Classification and processing of 24-hour wrist accelerometer data

Date of submission of the second revision: Apr 23, 2018

Abstract

Background An important step in accelerometer data analysis is the classification of continuous, 24-hour, data into sleep, wake, and non-wear time. We compared classification times and physical activity metrics across different data processing and classification methods.

Methods Participants (n=576) from the Finnish Retirement and Aging Study (FIREA) wore an accelerometer on their non-dominant wrist for 7 days and nights and filled in daily logs with sleep and waking times. Accelerometer data were first classified as sleep or wake time by log, and Tudor-Locke, Tracy, and ActiGraph algorithms. Then, wake periods were classified as wear or non-wear by log, Choi algorithm and wear sensor. We compared time classification (sleep, wake, and wake wear time) as well as physical activity measures (total activity volume and sedentary time) across these classification methods.

Results Mean (SD) nightly sleep time was 467 (49) minutes by log and 419 (88), 522 (86) and 453 (74) minutes by Tudor-Locke, Tracy and ActiGraph algorithms, respectively. Wake wear time did not differ substantially when comparing Choi algorithm and the log. The wear sensor did not work properly in about 29% of the participants. Daily sedentary time varied by 8−81 minutes after excluding sleep by different methods and by 1−18 min after excluding non-wear time by different methods. Total activity volume did not substantially differ across the methods.

Conclusion The differences in wear and sedentary time were larger than differences in total activity volume. Methods for defining sleep periods had larger impact on outcomes than methods for defining wear time.

Key words: 24-h accelerometry, physical activity, sedentary, sleep

INTRODUCTION

Wearing the accelerometer on the (non-dominant) wrist is gaining popularity as an alternative to hip placement (Doherty et al., 2017; Schrack et al., 2016; Troiano et al., 2014). Wrist-worn accelerometers have been shown to be valid in estimating physical activity energy expenditure in free-living situations (Ellis et al., 2016; Staudenmayer et al., 2015; White et al., 2016) and they have four important advantages over the hip-worn accelerometers: increased participant compliance, increased comfort for 24-hour wear, enabling measurement of sleep duration and quality, and better detection of light activity related to daily tasks, which may be primarily upper body movements (Quante et al., 2015; Schrack et al., 2016; Troiano et al., 2014). However, wearing the device 24 hours/day creates new challenges to accelerometer data processing (McVeigh et al., 2016; Meredith-Jones et al., 2016; Tracy et al., 2014; van der Berg et al., 2016). Before being able to analyze either sleep or physical activity, one needs to separate non-wear, wake and sleep time (Kosmadopoulos et al., 2016; McVeigh et al., 2016). In particular, sedentary behavior, sleep and non-wear time are difficult to distinguish from each other based on the accelerometer readings alone, because they are all comprised of low intensity or no movement, resulting in the accelerometer registering predominantly zero counts (Kosmadopoulos et al., 2016; Quante et al., 2015).

Generally, sleep needs to be defined before wear time, and this can be done either by using participant logs or using sleep algorithms, such as those developed by Tudor-Locke (Tudor-Locke et al., 2014), Tracy (Tracy et al., 2014) and Van Hees (van Hees et al., 2015). Different methods have been used to separate wear time from non-wear time including algorithms, participant logs and wear sensors, all of which have their own strengths and weaknesses. Commonly used non-wear algorithms, such as Troiano (Troiano et al., 2008) and

Choi (Choi, Liu et al., 2011; Choi et al., 2012), define non-wear time based on the number of consecutive zero counts. Although participant logs are commonly used, they increase participant burden, often have missing values and are subject to biases due to recall and social desirability (Keadle et al., 2014; Quante et al., 2015; Shiroma et al., 2015). In addition, some accelerometers have wear sensors, which detect wear times based on capacitive coupling or skin temperature, but their utility remains largely unexplored to date (Intille et al., 2012; Zhou et al., 2015). As the non-wear algorithms cannot detect non-wear periods that are shorter than the minimum number of consecutive zero-counts (Winkler et al., 2012), wear sensors might improve wear estimations by detecting also the short non-wear periods. However, to our knowledge, the utility of the ActiGraph wear sensor in separating wear and non-wear time has not yet been reported.

Previous studies have assessed the effect of different non-wear algorithms on ActiGraph accelerometers' wear time and sedentary time, but mainly from accelerometers worn on the hip during wake time only (Evenson & Terry, 2009; Keadle et al., 2014; Masse et al., 2005; Peeters et al., 2013; Winkler et al., 2012). In addition, some studies have assessed impact of sleep algorithms on sleep and sedentary time in adults wearing the accelerometer on hip (McVeigh et al., 2016; Meredith-Jones et al., 2016) or alternating between hip and wrist placement (Jaeschke et al., 2017; Rosenberger et al., 2016; Zinkhan et al., 2014). However, no previous studies have examined the effects of sleep and non-wear algorithms on the classification of sleep, non-wear and sedentary time based on wrist measurement (Migueles et al., 2017; Schrack et al., 2016).

To address these gaps in the literature, we used accelerometers worn on wrist for 24 hours/day to compare classification times and physical activity metrics across different data processing and classification methods. There were two primary aims of this study. Aim 1) To compare estimates of sleep time from participant logs against three published algorithms (ActiGraph, 2017; Tracy et al., 2016; Tudor-Locke et al., 2014). Aim 2) To compare estimates of wake wear time and number of participants with ≥ 4 valid days using three methods: Choi non-wear algorithm (Choi et al., 2012), ActiGraph wear sensor or log-indicated wear time. Wake wear time estimates are dependent on first identifying and excluding sleep times, and then sequentially excluding non-wear time. Therefore, the three non-wear time detection methods tested in this study were applied to each of the four 'wake time data sets' generated by the algorithms tested in aim one, resulting in 18 pairwise sleep time and 12 pairwise non-wear time comparisons. Aim 3) was to compare common accelerometer metrics, including mean daily vector magnitude (VM) counts/60 seconds, as a measure of total activity volume, and mean sedentary time, when the different combinations of sleep and non-wear detection methods were applied to 24-hour accelerometer data.

METHODS

Participants

Finnish Retirement and Aging Study (FIREA) is an ongoing longitudinal cohort study of older adults in Finland established in 2013. The aim of the FIREA study is to determine how health behaviors and clinical risk factors change during the transition from working to statutory retirement among aging workers. The eligible population for the FIREA study cohort included all public-sector employees whose individual retirement date is between 2014 and 2019 and who were working in year 2012 in one of the 27 municipalities in Southwest Finland or in the 9 selected cities or 5 hospital districts around Finland. We first contacted participants 18 months prior to their estimated retirement date by sending a questionnaire. To those who responded to the questionnaire, we mailed an invitation to participate in the accelerometer sub-study. We mailed the accelerometers to all those participants who returned the signed informed consent and who were still working. The FIREA study was conducted in accordance with the Helsinki declaration, and was approved by the Ethics Committee of Hospital District of Southwest Finland.

For the current study, we included baseline data from the first 604 participants of the accelerometer sub-study who wore the accelerometer between September 20, 2014 and February 18, 2017. We excluded those who returned the accelerometer unused (n = 20) and those who had less than 2 days and 2 nights of recording with log entries on bed time and time of waking up (n = 8). This resulted in 576 participants (95% of the original sample) in the analytic sample, of whom 22 (4%) had night shifts during the measurement period.

Measurements

Protocol We mailed a triaxial ActiGraph wActiSleep-BT accelerometer (ActiGraph, Pensacola, Florida, US) to the participants and asked them to wear the device on their non-dominant wrist starting from the Saturday following receiving the device and continuing until the morning of next Saturday, i.e. 7 days and nights. Participants were instructed to wear the device at all times, including during water-based activities such as swimming, but to remove it for sauna. Participants were provided a daily log, where they were asked to record the date, bedtime and waking time in a log for each day that they wore the device. In some cases, participants also recorded the time they put the device on the first time, and time when they finished the measurement, but these were not requested in the log. After the one-week measurement, participants mailed the devices and logs back to the research office in a pre-paid envelope.

Identification of sleep and waking periods. Figure 1 shows the methods used for detecting sleep, wake and non-wear times as well as outcomes produced with different methods. We used four methods to separate wake and sleep periods: participant logs, algorithms developed by Tudor-Locke and colleagues (Tudor-Locke et al., 2014), and Tracy and colleagues (Tracy et al., 2014) and an algorithm available in ActiGraph's ActiLife software (ActiGraph, 2017).

First, using the *participant logs*, we defined waking period as times between waking and bed times during the same day (or the following day, if the time was past midnight) and sleep period as times between bed time and waking time on the following day (or the same day if the bed time was past midnight). If data on individual date (but not time) were missing, the research assistants imputed the date when entering the log data in database. If data on wake or bed time was missing, the sleep and waking period data was also marked as missing starting from previous recorded time until the following recorded time.

The second sleep detection method was the *Tudor-Locke algorithm* (Tudor-Locke et al., 2014) which first uses Sadeh algorithm to define sleep and wake epochs (Sadeh et al., 1994) and then detects the in-bed and waking times based on the wake epochs. Tudor-Locke algorithm defines the in-bed time as the first five consecutive epochs of sleep, and waking time as the first 10 consecutive wake epochs following a sleep period. We ran the Tudor-Locke algorithm in ActiLife software, where it is available as an automatic sleep period detection option, but it is slightly modified from the original Tudor-Locke algorithm. The ActiLife implementation allows the user to select either Sadeh (Sadeh et al., 1994) or Cole-Kripke (Cole et al., 1992) algorithms to identify sleep and wake epochs (ActiGraph, 2017). In the current study we chose to use the Cole-Kripke algorithm because it was originally validated in adult population using wrist-worn accelerometers (Cole et al., 1992).

The third method, *bedrest algorithm by Tracy and colleagues* was developed with Actigraph GT1M accelerometers worn on the dominant wrist and validated in youth (Tracy et al., 2016). We used the algorithm modified for adult participants: first, the algorithm marks 45-min time blocks with mean axis 1 counts lower than 400 counts/min as sleep time. After this, the algorithms finds a transition time before the first 45-min sleep time block where counts below 300/min mark the first sleep minute and a transition time after the last sleep time block where counts above 800 mark the first waking minute (D. J. Tracy, personal communication, January 26, 2017).

The fourth method for detecting sleep periods was the *ActiGraph algorithm* available in ActiLife software which builds on Troiano’s wear time validation algorithm and defines non-wear times less than 24 hours and with minimum of 5 minutes of non-zero counts as sleep time (ActiGraph, 2017).

Identification of wake non-wear time. We used three methods to differentiate between wake wear and non-wear time: participant logs, Choi algorithm and wear sensor. First, using participant logs, we defined the whole time between start of the measurement (first date and time marked in the log, usually the wake time on the first morning) and end of the measurement (last date and time marked in the log, usually the last sleep time on the 7th day of measurement) as wear time. Thus, we did not remove any non-wear time between start and end time of measurement using the participant log. Second, we used the Choi algorithm which defines non-wear time as 90 consecutive minutes of vector magnitude zero counts, allowing for 2 minutes of non-zero counts, providing that there were 30 minutes of zero counts before or after the non-zero counts (Choi et al., 2011; Choi et al., 2012). The Choi algorithm has later been validated for 24-hour measurement by wrist-worn triaxial accelerometers (Choi et al., 2012). Third, we utilized the wear sensor in wActiSleep-BT that provides minute-by-minute information on wear time based on capacitive coupling (ActiGraph, 2016; Quante et al., 2015). We assessed the functioning of the sensor by visual inspection method, which is described in detail in Online Appendix 1. Visual inspection has previously been used as a reference method for separating wake wear time from sleep periods (McVeigh et al., 2016), and it has shown to be valid for identifying wear and non-wear days (Shiroma et al., 2015).

Statistical analysis

All the continuous variables were normally distributed; thus, linear mixed models were used. The results are shown as means and their Bonferroni corrected 95% confidence intervals. All models included participant and time as random effects, and the estimation method as a fixed effect. We only included nights when participants had recorded both the time when they went to bed and the time when they woke up. For sleep time (Aim 1), the comparisons were done between the sleep time estimation methods. For wake wear time (Aim 2), and physical activity measures (Aim 3), the comparisons were done between different methods to exclude sleep time (log, and Tudor-Locke, Tracy and ActiGraph algorithm) and non-wear time (Choi algorithm, log and sensor). For Aim 3, only valid days and only participants with valid data from ≥4 days were included. Sedentary time was defined as VM counts ≤1853/60sec (Koster et al., 2016).

To visualize the magnitude of the pairwise differences of the estimates obtained after applying the most often employed methods, we used Bland-Altman analysis for paired measurements of a varying true value (Bland & Altman, 2007). The results are shown as mean differences and 95% limits of agreement (LOA). For Aim 1, we compared the three sleep detection algorithms to logs, as participant logs are commonly used to define sleep in 24-h accelerometer measurements. For Aim 2, we compared wake wear time and for Aim 3 sedentary time a) after excluding sleep time by the algorithms to those obtained after excluding sleep time by the diary and b) after excluding non-wear time by participant log or wear sensor to those obtained after excluding non-wear time by Choi algorithm. We chose the Choi algorithm because it is the most commonly used method to identify non-wear time.

The accelerometer counts and wear sensor data were processed in ActiLife v6.13.3 software (ActiGraph, Pensacola, Florida, US) into 1 minute epochs and exported into .csv file. Sleep periods according to Tudor-Locke and ActiGraph algorithms and wear sensor information were also processed in ActiLife software. Sleep periods according to the Tracy algorithm and non-wear time according to the Choi algorithm were calculated in R program using packages “PhysActBedRest” (Tracy et al., 2016) and “PhysicalActivity” (Choi, Zhouwen et al., 2011), respectively. All other analyses were performed using SAS 9.4 statistical software (SAS Institute Inc, Cary, NC, USA).

RESULTS

Mean age of the participants was 62.6 years (standard deviation, SD 1.1), 97 (17%) of them were men. The 576 participants contributed to 3303 nights and 3908 days of data. Of the participants, 534 (91%) and 555 (96%) and had minimum of 6 nights and days, respectively, with log times available. Figure 2 shows an example of wear and wake time defined by different methods.

Effect of sleep algorithms on length of sleep periods (Aim 1). Mean (SD) nightly sleep period was 467 (49) min by participant log and 419 (88), 522 (86) and 453 (74) min by Tudor-Locke, Tracy and ActiGraph algorithms, respectively. Compared to the log, estimates of sleep by Tudor-Locke algorithm were 47 min lower and by Tracy algorithm 51 min higher, whereas estimates derived from ActiGraph algorithm were on average only 12 min lower (Figure 3 and Table 1). Based on the Bland-Altman plots, the differences between methods were slightly larger as the mean in-bed time increased. Figure 3 also shows that there were nights when Tudor-Locke algorithm or ActiGraph algorithm did not detect any sleep.

Effect of different sleep and non-wear detection methods on length of wake wear time and number of included participants (Aim 2). Mean wake time and wake wear time obtained by the different methods, and the number of participants with minimum of 4 valid days are shown in Appendix Table S1 and graphical representation of the results are shown in Appendix Figure S2. Differences in mean wake wear time, VM counts and sedentary time between sleep and non-wear detection methods are presented in Tables 2 and 3, respectively. Wake wear time differed widely, up to 98 min, when different methods were used for excluding sleep periods (Table 2 and Appendix Figure S3). For example, mean (SD) wake wear time was 949 (65), 991 (92), 903 (90) and 957 (83) min when defined by log, Tudor-Locke algorithm, Tracy algorithm and ActiGraph algorithm, respectively, while excluding non-wear time by Choi algorithm in all the cases (Appendix Table S1).

Based on the visual inspection, the wear sensors indicated non-wear during apparent wear time. This can be seen from Appendix Figure S3, panel E, where the mean difference in wake wear time between the sensors and Choi algorithm is -85 min/day. Thus, in the following results we only include data from the functioning sensors (n=409). Mean daily differences in wake wear time derived from different methods to exclude non-wear time varied only up to 24 min (Table 3 and Appendix Figure S3). For example, mean (SD) wake wear time was 949 (65) min for Choi algorithm, 964 (49) min for log and 939 (66) min for wear sensor when sleep periods were excluded based on the logs (Appendix Table S1).

Number of participants with minimum of 4 days of valid data differed only by maximum of 6 participants (1% of the sample) after excluding sleep periods by different methods and non-wear time by Choi algorithm or participant log (Appendix Table S1 and Appendix Figure S2). However, data from 167 participants (29%) were excluded because of non-functioning wear sensors.

Effect of different sleep and non-wear detection methods on vector magnitude counts and sedentary time (Aim 3). Mean VM counts did not vary markedly between the methods when only those with minimum of 4 valid days were included in the analysis: excluding sleep by Tudor-Locke algorithm generally resulted in 2-3% smaller and Tracy 4-5% higher counts than when excluding sleep by log or ActiGraph algorithms (Table 1 and Appendix Table S2). On the contrary, sedentary time varied widely, especially between different methods used to exclude sleep time, shown by wide LOAs in Figure 4, panels A, B and C. Excluding sleep by different methods resulted in 8−81 min differences in daily sedentary time (Appendix Table S2) while using different methods to exclude non-wear time resulted only in 1−18 minute differences in sedentary time (Appendix Table S3).

DISCUSSION

In this study, we compared 24-hour classification times and physical activity metrics derived with four different methods for defining sleep periods and three different methods for defining non-wear time. Our results highlight the large impact of sleep algorithms on estimated sleeping time and resulting sedentary time during waking hours. Compared to the participant log, defining sleep time by the algorithm available in the ActiLife software resulted in only 10–15 min differences in sleep, wake wear and sedentary time, and thus it could be a method of choice when participant logs are not available. The differences between participant log and Choi algorithm in detecting non-wear time were also small and both methods are suitable for excluding non-wear time. Major uncertainty in the functioning of wear sensor and the resulting exclusion of large part of data lead us not to recommend use of ActiGraph's wear sensor for non-wear time detection.

Sleep and waking periods (Aim 1). Although 24-hour measurement with ActiGraph accelerometers using wrist positioning is gaining popularity, the methods for sleep detection using only accelerometer data are not established. We found considerable differences in sleep times identified by different methods, which is in concordance with some previous studies comparing sleep estimations between different methods (Hjorth et al., 2012; McVeigh et al., 2016; Meredith-Jones et al., 2016; Zinkhan et al., 2014). In previous studies Cole-Kripke algorithm, which we used as a basis for Tudor-Locke algorithm, resulted in half an hour more sleep in adults (Zinkhan et al., 2014) but >1 hour less sleep in children (Hjorth et al., 2012) when compared to self-reported sleep in 24-hour wrist measurement. In our sample, Tudor-Locke algorithm indicated about 47 min less sleep than the participant logs. Part of this might be explained by the latency between going to bed and falling asleep, as the participants were asked to fill in the time when they went to bed, which does not necessarily correspond to the time when they fall asleep. In our study, the algorithm developed by Tracy et al. estimated longer sleep periods than the logs. The algorithm allows classifying short sleep periods as naps. Using this option would probably have improved the estimations, however, we decided to assess the original algorithm. The algorithm available in ActiLife software, on the other hand, resulted in sleep period lengths that were close to self-reported sleep lengths. To our knowledge, this is the first study to compare this algorithm to other sleep detection methods. Both the Tudor-Locke and ActiGraph algorithms were originally developed for waist-worn devices, which might in part explain why they indicated longer sleep periods than participant logs: small movement during sleep might be classified as wake time because the counts from accelerometers worn on wrist are inherently higher than those worn on waist.

Wake wear time and number of participants (Aim 2). After using different methods for excluding sleep periods, wake wear time varied widely but number of participants with ≥4 days of data remained almost similar. Although we did not exclude any non-wear time based on the participant log, the differences in wake wear time between different methods for excluding non-wear time were not large. As the Choi algorithm and wear sensors only excluded 15–25 min/day non-wear time from the log-indicated measurement period, it seems that compliance for wearing the device among our participants was very good. The daily non-wear of 15–25 min would be in concordance with short non-wear periods daily caused by removing the device for shower or sauna and it is also similar to daily non-wear time found in a previous study using 24-hour accelerometer measurements (Jaeschke et al., 2017).

To our knowledge, this is the first study to evaluate the functionality of ActiGraph wear sensor. Because almost 30% of the sensors were not functioning properly but indicated non-wear during apparent wear time, we cannot recommend ActiGraph’s wear sensor as a reliable method for defining wear time. However, in cases where the sensor was functioning properly, it detected even short non-wear periods (data not shown), which resulted in few minutes less wear time than was detected by Choi algorithm. Therefore, it can be the method of choice for studies where the researchers can visually inspect the data and confirm that the sensors are working properly to avoid unreliable wear time estimations.

Vector magnitude counts and sedentary time (Aim 3). Our results echoed a previous study that used 24-hour measurement with hip-worn accelerometers in finding that mean VM counts differ only slightly after excluding sleep and non-wear time by different algorithms or participant log (Meredith-Jones et al., 2016). In our study, sedentary time was greatly affected by the data processing decisions, especially the method for defining in-bed time. Depending on the method used for excluding sleep and non-wear time, the participants in our study spent 7.6–9.2 hours/day being sedentary, which is slightly less than 8.5–10.4 hours/day in older adults found in previous studies using objective methods (Harvey et al., 2015). In a previous study, estimations of sedentary time varied up to 50 min/day between sleep detection algorithms and logs in children (Meredith-Jones et al., 2016). Our results were fairly similar, with 8−81 min differences in sedentary time between sleep detection methods.

The strengths of our study include large number of participants from both men and women, with diverse occupational backgrounds, and high compliance in both wearing the monitors and filling out the participant log. In addition, we had participants with wide variety of activity-rest patterns in our sample, including people with night shifts. As a weakness, we did not have a criterion measure ("gold standard"), such as polysomnography, for detecting sleep and thus we were not able to define the validity of different methods for defining sleep time. Previous research shows that participant logs can either overestimate (Silva et al., 2007) or underestimate (Zinkhan et al., 2014) total sleep time compared to polysomnography. However, to facilitate comparisons between different studies that do not include participant logs or the logs are poorly filled in, we provided estimates of different sleep detection algorithms in comparison to participant logs. Participant logs are a more feasible method for assessing sleep time in large scale studies than polysomnography and thus widely used. Participant logs have also been used as a criterion method to which other sleep detection methods are compared (Barreira et al., 2015; van Hees et al., 2015; Zhou et al., 2015). As a limitation, our results might not be generalized to populations with markedly poorer compliance of wearing the device.

In conclusion, we found that data processing decisions have large impact on estimations of sleep, waking wear, and sedentary time in accelerometer measurements with wrist placement and 24-hour measurement protocol. The impact on physical activity volume was smaller. Sleep detection methods had generally larger impact on outcomes than methods for detecting non-wear time, apart from wear sensor which gave unreliable estimates among about 30% of the participants. In studies with no log information, we recommend using ActiGraph algorithm for detecting sleep periods and Choi algorithm for detecting non-wear time.

List of abbreviations

FIREA: Finnish Retirement and Aging Study; LOA: limits of agreement; SD: standard deviation; VM: vector magnitude

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by Juho Vainio Foundation, Finland (to SS); the Academy of Finland (Grants 286294 and 294154 to SS); and Finnish Ministry of Education and Culture (to SS).

Acknowledgements

We are grateful to all the participants of FIREA for taking part in this study. We thank Mrs Saana Myllyntausta, MA, for her help in the visual inspection.

Authors' contributions

SS, AP, EJS and TBH conceptualized and designed this study. SS designed the data collection. AP analyzed the data and AP, EJS, JP, JV and SS contributed to the interpretation of the data. AP drafted the manuscript, with critical revisions from EJS, TBH, JP, JV and SS. All authors approved the final manuscript.

References

ActiGraph. (2016). wGT3X-BT and wActiSleep-BT wear sensor. Retrieved from https://actigraph.desk.com/customer/en/portal/articles/2515554-wgt3x-bt-and-wactisleep-bt-wear-sensor

ActiGraph. (2017). What does the "detect sleep periods" button do and how does it work? Retrieved from https://actigraph.desk.com/customer/en/portal/articles/2515836-what-does-the-%22detect-sleep-periods%22-button-do-and-how-does-it-work-

Barreira, T. V., Schuna, J. M.,Jr, Mire, E. F., Katzmarzyk, P. T., Chaput, J. P., Leduc, G., & Tudor-Locke, C. (2015). Identifying children's nocturnal sleep using 24-h waist accelerometry. *Medicine and Science in Sports and Exercise*, *47*(5), 937-943.

Bland, J. M., & Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics, 17*(4), 571-582.

Choi, L., Liu, Z., Matthews, C. E., & Buchowski, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine & Science in Sports & Exercise, 43*(2), 357-364.

Choi, L., Zhouwen, L., Matthews, C. E. & Buchowski, M. S. (2011). PhysicalActivity: Process physical activity accelerometer data. Retrieved from https://CRAN.R-project.org/package=PhysicalActivity

Choi, L., Ward, S. C., Schnelle, J. F., & Buchowski, M. S. (2012). Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. *Medicine & Science in Sports & Exercise, 44*(10), 2009-2016.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep, 15*(5), 461-469.

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., Preece, S. J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N. J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *Plos One, 12*(2), e0169649.

Ellis, K., Kerr, J., Godbole, S., Staudenmayer, J., & Lanckriet, G. (2016). Hip and wrist accelerometer algorithms for free-living behavior classification. *Medicine and Science in Sports and Exercise, 48*(5), 933-940.

Evenson, K. R., & Terry, J. W. (2009). Assessment of differing definitions of accelerometer nonwear time. *Research Quarterly for Exercise and Sport, 80*(2), 355-362.

Harvey, J. A., Chastin, S. F., & Skelton, D. A. (2015). How sedentary are older people? A systematic review of the amount of sedentary behavior. *Journal of Aging and Physical Activity, 23*(3), 471-487.

Hjorth, M. F., Chaput, J., Damsgaard, C. T., Dalskov, S., Michaelsen, K. F., Tetens, I., & Sjödin, A. (2012). Measure of sleep and physical activity by a single accelerometer: Can a waist-worn actigraph adequately measure sleep in children? *Sleep and Biological Rhythms, 10*(4), 328-335.

Intille, S. S., Lester, J., Sallis, J. F., & Duncan, G. (2012). New horizons in sensor development. *Medicine & Science in Sports & Exercise, 44*(Suppl 1), S24-31.

Jaeschke, L., Luzak, A., Steinbrecher, A., Jeran, S., Ferland, M., Linkohr, B., Schulz, H., & Pischon, T. (2017). 24 H-accelerometry in epidemiological studies: Automated detection of non-wear time in comparison to diary information. *Scientific Reports, 7*(1), 2227-017-01092-w.

Keadle, S. K., Shiroma, E. J., Freedson, P. S., & Lee, I. (2014). Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. *BMC Public Health, 14*(1), 1210.

Kosmadopoulos, A., Darwent, D., & Roach, G. D. (2016). Is it on? an algorithm for discerning wrist-accelerometer non-wear times from sleep/wake activity. *Chronobiology International, 33*(6), 599-603.

Koster, A., Shiroma, E. J., Caserotti, P., Matthews, C. E., Chen, K. Y., Glynn, N. W., & Harris, T. B. (2016). Comparison of sedentary estimates between activPAL and hip- and wrist-worn ActiGraph. *Medicine and Science in Sports and Exercise, 48*(8), 1514-1522.

Masse, L. C., Fuemmeler, B. F., Anderson, C. B., Matthews, C. E., Trost, S. G., Catellier, D. J., & Treuth, M. (2005). Accelerometer data reduction: A comparison of four reduction algorithms on select outcome variables. *Medicine & Science in Sports & Exercise, 37*(Suppl 11), 544-554.

McVeigh, J. A., Winkler, E. A. H., Healy, G. N., Slater, J., Eastwood, P. R., & Straker, L. M. (2016). Validity of an automated algorithm to identify waking and in-bed wear time in hip-worn accelerometer data collected with a 24 h wear protocol in young adults. *Physiological Measurement, 37*(10), 1636.

Meredith-Jones, K., Williams, S., Galland, B., Kennedy, G., & Taylor, R. (2016). 24 h accelerometry: Impact of sleep-screening methods on estimates of sedentary behaviour and physical activity while awake. *Journal of Sports Sciences, 34*(7), 679-685.

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine (Auckland, N.Z.), 47*(9), 1821-1845.

Peeters, G., van Gellecum, Y., Ryde, G., Farías, N. A., & Brown, W. J. (2013). Is the pain of activity log-books worth the gain in precision when distinguishing wear and non-wear time for tri-axial accelerometers? *Journal of Science and Medicine in Sport, 16*(6), 515-519.

Quante, M., Kaplan, E. R., Rueschman, M., Cailler, M., Buxton, O. M., & Redline, S. (2015). Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. *Sleep Health, 1*(4), 275-284.

Rosenberger, M. E., Buman, M. P., Haskell, W. L., Mcconnell, M. V., & Carstensen, L. L. (2016). Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Medicine & Science in Sports & Exercise, 48*(3), 457-465.

Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: An empirical test of methodological issues. *Sleep, 17*(3), 201-207.

Schrack, J. A., Cooper, R., Koster, A., Shiroma, E. J., Murabito, J. M., Rejeski, W. J., Ferrucci, L., & Harris, T. B. (2016). Assessing daily physical activity in older adults:

Unraveling the complexity of monitors, measures, and methods. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 71*(8), 1039-1048.

Shiroma, E. J., Kamada, M., Smith, C., Harris, T. B., & Lee, I. (2015). Visual inspection for determining days when accelerometer is worn: Is this valid? *Medicine and Science in Sports and Exercise, 47*(12), 2558-2562.

Silva, G. E., Goodwin, J. L., Sherrill, D. L., Arnold, J. L., Bootzin, R. R., Smith, T., Walsleben, J. A., Baldwin, C. M., & Quan, S. F. (2007). Relationship between reported and measured sleep times: The sleep heart health study (SHHS). *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine, 3*(6), 622-630.

Staudenmayer, J., He, S., Hickey, A., Sasaki, J., & Freedson, P. (2015). Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of Applied Physiology, 119*(4), 396-403.

Tracy, D. J., Xu, Z., Choi, L., Acra, S., Chen, K. Y., & Buchowski, M. S. (2014). Separating bedtime rest from activity using waist or wrist-worn accelerometers in youth. *Plos One, 9*(4), e92512.

Tracy, D. J., Xy, Z., Acra, S., Chen, K. Y. & Buchowski, M. S. (2016). PhysActBedRest: Marks periods of 'bedrest' in actigraph accelerometer data. Retrieved from https://CRAN.R-project.org/package=PhysActBedRest

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the united states measured by accelerometer. *Medicine & Science in Sports & Exercise, 40*(1), 181-188.

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine, 48*(13), 1019-1023.

Tudor-Locke, C., Barreira, T. V., Schuna, J. M., Mire, E. F., & Katzmarzyk, P. T. (2014). Fully automated waist-worn accelerometer algorithm for detecting children's sleep-period time separate from 24-h physical activity or sedentary behaviors. *Applied Physiology, Nutrition, and Metabolism, 39*(1), 53-57.

van der Berg, J., Willems, P. J. B., van der Velde, J. H. P. M., Savelberg, H. H. C. M., Schaper, N. C., Schram, M. T., Sep, S. J. S., Dagnelie, P. C., Bosma, H., Stehouwer, C. D. A., & Koster, A. (2016). Identifying waking time in 24-h accelerometry data in adults using an automated algorithm. *Journal of Sports Sciences, 34*(19), 1867-1873.

van Hees, V. T., Sabia, S., Anderson, K. N., Denton, S. J., Oliver, J., Catt, M., Abell, J. G., Kivimäki, M., Trenell, M. I., & Singh-Manoux, A. (2015). A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *Plos One, 10*(11), e0142533.

White, T., Westgate, K., Wareham, N. J., & Brage, S. (2016). Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *Plos One, 11*(12), e0167472.

Winkler, E. A. H., Gardiner, P. A., Clark, B. K., Matthews, C. E., Owen, N., & Healy, G. N. (2012). Identifying sedentary time using automated estimates of accelerometer wear time. *British Journal of Sports Medicine, 46*(6), 436-442.

Zhou, S., Hill, R. A., Morgan, K., Stratton, G., Gravenor, M. B., Bijlsma, G., & Brophy, S. (2015). Classification of accelerometer wear and non-wear events in seconds for monitoring free-living physical activity. *BMJ Open*, *5*(5)

Zinkhan, M., Berger, K., Hense, S., Nagel, M., Obst, A., Koch, B., Penzel, T., Fietze, I., Ahrens, W., Young, P., Happe, S., Kantelhardt, J. W., Kluttig, A., Schmidt-Pokrzywniak, A., Pillmann, F., & Stang, A. (2014). Agreement of different methods for assessing sleep characteristics: A comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography. *Sleep Medicine*, *15*(9), 1107-1114.

Table 1 Differences in sleep time between the sleep detection methods

Sleep detection methods being compared	Mean difference (95% confidence interval) in sleep time, min[1]
Log vs. Tudor-Locke algorithm	47 (41 to 52)
Log vs. Tracy algorithm	-51 (-57 to -67)
Log vs. ActiGraph algorithm	12 (6 to 18)
Tudor-Locke algorithm vs. Tracy algorithm	-98 (-104 to -92)
Tudor-Locke algorithm vs. ActiGraph algorithm	-35 (-40 to -29)
Tracy algorithm vs. ActiGraph algorithm	63 (58 to 69)

[1] The confidence intervals are Bonferroni corrected

Table 2 Differences in wake wear time, mean vector magnitude counts, and mean sedentary time between sleep detection methods

	Non-wear detection method		
Sleep detection methods being compared	Choi algorithm Mean difference (95% CI) [1]	Log Mean difference (95% CI) [1]	Wear sensor Mean difference (95% CI) [1]
Wake wear time, minutes			
Log vs. Tudor-Locke algorithm	-46 (-52 to -40)	-36 (-43 to -29)	-50 (-57 to -42)
Log vs. Tracy algorithm	43 (36 to 49)	62 (54 to 69)	38 (31 to 46)
Log vs. ActiGraph algorithm	-12 (-18 to -6)	-7 (-14 to 0.2)	-15 (-22 to -7)
Tudor-Locke vs. Tracy algorithm	88 (82 to 94)	98 (90 to 105)	88 (81 to 95)
Tudor-Locke vs. ActiGraph algorithm	34 (28 to 40)	29 (21 to 36)	35 (28 to 42)
Tracy vs. ActiGraph algorithm	-55 (-61 to -48)	-69 (-76 to -62)	-53 (-60 to -46)
Mean vector magnitude counts/60s[2]			
Log vs. Tudor-Locke algorithm	63 (52 to 75)	43 (31 to 56)	74 (60 to 88)
Log vs. Tracy algorithm	-102 (-114 to -90)	-128 (-141 to -116)	-92 (-106 to -77)
Log vs. ActiGraph algorithm	18 (6 to 30)	8 (-5 to 20)	25 (11 to 39)
Tudor-Locke vs. Tracy algorithm	-166 (-178 to -154)	-172 (-185 to -159)	-166 (-180 to -152)
Tudor-Locke vs. ActiGraph algorithm	-46 (-57 to -34)	-36 (-49 to -23)	-49 (-63 to -35)
Tracy vs. ActiGraph algorithm	120 (108 to 132)	136 (123 to 149)	117 (103 to 131)
Mean sedentary time, minutes[2]			
Log vs. Tudor-Locke algorithm	-38 (-43 to -32)	-30 (-35 to -24)	-42 (-48 to -36)
Log vs. Tracy algorithm	40 (-35 to 46)	50 (44 to 56)	36 (30 to 43)
Log vs. ActiGraph algorithm	-11 (-17 to -6)	-8 (-14 to -2)	-14 (-21 to -8)
Tudor-Locke vs. Tracy algorithm	78 (73 to 83)	80 (75 to 86)	78 (72 to 85)
Tudor-Locke vs. ActiGraph algorithm	26 (21 to 32)	22 (16 to 28)	28 (21 to 34)
Tracy vs. ActiGraph algorithm	-52 (-57 to -46)	-58 (-64 to -52)	-51 (-57 to -44)

[1] The confidence intervals are Bonferroni corrected

[2] Includes only participants with 4 valid days

Table 3 Differences in wake wear time, mean vector magnitude counts, and mean sedentary time between non-wear methods

Non-wear detection methods being compared	Sleep detection method			
	Log	Tudor-Locke algorithm	Tracy algorithm	ActiGraph algorithm
	Mean difference (95% CI) [1]	Mean difference (95% CI) [1]	Mean difference (95% CI) [1]	Mean difference (95% CI) [1]
Wake wear time, minutes				
Choi vs. log	-14 (-17 to -12)	-7 (-10 to -4)	2 (1 to 4)	-12 (-16 to -8)
Choi vs. sensor	10 (7 to 13)	9 (5 to 13)	8 (6 to 9)	8 (4 to 13)
Log vs. sensor	24 (21 to 27)	16 (12 to 20)	5 (4 to 6)	20 (16 to 25)
Mean vector magnitude counts/60s[2]				
Choi vs. log	24 (19 to 28)	6 (3 to 9)	-3 (-3 to -1)	14 (9 to 18)
Choi vs. sensor	-23 (-28 to -18)	-20 (-24 to -16)	-18 (-20 to -17)	-18 (-24 to -13)
Log vs. sensor	-47 (-52 to -42)	-26 (-29 to -22)	-15 (-17 to -14)	-32 (-37 to -26)
Mean sedentary time, minutes[2]				
Choi vs. log	-9 (-11 to -7)	-2 (-3 to -0)	1 (1 to 2)	-5 (-7 to -3)
Choi vs. sensor	9 (7 to 10)	8 (7 to 9)	6 (6 to 7)	7 (5 to 9)
Log vs. sensor	18 (16 to 19)	10 (9 to 11)	5 (4 to 6)	12 (10 to 14)

[1] The confidence intervals are Bonferroni corrected

[2] Includes only participants with 4 valid days

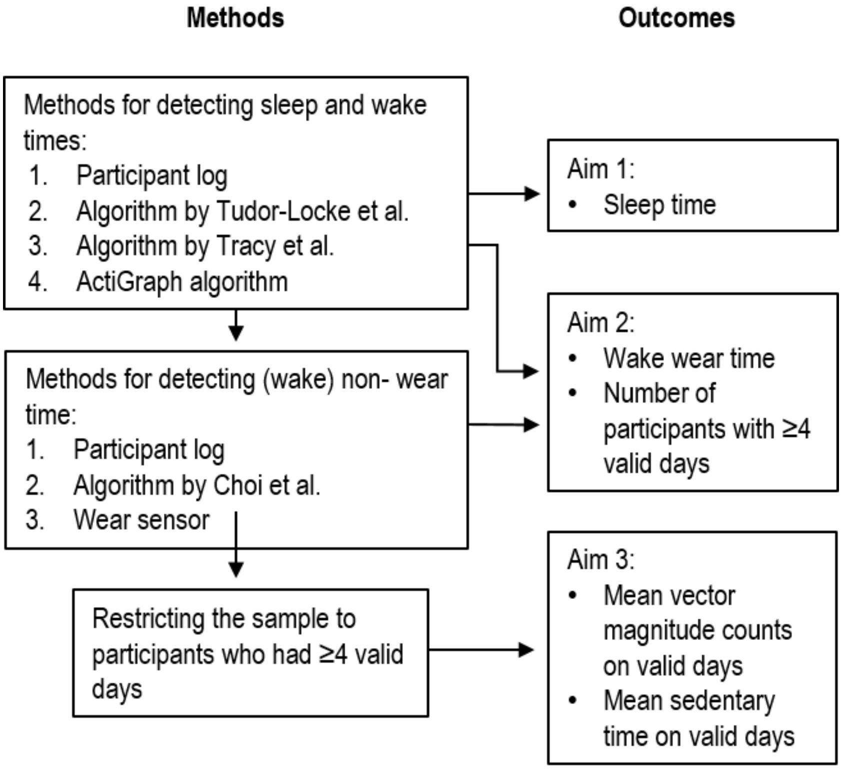

Figure 1 Flow chart of the analyses.

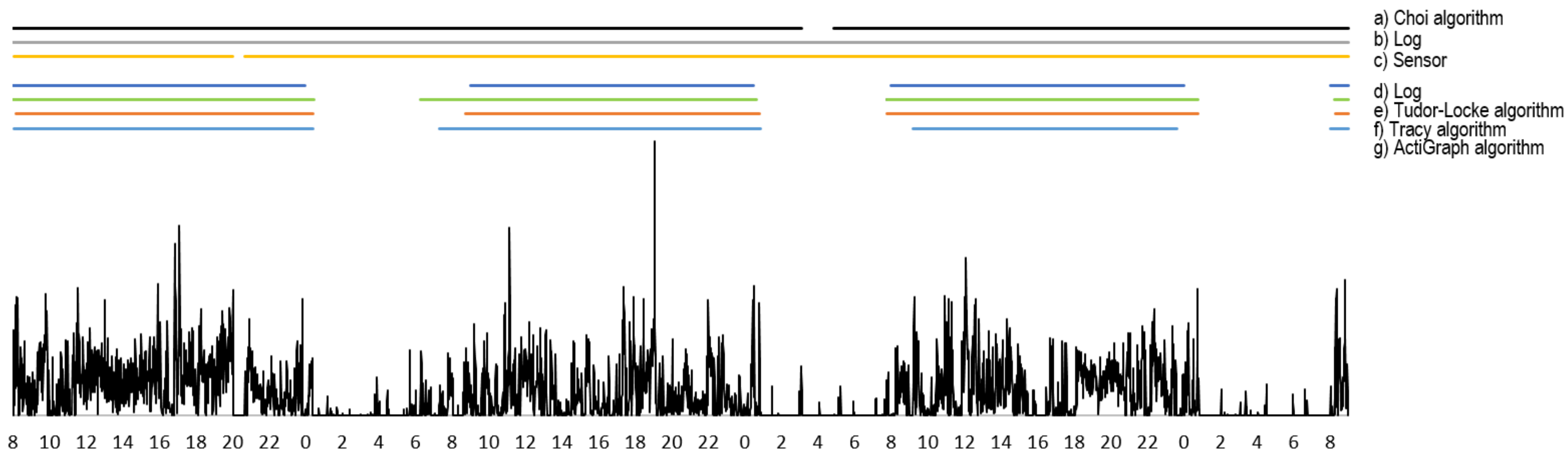

Figure 2 Graphical presentation of wear time (a-c) and wake time (d-g) defined by the different methods.

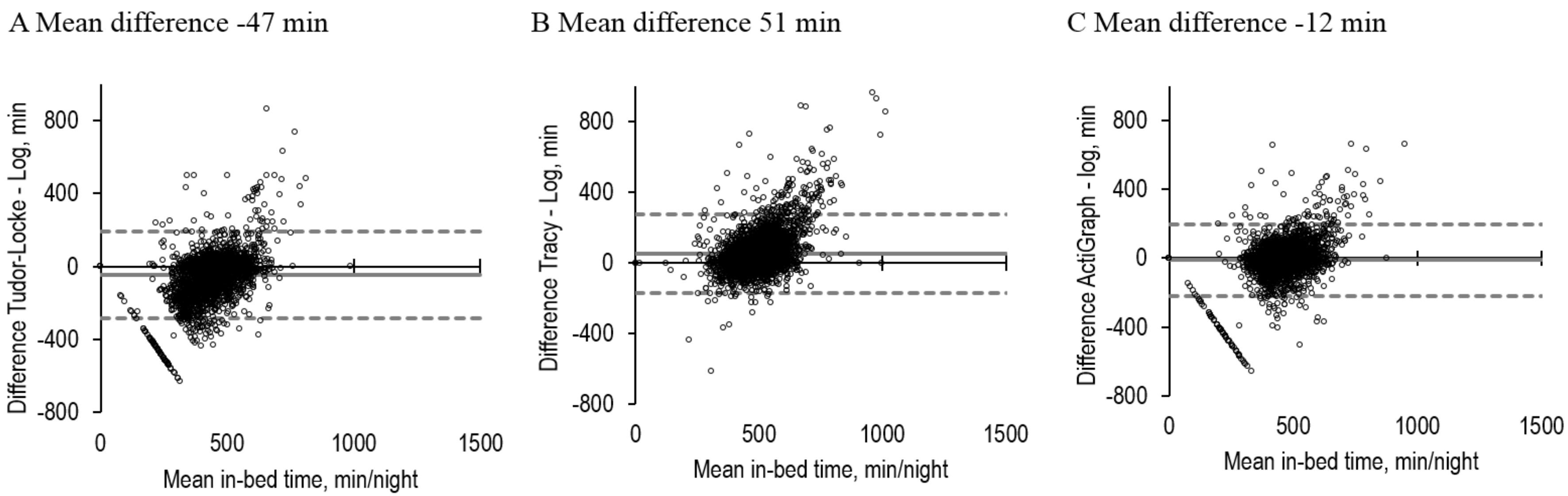

Figure 3 The Bland-Altman plots describing the level of agreement in sleep time defined by the different methods.

Legend for Figure 3: The difference in sleep times between the participant log and a) Tudor-Locke, b) Tracy and c) Actigraph algorithm is plotted against the mean of sleep time obtained by these two methods. Zero bias line (solid gray line) represents the mean of the difference and 95% upper and lower limits are 95% limits of agreement (dashed lines).

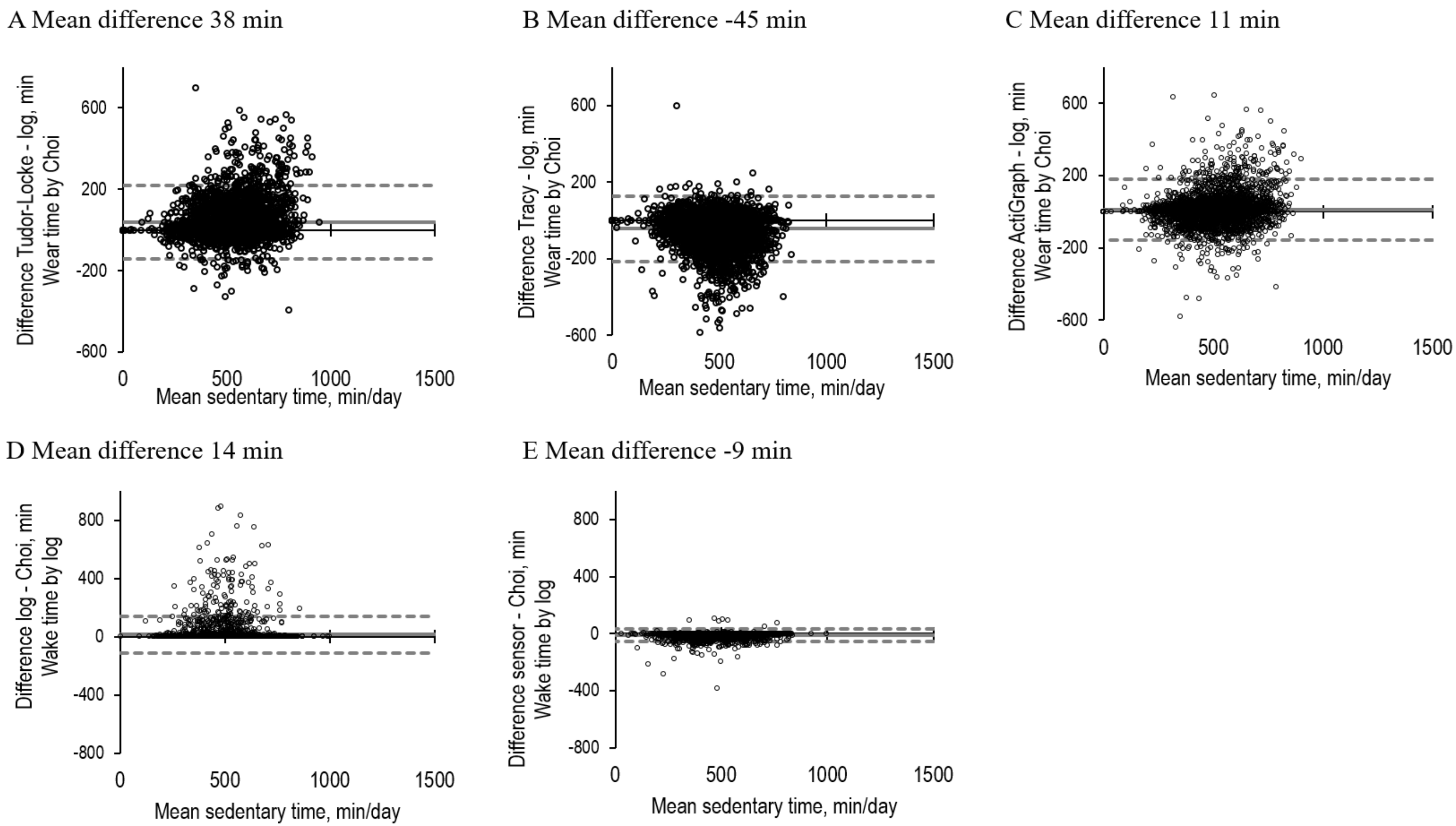

Figure 4 The Bland-Altman plots describing the level of agreement in sedentary time defined by the different methods.

Legend for Figure 4: The difference in sedentary times between the participant log and a) Tudor-Locke, b) Tracy and c) ActiGraph sleep algorithm is plotted against the mean of sedentary time obtained by these two methods, while non-wear time is excluded by the Choi algorithm. The difference in sedentary time between excluding non-wear time by Choi algorithm and d) participant log and e) functioning wear sensors is plotted against the mean of sedentary time obtained after excluding non-wear by these two methods, while waking time is defined by the participant log. Zero bias line (solid gray line) represents the mean of the difference and 95% upper and lower limits are 95% limits of agreement (dashed lines).