GGIR/GENEActiv Report

Amy R Hodgson

02/08/2019

Contents

The studies in this review were identified from the GGIR publication list (https://github.com/wadpac/GGIR/wiki/Publication-list). Articles not available due to paywall were not included.

1 Introduction

In this review I will begin by briefly describing the development of the GGIR package as a means of analysing raw accelerometry data, how the data is processed with this method, and how key variables are derived. Then I will describe how the method has been used by others (e.g. the variables extracted, the set-up of the device), and explain recent developments and additions within GGIR. Following this I will discuss how the results from GENEActiv devices, with variables derived via GGIR, may differ from other methods. Finally, I will describe the use of GGIR in different populations, in which the method may not have been validated.

2 Raw accelerometry and the GGIR package

Physical activity and sleep can be measured objectively by accelerometry from participants in natural living conditions. Previously, acceleration would be expressed as 'count' values which are manufacturer-specific units determined by a proprietary algorithm. This limits the comparability of data across studies, and limits the researcher's control over data processing. However in recent years, technological advancements have allowed collection of data in units of gravitational acceleration, which is a universal measure that is standard across devices and accelerometer brands (Van Hees et al. 2010). This is referred to as raw accelerometry.

2.1 GGIR

GGIR is an R package designed to process raw accelerometry data and extract variables related to physical activity and sleep. The signal processing perfomed by GGIR includes calibration, detection of abnormally high values and non-wear, and calculating the magnitude of acceleration. This information is then used to assess levels of physical activity, inactivity, and sleep. GGIR has been cited by over 100 studies so far, and this method has been employed in large population studies such as the Whitehall II cohort and UK Biobank. A full list of the insitutions using GGIR is available here (https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html#who-has-been-using-ggir), and a list of publications here (https://github.com/wadpac/GGIR/wiki/Publication-list).

2.2 GGIR processing

The processing performed by GGIR (auto-calibration, non-wear detection and imputation, choice of metric for acceleration) is described by the package author in the vignette (https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html). Further detail on the acceleration metric can be found in Hees et al. (2013) and on auto-calibration in Van Hees et al. (2014). As these aspects have been explained adequately elsewhere, I will not describe these in detail in this review. Detection of sleep is not explained in the vignette, and so I will describe the method and its validation briefly below.

2.3 Sleep detection in GGIR

2.3.1 The method

A paper by the developer of the package (Van Hees et al. 2015) describes the algorithm used to detect sleep in GGIR. This method is also tested against a sleep log in a sample of 4094 participants, and against PSG in 28 participants. While some algorithms had been developed to extract physical activity variables from raw accelerometry data, there was little work done on deriving sleep variables. Detecting sleep from acceleration alone was challenging, as inactivity could represent either sleep or daytime sedentary behaviour. However, the algorithm developed by van Hees and colleagues also uses the angle of the arm wearing the accelerometer to help distinguish periods of sleep from daytime inactivity. Sleep is characterised by reduced changes in arm angle relative to wakefulness.

$$angle = \left(\tan^{-1} \frac{a_z}{\sqrt{a_x^2 + a_y^2}} \right) \cdot 180/\pi$$

Figure 1: The formula for estimating arm angle. a_x, a_y and a_z are the median values of the three orthogonally positioned raw acceleration sensors in g-units (1g = 1000 mg) derived based on a rolling five second time window.

The figure above shows the formula for estimating arm angle. Arm angle was averaged over 5 second epochs, which then allows change between successive epochs to be assessed. When there was no change in arm angle greater than 5 degrees over at least 5 minutes, this was classified as a bout of sustained inactivity, and therefore a potential sleep period. In this validation study, sustained inactivity was interpreted as sleep if it occurred within the window defined as sleep by a sleep log. 'Time in bed' was defined as the difference between the first bout of sustained inactivity within this window (sleep onset) and the end of the final inactivity bout within this window (wake time). Total sleep duration was the sum of all inactivity bouts within this window.

2.3.2 How well does it agree with self report?

The kappa coefficients indicate fair agreement between self-report and accelerometer measures of time in bed, and moderate agreement for total sleep duration. Total sleep duration assessed by accelerometer tended to be higher than self-report in short sleepers and lower in long sleepers. Agreement between the two methods for 'time in bed' was lower in women, depressed participants, people with higher insomnia symptoms, and on weekend days. These differences were not found for total sleep duration.

2.3.3 How well does it agree with PSG?

Agreement between accelerometer estimates of sleep and PSG were good. When a 5 minute window and 5 degree angle threshold were used, on average there was a 31 minute overestimation of sleep duration, and 83% accuracy.

3 How accelerometers and GGIR have been used

3.1 Accelerometer settings (i.e. Hz)

There appears to be no real consensus on the ideal frequency to use when setting up the devices, as is noted by Rowlands et al. (2016). Generally researchers choose the highest frequency setting possible, considering the length of the study and battery life of the device. However, even a 'low' frequency setting such as 30Hz should provide rich data, as this means that a measurement is taken approximately 30 times per second (mentioned in e.g. Bakrania et al. 2016; Menai et al. 2017). I am not aware of any study that compares data collected using different frequency settings, but I would not expect this to have a significant impact upon results. GGIR works with values aggregated to epochs (of e.g. 1 or 5 seconds) even though much more detailed data is collected, and this has the effect of helping to standardise for differences in sampling frequency. The reasons for this are explained in more detail here: https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html#why-collapse-information-to-epoch-level.

3.2 Variables extracted with GGIR

The majority of studies using GGIR are focused on measuring physical activity rather than sleep. GGIR provides an extensive range of physical activity variables, such as the amount of time spent in sedentary, light, moderate, and vigorous physical activity, and data on acceleration levels in milli-gravity (mg). These variables, like most variables in GGIR, are provided for for all days, just week days, and just weekend days.

There are also a number of studies using GGIR to assess sleep. The sleep variables generally derived were sleep onset, wake time, sleep duration, time in bed, and sleep efficiency (defined as total sleep time / time in bed), (e.g. Bradley et al. 2017; Duncan et al. 2016, 2018; Fairclough et al. 2017; Gubelmann et al. 2018; Horne et al. 2018; Jones et al. 2019; Koopman-Verhoeff et al. 2019; Lane et al. 2019; Papandreou et al. 2019; Richmond et al. 2019; Shepherd et al. 2018; Warehime et al. 2018). 'Sleep variability' has beeen measured as

the standard deviation of sleep duration, and 'sleep fragmentation' as the number of sleep episodes within the sleep period time window identified by GGIR [Jones et al. (2019); these variables are available in the GGIR output]. Sleep onset latency does not seem to be used, as there is no functionality to use the event marker (button) on the GENEActiv devices to indicate bed time, though this variable could presumably be calculated if the participant provided data on their bed time as part of a sleep diary. GGIR provides data on the number and duration of night time awakenings (this variable is named 'nightwak' in the part 5 GGIR output) which has also been used in published studies (e.g. Duncan et al. 2016).

Circadian measures can also be derived from GGIR, though these have been less commonly used so far. The timing and acceleration levels (mg) of the most and least active hours of the day are included in the part 5 GGIR output. Timing of the least active 5 hours (L5) has been used several times as an indication of sleep timing (Jones et al. 2019; Richmond et al. 2019). One study investigating the genetics of chronotype (Jones et al. 2019) used L5 timing, the midpoint of sleep, and the midpoint of the most active 10 hours (M10) of the day as objective measures of sleep timing. Another study (Bradley et al. 2017) used the mean acceleration recorded in the L5 and M10 to calculate the relative amplitude between day and night activity. In this study they used a combination of sleep estimates and the actograms produced in the GGIR output to classify participants by length of sleep (normal, short, long) and circadian timing (delayed, advanced, irregular, non24). The actograms were also used by Horne et al. (2018) to identify sleep-wake patterns in patients in a psychiatric unit. I have not found any papers including the IV or IS, though these variables are also provided in the part 2 GGIR output.

The light sensor on the devices is generally not used, primarily due to concerns about the reliability of the data being affected by the sensor being covered by sleeves during the day or by bedding at night (Van Hees et al. 2015). However, in one case it was used in a study on patients in an inpatient psychiatric unit to identify a relationship between night-time observation of patients by unit staff and disturbance of patient sleep (Horne et al. 2018).

3.3 GGIR Settings

3.3.1 Exclusion/inclusion criteria for data

Typically, days when the participants have worn the devices for less than 16 hours are excluded from the analysis. This is the default setting in GGIR, but this can be changed if desired. Participants who do not wear the device for a sufficient length of time for a minimum number of days (depending on length of the study) are generally also excluded. There is also an option in GGIR to exclude the first and last days from the analysis, which has been used to remove the influence of changes in the participant's behaviour after being fitted with the device.

3.3.2 Sleep log

Providing GGIR with the data from a simple sleep log (indicating sleep and wake times) can be used to improve sleep estimates (Van Hees et al. 2015). If this information is not provided, the 'sleep window' will be estimated by GGIR's algorithm, which has a fairly high level of accuracy but may make errors, particularly if participants have sleep disorders. For this reason it is preferred to have the data from a simple sleep log, which can be used to distinguish sleep from other sedentary time.

3.3.3 Physical activity thresholds

By default GGIR uses a certain set of thresholds (in mg) for light, moderate, moderate-to-vigorous, and vigorous levels of activity, which were determined in a study by Hildebrand et al. (2014). These are typically not altered, though if required (for example, if the study is being conducted in a population in which these thresholds were not validated) these threshold values can easily be altered in the GGIR shell script, or multiple sets of thresholds can be specified.

4 Recent developments in GGIR

4.1 Detecting a 'sleep period window' without sleep diary

A 2018 paper by van Hees (developer of the GGIR package) describes an algorithm to detect a sleep period time window (SPT-window) in the absence of sleep diary data. The SPT window was defined as 'the time window starting at sleep onset and ending when waking up after the last sleep episode of the night'. Once this window is detected, the previously published method (explained above) can be used to detect sleep within this window. The algorithm was then tested in its ability to estimate sleep variables (waking time, sleep onset time) against a sleep diary in 3752 participants, and against PSG in both sleep clinic patients (n = 28) and healthy sleepers (n = 22).

When the SPT algorithm was used, estimates of sleep onset were 12.5 minutes earlier than sleep diary on average for men, and 7.5 minutes earlier in women. It also estimated sleep onset 3.9 minutes earlier per 10 years of participant age relative to mean age, and 3.0 minutes earlier for weekend days. This performance is significantly better than the previous algorithm used to detect 'time in bed' in the absence of a sleep log. The SPT window was 10.9 and 2.9 minutes longer compared to sleep diary in men and women respectively.

No significant differences were found between accelerometer-derived and PSG measures of sleep onset, wake time, SPT-window duration or sleep duration within the SPT window. The SPT durations only differed by more than 2 hours in 7 participants, six of whom had sleep disorders. The C-statistic (concordance statistic) compared to PSG was 0.86 in sleep clinic patients and 0.83 in healthy sleepers, and the average sensitivity to detect sleep within this window was >91%.

Compared to sleep diary, the algorithm's estimates differed on average by less than 15 minutes, supporting the algorithm's accuracy against a sleep diary. Poor classification was more likely to occur in participants showing no deep sleep or long periods of wakefulness in the night. However the authors note that only 2.4% of all nights in a daily life setting were affected by >1hr. This algorithm may prove useful in cases where sleep log information is not available. It is available in GGIR version 1.5-23 onwards as the in the part 4 output (variables SptDuration and SleepDurationinSpt), and has since been used in published studies (e.g. Jones et al. 2019).

4.2 Machine learning

Another recent study involving van Hees (Van Kuppevelt et al. 2019) investigated how unsupervised machine learning could be used to segment accelerometer-measured activity, and compared this to traditional cut-points approaches. Typically, activity between certain thresholds is registered as either sedentary, light activity, etc. However, there is a complex relationship between acceleration and energy expenditure, which may differ by population, study design, activity type, and other factors. This means that cut-points may overfit to the conditions in which they were determined (usually laboratory-based activities in a specific population). Likewise, supervised machine learning approaches (e.g. Kerr et al. 2017; Montoye et al. 2018) which involve a classifier being trained to distinguish activity type based on a training dataset, may overfit to the experimental conditions under which they were trained.

Unsupervised machine learning methods allow for identification of 'states' within the data, and it is suggested that these approaches could avoid the need for expensive calibration studies. The authors suggest that this data-driven approach may be able to identify physiologically meaningful segments in the data, as it is only provided with data with a known physiological meaning (i.e. magnitude of acceleration). Hidden Semi-Markov Models (HSMM) segment data into time periods that can be clustered into segments with similar behaviour. They invesigated this approach by attempting to relate activity intensity states identified by a HSMM with cut-points categories and time-use diaries in a 500 UK teenagers. The cut-points used for comparison were determined in a previous study, and are those used by default in GGIR (Hildebrand et al. 2014).

HSMM states were found to be related to cut-points categories, but not highly correlated. Although HSMM may be able to identify informative states within the data, it should not be particularly suprprising that they

differ from cut-points approaches. Advantages of HSMM include the fact that it is unbiased by self-report data, avoids issues of generalisability that come with cut-points and supervised approaches, and can account for multi-variate input, which may offer a better description of physical activity. The authors argue that this method may allow for a data-driven approach to understanding how variation in activity (which presumably also includes sleep) relate to health and disease.

It seems to me that the machine learning studies in this area have some way to go before these methods are widely adopted and validated. The authors of this study note that future research is needed to better understand the application and the value of unsupervised machine learning with physical activity data. However, machine learning approaches may have advantages: for example, one study found that supervised machine learning models outperformed the threshold-based method currently used in GGIR to classify activity (Montoye et al. 2018).

5 How does the GENEActiv accelerometer compare to other brands (using GGIR)?

A relatively early study (Hildebrand et al. 2014) compared the Actigraph GT3X+ with the GENEActiv for measuring different levels of physical activity in both children and adults. They found that when attached to the same body position, accelerometer output from the GENEActiv and the Actigraph are comparable in adults. However in children, differences in classification between the two brands are apparent, potentially limiting the comparability of physical activity estiates between brands in this age group.

Another study (Rowlands et al. 2016) has investigated how GENEActiv accelerometers compare to the Actigraph when using GGIR to assess physical activity and sleep. It has been suggested that although the two devices are nearly equivalent, the magnitude of acceleration measured by GENEActiv is slightly greater than that measured by Actigraph (Rowlands et al. 2015, 2018). In this study too, total magnitude of acceleration across 24 hours and the magnitude of the least active 5 hours were significantly greater with the GENEActiv than Actigraph. This meant that the number of minutes between 0 and 40 mg was greater for the Actigraph (by ~12 minutes), and minutes between 40 and 80 mg was greater for the GENEActiv (~5 minutes; the two categories represent sedentary and light activity). However, there were no other significant differences between the devices, for example in their recording of moderate-vigorous physical activity (MVPA) or sleep.

A review by Konstabel et al. (2019) suggested that ActiGraph monitors may underestimate vigorous intensity activity due to firmware features in the devices. As a result, some of the discrepancies between the ActiGraph and the GENEActiv may be due to the greater accuracy of the GENEActiv device for recording higher intensity activity levels.

As for comparing raw accelerometry with the older count-based methods, there is some evidence that the physical activity estimates from the two methods may not be comparable. Buchan and McLellan (2019) compared estimates of moderate, vigorous, and moderate-to-vigorous physical activity across processing methods, and therefore activity cut-points, in the ActiGraph GT3X+ accelerometer. They claimed that due to the large differences in estimates and lack of agreement between methods, the conclusions regarding activity are influenced by the type of data being processed. Another study by Kim et al. (2017) similarly found that estimates from the ActiGraph of sedentary time and MVPA were significantly different across raw and count-based methods. Although these studies were only conducted with one type of accelerometer and examined physical activity rather than sleep, it may be worthwhile bearing in mind that the old and new methods of analysis may not produce comparable results in some respects.

6 Use of GGIR in different populations

In addition to use in healthy adults, GGIR has been used to process and analyse data from a range of other populations, such as children (Koopman-Verhoeff et al. 2019), inpatients in a psychiatric facility (Horne et al.

2018), and patients who have had a stroke (Shepherd et al. 2018). As the processing and sleep detection in GGIR was validated in a population of healthy adults, I thought it may be interesting to examples of its use outside of this group, and identify any issues raised by the authors about this.

6.1 Children

Koopman-Verhoeff et al. (2019) used subjective measures in addition to GENEActiv accelerometers and GGIR to assess sleep duration, sleep efficiency, sleep onset and wake time in children aged 11, who also kept a sleep diary. They found that preschool family irregularity was associated with shorter sleep duration and later sleep onset as measured by the accelerometry, as well as reported sleep problems. This consistency seems to support the accuracy of GGIR sleep measures. However, child psychopatholgy mediated this association only when using reported measures of sleep. The authors attribute this to differences in the aspects of sleep assessed by the two methods.

6.2 Psychological disorders

GENEActiv accelerometers and GGIR were used by Horne et al. (2018) in a study of sleep disturbance on inpatient psychiatric units. They measured total sleep time, daytime naps and sleep fragmentation (frequent night wakening). The actograms produced by GGIR were used in conjunction with measures such as sleep duration to understand the different patterns of sleep and activity shown by different patients with different diagnoses. Significant sleep fragmentation and night time awakening was indentified in this population, consistent with researchers' predictions that night-time checks on patients would lead to sleep disturbance. The light sensors on the device were used to support the relationship between sleep disturbance and night-time observations. The conclusions drawn from accelerometry were consistent with patient self-report of disturbed sleep. The authors comment that 'Accelerometry was reasonably well tolerated but still only worn continuously for 20 out of 32 participants'.

Bradley et al. (2017) evaluated sleep and circadian rhythms in bipolar disorder using both objective (accelerometry, melatonin levels) and subjective methods. The sleep variables derived from accelerometry were sleep onset and offset, total sleep time, time in bed, sleep efficiency, and mean 24h sleep duration. Intra-subject variability in sleep measures was defined as between night standard deviation in the variable over the study period. The relative amplitude between day and night activity was calculated from the least active 5 hours and most active 10 hours (van Someren et al., 1999). This data, alongside the actograms provided by GGIR, was used to classify participants as normal, short, or long sleepers, and circadian rhythm disturbance was indicated by delayed, advanced, irregular, or non24 sleep-wake patterns.

Compliance with the accelerometry was very high, and a range of differences in sleep and circadian rhythms were identified between people with BD and controls, including more fragmented sleep and unstable sleep-wake patterns. Participants whose sleep was rated as 'abnormal' by accelerometry also rated their own sleep quality as lower on the PSQI. Participants with BD and 'normal' sleep showed significantly greater 24 hour melatonin secretion than those classified as 'abnormal', but did not differ from controls with 'normal' sleep. This suggests agreement between accelerometry and other measures of sleep and circadian rhythms, at least in a general sense.

A second analysis of the data from these BD participants was recently published (Bradley et al. 2019), this time analysing the association between sleep and cognitive abnormalities. Previously it has been found that participants with BD show deficits in cognitive functions across a range of cognitive domains. Interestingly, in this study they found that although the BD group showed significant differences to controls on a range of cognitive assessments, such differences were almost entirely due to BD participants with abnormal sleep (defined by accelerometry). BD participants with normal sleep showed very few differences to controls.

6.3 Insomnia

A study on the genetics of insomnia by Lane et al. (2019) identified 57 loci for self-reported insomnia in the UK Biobank. Their effects were subsequently confirmed on self-reported insomnia in a different study, diagnosed insomnia in another, and on accelerometer-derived sleep efficiency and duration in the UK Biobank. The Biobank participants underwent 7 days of wrist-worn accelerometry, from which a range of sleep variables was derived using GGIR. Genetic risk for insomnia was associated with a higher number of sleep episodes, lower sleep efficiency, and greater variability in sleep duration.

6.4 Stroke

Shepherd et al. (2018) used GENEActiv accelerometers and GGIR to measure sleep, physical activity and fatigue in stroke survivors living in the community. They aimed first to assess the utility and acceptability of wrist worn accelerometers in this population, and to then assess their physical activity (total, light, moderate, vigorous), sleep duration and sleep efficiency. Accelerometers were fitted to the participants' non-affected side for 7 days on three occasions. The research team offered assiance in fitting the devices when patients were unable to do so themselves.

Authors reported that compliance with wrist worn accelerometers was very high, with 93% of participants meeting the wear time criteria at each measurement stage. They also claim that the method was able to provide a robust and detailed analysis of sleep and physical activity in stroke survivors, and the feasibility and acceptability of the method was supported. However, it was noted that there are no established cut-points to delineate intensity of physical activity in stroke survivors, and it is therefore uncertain whether the activity cut-points used (validated in a healthy population) were appropriate for classifying activity in stroke survivors. Further research would therefore be needed to establish cut-points to best classify activity in this population. Similarly, larger studies may be needed to assess the efficacy of this method to measure sleep in stroke survivors.

7 Summary

The GGIR package allows for processing and classification of raw accelerometry, and is the primary choice for researchers working with this type of data. It also contains an algorithm for identifying sleep, which has been validated against sleep diaries and PSG. Although GGIR has been developed fairly recently, the sleep and circadian variables derived from GGIR (as well as the physical activity esimates) have been used in dozens of studies and this number has increased significantly every year. Furthermore, GGIR is continually developing, with more useful features being added such as the detection of the 'SPT-window', and research being carried out into the application of machine learning approaches for classifying activity.

The estimates of physical activity and sleep should be broadly similar across accelerometer brands when using GGIR, which processes raw accelerometry data from all devices in the same way. While there may be some differences in estimates of higher levels of physical activity, I have not seen any evidence that sleep estimates will differ. Estimates of physical activity across older count-based methods and newer methods using raw accelerometry may be less comparable, but this would only be relevant when comparing current data to previous results.

It seems that the sleep estimates from GGIR can and have been used successfully in a range of different study populations. However, it must be remembered that the classification of sleep was developed and validated in healthy older adults (Van Hees et al. 2015) and thresholds for physical activity measures were developed in a small group of healthy children and adults (Hildebrand et al. 2014). GGIR allows the thresholds for physical activity to be set by the user, so these can be adjusted if they are found to be inappropriate.

Devices providing raw accelerometry data (such as GENEActiv) are increasingly preferred due to the standardised units of measurement, the greater control available to the researcher over data processing and

analysis, and therefore the increased ability to compare findings across studies. GGIR appears to be a reliable method for processing this data and extracting variables of interest. Importantly, it is free and open source, which means greater transparency in its methods and availability to researchers.

References

Bakrania, Kishan, Thomas Yates, Alex V. Rowlands, Dale W. Esliger, Sarah Bunnewell, James Sanders, Melanie Davies, Kamlesh Khunti, and Charlotte L. Edwardson. 2016. "Intensity thresholds on raw acceleration data: Euclidean norm minus one (ENMO) and mean amplitude deviation (MAD) approaches." Edited by Maciej Buchowski. *PLoS ONE* 11 (10): e0164045. https://doi.org/10.1371/journal.pone.0164045.

Bradley, A. J., K. N. Anderson, P. Gallagher, and R. H. McAllister-Williams. 2019. "The association between sleep and cognitive abnormalities in bipolar disorder." *Psychological Medicine*, January, 1–8. https://doi.org/10.1017/S0033291718004038.

Bradley, A. J., R. Webb-Mitchell, A. Hazu, N. Slater, B. Middleton, P. Gallagher, H. McAllister-Williams, and K. N. Anderson. 2017. "Sleep and circadian rhythm disturbance in bipolar disorder." *Psychological Medicine* 47 (9): 1678–89. https://doi.org/10.1017/S0033291717000186.

Buchan, Duncan S., and Gillian McLellan. 2019. "Comparing physical activity estimates in children from hip-worn Actigraph GT3X+ accelerometers using raw and counts based processing methods." *Journal of Sports Sciences* 37 (7): 779–87. https://doi.org/10.1080/02640414.2018.1527198.

Duncan, Mitch J., Wendy J. Brown, Tracy L. Burrows, Clare E. Collins, Sasha Fenton, Nicholas Glozier, Gregory S. Kolt, et al. 2018. "Examining the efficacy of a multicomponent m-Health physical activity, diet and sleep intervention for weight loss in overweight and obese adults: Randomised controlled trial protocol." *BMJ Open* 8 (10): 26179. https://doi.org/10.1136/bmjopen-2018-026179.

Duncan, Mitch J., Corneel Vandelanotte, Stewart G. Trost, Amanda L. Rebar, Naomi Rogers, Nicola W. Burton, Beatrice Murawski, Anna Rayward, Sasha Fenton, and Wendy J. Brown. 2016. "Balanced: A randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults." *BMC Public Health* 16 (1): 670. https://doi.org/10.1186/s12889-016-3256-x.

Fairclough, Stuart J., Dorothea Dumuid, Sarah Taylor, Whitney Curry, Bronagh McGrane, Gareth Stratton, Carol Maher, and Timothy Olds. 2017. "Fitness, fatness and the reallocation of time between children's daily movement behaviours: An analysis of compositional data." *International Journal of Behavioral Nutrition and Physical Activity* 14 (1): 64. https://doi.org/10.1186/s12966-017-0521-z.

Gubelmann, Cédric, Raphael Heinzer, José Haba-Rubio, Peter Vollenweider, and Pedro Marques-Vidal. 2018. "Physical activity is associated with higher sleep efficiency in the general population: The CoLaus study." *Sleep* 41 (7). https://doi.org/10.1093/sleep/zsy070.

Hees, Vincent T. van, Lukas Gorzelniak, Emmanuel Carlos Dean León, Martin Eder, Marcelo Pias, Salman Taherian, Ulf Ekelund, et al. 2013. "Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity." Edited by Michael Müller. *PLoS ONE* 8 (4): e61691. https://doi.org/10.1371/journal.pone.0061691.

Hildebrand, Maria, Vincent T. Van Hees, Bjorge Hermann Hansen, and Ulf Ekelund. 2014. "Age group comparability of raw accelerometer output from wrist-and hip-worn monitors." *Medicine and Science in Sports and Exercise* 46 (9): 1816–24. https://doi.org/10.1249/MSS.0000000000000289.

Horne, Sam, Katherine Hay, Stuart Watson, and Kirstie N. Anderson. 2018. "An evaluation of sleep disturbance on in-patient psychiatric units in the UK." *Psychiatrist* 42 (5): 193–97. https://doi.org/10.1192/bjb.2018.42.

Jones, Samuel E., Jacqueline M. Lane, Andrew R. Wood, Vincent T. van Hees, Jessica Tyrrell, Robin N. Beaumont, Aaron R. Jeffries, et al. 2019. "Genome-wide association analyses of chronotype in 697,828

individuals provides insights into circadian rhythms." *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-018-08259-7.

Kerr, Jacqueline, Catherine R. Marinac, Katherine Ellis, Suneeta Godbole, Aaron Hipp, Karen Glanz, Jonathan Mitchell, Francine Laden, Peter James, and David Berrigan. 2017. "Comparison of Accelerometry Methods for Estimating Physical Activity." *Medicine and Science in Sports and Exercise* 49 (3): 617–24. https://doi.org/10.1249/MSS.0000000000001124.

Kim, Youngwon, Paul Hibbing, Pedro F. Saint-Maurice, Laura D. Ellingson, Erin Hennessy, Dana L. Wolff-Hughes, Frank M. Perna, and Gregory J. Welk. 2017. "Surveillance of Youth Physical Activity and Sedentary Behavior With Wrist Accelerometry." *American Journal of Preventive Medicine* 52 (6): 872–79. https://doi.org/10.1016/j.amepre.2017.01.012.

Konstabel, Kenn, Swati Chopra, Robert Ojiambo, Borja Muñiz-Pardos, and Yannis Pitsiladis. 2019. "Accelerometry-Based Physical Activity Assessment for Children and Adolescents." In, 135–73. https://doi.org/10.1007/978-3-319-98857-3_7.

Koopman-Verhoeff, Maria Elisabeth, Fadila Serdarevic, Desana Kocevska, F. Fenne Bodrij, Viara R. Mileva-Seitz, Irwin Reiss, Manon H.J. Hillegers, et al. 2019. "Preschool family irregularity and the development of sleep problems in childhood: a longitudinal study." *Journal of Child Psychology and Psychiatry and Allied Disciplines*, April, jcpp.13060. https://doi.org/10.1111/jcpp.13060.

Lane, Jacqueline M., Samuel E. Jones, Hassan S. Dashti, Andrew R. Wood, Krishna G. Aragam, Vincent T. van Hees, Linn B. Strand, et al. 2019. "Biological and clinical insights from genetics of insomnia symptoms." *Nature Genetics* 51 (3): 387–93. https://doi.org/10.1038/s41588-019-0361-7.

Menai, Mehdi, Vincent T. Van Hees, Alexis Elbaz, Mika Kivimaki, Archana Singh-Manoux, and Séverine Sabia. 2017. "Accelerometer assessed moderate-To-vigorous physical activity and successful ageing: Results from the Whitehall II study." *Scientific Reports* 8 (1): 45772. https://doi.org/10.1038/srep45772.

Montoye, Alexander H.K., Bradford S. Westgate, Morgan R. Fonley, and Karin A. Pfeiffer. 2018. "Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer." *Journal of Applied Physiology* 124 (5): 1284–93. https://doi.org/10.1152/japplphysiol.00760.2017.

Papandreou, Christopher, Mónica Bulló, Andrés Díaz-López, Miguel A. Martínez-González, Dolores Corella, Olga Castañer, Jesus Vioque, et al. 2019. "High sleep variability predicts a blunted weight loss response and short sleep duration a reduced decrease in waist circumference in the PREDIMED-Plus Trial." *International Journal of Obesity*. https://doi.org/10.1038/s41366-019-0401-5.

Richmond, Rebecca C., Emma L. Anderson, Hassan S. Dashti, Samuel E. Jones, Jacqueline M. Lane, Linn Beate Strand, Ben Brumpton, et al. 2019. "Investigating causal relations between sleep traits and risk of breast cancer in women: Mendelian randomisation study." *BMJ (Online)* 365: l2327. https://doi.org/10.1136/bmj.l2327.

Rowlands, Alex V., FranÇois Fraysse, Mike Catt, Victoria H. Stiles, Rebecca M. Stanley, Roger G. Eston, and Tim S. Olds. 2015. "Comparability of Measured Acceleration from Accelerometry-Based Activity Monitors." *Medicine & Science in Sports & Exercise* 47 (1): 201–10. https://doi.org/10.1249/MSS.0000000000000394.

Rowlands, Alex V., Evgeny M. Mirkes, Tom Yates, Stacey Clemes, Melanie Davies, Kamlesh Khunti, and Charlotte L. Edwardson. 2018. "Accelerometer-assessed physical activity in epidemiology: Are monitors equivalent?" *Medicine and Science in Sports and Exercise* 50 (2): 257–65. https://doi.org/10.1249/MSS.0000000000001435.

Rowlands, Alex V., Tom Yates, Melanie Davies, Kamlesh Khunti, and Charlotte L. Edwardson. 2016. "Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter?" *Medicine and Science in Sports and Exercise* 48 (10): 1935–41. https://doi.org/10.1249/MSS.0000000000000978.

Shepherd, Anthony I., Richard Pulsford, Leon Poltawski, Anne Forster, Rod S. Taylor, Anne Spencer, Laura Hollands, et al. 2018. "Physical activity, sleep, and fatigue in community dwelling Stroke Survivors." *Scientific Reports* 8 (1): 7900. https://doi.org/10.1038/s41598-018-26279-7.

Van Hees, Vincent T., Zhou Fang, Joss Langford, Felix Assah, Anwar Mohammad, Inacio C.M. Da Silva, Michael I. Trenell, Tom White, Nicholas J. Wareham, and Søren Brage. 2014. "Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: An evaluation on four continents." *Journal of Applied Physiology* 117 (7): 738–44. https://doi.org/10.1152/japplphysiol.00421.2014.

Van Hees, Vincent T., Marcelo Pias, Salman Taherian, Ulf Ekelund, and Søren Brage. 2010. "A method to compare new and traditional accelerometry data in physical activity monitoring." *2010 IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks", WoWMoM 2010 - Digital Proceedings*. https://doi.org/10.1109/WOWMOM.2010.5534986.

Van Hees, Vincent T., Séverine Sabia, Kirstie N. Anderson, Sarah J. Denton, James Oliver, Michael Catt, Jessica G. Abell, Mika Kivimäki, Michael I. Trenell, and Archana Singh-Manoux. 2015. "A novel, open access method to assess sleep duration using a wrist-worn accelerometer." Edited by Delphine S. Courvoisier. *PLoS ONE* 10 (11): e0142533. https://doi.org/10.1371/journal.pone.0142533.

Van Kuppevelt, Dafne, Joe Heywood, Mark Hamer, Séverine Sabia, Emla Fitzsimons, and Vincent Van Hees. 2019. "Segmenting accelerometer data from daily life with unsupervised machine learning." *PLoS ONE* 14 (1). https://doi.org/10.1371/journal.pone.0208692.

Warehime, Shane, Danae Dinkel, Kailey Snyder, and Jung Min Lee. 2018. "Postpartum physical activity and sleep levels in overweight, obese and normal-weight mothers." *British Journal of Midwifery* 26 (6): 400–408. https://doi.org/10.12968/bjom.2018.26.6.400.