Comparison of a thigh worn accelerometer algorithm with diary estimates of time in bed and time asleep: the 1970 British Cohort Study

19th of November, 2020

ABSTRACT

Background: Thigh-worn accelerometers have established reliability and validity for measurement of free-living physical activity-related behaviours. However, comparisons of methods for measuring sleep and time in bed using the thigh worn accelerometer are rare. We compared the thigh-worn accelerometer algorithm that estimates time in bed with the output of a sleep diary (time in bed and time asleep).

Methods: Participants (n=5498), from the 1970 British Cohort Study (BCS70) wore an activPAL device on their thigh continuously for seven days and completed a sleep diary. Bland-Altman plots and Pearson correlation coefficients were used to examine associations between the algorithm derived and diary time in bed and asleep.

Results: Algorithm estimated acceptable levels of agreement with time in bed when compared to diary time in bed (mean bias of -11.4 min; LoA -264.6 to 241.8). The algorithm-derived time in bed overestimated diary sleep time (mean bias of 55.2 min; LoA -204.5 to 314.8 min). Algorithm and sleep diary are reasonably correlated (ρ=0.48, 95% CI: 0.45, 0.52 for women and ρ=0.51, 95% CI: 0.47, 0.55 for men) and provide broadly comparable estimates of time in bed but not for sleep time.

Conclusions: The algorithm showed acceptable estimates of time in bed compared to diary at the group level. However, about half of the participants were outside of the ±30 min difference of a clinically relevant limit at an individual level.

Keywords: accelerometer, activPAL, sleep, sleep diary

Introduction

Lifestyle behaviours are associated with a multitude of health outcomes, including cardiovascular diseases and mortality (Hoevenaar-Blom et al., 2014; Xiao et al., 2014). Among them, the potential health impacts of sleep, as reflected by sleep duration, quality, and timing, are less well explored (Barbaresko et al., 2018), possibly due to the difficulties with robustly measuring sleep-related exposures, including sleep duration. Self-reported sleep duration (short or long sleep duration) is linked to adverse health outcomes including obesity, diabetes, cardiovascular diseases, mood disorders and mortality (Grander, 2017). Although laboratory-based polysomnography is the gold standard of objective sleep measurement, it is impractical in free-living epidemiological studies considering the cost, professional monitoring and large resource demands due to its specialized equipment (Van de Water et al., 2011). Diaries are common low-cost/low-tech alternatives for sleep monitoring in population research. However, diary-based methods could be burdensome for participants and subject to recall bias (Tonetti et al., 2016), among other limitations (Riemann, 2012). 24-h device-based measurement methods might be a less burdensome option to estimate sleep duration in large scale epidemiological studies with the added advantage of not being subject to recall bias.

Wearable devices, non-invasive and inexpensive methods for use in non-laboratory settings, have been increasingly used to estimate sleep, and several studies have examined the agreement between self-reported measures and accelerometer data in different populations (Girschik et al., 2012; Arora et al., 2013; McCrae et al., 2005). Although the thigh-worn accelerometer is considered as the gold standard for free-living measurements of sitting time and posture (Lyden et al., 2017; Dahlgren et al., 2010; Oliver et al., 2011) their uptake in sleep measurement studies is limited. Winkler et al. (2016) and Van der Berg et al. (2016) have recently developed automated algorithms to isolate adults' valid waking wear periods from thigh worn activPAL data collected with a continuous wear protocol. In addition to these algorithms, other time in bed estimation algorithms exist, e.g. the "CREA" algorithm built into the activPAL software (PAL Technologies) which considers 24-h wear time, classifies lying time as primary (e.g. during the night) or secondary (e.g. during the day) and automatically excludes sleeping time. Recent studies have used Winkler et al's (2016) algorithm to calculate sleep duration (Biddle et al., 2018; Ezeugwu, & Manns, 2017). Biddle et al. (2018) have suggested an agreement between results of algorithm-derived and diary-based sleep duration in their study; for example, the algorithm-derived and diary-based sleep time association estimates with fasting glucose were nearly identical (1.01, 95% CI 0.95 to 1.07 and 1.02, 95% CI 0.96; 1.08, respectively).

Despite the increasingly frequent (Biddle et al., 2018; Ezeugwu, & Manns, 2017) use of this algorithm for estimating time in bed, very few studies have compared algorithm derived time in bed with other common measures such as diaries (Winkler et al., 2016; Van der Berg et al., 2016) of sleep time. These studies generally show a correlation between algorithm derived time in bed and sleep diary derived time in bed (Winkler et al., 2016; Van der Berg et al., 2016), although none of them made direct comparisons with sleep time. The aim of this study was to compare Winkler et al's (2016) algorithm-based method that uses the thigh worn accelerometer data with diary estimates of time in bed and time asleep in a large and established population birth cohort from Britain.

Methods

Participants and design

These secondary analyses of available aggregate data have been conducted using the 1970 British Cohort Study (BCS70) data. The BCS70 is an observational prospective population-based cohort study, following the lives of 17,287 people born in a single week of 1970 in England, Scotland, and Wales. In 2016-18, a new wave of data collection was conducted when participants were aged 46-48 years. This comprised of computer assisted personal interviewing to collect the self-reported information via interviews during the home visit (1970 British Cohort Study). Nurses conducted physical examinations and placed the activity monitor on participants. The rationale and sampling methods used in the BCS70 are described in detail elsewhere (1970 British Cohort Study; Elliott, & Shepherd, 2006). All participants gave written informed consent, and the age-46 biomedical survey received ethics from NRES Committee South East Coast - Brighton & Sussex (Ref 15/LO/1446).

Measurements

Computer assisted personal interviews (CAPI) collected data on participants' self-rated general health, disability/limitations, smoking, and occupation. The disability/physical limitation was assessed using the European Statistics on Income and Living Conditions (EU-SILC) (Arora et al., 2015). During the home visit, a nurse took participant's anthropometric measurements including height and weight. Body Mass Index (BMI) was calculated as weight (kg) divided by height squared (m^2).

Participants were asked to wear an activPAL3 device (PAL Technologies, Glasgow, UK) on their thigh for the seven days following their nurse visit. At the end of the visit, nurses placed and attached the devices to the thigh using a medical dressing. The device is a triaxial accelerometer that provides estimated body posture (sitting/reclining/lying, standing) and stepping speed (cadence) based on 3d-acceleration information with a sampling frequency of 20 Hz.

Devices were waterproofed to allow for continuous wear 24 hours (h)/day. Participants were asked to wear the device for seven consecutive days without removing it at any time. After the device was returned, data were downloaded and processed using an open-source program that incorporates the Winkler et al. (2016) algorithm to quantify valid waking wear times by the custodians (Winkler et al., 2016). The Winkler et al. (2016) algorithm was set up to identify time as either a) time in bed or non-wear on a valid day b) waking wear time on a valid day c) any time on an invalid day. This algorithm was developed for use with 24-h wear protocols in adults to classify activity bouts recorded in activPAL `Events` files as `sleep`/ non-wear (or not) and on a valid day (or not). This automated approach excludes long periods without posture change/movement, adjacent low-active periods, and days with minimal movement and wear based on a simple algorithm. Briefly, development of an algorithm to estimate valid waking wear protocols has 4 steps including identifying bouts, examining surrounding bouts, identifying other invalid data and quality control such as checking and error correction. The algorithm was validated based on a minimum of four valid wear days with at least 10 h of waking wear data and > 500 steps (Winkler et al., 2016). The algorithm aimed to measure in bed and non-wear time versus waking wear time. We used in bed and non-wear time together in our analysis. We excluded the first day of data and defined subsequent days as the 24 h between consecutive midnights. Participants providing at least one valid day, defined as waking wear time of more than 10 h per day, were included in the core analysis (van der Velde JH et al., 2018). Figure 1 illustrates each stage of the algorithm.

[Figure 1 here]

Participants were also asked to complete a sleep diary for each day that they wore the monitor. The diary recorded some key information including the exact times (hh:mm) they went to bed, fell asleep, woke up, and got out of bed. There were also separate entries on how many times participants got up during the night and self-rated sleep quality. The full diary is shown in Supplemental Appendix 1.

Data handling

Since the algorithm was designed to distinguish waking wear time from time in bed (Winkler et al., 2016), the algorithm-derived time in bed is computed by subtracting valid waking wear time from 24 h (Biddle et al., 2018). We made both day to day comparisons and mean value comparisons of valid days for sleep diary time in bed and sleep time data as well as algorithm-derived time in bed data. As PSG can measure the sleep accurately, we defined diary-reported sleep time as the time participants wake-up minus the time they fall asleep. Diary time in bed was defined as the time participants get out of bed minus the time they go to bed. Both algorithm data and sleep diary data were calculated as minutes.

Statistical analysis

Statistical analysis was performed using SPSS Version 26.0 (IBM, Chicago, IL, USA) and MedCalc software (Ostend, Belgium). The accelerometer and sleep diary variables were analysed as continuous variables. We ran the $\chi 2$ test and analysis of variance to examine differences by sex for categorical and continuous variables, respectively. Descriptive statistics were calculated for demographic characteristics of participants. We compared algorithm derived time in bed versus-bed time and sleep duration of mean minute/day from diary using the paired sample t-test. We also used Schuirmann's (1987) two one-sided tests (TOST) approach to test equivalence with a specified confidence level between observations. We defined *a priori* differences between algorithm and diary of ±30 min as satisfactory for time in bed and sleep time and calculated 90%

confidence interval (CI). If the entire range of 90% CI of the mean difference lay within the rage of ±30 min, we concluded that the two observations were equivalent. Pearson correlation coefficients were calculated to test the association between the algorithm derived and diary time in bed and sleep time. We used bootstrapping methods to calculate 95% confidence intervals. Differences in correlations across subgroups were tested using Fisher's z test. In addition to Pearson correlation coefficients, we calculated absolute intraclass correlation coefficients (ICC) to assess reliability of repeated measurements among different days as a sensitivity analysis. In stratified analyses we examined the correlation (Pearson coefficients) between the algorithm derived and diary time in bed and asleep across different education groups and health statuses. We conducted Bland-Altman plot with multiple measurements per subject for daily data to examine the agreement of the algorithm-derived and diary-derived times in bed and times asleep. Limits of agreement (LoA) were calculated as bias ± 1.96 SD of the difference. A positive value of the mean difference between algorithm and diary indicates that algorithm overestimates diary data, whereas a negative value indicates that algorithm underestimates diary data. Similar to previous studies (de Zambotti, Baker, & Colrain, 2015; Short et al., 2017), we performed additional Bland-Altman plots where we defined *a priori* differences between algorithm and diary of ±30 min as satisfactory for time in bed and sleep time. The percentage of participants falling within this range is provided. We also performed Bland-Altman plot to examine the agreement of the algorithm-derived and diary-derived times in bed and times asleep for each day separately as sensitivity analysis. Additionally, we used the Bland-Altman plot to examine the agreement of the algorithm-derived and diary-derived times in bed and times asleep. We performed linear regression analysis to evaluate proportional bias. We specified the difference between algorithm and diary as the dependent variable, and the mean of the algorithm and diary as the independent variable. In this analysis, a $P < 0.05$ model coefficient value indicated the presence of proportional bias. In a sensitivity analysis we repeated the above Bland-Altman plots but included participants with at least > 4 d of valid data and at least 20 h per day wear. We also calculated Pearson correlation coefficients for each day. All statistical tests were two-tailed, and values are reported as mean and 95% confidence intervals.

Results

Table 1 shows participants' demographic, health status, and lifestyle health behaviours characteristics. There were no appreciable differences in the accelerometer 24-h waking wear time between men and women (mean 24-h waking wear time 16.0 h/day, standard deviation 1.3 h/day and mean 24-h waking wear time 15.7 h/day, standard deviation 1.3 h/day, respectively). The mean non-wear time for the valid days was 486.0 min (23.4 min -776.4 min).

[Table 1 here]

Table 2 compares absolute accelerometer time in bed with the diary-reported time in bed and asleep. The differences between algorithm-derived and diary time in bed by sex were statistically significant but practically small (mean -15.5 min per day for women, -3.5 min for men). The differences between algorithm-derived time in bed and diary sleep time were larger and also statistically significant (mean 56.0 min for women, 57.4 min for men). According to the TOST approach, we found that algorithm-derived and diary time in bed was equivalent for total, women and men (%90 CI of difference: -11.3, -8.2; %90 CI of difference: -17.6, -13.3; and %90 CI of difference: -5.7, -1.2, respectively) whereas the entire range of 90% CI of difference for algorithm-derived time in bed and diary sleep time was not in the *a priori* defined limits of ±30 min. Table 3 presents Pearson's correlation coefficient between algorithm-derived time in bed and diary time in bed and sleep time (Supplementary Figure 1). The correlation coefficients between algorithm-derived time in bed and diary time in bed were 0.48 in women (95% CI: 0.45, 0.52) and 0.51 in men (95% CI: 0.47, 0.55). The correlation coefficients between algorithm-derived time in bed and diary sleep time were lower for both women (ρ=0.34, 95% CI: 0.30, 0.38) and men (ρ=0.39, 95% CI: 0.35, 0.43). Pearson's correlation coefficient for the association between algorithm-derived time in bed and diary time in bed and asleep for each day also produced similar results to the main analysis (Supplementary Table 1). The ICC correlation coefficients were low for both women and men (Supplementary Table 2). We also compared absolute accelerometer time in bed with the diary-reported time in bed and time asleep by health status (Supplementary Table 3). The mean difference between algorithm derived and diary derived time in bed was lowest for healthy participants (-8.3 min), whereas it was the highest in participants who were severely hampered in activities because of health problems (-28.6 min). We presented correlations between algorithm and diary time in bed and asleep by health status in Supplementary Table 4. We also presented absolute differences between mean amounts of time in bed from the sleep diary and the accelerometer data by education level in Supplementary Table 5 and correlation between algorithm and diary for time in bed and asleep by education level in Supplementary Table 6. The correlations coefficient between algorithm-derived time in bed and diary time in bed were 0.51 in the lowest education level and 0.55 in the highest education level.

[Table 2 here]

[Table 3 here]

The limits of agreement between algorithm-derived time in bed and diary time in bed were shown in the repeated measures of Bland-Altman plots (Figure 2) which depicts a systematic error (with a mean bias of -6.1 min; LoA -260.4 to 248.2 min for men and with a mean bias of -16.1

min; LoA -268.0 to 235.8 min for women). Assuming the diary is the reference method, the algorithm underestimated the time spent in bed for both women and men. Furthermore, linear regression analysis showed that this underestimation was statistically significant (95% CI: 0.398, 0.426, $P < 0.001$) (Table 4). According to *a priori* defined limits, 36.9% of the measurements were in the range of ±30 min difference. The repeated measures of Bland-Altman method for comparison between algorithm-derived time in bed and diary sleep time showed a mean bias of 55.0 min (LoA -205.0 to 315.1 min) for men and a mean bias of 55.3 min (LoA -204.0 to 314.6 min) for women, indicating that algorithm-derived time in bed overestimated sleep time in both sexes (Figure 3). The proportion of the total measurements in the range of clinically acceptable limits (±30 min difference) was 23.5%. Linear regression analysis also demonstrated the presence of proportional bias between algorithm-derived and diary for both sexes indicating that algorithm-derived time in bed underestimated the diary time in bed whereas algorithm-derived time in bed overestimated the diary sleep time. Overall, based on the observed magnitude of the regression coefficients, proportional bias was greater in women than men for all variables (Table 4). The sensitivity analysis of day to day Bland-Altman method for comparison between algorithm-derived time in bed with diary sleep time and time in bed also produced similar results with the main analysis of repeated measures of Bland-Altman plots (Supplementary Table 7). In addition, Bland-Altman agreement between algorithm time in bed and diary time in bed and sleep time showed similar results with the main analysis (Supplementary Figure 2 and 3). For instance, Bland-Altman method for comparison between algorithm derived time in bed and diary time in bed showed a mean bias of -3.5 min (LoA -139.1 to 132.1 min) for men and a mean bias of -15.5 min (LoA -152.7 to 121.8 min) for women indicating an underestimation of time spent in bed for both men and women (Supplementary Figure 2). Algorithm derived time in bed also overestimated the diary sleep time in the Bland Altman plot with a mean bias of 57.4 min (LoA -90.1 to 204.8 min) for men and 56.0 min (LoA -96.1 to 208.0 min) for women (Supplementary Figure 3).

[Table 4 here]

[Figure 2 here]

[Figure 3 here]

In sensitivity analyses we separately examined participants with high (>4d and >20hr) and low (<4d and <20 hr) wear compliance; Bland-Altman plots are shown in Supplementary Figure 4. We also reported participants' time in bed and sleep time according to valid wear days and valid h in Supplementary Table 8and 9. According to Bland-Altman plots, algorithm-derived time in bed overestimated the diary sleep time for both in participants who had more than 4 d activPAL wearing days (a mean bias of 55.3 min, LoA -82.1 to 192.6 min) and less than 4 d activPAL wearing days (a mean bias of 69.9 min, LoA -169.3 to 309.1 min).

Discussion

To our knowledge, this study is the largest population cohort to compare the thigh worn accelerometer algorithm that estimates time in bed with a sleep diary. The findings suggest that the algorithm estimates acceptable diary time in bed on a group level. However, about 36.9% of the measurements and about half of the participants were in the range of ±30 min difference of a clinically relevant limit at individual level. As expected, the correlations with diary estimated sleep time data were lower concluding that the algorithm estimates longer sleep time on a group level compared to diary.

Average absolute differences in time in bed between the two methods were generally small, e.g. the algorithm-derived time in bed was 9.8 minutes less than diary time in bed which was within the clinically acceptable range of ≤ 30 min difference. Although Winkler et al found a good correlation (Pearson correlation coefficient = 0.67) between the algorithm and sleep diary waking times, the algorithm overestimated waking wear time relative to the diary thus resulting in underestimation of diary time in bed (Winkler et al., 2016). van der Berg et al. (2018) developed another algorithm for the assessment of time in bed from activPAL data which was based on the number and duration of sedentary periods to identify time in bed, and on the number and duration of active periods (standing or stepping) to identify wake times. They showed that the algorithm estimates of time in bed differed on average by less than 25 minutes compared to the self-reported bedtimes. Their algorithm was strongly associated with self-reported wake and time in bed (Intraclass correlation coefficient = 0.79). We also found an acceptable agreement on estimation of algorithm-derived time in bed and diary time in bed on a group level although the LoAs were relatively wide.

The main potential use of a future activPAL algorithm will be to evaluate associations between sleep duration and health outcomes. It is therefore important to evaluate the capacity of the algorithm to produce consistent results. For instance, Biddle et al. (2018) examined the association between physical behaviours (sleep, sitting, standing, and stepping) and markers of metabolic health including fasting glucose and insulin, 2-h glucose and insulin. Sleep time was estimated with both the Winkler et al.'s (2016) algorithm and sleep diary. It is encouraging that Biddle et al. (2018) found that the results were materially the same when associations of self-reported sleep time and cardiometabolic outcomes were compared to those of algorithm-derived time in bed. In our recent BCS70 analysis on the associations between different sleep indicators and a range of cardiometabolic outcomes, we found that there was no material difference between the algorithm derived time in bed and diary time in bed or sleep time (Huang B, et al., 2020). We found that algorithm-derived time in bed was higher than diary sleep time by approximately 1 h in both men and women. Further development in algorithms to estimate sleep duration from the thigh-worn accelerometer data is needed. Studies that determine the time in bed and sleep differences between diary and wrist actigraphy produce different results. For instance, while some studies showed that diary overestimated the total sleep time compared to actigraphy (Campanini et al., 2017), other studies showed an underestimation (Liu et al., 2019). Yet, the conclusion from these studies were that the levels of disagreement are reasonable for the devices to be used interchangeably (Campanini et al., 2017; Liu et al., 2019).

We showed that 23.5% of the participants were in the clinically acceptable range (±30 min difference) for sleep time. Unlike our study, the wrist-worn accelerometer study showed that 88% of the participants were in the clinically satisfactory ranges for total sleep time (de Zambotti et al., 2019). The reason for these different findings can be attributed to properties of the devices. The thigh-worn accelerometer can detect sitting/lying time, upright time, sitting/lying to upright transitions and reduction in sitting as well as distinguish standing from stepping (Edwardson et al., 2017). However, the wrist-worn accelerometer measures cannot use postural information when estimating sleep or waking state, habitual physical activity and energy expenditure (Doherty et al., 2017).

Although wrist actigraphy has excellent concordance with the PSG in the measurement of sleep time in healthy people (Martin & Hakim, 2011), wrist actigraphy is prone to overestimating sleep time in different health conditions compared to PSG (Blackwell at al., 2011) and is prone to underestimating sleep time compared to diary (Moore et al., 2015). We also found that the health status of participants influenced the comparisons of time in bed and sleep time assessed by the thigh-worn accelerometer and diary. We showed that the differences between algorithm and diary for both time in bed and sleep time were higher in participants with a long-standing health condition compared to those who did not have a long-standing health condition. Theoretically at least, this observation opens up the possibility that either healthy people report sleep time more accurately or the algorithm works better for healthy people. This finding requires further attention in future research.

Our study has several notable strengths, including the large sample size and the population-based sample that increases generalisability of our findings. Another strength was that 63.5% of participants provided at least 6 valid days of sleep diary and the accelerometer data. Also, it is a strength that our participants were asked for bedtime and wake-up time rather than for the number of hours slept. The latter would entail a calculation by the respondents and thus an increased risk of reporting error of sleeping times.

Our study has limitations also. The sleep diary as a method to measure time in bed and sleep time may be subject to recall limitations or incompleteness. Because of the way participants were instructed to complete the diary (on the following day), we expect that recall bias is less pronounced than recall questionnaires utilising a specific time frame or inquire about "usual" sleep duration. There is a need for studies that compare the algorithm we used with the gold standard polysomnography. In addition, because we only had the accelerometer data for Winkler et al. (2016)'s algorithm, we could not make any comparison with other algorithms. We were not able to evaluate the validity on daily total time. Another limitation is that we had algorithm data for duration of sleep and time in bed, not for the time participants go to bed and the time they wake up.

CONCLUSION

In summary, the algorithm we tested showed acceptable estimates of time in bed compared to diary at the group level. As such, the algorithm is appropriate for use in large-scale population studies to estimate time in bed at a group level. However, despite the limited bias between algorithm and diary, the broad 95% limits of agreement suggest that there may still be disparities between these measurement modalities at individual participant levels in estimating time in bed and sleep time. This was especially true for sleep time; as the average value for sleep time increased, there appeared to be less agreement between the measurements. The limited research to date, suggests that such disagreement has limited impact on estimates of association between sleep and health outcomes. With the increasing use of thigh worn monitors in the field of physical activity, sedentary behaviour and sleep, automated estimation of sleep behaviour parameters has several practical advantages and can maximise use of the accelerometer data. More research is required to further refine the different algorithms that estimate sleep duration from the thigh-worn accelerometer data.

References

Arora, T., Broglia, E., Pushpakumar, D., Lodhi, T., & Taheri, S. (2013). An investigation into the strength of the association and agreement levels between subjective and objective sleep duration in adolescents. *PloS one*, *8*(8).

Arora, V. S., Karanikolos, M., Clair, A., Reeves, A., Stuckler, D., & McKee, M. (2015). Data resource profile: the European Union statistics on income and living conditions (EU-SILC). *International journal of epidemiology*, *44*(2), 451-461.

Barbaresko, J., Rienks, J., & Nöthlings, U. (2018). Lifestyle indices and cardiovascular disease risk: a meta-analysis. *American journal of preventive medicine*, *55*(4), 555-564.

Biddle, G. J., Edwardson, C. L., Henson, J., Davies, M. J., Khunti, K., Rowlands, A. V., & Yates, T. (2018). Associations of physical behaviours and behavioural reallocations with markers of metabolic health: a compositional data analysis. *International journal of environmental research and public health*, *15*(10), 2280.

Blackwell, T., Ancoli-Israel, S., Redline, S., Stone, K. L., & Osteoporotic Fractures in Men (MrOS) Study Group. (2011). Factors that may influence the classification of sleep-wake by wrist actigraphy: the MrOS Sleep Study. *Journal of Clinical Sleep Medicine*, *7*(4).

Campanini, M. Z., Lopez-Garcia, E., Rodríguez-Artalejo, F., González, A. D., Andrade, S. M., & Mesas, A. E. (2017). Agreement between sleep diary and actigraphy in a highly educated Brazilian population. *Sleep medicine*, *35*, 27-34.

Dahlgren, G., Carlsson, D., Moorhead, A., Häger-Ross, C., & McDonough, S. M. (2010). Test–retest reliability of step counts with the ActivPAL™ device in common daily activities. *Gait & posture*, *32*(3), 386-390.

de Zambotti, M., Baker, F. C., & Colrain, I. M. (2015). Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep, 38*(9), 1461-1468.

de Zambotti, M., Rosas, L., Colrain, I. M., & Baker, F. C. (2019). The sleep of the ring: comparison of the ŌURA sleep tracker against polysomnography. *Behavioral sleep medicine, 17*(2), 124-136.

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., ... & Preece, S. J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. *PloS one, 12*(2), e0169649.

Edwardson, C. L., Winkler, E. A., Bodicoat, D. H., Yates, T., Davies, M. J., Dunstan, D. W., & Healy, G. N. (2017). Considerations when using the activPAL monitor in field-based research with adult populations. *Journal of Sport and Health Science, 6*(2), 162-178.

Elliott, J., & Shepherd, P. (2006). Cohort profile: 1970 British birth cohort (BCS70). *International journal of epidemiology*, *35*(4), 836-843.

Ezeugwu, V. E., & Manns, P. J. (2017). Sleep duration, sedentary behavior, physical activity, and quality of life after inpatient stroke rehabilitation. *Journal of Stroke and Cerebrovascular Diseases*, *26*(9), 2004-2012.

Girschik, J., Fritschi, L., Heyworth, J., & Waters, F. (2012). Validation of self-reported sleep against actigraphy. *Journal of epidemiology, 22*(5), 462-468.

Grandner, M. A. (2017). Sleep, health, and society. *Sleep medicine clinics*, *12*(1), 1-22.

Hoevenaar-Blom, M. P., Spijkerman, A. M., Kromhout, D., & Verschuren, W. M. (2014). Sufficient sleep duration contributes to lower cardiovascular disease risk in addition to four traditional lifestyle factors: the MORGEN study. *European journal of preventive cardiology, 21*(11), 1367-1375.

Huang, B. H., Inan-Eroglu, E., Hamer, M., & Stamatakis, E. (2020). Joint associations of device-measured physical activity and sleep duration with cardiometabolic health in the 1970 British Cohort Study. *Journal of Science and Medicine in Sport*. [DOI: 10.1016/j.jsams.2020.07.012]

Liu, J., Wong, W. T., Zwetsloot, I. M., Hsu, Y. C., & Tsui, K. L. (2019). Preliminary agreement on tracking sleep between a wrist-worn device fitbit alta and consensus sleep diary. *Telemedicine and e-Health, 25*(12), 1189-1197.

Lyden, K., Keadle, S. K., Staudenmayer, J., & Freedson, P. S. (2017). The activPAL™ accurately classifies activity intensity categories in healthy adults. *Medicine and science in sports and exercise*, *49*(5), 1022.

Martin, J. L., & Hakim, A. D. (2011). Wrist actigraphy. *Chest*, *139*(6), 1514-1527.

McCrae, C. S., Rowe, M. A., Tierney, C. G., Dautovich, N. D., DeFinis, A. L., & McNamara, J. P. (2005). Sleep complaints, subjective and objective sleep patterns, health, psychological adjustment, and daytime functioning in community-dwelling older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*(4), P182-P189.

Moore, C. M., Schmiege, S. J., & Matthews, E. E. (2015). Actigraphy and sleep diary measurements in breast cancer survivors: discrepancy in selected sleep parameters. *Behavioral Sleep Medicine, 13*(6), 472-490.

Oliver, M., Badland, H. M., Shepherd, J., & Schofield, G. M. (2011). Counting steps in research: a comparison of accelerometry and pedometry. *Open Journal of Preventive Medicine, 1*(01), 1.

Riemann, D. (2012). Insomnia is coming of age. *Sleep*, *35*, 175

Short, M. A., Gradisar, M., Lack, L. C., Wright, H., & Carskadon, M. A. (2012). The discrepancy between actigraphic and sleep diary measures of sleep in adolescents. *Sleep medicine*, *13*(4), 378-384.

Tonetti, L., Mingozzi, R., & Natale, V. (2016). Comparison between paper and electronic sleep diary. *Biological Rhythm Research*, *47*(5), 743-753.

University of London IoE, Centre for Longitudinal Studies. 1970 British Cohort Study: Forty-Six- Year Follow-Up, 2016-2018 (http://doi.org/10.5255/UKDA-SN-8547-1): UK Data Service; 2019.

Van De Water, A. T., Holmes, A., & Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review. *Journal of sleep research*, *20*(1pt2), 183-200.

van der Berg, J. D., Willems, P. J., van der Velde, J. H., Savelberg, H. H., Schaper, N. C., Schram, M. T., ... & 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 der Velde, J. H., Schaper, N. C., Stehouwer, C. D., van der Kallen, C. J., Sep, S. J., Schram, M. T., ... & Savelberg, H. H. (2018). Which is more important for cardiometabolic health: sedentary time, higher intensity physical activity or cardiorespiratory fitness? The Maastricht Study. *Diabetologia*, *61*(12), 2561-2569.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics, 15*(6), 657-680.

Winkler, E. A., Bodicoat, D. H., Healy, G. N., Bakrania, K., Yates, T., Owen, N., ... & Edwardson, C. L. (2016). Identifying adults' valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol. *Physiological Measurement*, *37*(10), 1653.

Xiao, Q., Keadle, S. K., Hollenbeck, A. R., & Matthews, C. E. (2014). Sleep duration and total and cause-specific mortality in a large US cohort: interrelationships with physical activity, sedentary behavior, and body mass index. *American journal of epidemiology*, *180*(10), 997-1006.

.

Figure 1. Flow chart for the steps of algorithm

Figure 2(A-C). The repeated measures of Bland-Altman agreement between algorithm and diary for the duration of time in bed. Mean of the differences between algorithm and diary time in bed and lower and upper agreement limits (mean difference ± 1.96 standard deviation) are displayed for each Bland-Altman plot. The green lines represent the upper and lower a priori-set clinically satisfactory limits (± 30 min from the zero line).

Figure 3(A-C). The repeated measures of Bland-Altman agreement between algorithm and diary for the duration of time in bed and sleep time. Mean of the differences between algorithm time in bed and diary sleep time and lower and upper agreement limits (mean difference ± 1.96 standard deviation) are displayed for each Bland-Altman plot. The green lines represent the upper and lower a priori-set clinically satisfactory limits (± 30 min from the zero line).

Table 1. Characteristics of Participants

	Total* (n=5498)				Women (n=2879)				Men (n=2619)			
	%	95% CI	Mean	95% CI	%	95% CI	Mean	95% CI	%	95% CI	Mean	95% CI
Employed	84.0	83.0, 85.0			79.2	77.7, 80.6			89.3	88.1, 90.6		
Self-employed	2.4	2.0, 2.8			2.6	2.1, 3.3			2.1	1.6, 2.7		
Shift-workers	15.3	14.2, 16.3			13.9	12.6, 15.3			16.7	15.2, 18.2		
BMI[a]			28.2	28.02, 28.30			27.9	27.67, 28.09			28.5	28.29, 28.66
Current Smoker	13.3	12.3, 14.2			12.9	11.7, 14.0			13.7	12.4, 15.0		
General health score			68.9	68.6, 69.9			69.6	68.94, 70.72			68.2	67.81, 69.42
No long-standing health condition	84.4	83.4, 85.4			82.1	80.7, 83.5			87.0	85.6, 88.3		
Total time to fall asleep (0-15 min)	51.5	50.3, 52.9			47.6	45.6, 49.5			55.8	53.7, 57.6		
Getting enough sleep (most of the time)	33.5	32.3, 34.8			32.5	30.9, 34.1			34.6	32.8, 36.6		
Times out of bed during night			0.7	0.71, 0.76			0.8	0.70, 0.76			0.7	0.71, 0.77
Self-rated sleep quality (1 to 10)			6.9	6.91, 7.00			6.9	6.90, 7.01			6.9	6.88, 7.01
Sleep hours over the last 4 weeks			6.8	6.78, 6.84			6.9	6.86, 6.96			6.7	6.67, 6.76

Number of valid days for sleep time (diary)	6.5	6.47, 6.54	6.5	6.47, 6.56	6.5	6.45, 6.54
Number of valid days for time in bed (diary)	6.5	6.51, 6.58	6.6	6.52, 6.61	6.5	6.48, 6.57
Number of valid days (Accelerometer)	6.2	6.12, 6.20	6.2	6.15, 6.25	6.1	6.05, 6.18
Accelerometer wear time, h/day[b]	15.8	15.80, 15.88	15.7	15.65, 15.75	16.0	15.94, 16.05

Abbreviations: CI, confidence interval.

* Age is identical for the whole sample.

[a] Weight (kg)/height (m)^2.

[b] Average accelerometer wear time per day, where non-wear was defined by intervals of at least 60 minutes of zero activity counts, with allowance for up to 2 consecutive minutes of 1–100 counts/minute.

Table 2. Absolute Differences Between Mean Amounts of Time in Bed From The Sleep Diary and Accelerometer Data, by Sex

	Time in Bed (min/day)				Sleep Time (min/day)		Difference[a] (min/day)		Difference[b] (min/day)		P Value[c]	P Value[d]
	Algorithm		Diary		Diary							
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI		
Women (n=2879)	497.8	494.8, 500.2	513.1	510.9, 515.2	441.6	439.5, 443.7	-15.5	-18.1, -12.9	56.0	53.0, 58.9	<0.001	<0.001
Men (n=2619)	480.0	477.1, 483.2	483.6	481.1, 485.8	422.8	420.4, 425.0	-3.5	-6.3, -0.7	57.4	54.3, 60.4	0.011	<0.001
Total (n=5498)	489.4	487.3, 491.5	499.0	497.4, 500.8	432.6	431.0, 434.2	-9.8	-11.6, -8.0	56.6	54.6, 58.6	<0.001	<0.001

Abbreviations: CI, confidence interval.

[a] Difference between time in bed (algorithm) and time in bed (diary)

[b] Difference between time in bed (algorithm) and sleep time (diary)

[c] *P* value for the difference between algorithm-derived data (time in bed) and diary (time in bed), according to the Paired Sample t-test

[d] *P* value for the difference between algorithm-derived data (time in bed) and diary (sleep time), according to the Paired Sample t-test

Table 3. Correlation Between Algorithm and Diary for Time in Bed and Sleep Time, by Sex

Algorithm,	Women (n=2812)		Men (n=2544)		P value [a]
Diary	ρ [b]	95% CI [c]	ρ [b]	95% CI [c]	
Time in Bed	0.48*	0.45, 0.52	0.51*	0.47, 0.55	0.207
Sleep Time	0.34*	0.30, 0.38	0.39*	0.35, 0.43	0.023

Abbreviations: CI, confidence interval

* P <0.001

[a] *P* value for the difference between Pearson's ρ for women and men, calculated using Fisher's z test.

[b] Pearson's correlation coefficient.

[c] Confidence intervals were computed using a bootstrapping procedure.

Table 4. Linear Regression Coefficients Between the Mean of the Algorithm and Diary Derived Time, and the Difference Between the Algorithm and Diary Derived Time

	Coefficient	95% CI	*P*
Women			
Time in Bed (Algorithm and Diary)	0.341	0.442, 0.481	<0.001
Time in Bed (Algorithm) and Sleep Time (Diary)	0.357	0.517, 0.560	<0.001
Men			
Time in Bed (Algorithm and Diary)	0.294	0.371, 0.412	<0.001
Time in Bed (Algorithm) and Sleep Time (Diary)	0.327	0.454, 0.499	<0.001
Total			
Time in Bed (Algorithm and Diary)	0.310	0.398, 0.426	<0.001
Time in Bed (Algorithm) and Sleep Time (Diary)	0.341	0.487, 0.518	<0.001