

Article

Sleep Measurement Using Wrist-Worn Accelerometer Data Compared with Polysomnography

John D. Chase ¹, Michael A. Busa ², John W. Staudenmayer ³ and John R. Sirard ^{1,*}

¹ Department of Kinesiology, University of Massachusetts Amherst, Amherst, MA 01003, USA; jchase@umass.edu

² Institute for Applied Life Sciences, University of Massachusetts Amherst, Amherst, MA 01003, USA; mbusa@umass.edu

³ Department of Mathematics & Statistics, University of Massachusetts Amherst, Amherst, MA 01003, USA; jstauden@math.umass.edu

* Correspondence: jsirard@kin.umass.edu; Tel.: +1-413-545-7898

Abstract: This study determined if using alternative sleep onset (SO) definitions impacted accelerometer-derived sleep estimates compared with polysomnography (PSG). Nineteen participants (48%F) completed a 48 h visit in a home simulation laboratory. Sleep characteristics were calculated from the second night by PSG and a wrist-worn ActiGraph GT3X+ (AG). Criterion sleep measures included PSG-derived Total Sleep Time (TST), Sleep Onset Latency (SOL), Wake After Sleep Onset (WASO), Sleep Efficiency (SE), and Efficiency Once Asleep (SE_ASLEEP). Analogous variables were derived from temporally aligned AG data using the Cole–Kripke algorithm. For PSG, SO was defined as the first score of ‘sleep’. For AG, SO was defined three ways: 1-, 5-, and 10-consecutive minutes of ‘sleep’. Agreement statistics and linear mixed effects regression models were used to analyze ‘Device’ and ‘Sleep Onset Rule’ main effects and interactions. Sleep–wake agreement and sensitivity for all AG methods were high (89.0–89.5% and 97.2%, respectively); specificity was low (23.6–25.1%). There were no significant interactions or main effects of ‘Sleep Onset Rule’ for any variable. The AG underestimated SOL (19.7 min) and WASO (6.5 min), and overestimated TST (26.2 min), SE (6.5%), and SE_ASLEEP (1.9%). Future research should focus on developing sleep–wake detection algorithms and incorporating biometric signals (e.g., heart rate).

Keywords: accelerometer; polysomnography; algorithm; sleep; Cole–Kripke



Citation: Chase, J.D.; Busa, M.A.; Staudenmayer, J.W.; Sirard, J.R. Sleep Measurement Using Wrist-Worn Accelerometer Data Compared with Polysomnography. *Sensors* **2022**, *22*, 5041. <https://doi.org/10.3390/s22135041>

Academic Editor: Marc Parrilla

Received: 20 May 2022

Accepted: 27 June 2022

Published: 4 July 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sleep regulates all major physiological systems and can have a clinically meaningful impact on metabolic and cognitive health [1–8]. The accurate and reliable measurement of sleep metrics is, therefore, essential due to the mediating effect of sleep on numerous health outcomes [1–8]. Polysomnography (PSG) is the gold standard tool for understanding physiological processes related to sleep. However, PSG is often cost-prohibitive, collected in clinical settings unfamiliar to participants, and requires trained technicians to operate and process data. The cumbersome instrumentation used with PSG can shorten sleep duration and decrease sleep quality, particularly during the first night experiencing PSG [9]. Wrist-worn accelerometers provide an inexpensive alternative for the multi-night assessment of several sleep characteristics. Accelerometers can be deployed in ecological settings and provide a less obtrusive method of sleep measurement with the dual capability of measuring other human behaviors, such as physical activity and sedentary behaviors.

Consumer wearable accelerometers continue to grow in popularity, partially due to increased accessibility, social trends, and technological advancement [10]. Since consumers can rely heavily on wearable devices to provide behavioral information, including sleep characteristics, focusing on metric accuracy will allow for improved public health recommendations. As consumer-grade devices do not provide access to raw data, evaluating

candidate algorithms from research-grade device data is a necessary step in understanding metric performance before applying the algorithms in a wider context.

Previous studies in healthy adults have detailed the tendency for accelerometers to overestimate sleep and underestimate waking, compared with polysomnography [11–14]. These systematic errors often occur when accelerometers poorly distinguish between sleep and sedentary supine waking periods (e.g., lying recumbent while reading or watching television) [12,15–17]. The Sleep Onset Latency (SOL) period is a primary example of a sedentary supine waking period susceptible to high misclassification error. Varying methods exist for defining SOL, and the method is often dictated by either the algorithm [12,18,19] or user-defined epoch scoring settings [20–22]. For instance, the widely used Cole–Kripke algorithm [23] requires 60 s epochs for sleep–wake scoring, while the Oakley algorithm can be applied with 15, 30, 60, or 120 s epochs at low, medium, and high sensitivity count thresholds [20–22]. Alongside algorithm selection, researchers can define how sleep onset is scored. For instance, sleep onset may be defined as the first ‘sleep’ epoch, while other operational definitions of sleep onset require multiple consecutive epochs of ‘sleep’ [19,24]. Accordingly, recent efforts have been made to improve the accuracy of accelerometer classification during the SOL period in patients with Obstructive Sleep Apnea and/or Periodic Limb Movement Disorder [15], as well as children and adolescents [25].

Still, few studies have addressed accelerometer errors due to the misclassification of epochs as ‘Sleep’ or ‘Wake’ during periods of the wake-to-sleep transition, particularly during SOL, despite the widespread use of accelerometers in sleep analyses. While optimal sleep onset scoring lengths have been proposed for adults with Obstructive Sleep Apnea and Periodic Limb Movement in Sleep (5 min) [15], as well as children (3 min) and adolescents (20 min) [25], there is no established sleep onset ruling for healthy adults. Sleep onset periods of 1, 5, and 10 min are commonly available settings. The optimization of sleep–wake detection algorithms, particularly during periods of high misclassification, would enhance the utility of accelerometers to provide a more accurate assessment of avital health behavior. Therefore, the purpose of this study was to determine if applying alternative sleep onset (SO) definitions (1, 5, 10 min of immobility), that have been most commonly used in other populations, would improve the accuracy of the commonly used Cole–Kripke algorithm in the classification of accelerometer-derived sleep measurements in healthy adults, compared with PSG.

2. Materials and Methods

2.1. Participants

Twenty healthy participants (50%F; Age = 24.6 ± 2.7 years; BMI (mean \pm SD = 24.3 ± 3.6 kg/m² [Range = 19.4–31.7 kg/m²])) completed a 48 h visit within a laboratory that simulated an apartment-style home. This was a secondary data analysis from a study designed to validate home-based health monitoring technologies. Participants had no known sleep disorders. Due to the “first-night-effect” (decreased sleep duration and quality during initial PSG night) [9], only the second night was used in the present analysis. PSG data were recorded for twenty participants, but one participant was excluded due to PSG equipment malfunction; therefore, data from 19 participants were included for these analyses. Informed consent was obtained from all participants prior to study participation. The study protocol was approved by the University of Massachusetts Amherst IRB.

2.2. Study Details

2.2.1. Polysomnography

Participants notified the technicians when they were ready to attempt night-time sleep. A standard PSG montage was used, following International 10-20 System guidelines. An Embletta MPR ST+ Proxy (Embla Systems, Natus Neurology, Middleton, WI, USA) was used to sample data at 500 Hz. The MPR was equipped with 16-channel electroencephalography leads (EEG; F3, F4, C3, C4, O1, O2, M1, M2, GRD/REF), electromyography (EMG; 3 sub-mental leads), and electrocardiography (ECG; R/L Arm). Participants also wore

thoracic and abdominal respiratory belts, a nasal cannula, and thermistor, all of which were synced to the MPR unit. All impedances were under 10 k Ω upon completion of PSG setup. PSG and accelerometer data were temporally aligned for epoch-by-epoch comparison and merged by participant and timestamp. All devices were initialized on the same computers to ensure clock synchronization.

2.2.2. Accelerometers

Participants wore an ActiGraph GT3X+ (ActiGraph, Pensacola, FL, USA) tri-axial accelerometer (AG) on their non-dominant wrist. The accelerations were sampled at 80 Hz by a 12-bit analog to digital converter with a dynamic range of $\pm 6G$. Upon study completion, data were downloaded and extracted from the AG non-volatile memory. The accelerations from 80 Hz data were collapsed in 1 s epochs using ActiLife software (v. 6.13.4). The 1 s count data were then collapsed into 60 s epochs using a custom R Script (R Studio Software v. 1.2.5042, RStudio, Boston, MA, USA) as per the requirements for the Cole–Kripke sleep scoring algorithm ('actigraph.sleepr'). The Cole–Kripke algorithm uses a sliding weighted average window to score 60 s epochs as 'Sleep' or 'Wake' based upon x-axis (anteroposterior body position; parallel to forearm) count data [26]. The data and custom R script are available upon request.

2.3. Data Processing

Night-time sleep was recorded on both nights by an American Academy of Sleep Medicine (AASM)-certified technician. Sleep stages were scored in 30 s epochs by an AASM-certified technician according to AASM standard procedures [27]. Sleep technicians noted 'Lights Out' and 'Awake' time. Awake was identified as the last PSG epoch scored as 'Sleep'. Time in Bed (TIB) was identified as the time between 'Lights Out' and 'Awake'.

From the included participants ($n = 19$), portions of PSG data were unable to be scored from $n = 6$ participants (range = 17 to 246 min missing); these missing epochs were excluded from further processing and analyses, in both the PSG and ActiGraph data. The majority of disrupted PSG data instances ($n = 4$) occurred at the end of the night, which simply truncated the data from those participants. When PSG data were interrupted in the middle of the night, the corresponding accelerometer and unscored PSG data were removed. Data from the beginning and end of the night were then temporally synchronized between devices. The first 30 s epoch of every minute of PSG data was taken as the representative value for that minute and time aligned with the corresponding AG minute, to satisfy the Cole–Kripke algorithm requirement (60 s epoch) and eliminate any intra-minute conflicts in the PSG data.

Sleep Onset (SO) was defined by three commonly applied rules based on 1, 5, and 10 consecutive minutes of 'Sleep' with no 'Wake' score (SO_{1-min} [28,29], SO_{5-min} [15], SO_{10-min} [18,30,31], respectively; Figure 1). The following sleep measures were extrapolated from the PSG data using the three SO rules: Total Sleep Time (TST; total minutes scored as 'Sleep' within TIB range), Sleep Onset Latency (SOL; total minutes scored as 'Wake' between 'Lights Out' and 'Sleep Onset'), Sleep Efficiency (SE (%)) = (TST/TIB) \times 100, Efficiency Once Asleep (SE_ASLEEP (%)) = TST/(TIB-SOL) \times 100, and Wake After Sleep Onset (WASO; total minutes scored as 'Wake' between 'Sleep Onset' and 'Awake'). Analogous sleep measures were calculated from AG data.

2.4. Statistical Analysis

Agreement between devices, at the 60 s epoch level, was calculated as the percentage of correctly identified 'Sleep' and 'Wake' epochs by AG, using the PSG scored data as the criteria. Sensitivity was calculated as the percentage of correctly identified 'Sleep' epochs. Specificity was calculated as the percentage of correctly identified 'Wake' epochs. Cohen's d effect sizes were used to assess differences in sleep measures, according to the device (PSG vs. AG) and SO rule (SO_{1-min}, SO_{5-min}, and SO_{10-min}). To allow comparability with previous studies, standard Cohen's d effect size ranges were used to determine the magnitude of

effect (negligible = < 0.2; small = 0.2–0.49; moderate = 0.5–0.79; large ≥ 0.8) [32]. Linear mixed-effects regression models were used to further corroborate the calculated effect sizes. Participant was used as the random effect, while ‘Device’ and ‘Sleep Onset Rule’ were fixed effects. Full and reduced models were compared using a likelihood ratio test to determine if there was an interaction between main effects (‘Device’ × ‘Sleep Onset Rule’). The interactions were not statistically significant at the alpha = 0.05 level. As a result, we used reduced models that only included main effects of ‘Device’ and ‘Sleep Onset Rule’ for all sleep measures.

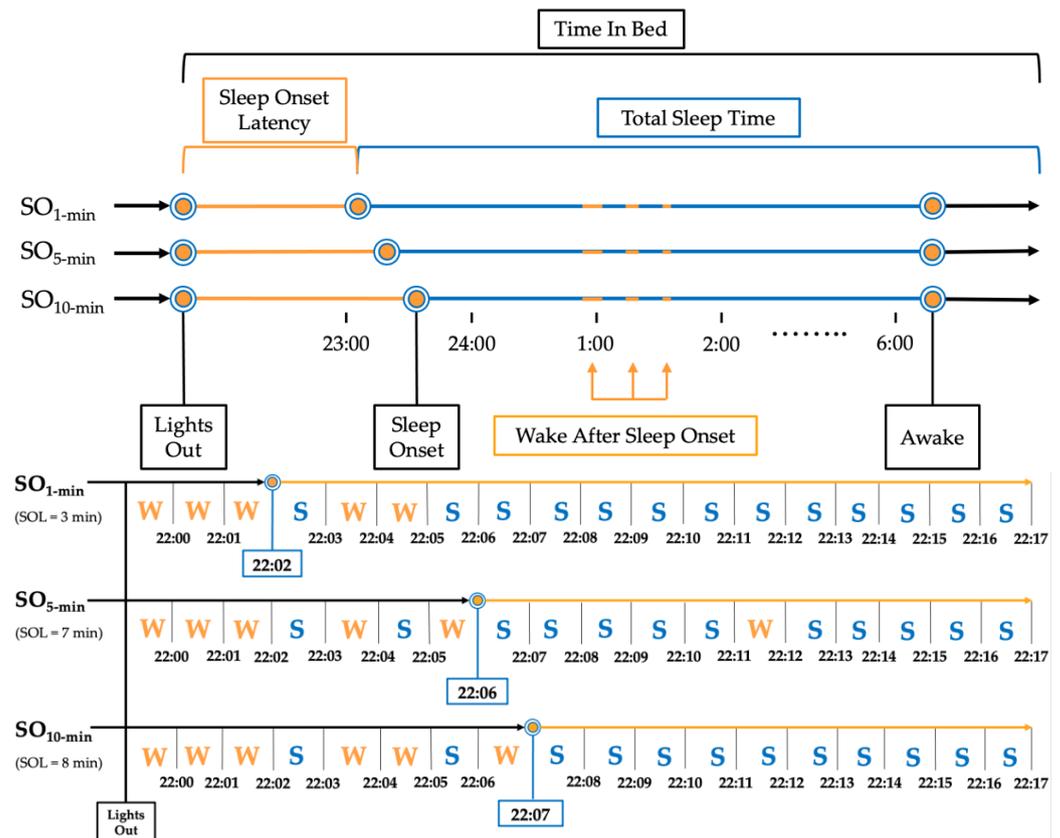


Figure 1. Example of implementation of sleep onset rules for a single participant. Note: Representation of a single night processed using the three sleep onset rules for a mock participant. $SO_{1\text{ min}}$ = Sleep Onset defined as first epoch scored as ‘Sleep’; $SO_{5\text{ min}}$ = Sleep Onset defined as first 5 consecutive epochs scored as ‘Sleep’; $SO_{10\text{ min}}$ = Sleep Onset defined as first 10 epochs scored as ‘Sleep’.

3. Results

The median and interquartile range for each sleep measure is presented in Table 1.

Table 1. Descriptive statistics of sleep measures for device and sleep onset rule.

Sleep Measure	Device Median (IQR)			Sleep Onset Rule Median (IQR)	
	PSG	AG	1	5	10
TST (min)	411.0 [321.0–447.0]	431.0 [348.0–465.0]	427.5 [347.3–456.3]	426.0 [347.0–456.3]	426.0 [332.8–453.3]
SOL (min)	19.0 [10.0–34.0]	1.0 [0.0–8.0]	7.5 [0.3–18.3]	9.5 [1.3–18.3]	11.0 [5.3–23.3]

Table 1. Cont.

Sleep Measure	Device Median (IQR)		Sleep Onset Rule Median (IQR)		
	PSG	AG	1	5	10
WASO (min)	25.0 [13.0–32.0]	16.0 [10.0–29.0]	23.5 [13.5–31.3]	20.5 [13.0–31.3]	20.5 [13.0–31.3]
SE (%)	89.0 [92.1–96.0]	94.7 [91.5–97.2]	91.7 [89.0–95.5]	91.7 [88.9–95.2]	91.5 [87.6–94.9]
SE_ASLEEP (%)	93.5 [92.6–96.1]	96.1 [93.2–97.2]	95.5 [93.1–97.0]	96.5 [93.2–97.0]	96.5 [93.4–97.0]

3.1. Epoch-by-Epoch Agreement, Sensitivity, and Specificity

Binary epoch-by-epoch sleep/wake scores were used to calculate the agreement, sensitivity, and specificity of AG, compared with PSG (Table 2). ‘Sleep Onset Rule’ did not have a meaningful impact on agreement, sensitivity, or specificity. Sleep/wake agreement and sensitivity were high (89.0–89.5% and 97.2%, respectively), while specificity was low (23.6–25.1%).

Table 2. Epoch-by-epoch agreement, sensitivity, and specificity.

	Agreement (%)	Sensitivity (%)	Specificity (%)
1	89.0	97.2	25.1
5	89.2	97.2	23.7
10	89.5	97.2	23.6

3.2. Analysis of ‘Sleep Onset Rule’ and ‘Device’

There was no effect of ‘Sleep Onset Rule’ for any sleep measure. According to the effect size analyses (Table 3), the AG significantly underestimated SOL ($d = 1.09$ to 1.46) and WASO ($d = 0.31$ to 0.42), compared with PSG (1 min). Conversely, AG significantly overestimated TST ($d = -0.28$ to -0.31), SE ($d = -1.18$ to -1.37), and SE_ASLEEP ($d = -0.53$ to -0.64).

Table 3. Effect of sleep onset rule on sleep measures.

Sleep Measure	Effect Size Cohen’s d (95% CI)				
	PSG		AG		
	5	10	1	5	10
TST (min)	0.03 (−0.64, 0.69)	0.01 (−0.68, 0.69)	−0.31 * (−1.03, 0.37)	−0.30 * (−0.94, 0.37)	−0.28 * (−1.01, 0.4)
SOL (min)	−0.07 (−0.75, 0.60)	−0.13 (−0.78, 0.55)	1.46 † (1.11, 2.59)	1.28 † (0.92, 2.18)	1.09 † (0.72, 1.84)
WASO (min)	0.02 (−0.64, 0.69)	0.02 (−0.64, 0.74)	0.31 * (−0.40, 0.93)	0.38 * (−0.26, 1.02)	0.42 * (−0.24, 1.00)
SE (%)	0.05 (−0.64, 0.69)	0.11 (−0.52, 0.77)	−1.37 † (−2.28, −0.82)	−1.32 † (−2.15, −0.74)	−1.18 † (−1.19, −0.60)
SE_ASLEEP (%)	−0.02 (−0.70, 0.64)	−0.03 (−0.65, 0.62)	−0.53 † (−1.22, 0.14)	−0.61 † (−1.29, 0.00)	−0.64 † (−1.37, −0.04)

PSG with 1 min sleep onset rule was used as the criterion measure for all effect size comparisons. * = small effect; † = moderate effect; ‡ = large effect.

The linear mixed model estimates for ‘Device’ and ‘Sleep Onset Rule,’ along with their 95% confidence intervals, are represented in Figure 2. There were no significant interactions or main effects of ‘Sleep Onset Rule’ for any sleep measure. Consistent with the effect size

analyses, there was a main effect of ‘Device’ where AG overestimated TST (26.2 min), SE (6.5%), and SE_ASLEEP (1.9%), but underestimated SOL (19.7 min) and WASO (6.5 min).

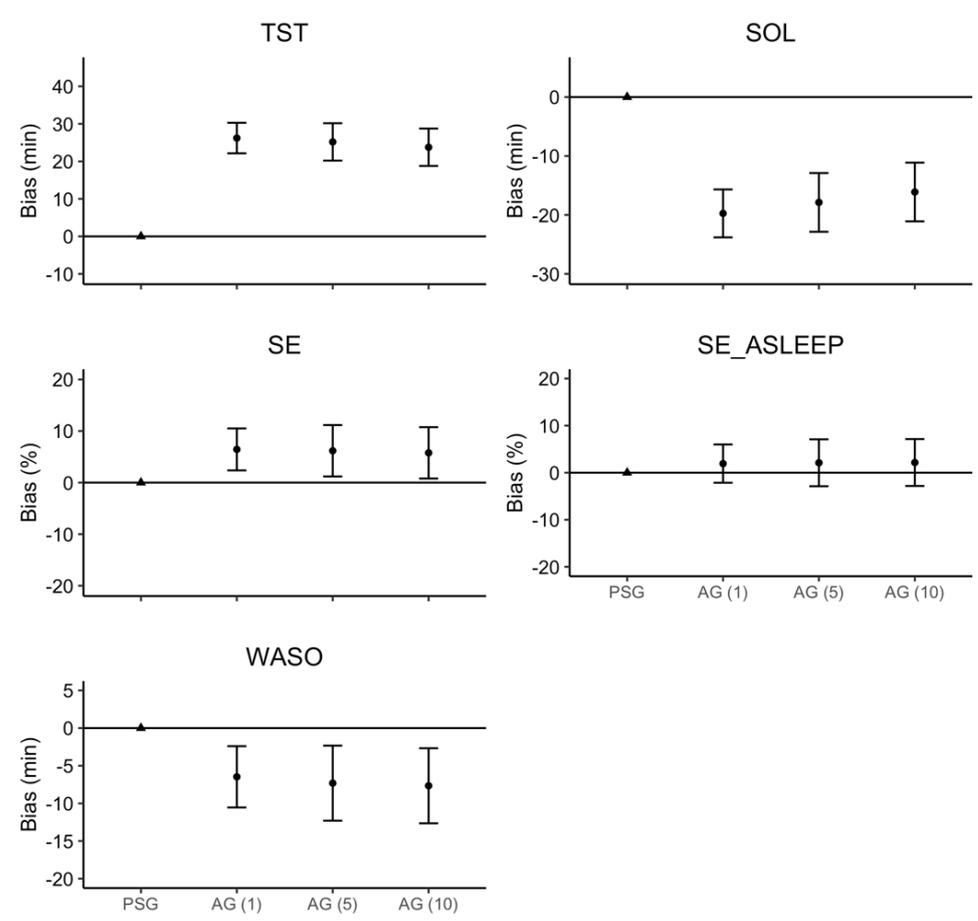


Figure 2. Regression model bias estimates for AG-derived sleep measure. Note: Error bars represent the 95% confidence interval around the bias for each estimate.

4. Discussion

The purpose of this study was to determine if applying alternative sleep onset (SO) definitions, which have been beneficial in improving measurement accuracy in other populations, would improve the accuracy of the commonly used Cole–Kripke algorithm in the classification of accelerometer-derived sleep measurements in healthy adults, compared with PSG. The AG underestimation of SOL was the primary source of the overestimation of TST. However, alternative SO rules ($SO_{1\text{-min}}$, $SO_{5\text{-min}}$, $SO_{10\text{-min}}$) did not impact agreement statistics and there was no main effect of SO rule for any sleep metric.

4.1. Agreement, Sensitivity, and Specificity

In the present study, agreement and sensitivity were high, whilst specificity was low. Our results are consistent with most studies demonstrating moderate-to-high agreement and sensitivity between AG and PSG at the expense of specificity [33]. The high agreement may be partially due to the population studied. When Sleep Efficiency is high (>80%), as seen in typical healthy adults [34,35], accelerometers tend to overestimate sleep epochs and underestimate waking epochs [12,36,37]. The current study is consistent with this previous research; agreement between devices was high, and AG-derived TST and SE were both overestimated. These data are congruent with several similar studies on healthy or athletic populations [12,36,37].

4.2. Total Sleep Time

The AG overestimated TST, compared to PSG. The tendency for accelerometers to overestimate TST has been well-documented using a variety of devices and algorithms [12,19,28–30,36,38,39]. Relatively few studies have reported the underestimation of TST [40,41] or null findings [19,41–43]. Regardless of the SO rule applied in the current study, the AG overestimated TST by 26.2 min, which is within the range of previously reported overestimations of TST using the same device and algorithm (8.1 to 81.1 min) [19,28,36].

In the present study, we used the ActiGraph GT3X+ with the Cole–Kripke algorithm. Despite the widespread use of the GT3X+ device to measure patterns of activity and inactivity, few studies have used the same device to estimate sleep measures in healthy adults [19,28,36]. Two of these studies reported relatively large TST overestimations (62.3, 81.1 min) [28,36], compared with the present study (26.2 min). One previous study reported no significant difference in TST between AG and PSG (14.0 min) [19]. Interestingly, the two studies demonstrating large overestimations in TST included middle-aged participants with wide Body Mass Index (BMI) ranges (20–45, 17.7–45.2) [28,36]. The study reporting no significant differences in TST was performed in a healthy young adult population, with a cohort of similar age and anthropometrics to the present study [19]. These data suggest age and health status (e.g., BMI) may impact estimates of sleep metrics, including TST. Together, the ActiGraph GT3X+ accelerometer and Cole–Kripke algorithm appears to be better suited for estimating TST in healthy young adults than middle aged and older adults with high BMI.

4.3. Sleep Onset Latency

The AG underestimated SOL by 19.7 min, compared to PSG, which is similar to previous studies on healthy adults, regardless of the device or algorithm used [12,21,28,29,36–38,41–48]. Studies using the ActiGraph GT3X+ and Cole–Kripke algorithm have reported underestimations of SOL ranging from 6.4 to 15.1 min [19,28,36,46]. To the best of our knowledge, no study to date has demonstrated an overestimation of SOL in healthy adults.

The underestimation of SOL can be attributed to the poor ‘Wake’ epoch detection capability of accelerometers, supported by the low specificity observed in the present study. The underestimation of SOL and low specificity are well-documented issues when using accelerometers to measure sleep [15,17,33]. Studies on children and sleep-disordered adults have aimed to address the issue of SOL underestimation by applying various rules to define sleep onset. For instance, Chae and colleagues used sleep onset definitions of 4, 5, 6, and 15 min immobility with no more than 1 min interrupted by a ‘wake’ score, with sleep-disordered participants (Obstructive Sleep Apnea, Periodic Limb Movement Disorder) [15]. The 5 min rule outperformed other sleep onset definitions [15]. We adopted an approach similar to studies that demonstrated improvements in SOL estimation by applying three Sleep Onset Rules (1 min, 5 min, and 10 min) [15,25]. We did not observe differences for any sleep measures among sleep onset rules, suggesting the need for alternative methods to address SOL underestimation errors due to poor sleep–wake detection.

4.4. Wake after Sleep Onset

Our results are similar to previous studies demonstrating an underestimation of WASO, with biases ranging from 1.4 to 60.2 min [12,19,21,28,29,31,36,37,43,46,49,50]. Studies that used the GT3X+ and Cole–Kripke algorithm reported some of the lowest (1.4 min) [19] and highest (46.8–60.2 min) [28,36] underestimations of WASO. The source of WASO underestimation is likely the same as SOL underestimation, which is the poor ‘Wake’ epoch detection of accelerometers. Supporting this claim, the only other study to report specificity while using the GT3X+ and Cole–Kripke algorithm found a low specificity (35%) [19]. While the authors did not report SOL, there was a slight underestimation of WASO (1.4 min) within their sub-sample of healthy young adult men [19].

4.5. Sleep Efficiency

In the present study, we observed an overestimation of AG-derived SE, compared with PSG (1.9%), which can be attributed to the inaccurate prediction of sleep during still waking periods (underestimating SOL and WASO). Our results are consistent with other studies using the ActiGraph GT3X+ and Cole–Kripke algorithm, with overestimations of 1.3% [19], 12.6% [28], and 18.3% [36], respectively.

As an exploratory aim, we investigated the influence of removing waking epochs, from the PSG- and AG-derived data, at the beginning of the night, to determine if SE estimates would improve. The effect size for SE_ASLEEP was attenuated, compared with SE, and the regression analysis indicated that SE_ASLEEP was not significantly different from PSG. With the elimination of the wake-to-sleep transition time (SOL; a time that is known to be a source of misclassification), the underestimation of WASO by the ActiGraph improved the SE_ASLEEP accuracy. To the best of our knowledge, no other studies to date have explored the elimination of the wake-to-sleep transition when comparing AG versus PSG-derived SE. The consistent overestimation of SE with the GT3X+ and Cole–Kripke algorithm is likely due to the overestimation of TST, since two of the studies using this device and algorithm also reported the overestimation of TST [28,36]. Supporting this notion, the study reporting the highest overestimation of TST was the same to report the highest overestimation of SE [36].

Although the overestimation of ‘sleep’ epochs (e.g., TST and SE) adversely affect the agreement between AG and PSG, the primary source of error is the underestimation of ‘wake’ epochs (e.g., WASO and SOL). Previous studies reporting the overestimation of ‘sleep’ epochs and the underestimation of ‘wake’ epochs also report disproportionately high sensitivities (96.7–99.0%) and low specificities (14.3–48%) [12,29,37,44,46]. Importantly, these previous studies utilized a variety of devices, algorithms, and user-defined settings, highlighting the universal importance of improving ‘wake’ epoch detection to improve wearable device accuracy.

5. Summary and Conclusions

The present study contributes at least two novel additions to the study of accelerometer-derived sleep measures: (1) the exploration of different sleep onset rules to improve sleep–wake detection, a period of the wake-to-sleep transition with previously identified issues of low specificity; and (2) the use of healthy young adults. Previous studies that have explored the use of alternative sleep onset rules have only been conducted on children, adolescents, and sleep-disordered adults [15,51]. Furthermore, the present study accounted for the PSG ‘first-night-effect’ [9] by collecting data for two nights, but only analyzing data from the second night. The utilization of a linear mixed-effects regression represents a methodological advance, allowing us to appropriately address the non-normality in the data and provide directly interpretable metrics of change to the sleep-onset rules.

The primary limitation of the present study was the use of a single algorithm and device with one-wear location. This decision was made to simplify the aims of the present study. Future studies may consider comparing multiple algorithms across several devices and wear locations, which would require a larger sample size for adequate statistical power. Within this narrow scope, our findings indicate that, regardless of sleep onset rule, the Cole–Kripke algorithm applied to AG data misclassifies ‘wake’ periods while a person is lying in bed as ‘sleep’.

More sophisticated approaches, including machine learning techniques, are currently being developed [52,53] as alternative methods that could be used for improving the wake-to-sleep transitions in healthy and diseased populations. Additionally, incorporating biometric signals (e.g., heart rate, respirometry) related to physiological changes that occur as one falls asleep, and are indicative of sleep, may be combined with the same actigraphy-based device to increase the specificity of sleep measures. The development of validated accelerometer data processing methods to improve sleep/wake detection will enhance the portability and accessibility of ambulatory sleep monitoring in clinical populations.

Improvements in accelerometer sleep/wake scoring, when developed in conjunction with devices that incorporate biometric signals, will assist in minimizing the financial and physical burdens of laboratory-based polysomnography tests. Improving sleep/wake detection, and doing so in free-living environments, will allow researchers and clinicians to scale-up sleep and health intervention efforts.

Author Contributions: Conceptualization, J.D.C., M.A.B. and J.R.S.; methodology, J.D.C., M.A.B. and J.R.S.; software, J.D.C. and M.A.B.; validation, J.D.C., M.A.B. and J.R.S.; formal analysis, J.D.C. and J.W.S.; investigation, J.D.C. and M.A.B.; resources, J.D.C., M.A.B. and J.R.S.; data curation, J.D.C., M.A.B. and J.R.S.; writing—original draft preparation, J.D.C., M.A.B., J.W.S. and J.R.S.; writing—review and editing, J.D.C., M.A.B., J.W.S. and J.R.S.; visualization, J.D.C., M.A.B., J.W.S. and J.R.S.; supervision, M.A.B., J.W.S. and J.R.S.; project administration, J.D.C., M.A.B. and J.R.S.; funding acquisition, M.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This project was completed with the generous funding of the Novartis Institute of Biomedical Research.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the University of Massachusetts Amherst (ID: 2018-4642, 6/4/2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: Data for this project was generated at the University of Massachusetts Amherst Center for Human Health & Performance and Sleep Monitoring Core. The authors would also like to thank Jamie Allfisher for his contributions as the lead polysomnographic technologist on the current project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Watson, N.F.; Safwan Badr, M.; Belenky, G.; Bliwise, D.L.; Buxton, O.M.; Buysse, D.; Dinges, D.F.; Gangwisch, J.; Grandner, M.A.; Kushida, C.; et al. Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the Recommended Amount of Sleep for a Healthy Adult: Methodology and Discussion Consensus. *Sleep* **2015**, *38*, 1161–1183. [[CrossRef](#)] [[PubMed](#)]
2. Magee, C.A.C.A.; Caputi, P.; Iverson, D.C.D.C. Relationships between self-rated health, quality of life and sleep duration in middle aged and elderly Australians. *Sleep Med.* **2011**, *12*, 346–350. [[CrossRef](#)] [[PubMed](#)]
3. Buxton, O.M.; Marcelli, E. Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States. *Soc. Sci. Med.* **2010**, *71*, 1027–1036. [[CrossRef](#)] [[PubMed](#)]
4. Sonni, A.; Spencer, R.M.C. Sleep protects memories from interference in older adults. *Neurobiol. Aging* **2015**, *36*, 2272–2281. [[CrossRef](#)]
5. Rogers, N.L.; Dorrian, J.; Dinges, D.F. Sleep, waking and neurobehavioural performance. *Front. Biosci.* **2003**, *8*, S1056–S1067. [[CrossRef](#)]
6. Robertson, M.D.; Russell-Jones, D.; Umpleby, A.M.; Dijk, D.J. Effects of three weeks of mild sleep restriction implemented in the home environment on multiple metabolic and endocrine markers in healthy young men. *Metabolism* **2013**, *62*, 204–211. [[CrossRef](#)]
7. Leproult, R.; Van Reeth, O.; Byrne, M.M.; Sturis, J.; Van Cauter, E. Sleepiness, Performance, and Neuroendocrine Function during Sleep Deprivation: Effects of Exposure to Bright Light or Exercise. *J. Biol. Rhythms* **1997**, *12*, 245–258. [[CrossRef](#)]
8. Banks, S.; Dinges, D.F. Behavioral and physiological consequences of sleep restriction. *J. Clin. Sleep Med.* **2007**, *3*, 519–528. [[CrossRef](#)]
9. Byun, J.-H.; Kim, K.T.; Moon, H.-J.; Motamedi, G.K.; Cho, Y.W. The first night effect during polysomnography, and patients' estimates of sleep quality. *Psychiatry Res.* **2019**, *274*, 27–29. [[CrossRef](#)]
10. Piwek, L.; Ellis, D.A.; Andrews, S.; Joinson, A. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Med.* **2016**, *13*, e1001953. [[CrossRef](#)]
11. Zinkhan, M.; Kantelhardt, J.W. Sleep Assessment in Large Cohort Studies with High-Resolution Accelerometers. *Sleep Med. Clin.* **2016**, *11*, 469–488. [[CrossRef](#)] [[PubMed](#)]
12. De Souza, L.; Benedito-Silva, A.A.A.; Pires, M.L.N.; Poyares, D.; Tufik, S.; Calil, H.M. Further Validation of Actigraphy for Sleep Studies. *Sleep* **2003**, *26*, 81–85. [[CrossRef](#)] [[PubMed](#)]

13. Marino, M.; Li, Y.; Rueschman, M.N.; Winkelman, J.W.; Ellenbogen, J.M.; Solet, J.M.; Dulin, H.; Berkman, L.F.; Buxton, O.M. Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. *Sleep* **2013**, *36*, 1747–1755. [[CrossRef](#)] [[PubMed](#)]
14. Ancoli-Israel, S.; Cole, R.; Alessi, C.; Chambers, M.; Moorcroft, W.; Pollak, C.P. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* **2003**, *26*, 342–392. [[CrossRef](#)]
15. Chae, K.Y.; Kripke, D.F.; Poceta, J.S.; Shadan, F.; Jamil, S.M.; Cronin, J.W.; Kline, L.E. Evaluation of immobility time for sleep latency in actigraphy. *Sleep Med.* **2009**, *10*, 621–625. [[CrossRef](#)]
16. Pollak, C.P.; Tryon, W.W.; Nagaraja, H.; Dzwonczyk, R. How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep* **2001**, *24*, 957–965. [[CrossRef](#)]
17. Tryon, W.W. Issues of Validity in Actigraphic Sleep Assessment. *Sleep* **2004**, *27*, 158–165. [[CrossRef](#)]
18. Paquet, J.; Kawinska, A.; Carrier, J. Wake detection capacity of actigraphy during sleep. *Sleep* **2007**, *30*, 1362–1369. [[CrossRef](#)]
19. Quante, M.; Kaplan, E.R.; Cailler, M.; Rueschman, M.; Wang, R.; Weng, J.; Taveras, E.M.; Redline, S. Actigraphy-based sleep estimation in adolescents and adults: A comparison with polysomnography using two scoring algorithms. *Nat. Sci. Sleep* **2018**, *10*, 13–20. [[CrossRef](#)]
20. Kushida, C.A.; Chang, A.; Gadkary, C.; Guilleminault, C.; Carrillo, O.; Dement, W.C. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med.* **2001**, *2*, 389–396. [[CrossRef](#)]
21. Fuller, K.L.; Juliff, L.; Gore, C.J.; Peiffer, J.J.; Halson, S.L. Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes. *J. Sci. Med. Sport* **2017**, *20*, 756–760. [[CrossRef](#)] [[PubMed](#)]
22. Oakley, N. *Validation with Polysomnography of the Sleep-Watch Sleep/Wake Scoring Algorithm Used by the Actiwatch Activity Monitoring System*; Mini Mitter Co., Inc.: Bend, OR, USA, 1997.
23. Cole, R.J.; Kripke, D.F.; Gruen, W.; Mullaney, D.J.; Gillin, J.C. Automatic Sleep/Wake Identification From Wrist Activity. *Sleep* **1992**, *15*, 461–469. [[CrossRef](#)] [[PubMed](#)]
24. Meltzer, L.J.; Westin, A.M.L. A comparison of actigraphy scoring rules used in pediatric research. *Sleep Med.* **2011**, *12*, 793–796. [[CrossRef](#)] [[PubMed](#)]
25. Meltzer, L.J.; Walsh, C.M.; Peightal, A.A. Comparison of Actigraphy Immobility Rules with Polysomnographic Sleep Onset Latency in Children and Adolescents. *Sleep Breath* **2015**, *19*, 1415–1423. [[CrossRef](#)] [[PubMed](#)]
26. Strackiewicz, M.; Glynn, N.W.; Harezlak, J. On Placement, Location and Orientation of Wrist-Worn Tri-Axial Accelerometers during Free-Living Measurements. *Sensors* **2019**, *19*, 2095. [[CrossRef](#)]
27. Berry, R.B.; Brooks, R.; Gamaldo, C.; Harding, S.M.; Lloyd, R.M.; Quan, S.F.; Troester, M.T.; Vaughn, B.V. AASM Scoring Manual Updates for 2017 (Version 2.4). *J. Clin. Sleep Med.* **2017**, *13*, 665. [[CrossRef](#)]
28. Dunican, I.C.; Murray, K.; Slater, J.A.; Maddison, K.J.; Jones, M.J.; Dawson, B.; Straker, L.M.; Caldwell, J.A.; Halson, S.L.; Eastwood, P.R. Laboratory and home comparison of wrist-activity monitors and polysomnography in middle-aged adults. *Sleep Biol. Rhythms* **2018**, *16*, 85–97. [[CrossRef](#)]
29. Markwald, R.R.; Bessman, S.C.; Reini, S.A.; Drummond, S.P.A.A. Performance of a portable sleep monitoring device in individuals with high versus low sleep efficiency. *J. Clin. Sleep Med.* **2016**, *12*, 95–103. [[CrossRef](#)]
30. Laakso, M.-L.L.; Leinonen, L.; Lindblom, N.; Joutsiniemi, S.-L.L.; Kaski, M. Wrist actigraphy in estimation of sleep and wake in intellectually disabled subjects with motor handicaps. *Sleep Med.* **2004**, *5*, 541–550. [[CrossRef](#)]
31. Peterson, B.T.; Chiao, P.; Pickering, E.; Freeman, J.; Zammit, G.K.; Ding, Y.; Badura, L.L. Comparison of actigraphy and polysomnography to assess effects of zolpidem in a clinical research unit. *Sleep Med.* **2012**, *13*, 419–424. [[CrossRef](#)]
32. Cohen, J. A Power Primer. *Psychol. Bull.* **1992**, *112*, 155–159. [[CrossRef](#)] [[PubMed](#)]
33. Sadeh, A. The role and validity of actigraphy in sleep medicine: An update. *Sleep Med. Rev.* **2011**, *15*, 259–267. [[CrossRef](#)] [[PubMed](#)]
34. Unruh, M.L.; Redline, S.; An, M.-W.; Buysse, D.J.; Nieto, F.J.; Yeh, J.-L.; Newman, A.B. Subjective and Objective Sleep Quality and Aging in the Sleep Heart Health Study. *J. Am. Geriatr. Soc.* **2008**, *56*, 1218–1227. [[CrossRef](#)] [[PubMed](#)]
35. Desjardins, S.; Lapierre, S.; Hudon, C.; Desgagné, A. Factors involved in sleep efficiency: A population-based study of community-dwelling elderly persons. *Sleep* **2019**, *42*, zsz038. [[CrossRef](#)] [[PubMed](#)]
36. Zinkhan, M.; Berger, K.; Hense, S.; Nagel, M.; Obst, A.; Koch, B.; Penzel, T.; Fietze, I.; Ahrens, W.; Young, P.; et al. Agreement of different methods for assessing sleep characteristics: A comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography. *Sleep Med.* **2014**, *15*, 1107–1114. [[CrossRef](#)]
37. O'Hare, E.; Flanagan, D.; Penzel, T.; Garcia, C.; Frohberg, D.; Heneghan, C. A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. *Sleep Breath.* **2015**, *19*, 91–98. [[CrossRef](#)]
38. Mikkelsen, K.B.; Ebajemito, J.K.; Bonmati-Carrion, M.A.; Santhi, N.; Revell, V.L.; Atzori, G.; Della Monica, C.; Debener, S.; Dijk, D.-J.; Sterr, A.; et al. Machine-learning-derived sleep-wake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy. *J. Sleep Res.* **2019**, *28*, e12786. [[CrossRef](#)]
39. Montgomery-Downs, H.E.; Insana, S.P.; Bond, J.A. Movement toward a novel activity monitoring device. *Sleep Breath.* **2012**, *16*, 913–917. [[CrossRef](#)]
40. Sargent, C.; Lastella, M.; Halson, S.L.; Roach, G.D. The validity of activity monitors for measuring sleep in elite athletes. *J. Sci. Med. Sport* **2016**, *19*, 848–853. [[CrossRef](#)]

41. Rupp, T.L.; Balkin, T.J. Comparison of Motionlogger Watch and Actiwatch actigraphs to polysomnography for sleep/wake estimation in healthy young adults. *Behav. Res. Methods* **2011**, *43*, 1152–1160. [[CrossRef](#)]
42. Sánchez-Ortuño, M.M.; Edinger, J.D.; Means, M.K.; Almirall, D. Home is where sleep is: An ecological approach to test the validity of actigraphy for the assessment of insomnia. *J. Clin. Sleep Med.* **2010**, *6*, 21–29. [[CrossRef](#)] [[PubMed](#)]
43. Pigeon, W.R.; Taylor, M.; Bui, A.; Oleynk, C.; Walsh, P.; Bishop, T.M. Validation of the Sleep-Wake Scoring of a New Wrist-Worn Sleep Monitoring Device. *J. Clin. Sleep Med.* **2018**, *14*, 1057–1062. [[CrossRef](#)] [[PubMed](#)]
44. Chakar, B.; Senny, F.; Poirrier, A.-L.; Cambron, L.; Fanielle, J.; Poirrier, R. Validation of midsagittal jaw movements to measure sleep in healthy adults by comparison with actigraphy and polysomnography. *Sleep Sci.* **2017**, *10*, 122–127. [[CrossRef](#)] [[PubMed](#)]
45. Shambroom, J.R.; Fábregas, S.E.; Johnstone, J. Validation of an automated wireless system to monitor sleep in healthy adults. *J. Sleep Res.* **2012**, *21*, 221–230. [[CrossRef](#)]
46. Slater, J.A.; Botsis, T.; Walsh, J.; King, S.; Straker, L.M.; Eastwood, P.R. Assessing sleep using hip and wrist actigraphy. *Sleep Biol. Rhythms* **2015**, *13*, 172–180. [[CrossRef](#)]
47. Tonetti, L.; Pasquini, F.; Fabbri, M.; Belluzzi, M.; Natale, V. Comparison of Two Different Actigraphs with Polysomnography in Healthy Young Subjects. *Chronobiol. Int.* **2008**, *25*, 145–153. [[CrossRef](#)]
48. Uchida, S.; Endo, T.; Suenaga, K.; Iwami, H.; Inoue, S.; Fujioka, E.; Imamura, A.; Atsumi, T.; Inagaki, Y.; Kamei, A. Sleep evaluation by a newly developed PVDF sensor non-contact sheet: A comparison with standard polysomnography and wrist actigraphy. *Sleep Biol. Rhythms* **2011**, *9*, 178–187. [[CrossRef](#)]
49. Kosmadopoulos, A.; Sargent, C.; Darwent, D.; Zhou, X.; Roach, G.D. Alternatives to polysomnography (PSG): A validation of wrist actigraphy and a partial-PSG system. *Behav. Res. Methods* **2014**, *46*, 1032–1041. [[CrossRef](#)]
50. Matsuo, M.; Masuda, F.; Sumi, Y.; Takahashi, M.; Yamada, N.; Ohira, M.H.; Fujiwara, K.; Kanemura, T.; Kadotani, H. Comparisons of Portable Sleep Monitors of Different Modalities: Potential as Naturalistic Sleep Recorders. *Front. Neurol.* **2016**, *7*, 110. [[CrossRef](#)]
51. Meltzer, L.J.; Hiruma, L.S.; Avis, K.; Montgomery-Downs, H.; Valentin, J. Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents. *Sleep* **2015**, *38*, 1323–1330. [[CrossRef](#)]
52. Sundararajan, K.; Georgievska, S.; te Lindert, B.H.W.; Gehrman, P.R.; Ramautar, J.; Mazzotti, D.R.; Sabia, S.; Weedon, M.N.; van Someren, E.J.W.; Ridder, L.; et al. Sleep classification from wrist-worn accelerometer data using random forests. *Sci. Rep.* **2021**, *11*, 24. [[CrossRef](#)] [[PubMed](#)]
53. Van Hees, V.T.; Sabia, S.; Jones, S.E.; Wood, A.R.; Anderson, K.N.; Kivimäki, M.; Frayling, T.M.; Pack, A.I.; Bucan, M.; Trenell, M.I.; et al. Estimating sleep parameters using an accelerometer without sleep diary. *Sci. Rep.* **2018**, *8*, 12975. [[CrossRef](#)] [[PubMed](#)]