

RESEARCH

Open Access



Validation of minute-to-minute scoring for sleep and wake periods in a consumer wearable device compared to an actigraphy device

Joseph Cheung^{1,2*} , Jamie M. Zeitzer^{1,2,3}, Haoyang Lu² and Emmanuel Mignot^{1,2*}

Abstract

Background: Actigraphs are widely used portable wrist-worn devices that record tri-axial accelerometry data. These data can be used to approximate amount and timing of sleep and wake. Their clinical utility is limited, however, by their expense. Tri-axial accelerometer-based consumer wearable devices (so-called fitness monitors) have gained popularity and could represent cost-effective research alternatives to more expensive devices. Lack of independent validation of minute-to-minute accelerometer data for consumer devices has hindered their utility and acceptance.

Methods: We studied a consumer-grade wearable device, Arc (Huami Inc., Mountain View CA), for which minute-to-minute accelerometer data (vector magnitude) could be obtained. Twelve healthy participants and 19 sleep clinic patients wore on their non-dominant wrist, both an Arc and a research-grade actigraph (Actiwatch Spectrum, Philips, Bend OR) continuously over a period of 48 h in free-living conditions. Time-stamped data from each participant were aligned and the Cole-Kripke algorithm was used to assign a state of “sleep” or “wake” for each minute-long epoch recorded by the Arc. The auto and low scoring settings on the Actiwatch software (Actiware) were used to determine sleep and wake from the Actiwatch data and were used as the comparators. Receiver operating characteristic curves were used to optimize the relationship between the devices.

Results: Minute-by-minute Arc and Actiwatch data were highly correlated ($r = 0.94$, Spearman correlation) over the 48-h study period. Treating the Actiwatch auto scoring as the gold standard for determination of sleep and wake, Arc has an overall accuracy of $99.0\% \pm 0.17\%$ (SEM), a sensitivity of $99.4\% \pm 0.19\%$, and a specificity of $84.5\% \pm 1.9\%$ for the determination of sleep. As compared to the Actiwatch low scoring, Arc has an overall accuracy of $95.2\% \pm 0.36\%$, a sensitivity of $95.7\% \pm 0.47\%$, and a specificity of $91.7\% \pm 0.60\%$ for the determination of sleep.

Conclusions: The Arc, a consumer wearable device in which minute-by-minute activity data could be collected and compared, yielded fundamentally similar sleep scoring metrics as compared to a commonly used clinical-grade actigraph (Actiwatch). We found high degrees of agreement in minute-to-minute data scoring for sleep and wake periods between the two devices.

Keywords: Consumer wearable, Validation, Sleep monitoring, Actigraphy

* Correspondence: cheungj@stanford.edu; mignot@stanford.edu

¹Stanford University Center for Sleep Sciences and Medicine, Palo Alto, CA, USA

Full list of author information is available at the end of the article



Background

Actigraphs are portable wrist-worn devices that record tri-axial accelerometry data (i.e., gross movement in three directions). By imputing sleep patterns from accelerometry data, actigraphs have been used for nearly 30 years to objectively quantify longitudinal sleep patterns in research studies (Ancoli-Israel et al. 2003). The premise of the algorithms that have been developed for such imputation is to assume that the wearer is asleep when not moving and to determine when gross body movements are large and/or long enough to suggest that the wearer is awake (Cole et al. 1992; Sadeh et al. 1991). More recently, actigraphs have been used in clinical practice, especially in the monitoring and treatment of insomnia-related disorders (Ancoli-Israel et al. 2003; Kushida et al. 2001; Morgenthaler et al. 2007). Wide-spread use has however been limited by the high cost of these devices.

There has been a massive increase in the use of accelerometers in recent years as they are found in most cell phones and wrist-worn fitness trackers. Many of these devices use the accelerometer to track movement for use in both sleep and exercise tracking. As these are consumer devices, the algorithms that translate ‘raw movement’ data into ‘sleep/wake’ activity are proprietary. Despite the raw data that is used to impute sleep and wake not being made available to researchers, the whole-night sleep measures of a few of these devices have been validated to varying degrees (de Zambotti et al. 2016; Bianchi 2017; Roomkham et al. 2018). In order to perform proper validation studies, however, an important criterion is to have access to minute-by-minute raw data, as is available in research/clinical-grade actigraphs.

The objective of this study was to examine the feasibility of using a low-cost consumer grade wearable device as an actigraph device for sleep monitoring (see Table 1 for device specifications). We identified a low-cost wearable device, the Amazfit Arc (Huami, Inc), in which minute-by-minute activity data could be obtained. To our knowledge, this is the first study comparing the raw minute-by-minute accelerometry data obtained from a low-cost consumer wearable device to that obtained

from a clinical-grade actigraph in estimating sleep parameters in free-living conditions.

Methods

Twelve community-dwelling participants without significant self-reported health issues or sleep disorders and twenty-two sleep clinic patients at the Stanford University sleep clinic were recruited to participate in this study. Three of the sleep clinic participants did not complete the study due to missing data: two had missing Actiwatch data and one did not return the devices. In all, 31 participants completed the study, 20 of whom were female and 11 male, with a mean (\pm SD) age of 40.1 ± 7.9 years (range, 19–72). Of the 19 participants recruited from the sleep clinic (mean BMI of 25.2 ± 0.9), 16 were later diagnosed with obstructive sleep apnea (OSA, mild to severe), three were diagnosed with hypersomnia (one patient was diagnosed with hypersomnia and OSA), one was diagnosed with delayed sleep-wake phase disorder, two have hypertension. All participants wore on their non-dominant wrist both an Arc and Actiwatch Spectrum continuously over a period of 48 h in free-living conditions outside of the sleep clinic (i.e., two nights of data). Participants completed a custom sleep diary concomitant with wearing the actigraphs. Arc devices (six devices) were purchased from Huami Inc. (Mountain View, CA). Actiwatch Spectrum devices (three devices) were purchased from Philips Respironics (Bend, OR). Both Arc and Actiwatch devices were configured to store data as the integral of activity occurring in 60 s segments. Time synchronization was performed across the Arc and Actiwatch devices at the beginning of each participant’s study period. A Samsung Android (version 7.1.1) smartphone installed with the Amazfit app (version 1.0.2) was used to communicate with Arc devices. The app was used to synchronize the Arc devices before and after the study period. Minute-by-minute accelerometer data were obtained from the Huami Inc’s cloud (<https://github.com/huamitech/rest-api/wiki>; last accessed May 7, 2018). Actiwatch data were retrieved using Philips Actiware (version 6.0.9).

Time stamps were used to align minute-by-minute data from both devices. Sleep diary data were used to set the time in bed window. Spearman’s correlations were used to compare the raw values of the Arc and Actiwatch devices on a minute-by-minute basis in each participant. Actiwatch data in Actiware were also converted into “sleep” and “wake” using the built-in algorithms on both “auto” and “low” settings. For the Arc device, data were cleaned by removing a series of default output values of “20” while device was inactive. To determine the occurrence of wake, we first determined a *Wake Threshold Value* = $(\sum \text{all activity during mobile time/mobile time}) * k$; such that k is a constant and

Table 1 Comparison of consumer- and research-grade actigraphs

	Amazfit Arc	Philips Actiwatch Spectrum
Cost ^a	\$70	\$900–1000
Accelerometers	MEMS tri-axial	Piezo-electric tri-axial
Accelerometer sampling rate	25 Hz	32 Hz
Weight	20 g	31 g
Recording time	30 days	60 days
Battery life (one charge)	20 days	60 days
Light sensor	None	Yes

^aManufacturer suggested retail price at the time of this study. g gram

mobile time is the total time of minute epochs where activity is ≥ 2 . We then used the Cole-Kripke algorithm (Cole et al. 1992) to derive a window adjusted activity value for each 1-min epoch: $Total\ Activity = E_0 + E_1 * 0.2 + E_{-1} * 0.2 + E_2 * 0.04 + E_{-2} * 0.04$; such that E_0 is the activity level in the one-minute epoch of interest, E_1 is one minute later and E_{-1} is one minute earlier, and so on. If the *Total Activity* in a given one-minute epoch is less than or equal to the *Wake Threshold Value*, the epoch is scored as sleep. If the *Total Activity* in a given one-minute epoch is greater than the *Wake Threshold Value*, the epoch is scored as wake. The Actiwatch uses $k = 0.88888$ in its auto scoring method. In Actiwatch's low scoring method, a *Wake Threshold Value* of 20 is used. A secondary algorithm (Kripke et al. 2010; Webster et al. 1982; Jean-Louis et al. 2001) was used to automatically determine sleep onset time and sleep offset time. The algorithm scans the initial minute-by-minute scoring of each time in bed window. Within each window, the beginning of the first five or more consecutive sleep minutes was defined as sleep onset time. Epochs that were initially scored as sleep, before such an onset time, were rescored as wake. Similarly, the end of the last five or more consecutive sleep minutes was defined as sleep offset time. Any epochs that were initially scored as sleep, after such an offset time, were rescored as wake.

Using a receiver operating characteristic (ROC) analysis, we explored a range of constants to select an optimal value for *Wake Threshold Value* determination in the Arc, using the results from the Actiwatch as the “gold standard”. To determine the relative accuracy of the Arc device, we compared minute-by-minute sleep and wake assignments in both devices and calculated the overall accuracy $[(True\ Positive\ (TP) + True\ Negative\ (TN))/total]$, sleep sensitivity $[TP / (TP + False\ Negative\ (FN))]$ (same as wake

specificity), sleep specificity $[TN / (TN + False\ Positive\ (FP))]$ (same as wake sensitivity), and wake precision $[TN / (TN + FN)]$. Summary results on total sleep time (TST) and wake after sleep onset (WASO) were calculated. Data are presented as mean \pm SEM except where noted.

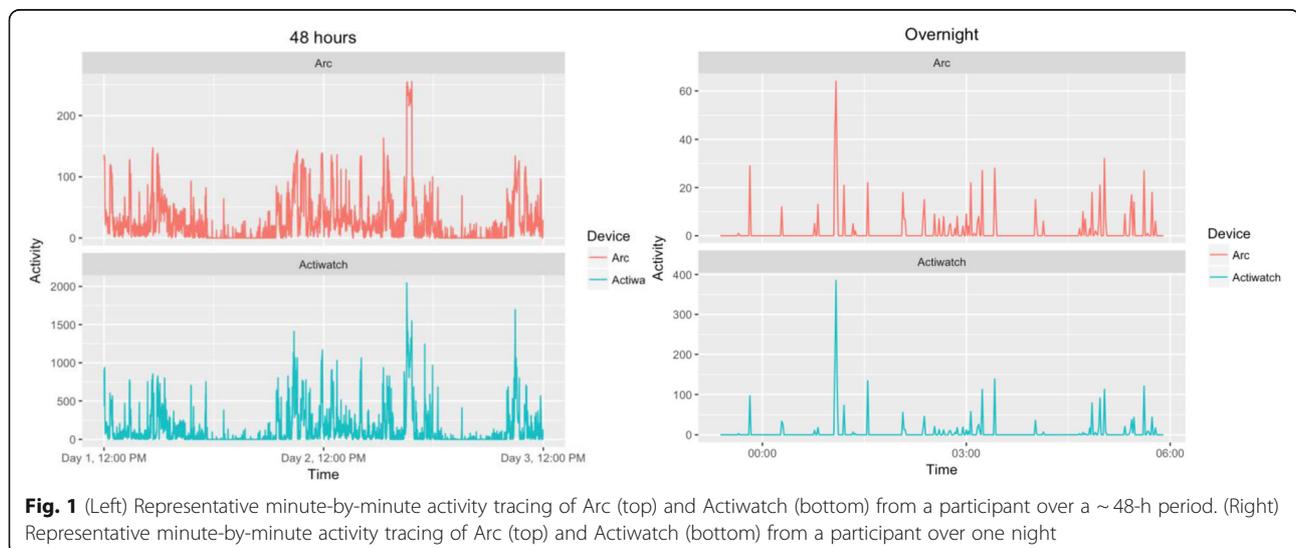
Results

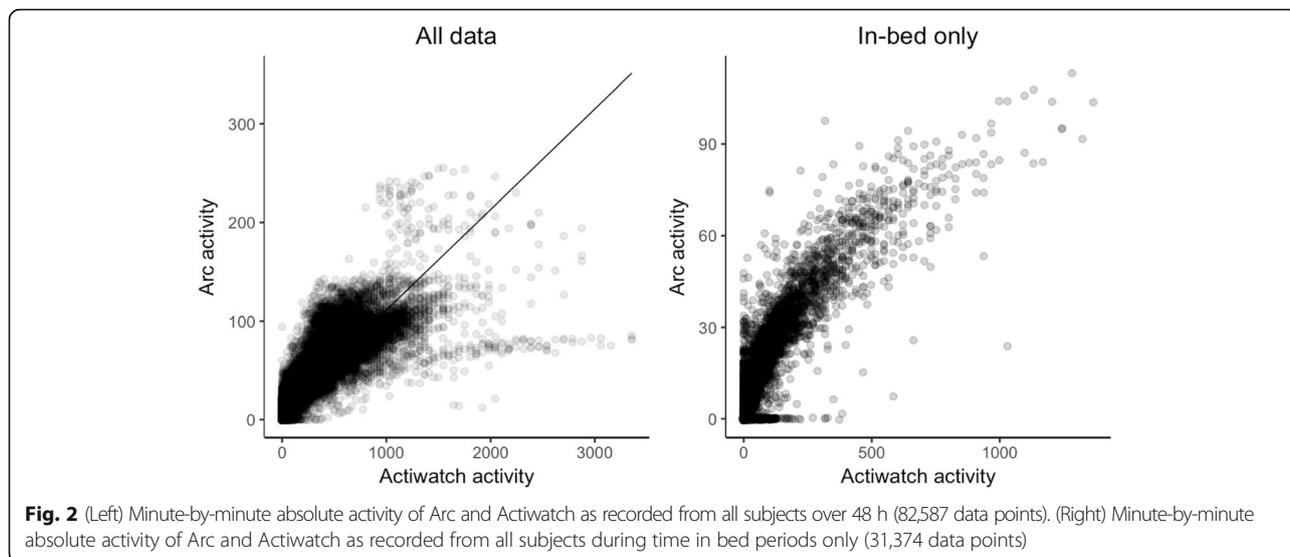
We compared minute-by-minute data obtained from both the Arc and Actiwatch devices over the 48-h study period from all 31 participants. The overall patterns observed between the Arc and Actiwatch appear to be quite similar (Fig. 1).

Within participants, absolute activity for the Actiwatch and Arc devices were highly correlated ($r = 0.94 \pm 0.005$, range: 0.87–0.98, $n = 31$; Spearman correlation). Movement data from in-bed periods were also well correlated ($r = 0.89 \pm 0.01$, range: 0.73–0.96, $n = 31$; Spearman correlation). The absolute difference in values obtained from the Actiwatch and Arc were approximately 9-fold different in magnitude (linear regression of all data, slope \pm SD = 0.11 ± 0.02) (Fig. 2).

To determine a *Wake Threshold Value* that would yield optimal correspondence between the minute-by-minute score of the Arc and Actiwatch, we compared sensitivity and specificity of a series of *Wake Threshold Values* using ROC analysis (Fig. 3). For the Actiwatch analysis in which the *Wake Threshold Value* was determined on auto setting, a k constant of 1.1 used for the Arc data was determined to produce an optimal alignment. For the Actiwatch analysis in which the *Wake Threshold Value* was determined on low setting (a high sensitivity with a threshold value of 20), a threshold value of 5 used for the Arc data produced an optimal alignment.

Using the *Wake Threshold Values* determined in the ROC analysis, we then examined the accuracy,





sensitivity, specificity, and precision of the imputed sleep/wake as determined by the Arc (Table 2). For the most part, there was good correspondence in the determination of sleep and wake by the Arc and Actigraph. Using the auto setting for scoring of the Actigraph data (corresponding to 1.1 on the Arc), there was a slight underscoring of wake with near perfect determination of sleep. Using the low setting for scoring of the Actigraph data (corresponding to 5 on the Arc), there was greater sensitivity for wake at the cost of a slight underscoring of sleep. We also split our data into those from healthy participants only ($n = 12$) and those from sleep patients ($n = 19$). The observed concordance between Arc and Actiwatch (auto setting) was similar, with an overall accuracy of 99.6% in the healthy group and 98.7% in the sleep patient group.

To examine the possibility of systematic bias in overall sleep parameter scoring, we generated Bland-Altman plots to visually inspect the level of agreement between Arc and Actiwatch derived results (Fig. 4). Comparing Arc (using k constant of 1.1) and Actiwatch auto setting, overall bias (discrepancy) in estimating TST was -0.44 min over one sleep period. The spread of the differences is observed to be even, with no bias in overestimation or underestimation of TST. For WASO, overall bias in estimating WASO over one sleep period was 0.35 min. In comparison to Actiwatch low setting (shown in Fig. 4), the overall bias in estimating TST was -4.5 min over one sleep period. In this case, it appears that using a threshold of 5 in Arc (compared to a threshold of 20 used in Actiwatch)

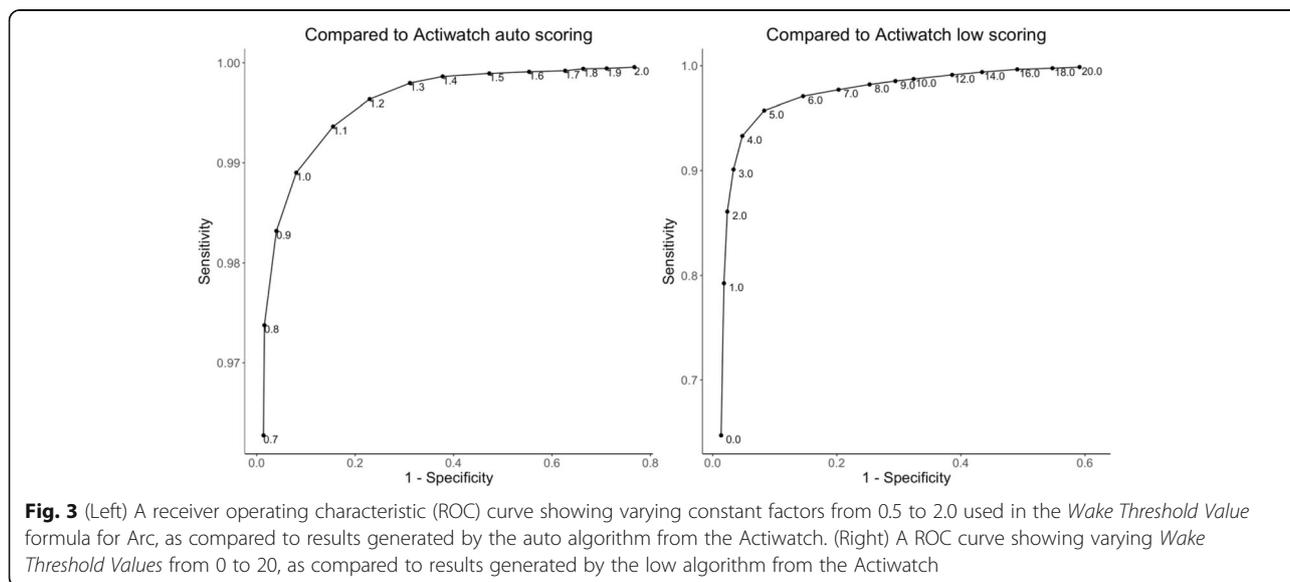


Table 2 Overall accuracy and comparative performance of Arc in detecting sleep/wake during the main sleep periods, in comparison to gold-standard determination of “sleep” and “wake” Actiwatch using the preset auto and low settings of the Actiwatch software

	Overall Accuracy	Sleep Sensitivity (wake specificity)	Sleep Specificity (wake sensitivity)	Wake Precision
“Auto” setting	99.0% ± 0.17%	99.4% ± 0.19%	84.5% ± 1.9%	81.2% ± 2.3%
“Low” setting	95.2% ± 0.36%	95.7% ± 0.47%	91.7% ± 0.60%	83.6% ± 1.2%

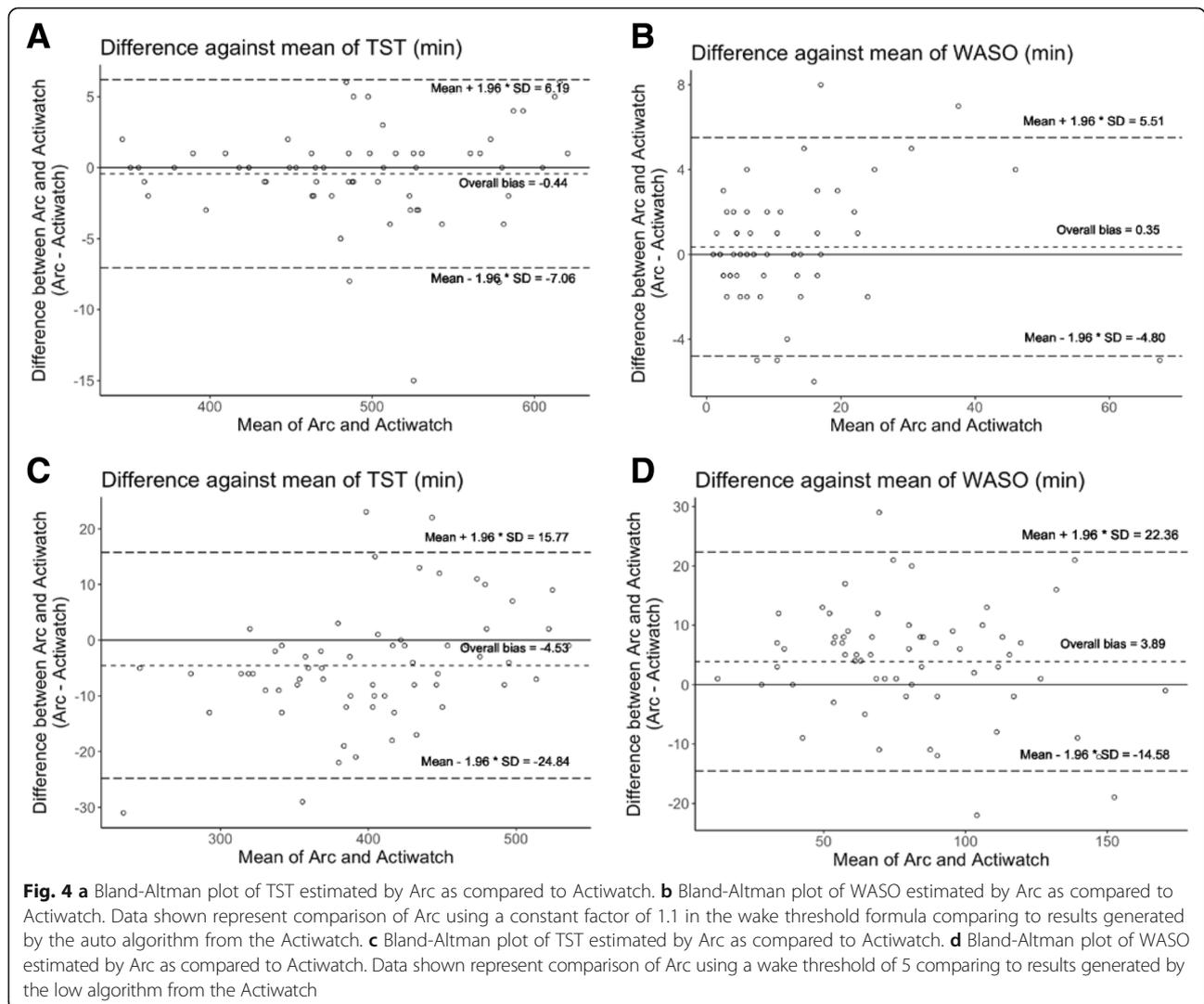
results in a slight underestimation of TST for the Arc device. In terms of WASO, overall bias in estimating WASO over one sleep period was 3.9 min, with a slight overestimation using the Arc device.

Discussion

In comparing the accuracy of Arc, a consumer wearable device, against a clinical/research-grade actigraphy device, Philips Actiwatch (Spectrum), we find that the consumer device performs similarly in the estimation of sleep parameters. Despite lower absolute (approximately 9-fold) value of activity recorded by the Arc, sufficient

signal-to-noise ratio was present to impute sleep and wake states. This is likely because the Cole-Kripke algorithm (Cole et al. 1992) is robust and utilizes *relative* movement data for the determination of sleep and wake. Using ROC analyses to objectively determine thresholds for the Arc device, we were also able to faithfully recapitulate the commonly used auto and low scoring settings on the Actiwatch device. The device performed similarly well in both a patient population (OSA, disrupted sleep) and a control population.

To our knowledge, this is the first validation study where minute by minute accelerometer data (vector



magnitude) from a consumer wearable device was compared to an actigraph in sleep monitoring. Previous studies have compared whole night summary data from wearables, including a recent study (Lee et al. 2017) comparing another consumer wearable (Fitbit Charge HR) with an actigraph (Actiwatch 2). These report good accuracy for sleep evaluation between the two devices, however, only sleep summary data were examined.

Besides the price difference, there are other differences between the Arc and the Actiwatch. While present on the Actiwatch, the Arc lacks a light sensor, a feature often useful in identifying bed and wake times. The Actiwatch is also capable of storing data at a higher average resolution (e.g., 15 s and 30s epochs) in comparison to the Arc. On the other hand, the Arc device is capable of recording raw accelerometer data at 25 Hz resolution. The Arc device also remotely uploads its data to a secure portal, eliminating the need for participants to come to the laboratory to have data from the actigraph downloaded, which is necessary with the Actiwatch. For longer duration longitudinal studies, this could be of significant benefit.

In comparing the Arc device to the Actiwatch, we use the latter as the “gold standard”. Future studies will need to compare Arc to polysomnography, as this is the true, current gold standard in determination of sleep and wake states. The current results do, however, support the potential use of Arc as an actigraphy device for the purpose of sleep monitoring.

Limitations

A limitation of any consumer device, including the Arc, is that the firmware or hardware could be changed without notification, which could make comparison of data between participants problematic. Furthermore, a degree of technical expertise is necessary to extract and convert the Arc data from the raw format to a more usable format, a process that is fairly seamless with the Actigraph and its associated software.

Future directions

Recently, a position statement on consumer sleep technology was published by the American Academy of Sleep Medicine (AASM) (Khosla et al., 2018). It supports that consumer technology including wearables should require rigorous testing against current gold standards and be FDA-cleared if the device or application is intended to render a diagnosis and/or treatment. We agree with this AASM position statement. At the time of this work, the Arc has not obtained FDA clearance, and therefore, should not replace existing clinical diagnostic procedure in the diagnosis of sleep conditions. However, we think that this work is a step forward in examining and validating a consumer wearable and provides supporting evidence for the Arc as an inexpensive

actigraphy tool for sleep research. Concomitant validation of the Actiwatch and of the Arc consumer-grade device against overnight polysomnography will be an important next step to determine full equivalence.

Conclusions

The Arc, a consumer wearable device, can be used as an actigraph for sleep monitoring and is able to produce sleep parameters that are comparable to a research-grade actigraph.

Abbreviations

OSA: Obstructive sleep apnea; PSG: Polysomnography; ROC: Receiver operating characteristic; TST: Total sleep time; WASO: Wake after sleep onset

Acknowledgements

We thank Ms. Eileen Leary for assistance in writing the IRB protocol. We thank Ms. Kary Newman and Ms. Polina Davidenko for assisting in patient recruitment and data collection. This work was supported by the Stanford Clinical and Translational Science Award to Spectrum (UL1 TR001085) and Stanford University Department of Psychiatry Small Grant Program. JC is supported by NIH K23NS101094.

Funding

This work was supported by Stanford Clinical and Translational Science Award to Spectrum (UL1 TR001085), and a Stanford University Department of Psychiatry Small Grant Program. JC is supported by NIH NINDS K23NS101094. The device maker Huami, Inc. was not involved in the design, conduct, funding, or analysis of the research, nor were they involved in the writing of this manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

JC, JMZ and EM designed and analyzed the study. HL assisted in data analyses. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All procedures were pre-approved by the Stanford University Institutional Review Board (IRB #36071). All participants signed a consent form prior to the initiation of the study procedures.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. None of the authors have any financial holdings or relationships with the device maker Huami, Inc.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Stanford University Center for Sleep Sciences and Medicine, Palo Alto, CA, USA. ²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. ³Mental Illness Research Education and Clinical Center, VA Palo Alto Health Care System, Palo Alto, CA, USA.

Received: 15 May 2018 Accepted: 24 July 2018

Published online: 02 September 2018

References

Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of Actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003;26(3):342–92.

- Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. *Metabolism Elsevier Inc.* 2017:1–10. <https://doi.org/10.1016/j.metabol.2017.10.008>.
- Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep.* 1992;15(5):461–9.
- de Zambotti M, Godino JG, Baker FC, Cheung J, Patrick K, Colrain IM. The boom in wearable technology: cause for alarm or just what is needed to better understand sleep? *Sleep.* 2016;39(9):1761–2. <https://doi.org/10.5665/sleep.6108>.
- Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD. Sleep detection with an accelerometer Actigraph: comparisons with polysomnography. *Physiol Behav.* 2001;72(1–2):21–8. [https://doi.org/10.1016/S0031-9384\(00\)00355-3](https://doi.org/10.1016/S0031-9384(00)00355-3).
- Khosla S, Deak MC, Gault D, Goldstein CA, Hwang D, Kwon Y, et al. Consumer sleep technology: an American Academy of sleep medicine position statement. *J Clin Sleep Med.* 2018;14(5):877–80. <https://doi.org/10.5664/jcsm.7128>
- Kripke DF, Hahn EK, Grizas AP, Wadiak KH, Loving RT, Steven Poceta J, Shadan FF, Cronin JW, Kline LE. Wrist Actigraphic scoring for sleep laboratory patients: algorithm development. *J Sleep Res.* 2010;19(4):612–9. <https://doi.org/10.1111/j.1365-2869.2010.00835.x>.
- Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of Actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med.* 2001;2(5):389–96. [https://doi.org/10.1016/S1389-9457\(00\)00098-8](https://doi.org/10.1016/S1389-9457(00)00098-8).
- Lee H-A, Lee H-J, Moon J-H, Lee T, Kim M-G, In H, Cho C-H, Kim L. Comparison of wearable activity tracker with Actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. *Psychiatry Investigation.* 2017;14(2):179. <https://doi.org/10.4306/pi.2017.14.2.179>.
- Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V, Boehlecke B, Brown T, et al. Practice parameters for the use of Actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep.* 2007;30:519–29.
- Roomkham S, Lovell D, Cheung J, Perrin D. Promises and Challenges in the Use of Consumer-Grade Devices for Sleep Monitoring. *IEEE Reviews in Biomedical Engineering.* 2018;XX(X):1–1. <https://doi.org/10.1109/RBME.2018.2811735>.
- Sadeh A, Lavie P, Scher A, Tirosh E, Epstein R. Actigraphic home-monitoring sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns. *Pediatrics.* 1991;87(4):494–9.
- Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. *Sleep.* 1982;5(4):389–99. <https://doi.org/10.1093/sleep/5.4.389>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

