Can an automated sleep detection algorithm for waist worn accelerometry replace sleep logs?

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Applied Physiology, Nutrition, and Metabolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>apnm-2017-0860.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>17-Mar-2018</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Barreira, Tiago; Syracuse University, Exercise Science
                                          Redmond, Jessica; Utica College
                                          Brutsaert, Tom; Syracuse University, Exercise Science
                                          Schuna, John; Oregon State University, School of Biological and Population Health Sciences
                                          Mire, Emily; Pennington Biomedical Research Center, Population Science
                                          Katzmarzyk, Peter; Pennington Biomedical Research Center
                                          Tudor-Locke, Catrine; University of Massachusetts |
| Keyword:                  | accelerometer, objectively measured, nocturnal |
| Is the invited manuscript for consideration in a Special Issue? | N/A                                           |
Title

Can an automated sleep detection algorithm for waist worn accelerometry replace sleep logs?

Tiago V. Barreira, Jessica G. Redmond, Tom D. Brutsaert, John M. Schuna Jr., Emily F. Mire, Peter T. Katzmarzyk, Catrine Tudor-Locke

aSchool of Education, Syracuse University, 820 Comstock Ave, Syracuse, New York, 13244, USA

bBiology Department, Utica College, 1600 Burrstone Rd, Utica, New York, 13505, USA

cCollege of Public Health and Human Sciences, Oregon State University, 1500 SW Jefferson St., Corvallis, Oregon, 97331, USA

dPennington Biomedical Research Center, 6400 Perkins Rd, Baton Rouge, Louisiana, 70808, USA

eDepartment of Kinesiology, University of Massachusetts Amherst, 30 Eastman Ln, Amherst, Massachusetts, 01002, USA

Corresponding Author:

Tiago Barreira, PhD,
Assistant Professor,
Exercise Science Department,
Syracuse University,
820 Comstock Ave,
Syracuse, NY 13210
Phone (315) 443-5588
Email tvbarrei@syr.edu
Abstract

The purpose of this study was to test whether estimates of bedtime, wake time, and sleep period time (SPT) were comparable between an automated algorithm (ALG) applied to waist-worn accelerometry data and a sleep log (LOG), in an adult sample. A total of 104 participants were asked to log evening bedtime and morning wake time and wear an ActiGraph wGT3X-BT accelerometer at their waist 24 h/days for 7 consecutive days. Mean difference (MD) and mean absolute difference (MAD) were computed. Pearson correlations and dependent sample t-tests were used to compare LOG-based and ALG-based sleep variables. Effect sizes were calculated for variables with significant mean differences. A total of 84 participants provided 2+ days of valid accelerometer and LOG data for a total of 368 days. There was no mean difference (p=.47) between LOG 472±59 min and ALG SPT 475±66 min (MAD=31±23 min, r=.81). There was no significant mean difference between bedtime (11:48 pm and 11:53 pm for LOG and ALG, respectively, p=.14, MAD=25±21 min, r=.92). However, there was a significant mean difference between LOG (7:41 am) and ALG (7:49 am) wake times (p=.01, d=0.11, MAD=24±21 min, r=.92). The LOG and ALG data were highly correlated and relatively small differences were present. The significant mean difference in wake time might not be practically meaningful (Cohen’s d=0.11) making the ALG useful for sample estimates. MAD, which gives a better estimate of the expected differences at the individual level, also demonstrated good evidence supporting ALG individual estimates.

Key words accelerometer; objectively measured; nocturnal;
Introduction

Accelerometers were used to estimate sleep (Kripke et al. 1978; Sadeh and Acebo 2002; Webster et al. 1982) prior to their introduction for physical activity (PA) measurement (Freedson et al. 1998; Janz 1994; Melanson and Freedson 1995); however, accelerometry-based sleep and physical activity measurement have historically been performed in independently of each another. Until recently, research protocols typically required participants to wear the devices exclusively during the nighttime to estimate sleep or during wake time to estimate PA. However, this recently changed as the National Health and Nutrition Examination Survey (NHANES) adopted a 24 h/day accelerometer monitoring protocol (Troiano et al. 2014). In response, other researchers have also moved to 24 h/day accelerometer-based data collection. This shift was prompted for 2 primary reasons: 1) the belief that participants would wear the devices for more hours during the day (Troiano et al. 2014), and 2) to study the influence of both sleep and PA on health parameters. Therefore, we chose to use a 24 h/day protocol with an accelerometer attached at the waist for the International Study of Childhood Obesity Lifestyle and the Environment (ISCOLE) (Chaput et al. 2015; Katzmarzyk et al. 2013; Katzmarzyk et al. 2015). The waist attachment was chosen because PA was our focus and at the time there was no consensus on the methodology to identify PA intensity using wrist-worn devices. The 24 h/day waist protocol we employed was helpful in increasing wear time compliance (Tudor-Locke et al. 2015). However, there was no methodology available to automatically identify sleep from a waist-worn accelerometer when worn 24 h/day, so we created an automated algorithm (ALG) to identify sleep with waist-worn accelerometry and tested its accuracy (Barreira...
et al. 2015; Tudor-Locke et al. 2014). The correct identification of sleep is important for the further identification and separation of sedentary time and non-wear because those can have similar acceleration signals. While we developed and tested the ALG accuracy with children aged 9-11 years (Barreira et al. 2015), other researchers (McVeigh et al. 2016) have tested the accuracy of our initial ALG (Tudor-Locke et al. 2014) but to our knowledge, the final version of the ALG (Barreira et al. 2015) has not been tested in other samples. Thus, the purpose of this study was to test whether estimates of bedtime, wake time, and sleep period time (SPT) were comparable between an automated algorithm (ALG) applied to waist-worn accelerometry data and a sleep log (LOG), in an adult sample.

**Material and Methods**

As part of a larger study, 104 participants were invited to take part in this additional data collection. Written informed consent was obtained from each participant and the Syracuse University Institutional Review Board approved the study protocol and procedures prior to any recruitment.

Each participant’s height and body mass were measured using a portable stadiometer (Seca 213; Seca GmbH & Co. KG., Hamburg, Germany) and a digital scale (SC-240; Tanita Corporation, Tokyo, Japan), respectively. Body mass index (BMI) was calculated as kg/m$^2$. Participant age was self-reported.

Accelerometer data were collected using the ActiGraph wGT3X-BT (ActiGraph, LLC, Pensacola, FL), which is a triaxial accelerometer capable of collecting raw acceleration data at sampling frequencies up to 100 Hz over a dynamic range of ±8g. To initialize and download data from the wGT3X-BT ActiLife software (ActiGraph LLC) was used. It
is possible to output a variety of data from ActiLife software, including raw acceleration and activity counts for each axis, vector magnitude from all three axes, steps, inclinometer position ("off," "lying," "sitting," or "standing"), and lux for different epochs. For this study, data were collected at 80 Hz and processed in 60 sec epochs using the low-frequency extension filter to increase sensitivity of detecting low-magnitude accelerations, as suggested by Hjorth et al. (2012). Identical to our previous work (Barreira et al. 2015; Tudor-Locke et al. 2014), participants were asked to wear an ActiGraph wGT3X-BT accelerometer at their waist 24 h/days for 7 consecutive days, unless coming into contact with water (swimming, bathing, etc.). For the LOG, participants were asked to write down the time they went to bed (bedtime) each night and the time they got out of bed (wake time) each morning for the same days of accelerometer monitoring. Participants were also asked to record details about any night wakening episodes and/or accelerometer removal. LOG was used to obtain self-reported bedtime and wake time. LOG sleep period time (SPT) was calculated as the elapsed time between self-reported bedtime and wake time.

We implemented our previously published (Barreira et al. 2015) and freely available algorithm (SAS syntax available at http://www.pbrc.edu/pdf/PBRCSleepEpisodeTimeMacroCode.pdf) using SAS 9.4 without any modifications. Briefly, the ALG begins by using the Sadeh algorithm (Sadeh et al. 1994) to classify each minute as wake or sleep, followed by changing any wake minute to sleep if the accelerometer was in the "off" position. Once each minute was classified, a set of rules were applied to identify bedtime and wake time. SPT was
calculated as the elapsed between bedtime and wake time. Although multiple SPTs were possible, no instance of multiples were included in the final analysis. Lastly, the SPT was examined for non-wear and if more than 90% of the minutes included in the SPT were classified as non-wear, the whole SPT was considered invalid and reclassified as non-wear.

We analyzed the data two ways: 1) complete data and 2) screened data. For the complete data we used all 466 days averaged between 95 participants. For the screened data, we paired all the valid LOG and ALG nights (466) and flagged every difference of 150+ minutes for bedtime, wake time, and SPT for closer examination. In total, 62 bed times, 32 wake times, and 83 SPT were examined. The examination included visual inspection of the 60 sec epoch accelerometer data and the LOG information. In most instances, a large SPT difference was indicative of a large difference in bed or wake time. If it was clear that the participant had made a mistake in LOG data entry all variables associated with that night were excluded from the analyses. For example, as shown in Figure 1, the ALG estimated wake time at 4:07 am while the participant log entry had a wake time at 7:03 am. Looking at steps/min, it is clear that the participant got out of bed close to 4:10 am and even went for a 30 min run starting at 4:52 am (additional examples provided supplementary figures S1 and S2 figures). In total 98 nights were excluded from the analysis, 77 were deleted due to dissimilar SPT, 7 due to different bed times, 3 due to different wake times, 1 due to different bed and wake times, and an additional 10 nights were excluded because of inconsistencies in most entries by the participant or just only a single valid day remained for a participant once all exclusions were made. If the difference was not due to a clear
mistake in log entry, data were retained (see Figure 2 for example). In this case, it appears again that the ALG has a better estimate of bedtime but it is possible that the participant went to bed around the indicated time in the LOG, had some restless sleep and briefly arose (possibly to go to the bathroom). In addition, participants' data were only included if they had 2+ days of valid data. In total 368 days were included. Participants' data were averaged across valid days, providing just one value per participant for each of the variables described above. In total, 85 participants had valid data. Similar to our previous publications (Barreira et al. 2015; Tudor-Locke et al. 2014), a number of statistical procedures were used to compare the multiple outputs by the 2 measurement tools. Pearson product–moment correlations were calculated to assess the magnitude of associations, and previously published standards (Safrit and Wood 1995) were used to classify the associations. MD and MAD were calculated to determine differences between methods. Paired t-tests were used to compare mean LOG SPT and mean ALG SPT. Cohen's d effect size was computed when significant mean differences were present. In addition, Bland–Altman plots were prepared. The same calculations were conducted to compare LOG bedtime and wake time to ALG variables. All statistical analyses were conducted using IBM SPSS Statistics (version 24; SPSS, Chicago, IL), and the level of significance was defined as P <.05.

Results

In total for the screened data, 84 participants (75% females, BMI 23.6±3.8 kg/m², age 23.8±6.0 years), provided 2+ days of valid accelerometer and LOG data for a total of 368 days. The results from both the screened and completed data are presented in Tables 1 and 2. For the screened data, there was no mean difference (t(83)=0.73,
p=.47) between LOG SPT min and ALG SPT. The Bland–Altman plot is presented in Figure 3. Only three data points were outside the 95% limits of agreement (±1.96 SD). There was no significant mean difference for bedtime. However, there was a small (Cohen’s d=0.11) but significant mean difference in wake time (t(83)=2.51, p=.01). For the complete data set, there were no significant mean differences (all p>.18) in any of the comparisons. However, absolute differences were 2-3 times as large as the absolute differences in the screened data, 25±21 vs 63±55, 24±21 vs 45±33, and 31±23 vs 85±63 for bed time, wake time, and SPT from screened and complete data respectively.

Discussion

We compared estimates of bedtime, wake time, and SPT between a LOG and the automated ALG and believe that waist worn accelerometry is a good alternative to LOG for estimation of those variables. LOG and ALG data were highly correlated and relatively small differences were present. Although a significant between-method mean difference of 8 min was found for wake time, this difference might not be meaningful in practice due to its small effect size. MAD, which gives a better estimate of the expected differences at the individual level, provided good validity evidence for the ALG use of individuals’ estimates. It is also important to note that during visual inspection of the data that showed large discrepancies between LOG and ALG, 20% of the LOG data were deemed erroneous. We presented both screened and complete data results but we believe that this clearly demonstrates one of the LOG limitations.

When compared to the results in our previous study with children (Barreira et al. 2015), the results in this study were positive. Overall, the correlations were higher and
differences were smaller. We showed a smaller MD (3 vs 20 min) but a similar MAD (31 vs 34 min) for SPT. The estimated LOG and ALG bedtimes were over 2 h later in this sample, and the wake time was 1 h later leading to a 1 h difference in SPT between the 2 samples. This adds to the validity of the ALG because young adults are known to sleep for shorter durations than children (Hayley et al. 2015).

Due to the limited research on the implementation of automated algorithms to detect sleep related estimates, comparison to other studies is difficult. To our knowledge, McVeigh et al. (2016) is the only study that has evaluated the accuracy of automated sleep algorithms in adults. However, in their study (McVeigh et al. 2016) the comparisons were made against visual inspection of the data and researcher determined in bed and out-of-bed times and not a comparison to a LOG like in this study. Due to the different data analysis between the studies, direct comparison is difficult; however in this study, we showed tighter limits of agreement than McVeigh et al. (2016) For example, SPT 95% limits of agreement in our study was -78-72 min while in-bed wear for McVeigh et al. was -309-228 min. This is likely caused by the different methods used to test the accuracy of the algorithm, visual inspection vs LOG.

As previously mentioned, there is a general belief that 24 h/day protocols can improve wear compliance, however there is a lack of data to either support or refute this notion. In our previous study with children, the 24 h/day waist accelerometry protocol was helpful in increasing wear time compliance (Tudor-Locke et al. 2015). Although it was not the purpose of this study, we did compute wake wear time. Once SPT was extracted and non-wear was estimated (Troiano et al. 2008), wake wear was estimated at 15.5±1.3 h/day. That was 1.3 h/day higher than the US representative sample from...
NHANES (Troiano et al. 2008). This suggests that the 24 h/day wear protocol has the potential to improve wear time.

Although carefully conducted, this study had some limitations. Approximately 20% of the LOG-ALG pairing days were not included in the screened analysis because their estimates differed by more than 150 min. However, the pairings were only excluded when it was clear that the LOG information was not plausible based on careful examination of the accelerometer data. Additional contact with participants to check LOG information could have been helpful in improving the quality of the LOGs and overall retention rate of data. The accuracy of the ALG was tested against a LOG and not polysomnography which is the gold standard for sleep assessment. However, the intention of the ALG is to replace the LOG and make it easier for researchers conducting large scale/epidemiological investigations where both sleep and physical activity are behaviors of interest. The ALG is not intended to replace clinical measures of sleep duration and sleep quality.

In conclusion, we found the ALG to be a valid substitute for a sleep LOG. Using a waist-worn accelerometer to estimate bedtime, wake time, and sleep period time has the potential to increase wear compliance, can lead to high wake wear time, and decrease the burden to the researcher and participant by eliminating the use of the sleep LOG.
Conflict of Interest

The authors have no conflicts of interest to report.

References


Katzmarzyk, P.T., Barreira, T.V., Broyles, S.T., Champagne, C.M., Chaput, J.-P.,

Fogelholm, M. et al. 2013. The International Study of Childhood Obesity, Lifestyle and
Katzmarzyk, P.T., Barreira, T.V., Broyles, S.T., Champagne, C.M., Chaput, J.P.,

Fogelholm, M. et al. 2015. Physical activity, sedentary time, and obesity in an


McVeigh, J.A., Winkler, E.A., Healy, G.N., Slater, J., Eastwood, P.R., and 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. Physiol.

Melanson, E.L., Jr., and Freedson, P.S. 1995. Validity of the Computer Science and

Rev. 6(2): 113-124.


and exercise science. St. Louis: Mosby.


Table 1. Bedtime, wake time, and sleep period time for screened and complete data set.

<table>
<thead>
<tr>
<th></th>
<th>Bedtime</th>
<th>Wake Time</th>
<th>Sleep Period Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screened Data</td>
<td>Algorithm Log</td>
<td>11:53 PM 7:49 AM</td>
<td>475±66</td>
</tr>
<tr>
<td>Complete Data</td>
<td>Algorithm Log</td>
<td>11:48 PM 7:56 AM</td>
<td>488±79</td>
</tr>
</tbody>
</table>

Table 2. Analysis results for the screened and complete data sets.

<table>
<thead>
<tr>
<th>Sleep Variable</th>
<th>Difference (min)</th>
<th>Correlation</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screened Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedtime</td>
<td>5±33</td>
<td>0.92</td>
<td>1.50</td>
<td>0.15</td>
</tr>
<tr>
<td>Wake Time</td>
<td>8±30</td>
<td>0.92</td>
<td>2.51</td>
<td>0.01</td>
</tr>
<tr>
<td>Sleep Period Time</td>
<td>3±38</td>
<td>0.81</td>
<td>0.73</td>
<td>0.47</td>
</tr>
<tr>
<td>Complete Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedtime</td>
<td>0±66</td>
<td>0.66</td>
<td>0.01</td>
<td>0.92</td>
</tr>
<tr>
<td>Wake Time</td>
<td>6±41</td>
<td>0.86</td>
<td>1.40</td>
<td>0.18</td>
</tr>
<tr>
<td>Sleep Period Time</td>
<td>6±82</td>
<td>0.34</td>
<td>0.75</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1 Example of a participant's self-recorded and accelerometer determined sleep related behaviors that was excluded from the analysis.

Figure 2 Example of a participant's self-recorded and accelerometer determined sleep related behaviors that was retained in the analysis.

Figure 3 Bland-Altman plot of Log and Algorithm sleep estimates.
Figure 1 Example of a participant’s self-recorded and accelerometer determined sleep related behaviors that was excluded from the analysis.

254x190mm (300 x 300 DPI)
Figure 2 Example of a participant’s self-recorded and accelerometer determined sleep related behaviors that was retained in the analysis.

254x190mm (300 x 300 DPI)
Figure 3 Bland-Altman plot of Log and Algorithm sleep estimates.

254x190mm (300 x 300 DPI)