubMed PMID: 112681.

37. John D, Sasaki J, Staudenmayer J, Mavilia M, Freedson PS. Comparison of raw acceleration from the GENEA and ActiGraph GT3X+ activity monitors. Sensors (Basel). 2013;13(11):14754-63. Epub 2013/11/02. doi: 10.3390/s131114754. PubMed PMID: 24177727; PubMed Central PMCID: PMCPMC3871080.

38. Rowlands AV, Fraysse F, Catt M, Stiles VH, Stanley RM, Eston RG, et al. Comparability of measured acceleration from accelerometry-based activity monitors. Med

Sci Sports Exerc. 2015;47(1):201-10. Epub 2014/05/30. doi: 10.1249/MSS.0000000000000394. PubMed PMID: 24870577.

39. Montoye AHK, Nelson MB, Bock JM, Imboden MT, Kaminsky LA, Mackintosh KA, et al. Raw and Count Data Comparability of Hip-Worn ActiGraph GT3X+ and Link Accelerometers. Med Sci Sports Exerc. 2018;50(5):1103-12. Epub 2017/12/29. doi: 10.1249/MSS.0000000000001534. PubMed PMID: 29283934.

Table 1: Non-wear Detection Algorithms

| Non-wear algorithm | Non-wear detection |
|---|---|
| VH_30/80 | 60 min sliding window with 45 min overlap<br>SD < 13 mg AND peak-to-peak < 150 mg for 2 out of 3 axes for 60 minutes<br><br>Wear-periods < 6 hours AND < 30% of bordering non-wear period is non-wear<br><br>Wear-periods < 3 hours and < 80% of bordering non-wear period is non-wear |
| SD_XYZ | 1 second sliding window<br>SD of x, y, and z axes < 13 mg for 30 consecutive minutes.<br><br>Wear periods < 30 minutes AND < 30% of bordering non-wear periods is non-wear |
| SD_VM | 1 second sliding window<br>SD of VM < 13 mg for 30 consecutive minutes.<br><br>Wear periods < 30 minutes AND < 30% of bordering non-wear periods is non-wear |
| SUM_HPF | 1 second sliding window<br>Apply 0.25 Hz High pass filter, and the Sum of x, y, and z axes = 0 mg for 30 consecutive minutes<br><br>Wear periods < 30 minutes AND < 30% of bordering non-wear periods is non-wear |
| TILT | 1 second sliding window<br>Change in tilt < 1.0° in all three axes for 30 consecutive minutes<br><br>Wear periods < 30 minutes AND < 30% of bordering non-wear periods is non-wear |

Table 2: Classification Performance of Non-wear Detection Algorithms Applied in Dataset 1

| **Algorithm** | **Agreement (%)** | **Se(%)** | **Sp(%)** | **AUC** |
|---|---|---|---|---|
| SD_XYZ | 95.8 | 97.1 | 92.0 | .95 |
| SD_VM | 95.9 | 97.2 | 91.9 | .95 |
| SUM_HPF | 95.6 | 96.6 | 92.3 | .94 |
| TILT | 96.1 | 97.4 | 91.9 | .95 |
| VH 30/80 | 90.0 | 92.2 | 81.7 | .86 |

Se(%), Sensitivity; Sp(%), Specificity; AUC, Area Under the Curve

Table 3: Performance of Non-wear Detection Algorithms for predicting daily non-wear time in Dataset 1

| Algorithm | Non-wear (SD) | Mean Bias (95% LOA) | RMSE | MAPE |
|---|---|---|---|---|
| Wear-time | 1089.8 (30.2) | | | |
| SD_XYZ | 1087.5 (61.8) | 2.3 (-19.4 – 24.1) | 67.2 | 5.0% |
| SD_VM | 1089.0 (61.6) | 0.8 (-20.9 – 22.4) | 65.9 | 4.9% |
| SD_HPF | 1081.5 (62.2) | 8.3 (-13.6 – 30.2) | 70.4 | 5.2% |
| Tilt | 1091.1 (59.8) | -1.3 (-22.4 – 19.8) | 63.0 | 4.7% |
| VH 30/80 | 1064.0 (172.9) | 25.8 (-37.8 – 89.4) | 179.6 | 12.9% |

Values indicate minutes/day; Mean Bias: Observed - Predicted

Table 4: Non-wear and Sleep Algorithm Applied in Parallel with Decision Rules in Dataset 2

| **Algorithm** | **Non-wear** | **Sleep** |
|---|---|---|
| Ground-Truth | 0.0 (0.0) | 417.3 (60.2) |
| Sleep | | 428.9 (84.2) |
| Decision Rules: | | |
| Override | 37.1 (43.6) | 391.8 (77.9) |
| 75% Rule: | 0.0 (0.0) | 428.9 (84.2) |

Values indicate Mean (SD) minutes/day