HHS Public Access

Author manuscript

Med Sci Sports Exerc. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

Med Sci Sports Exerc. 2019 July ; 51(7): 1538–1557. doi:10.1249/MSS.0000000000001947.

Wearable Sleep Technology in Clinical and Research Settings

Massimiliano de Zambotti[1], **Nicola Cellini**[2], **Aimee Goldstone**[1], **Ian M Colrain**[1,3], and **Fiona C Baker**[1,4]

[1]Center for Health Sciences, SRI International, Menlo Park, CA, US

[2]Department of General Psychology, University of Padova, Padova, Italy

[3]Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia

[4]Brain Function Research Group, School of Physiology, University of the Witwatersrand, Johannesburg, South Africa

Abstract

The accurate assessment of sleep is critical to better understand and evaluate its role in health and disease. The boom in wearable technology is part of the digital health revolution and is producing many novel, highly sophisticated and relatively inexpensive consumer devices collecting data from multiple sensors and claiming to extract information about users' behaviors, including sleep. These devices are now able to capture different bio-signals for determining, for example, heart rate and its variability, skin conductance, and temperature, in addition to activity. They perform 24/7, generating overwhelmingly large datasets (Big Data), with the potential of offering an unprecedented window on users' health. Unfortunately, little guidance exists within and outside the scientific sleep community for their use, leading to confusion and controversy about their validity and application. The current state-of-the-art review aims to highlight use, validation and utility of consumer wearable sleep-trackers in clinical practice and research. Guidelines for a standardized assessment of device performance is deemed necessary, and several critical factors (proprietary algorithms, device malfunction, firmware updates) need to be considered before using these devices in clinical and sleep research protocols. Ultimately, wearable sleep technology holds promise for advancing understanding of sleep health, however, a careful path forward needs to be navigated, understanding the benefits and pitfalls of this technology as applied in sleep research and clinical sleep medicine.

Keywords

Wearables; Polysomnography; Validation; Actigraphy; Digital health; Sleep

Corresponding author: Massimiliano de Zambotti, SRI International, 333 Ravenswood Avenue, Menlo Park, CA, 94025, Tel: (650) 859-2714; Fax: (650) 859-2743; massimiliano.dezambotti@sri.com; maxdeze@gmail.com.

Conflict of interest. Authors declare no conflict of interest. MdZ, FB, and IC have received research funding unrelated to this work from Ebb Therapeutics Inc., Fitbit Inc., and International Flavors & Fragrances Inc.

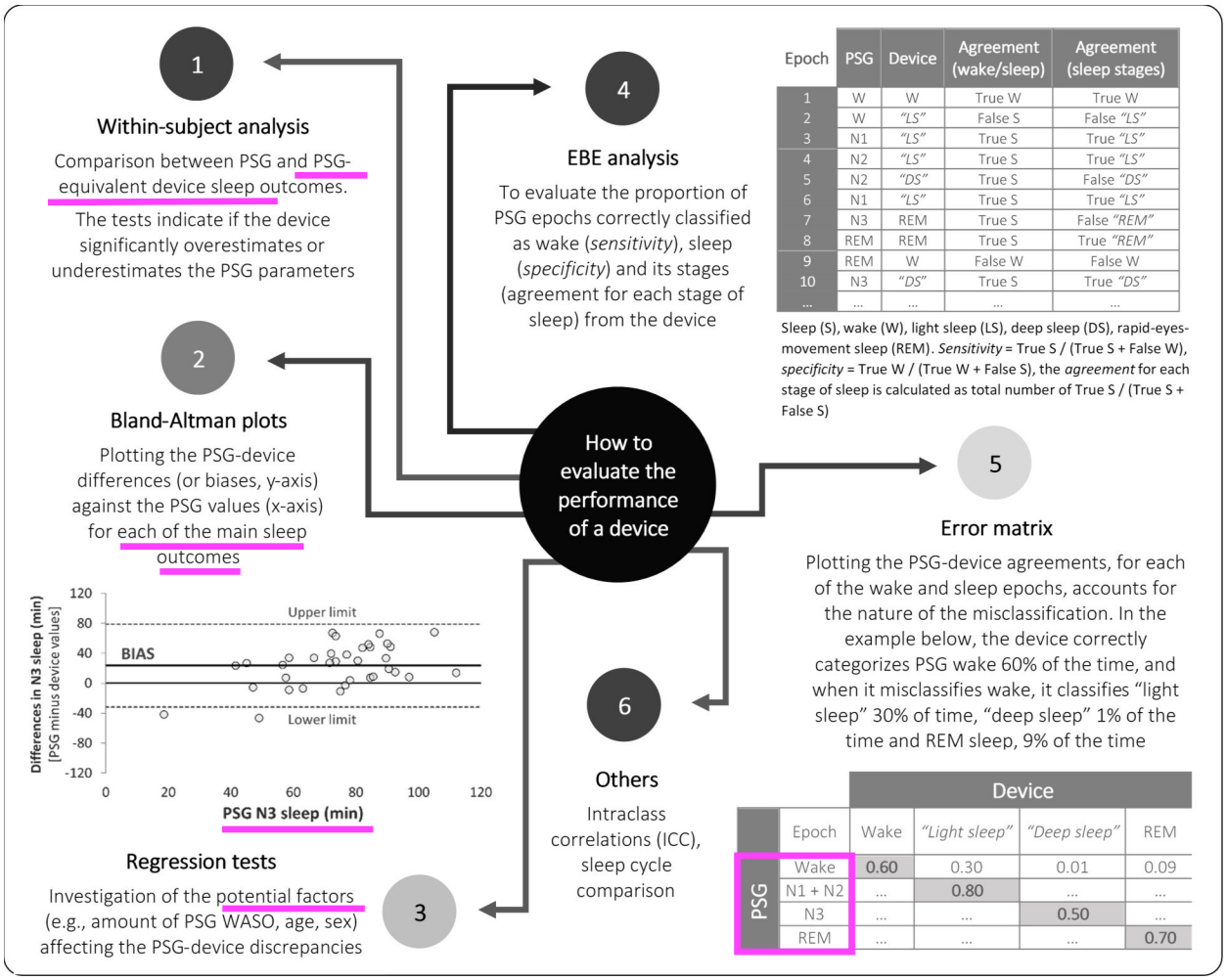

Figure 1.

Recommendations for the analysis and evaluation of the performance of a consumer wearable sleep tracker against polysomnography (PSG). EBE, epoch-by-epoch; WASO, wake after sleep onset

does not return sleep stages under various conditions. These include cases where the heartbeat signal (and hence the heart rate variability) is not cleanly detected throughout the night, if the total sleep duration is less than three hours, or if the battery runs out of power during the sleeping period". These criteria are based on different factors including a test of the integrity and amount of data they collect, which is not accessible to us. Thus, even when a sleep outcome is provided, we do not know specifically how much "reliable" information is used to provide that value.

7. The potential role of sleep wearables in clinical sleep disorders, intervention delivery and patient monitoring

Although the gold standard to evaluate the presence of sleep disorders is PSG, actigraphy has been commonly used in clinical practice to provide additional characterization of individuals with sleep disorders and to assess their treatment response (see 80). Nevertheless, so far only a few motion-based (first generation) consumer wearables have been tested in patients with clinical sleep disorders.

Two studies targeted children and adolescents with sleep disordered breathing (SDB). Meltzer et al. (48) showed that discrepancies between PSG and Fitbit Ultra changed as a function of SDB status and device sensitivity settings ("*normal*" or "*sensitive*"). Specifically, the study showed that despite Fitbit Ultra "*normal*" setting overestimated PSG TST and underestimated PSG WASO in both children with or without OSA, the PSG-device discrepancies were greater in mild OSA and further exacerbated in children with moderate/severe OSA. The authors also reported that most of the participants were outside the a priori-set "clinically satisfactory ranges" (i.e., TST <30 min and SE < 5%; see above for concerns about the use of these agreement limits). A reverse pattern was observed for the "*sensitive*" setting, characterized by greater PSG-device discrepancies in the no OSA category (TST underestimation and WASO overestimation), which progressively lessened in mild OSA and moderate/severe OSA categories (see Table 1 for details). Toon et al. (55) tested the Jawbone UP and showed no differences in PSG-Jawbone UP discrepancies in estimating TST, WASO, or SE as a function of SDB severity (i.e., primary snoring, mild or moderate-severe OSA). Moreover, the authors observed from the Bland-Altman plots that the Jawbone UP sleep outcomes were more consistent with PSG measures than were Actiwatch 2-PSG outcomes. Nevertheless, similar to Metzer et al. (48), the majority of the participants fell outside a priori-set "clinically satisfactory ranges". The authors indicated that, on the one hand, the Jawbone UP should be used as a diagnostic tool with caution; on the other hand, they observed that the Jawbone UP performance was, overall, similar to the Actiwatch 2.

Few studies have evaluated device performances in individuals with insomnia. Kang et at. (51) reported an overall good performance of the Fitbit Flex in the "*normal*" mode for good sleepers (no significant PSG-device differences for SOL, WASO, and SE, fair to excellent ICCs, and the majority of the participants fell inside the "satisfactory clinical agreement limits"). However, the Fitbit Flex showed more difficulties to assess sleep in the insomnia group. Specifically, the Fitbit Flex significantly overestimated PSG TST, SE and underestimate WASO in the insomnia group. Moreover, only 39.4% of the sample fell within

characterization of the relation (function) between demographics (e.g., age and sex) and PSG-device biases on a group level (see 28), could be used to adjust device outcomes at the individual level.

9. Conclusion

Sleep is fundamental for health (106). About one-third of the population is struggling with their sleep, a number that is estimated to increase. In our 24/7 sleepless society, sleep wearables may have a key role to better characterize and understand sleep and, within the framework of precision medicine, to ultimately improve health, safety and well-being for individuals and society. Collection of continuous data, day and night, could also lead to better understanding of links between sleep and daytime behaviors such as exercise.

Wearable sleep trackers are being increasingly adopted by both the general public and sleep researchers and clinicians. The second generation of multisensory sleep trackers opens a path for greater accuracy in measuring sleep, as compared to the motion-based approach to sleep/wake assessment. However, the proven theoretical advantage of the multisensory approach to sleep staging needs further empirical validation. Currently, these devices should be used cautiously, and interpretation of their outcomes should be carefully considered to avoid generating large inaccurate datasets leading to potential misleading scientific conclusions, assessment of sleep disturbances, and therapeutic decisions.

Further work is needed to investigate the potential use and performance, pros and cons, and limitations of these novel sleep trackers, particularly in sleep disorder populations. Keeping in mind the differential and overlapping motivations of various end-users' groups (e.g., research and clinical sleep community, wearable industry, consumers), partnership with industry is beneficial to combine excellence and speed in technological advancement from industry and advanced psychophysiological knowledge and scientific rigor from sleep science.

10. Acknowledgments

The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by the American College of Sports Medicine (ACSM).

Funding. This work was supported by the National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA; U01 AA021696 to IMC and FCB) and by the National Heart, Lung, and Blood Institute (NHLBI; R01HL139652 to MdZ). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. NC was supported by the University of Padua under the STARS Grants program, and the present work was carried out within the scope of the research program "*Dipartimenti di Eccellenza*", which is supported by a grant from MIUR to the Department of General Psychology, University of Padua.

11. References

1. Food and Drug Administration (FDA). Fostering Medical Innovation: A Plan for Digital Health Devices; Software Precertification Pilot Program. Avaliable at: https://www.fda.gov/MedicalDevices/DigitalHealth/UCM567265.

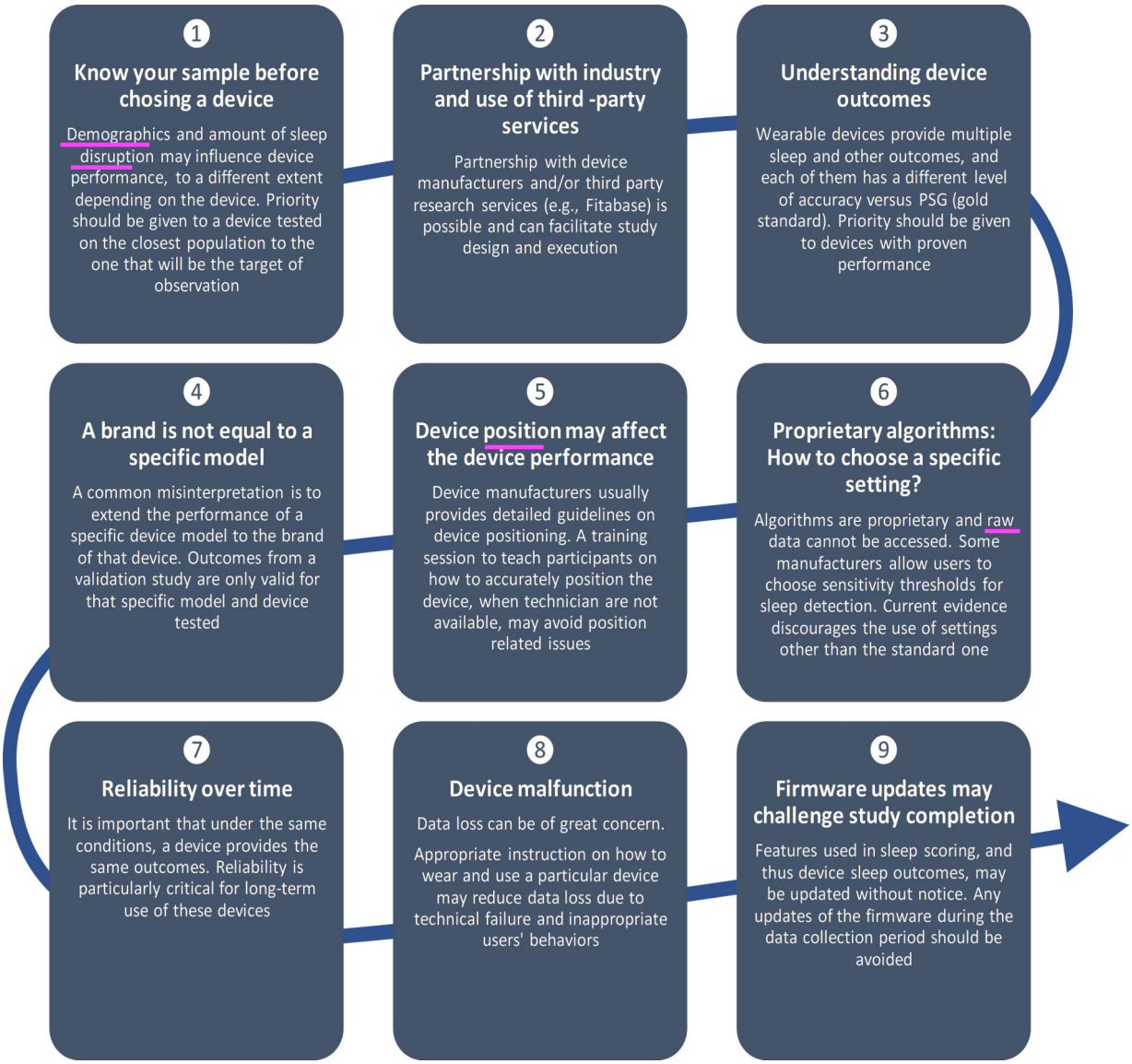

Figure 2.

Critical factors to consider when evaluating the potential use of a consumer wearable sleep trackers in research and clinical sleep settings

Table 1

Peer-reviewed journal articles evaluating the performance of wearable sleep trackers against standard polysomnography (PSG). Results about comparisons between overnight summary sleep outcomes from wearables (or actigraphy, when available) and PSG (PSG-device biases) are reported. When available, results from epoch-by-epoch (EBE) analysis are reported. Sample characteristic, type of devices, and amount of PSG sleep disruption are also provided for each study to allow better interpretation of the study results. When not specified, wearables data were collected using the default (normal) setting, and PSG records were scored in 30-s epochs.

							PSG- device biases		EBE analysis	
Study	**Authors**	**Sample characteristics**	**Age Range (or mean and SD)**	**Standard Actigraphy type**	**Wearable Device type**	**PSG SE (group mean)**	**Standard Actigraphy (mean, and SD of the biases when available)**	**Wearable Device (mean, and SD of the biases when available)**	**Standard Actigraphy (group mean)**	**Wearable Device (group mean)**
2012*In-lab*	Montgomery-Downs et al. (47)	24 healthy adults (10 female)	19 – 41 y	Actiwatch-64 (Minimitter, Inc.)	Fitbit "original" (Fitbit, Inc.)	< 85 %	Overestimated PSG SE (9.3 ± 9.7%) and TST (43.0 ± 46.6min)	Overestimated PSG SE (14.5 ± 10.7%) and TST (67.1 ± 51.3 min)	Sensitivity: 0.96 Specificity: 0.39	Sensitivity: 0.98 Specificity: 0.20
2015*In-lab*	Meltzer et al. (48)	63 children (32 female). 23% of the sample had 1.5 ≥ AHI ≤ 5 (mild OSA), and 16% of the sample had AHI > 5 (moderate OSA). The analyses were conducted on 49 children due to several device failures	3 – 17 y	AMI Motionlogger (Ambulatory Monitoring, Inc.) or Actiwatch Spectrum (Phillips Respironics) (analyses were conducted on sub-groups of 12 children for devices)	Fitbit Ultra (Fitbit, Inc.) using both "normal" and "sensitive" settings	< 85 %	No significant PSG-actigraphy biases	Overestimated PSG TST by 41 min and SE by 8%, and underestimated WASO by 32 min, using the "normal" setting. Discrepancies > 100 min for TST and WASO and > 20% using the "sensitive" setting. With increasing AHI (as well as with increasing in developmental age), the mean PSG-device discrepancies increased using the "normal" and decreased using the "sensitive" settings	No direct comparison with PSG was performed	Sensitivity of 0.87 for the "normal", and of 0.70 for the "sensitive' setting. Specificity of 0.52 for the "normal", and of 0.70 for the "sensitive" setting
2015*In-lab*	Toon et al. (55)	78 children (27 female). 41% of the sample had 1 > RDI ≤ 5 (mild OSA), 28% of the sample had RDI > 5 (moderate OSA), 6% had PLMI > 5, 31% had a diagnosis of primary snoring. In addition, 51% of the sample had other comorbidities (e.g., chronic inflammation, behavioral disorders) and 29% were under medication (e.g., methylphenidate)	3 – 18 y	Actiwatch 2 (Phillips Respironics)	Jawbone UP	< 85 %	Underestimated PSG SOL by an average of 21 min	No significant PSG-device biases. However, Jawbone UP underestimated PSG SOL in those participants with primary snoring (mean difference of 9.7 min). Also, biases for TST and SE changes from underestimating to overestimating, across developmental age. Differently, the bias for WASO changed from overestimating to underestimating, across developmental age	Sensitivity: 0.93 Specificity: 0.63	Sensitivity: 0.92 Specificity: 0.66
2015*In-lab*	de Zambotti et al. (28)	65 healthy adolescents (28 female)	12 – 22 y	-	Jawbone UP (Jawbone Inc.)	> 90 %	-	Overestimated PSG TST (10.0 ± 20.5 min) and SE (1.9 ± 4.2 %), and underestimated WASO (10.6 ± 14.7 min)	-	-
2015*In-lab*	de Zambotti et al. (56)	28 midlife women (12 of them meeting the DSM-IV criteria for insomnia)	44 – 60 y	-	Jawbone UP (Jawbone Inc.)	< 85 %	-	Overestimated PSG TST (26.6 ± 35.3 min) and SOL (5.2 ± 9.6 min), and underestimated WASO (31.2 ± 32.3 min). No differences in device performance according to disease status	-	Sensitivity: 0.96 Specificity: 0.37 No differences in device performance according to disease status
2016*At-home*	Mantua et al. (49)	40 healthy adults (19 female)	18 – 30 y	Actiwatch Spectrum (Phillips Respironics)	Basis Health (Intel, Corp.), Fitbit Flex (Fitbit, Inc.) Misfit Shine (Misfit, Inc.), Withings Pulse 02 (Withings, Inc.)	< 85 %	No significant PSG-actigraphy biases for TST and SE	Overestimation of PSG TST for both Misfit Shine (~ 75 min for the bias) and Withings Pulse O2 (~ 12 min for the bias), which also overestimated PSG SE with a bias > 5 %; Basis Health underestimated SE with a bias > 10 %. We decided not to report results for sleep staging due to the unusual aggregation of PSG N3 + REM, considered as "deep" sleep	-	-
2016*In-lab*	de Zambotti et al. (33)	32 healthy adolescents (15 female)	12 – 21 y	-	Fitbit Charge HR (Fitbit, Inc.)	> 90 %	-	Overestimated PSG TST (8.0 ± 21.0 min) and SE (1.8 ± 4.5 %), and underestimated PSG WASO (5.6 ± 14.3 min)	-	Sensitivity: 0.97 Specificity: 0.42

							PSG- device biases		EBE analysis	
Study	**Authors**	**Sample characteristics**	**Age Range (or mean and SD)**	**Standard Actigraphy type**	**Wearable Device type**	**PSG SE (group mean)**	**Standard Actigraphy (mean, and SD of the biases when available)**	**Wearable Device (mean, and SD of the biases when available)**	**Standard Actigraphy (group mean)**	**Wearable Device (group mean)**
2017*In-lab*	Cook et al. (50)	21 unmedicated adults (17 female) with DSM-IV major depressive disorder	26.5 ± 4.6 y	Actiwatch 2 (Phillips Respironics)	Fitbit Flex (Fitbit, Inc.) using both "normal" and "sensitive" settings	< 85 %	Overestimated PSG TST (by 40.6 min) and SE (by 7.0 %), and underestimated SOL (by 13.5 min) and WASO (by 27.1 min)	Overestimated PSG TST (by an average of 46.0 min) and SE (by an average of 8.1 %), and underestimated WASO (by an average of 44.0 min) in the "normal" setting. Wide PSG-device biases (> 60 min for TST and WASO, and > 15 % for SE) for the "sensitive" setting	Sensitivity: 0.97 Specificity: 0.31	Sensitivity of 0.98 for the "normal", and of 0.78 for the "sensitive" setting. Specificity of 0.35 for the "normal", and of 0.80 for the "sensitive" setting
2017*At-home*	Kang et a. (51)	33 drug-free individuals with (19 female) and 17 without (11 female) DSM-5 insomnia disorder	18 – 60 y	Actiwatch 2 (Phillips Respironics)	Fitbit Flex (Fitbit, Inc.) using both "normal" and "sensitive" settings	< 85 % in insomniac > 90 % in controls	Underestimated PSG TST (by an average of 17.8 min) and SE (by an average of 4.8 %) in controls. Underestimated PSG WASO (by an average of 21.6 min) in individuals with insomnia	Overestimated PSG TST (by an average of 6.5 min), using the "normal" setting in controls. Overestimated PSG TST (by an average of 32.9 min) and SE (by an average of 7.9%), and underestimated WASO (by an average of by 30.5 min), using the "normal" setting in individuals with insomnia. No data were provided for the "sensitive" setting	Sensitivity: 0.95 in controls and 0.96 in insomniacs Specificity: 0.61 in controls and 0.45 in insomniacs	Sensitivity of 0.97 (0.97 in insomniacs) for the "normal" and of 0.65 (0.64 in insomniacs) for the "sensitive" setting, in controls. Specificity of 0.36 (0.36 in insomniacs) for the "normal" and of 0.82 (0.89 in insomniacs) for the "sensitive" setting, in controls
2017*In-lab*	Maskevich et al. (52)	7 participants (6 female) carrying Huntington's gene, with disease severity ranging from presymptomatic (N = 4) to early symptomatic (N = 3)	54.1 ± 6.4 y	Actiwatch Spectrum Pro (Phillips Respironics)	Jawbone UP2 (Jawbone Inc.), and Fitbit One (Fitbit, Inc.)	Not provided. Sleep was scored in 1 min epochs	Overestimated PST TST (by 74.0 ± 54.4 min) and SE (by 14.8 ± 11.0 %)	Both Jawbone UP2 and Fitbit One overestimated PSG TST (by > 60 min) and SE (by > 15 %), and underestimated WASO (by > 30 min)	Sensitivity: 0.97 Specificity: 0.31	Sensitivity of 0.99 for both Jawbone UP2 and Fitbit One. Specificity of 0.34 for Jawbone UP2 and of 0.27 for Fitbit One
2017*At-home*	Gruwez et al. (57)[a]	15 healthy adults Demographics unclear for the final sample analyzed	18 – 40 y	SenseWear Pro (BodyMedia, Inc.)	Jawbone UP MOVE (Jawbone Inc.), and Withings Pulse 02 (Withings Inc.)	> 90 %	No significant PSG-actigraphy biases	Withings Pulse 02 overestimated PSG TST (by an average of 33 min), TIB (by an average of 16 min), and SE (by an average of 5 %). No significant biases were found for Jawbone UP MOVE. We decided to do not report results for sleep staging due to the unclear classification of "light" and "deep" sleep from the device manufacturers.	-	-
2017*In-lab*	de Zambotti et al. (43)	42 healthy adolescents (13 female)	14 – 22 y	-	ŌURA ring (Ouraring, Inc.)	> 90 %	-	Underestimated PSG N3 (19.6 ± 41.2 min) and overestimated REM (–17.2 ± 50.2 min)	-	Sensitivity of 0.96, and specificity of 0.48. Agreements for N1+N2 of 0.65, for N3 of 0.51, and for REM of 0.61
2018*In-lab*	de Zambotti et al. (42)	44 healthy adults (26 female). Separate analyses on 9 with PSG evidences of PLMS > 15/h	19 – 61 y	-	Fitbit Charge 2 (Fitbit Inc.)	> 85 % in both groups	-	Overestimated PSG TST (9 ± 24 min) and N1 + N2 (34 ± 34 min), and underestimated SOL (4 ± 9 min) and N3 (24 ± 28 min), in the main group. Underestimated PSG N3 (28 ± 35 min) in the PLMS group	-	Sensitivity of 0.96 for the main, and 0.95 for the PLMS group. Specificity of 0.61 for the main, and 0.62 for the PLMS group. Agreement for N1+N2 of 0.81 for the main, and 0.78 for the PLMS group. Agreement for N3 of 0.49 for the main, and 0.36 for the PLMS group. Agreement for REM of 0.74 for the main, and 0.62 for the PLMS group
2018*In-lab*	Sargent et at. (60)[b]	12 healthy elite athletes (sex not specified)	18.3 ± 1.0 y	-	Fitbit Charge HR (Fitbit, Inc.)	Not provided	-	No significant PSG-device biases for TST for Fitbit TST obtained in "automatic mode" for night-time sleep, as well as when bed timing was manually adjusted to match the bedtime opportunities for both night-time and day-	-	-

Study	Authors	Sample characteristics	Age Range (or mean and SD)	Standard Actigraphy type	Wearable Device type	PSG SE (group mean)	PSG- device biases		EBE analysis	
							Standard Actigraphy (mean, and SD of the biases when available)	Wearable Device (mean, and SD of the biases when available)	Standard Actigraphy (group mean)	Wearable Device (group mean)
								time sleep (see notes below about the protocol). However, PSG-device discrepancies in TST in "automatic mode" was > 240 min in 4 participants. Lack of details on how data were obtained and analysis performed (please refer to the original study (60))		
2018*At-home*	Pesonen and Kuula (59)[c]	17 healthy children (9 female) and 17 healthy adolescents (8 female)	9 – 13 y children, 14 – 20 y adolescents	Actiwatch 2 (Phillips Respironics)	Polar A370™ (Polar Electro, Inc.)	> 95 %	Overestimated PSG WASO in children (by 20.9 min) and adolescents (by 14.3 min). Underestimated PSG TST in children (by 43.6 min) and adolescents (by 26.8 min).	Overestimated PSG WASO in children (by 24.4 min) and adolescents (by 12.5 min). Underestimated PSG TST in children (by 28.9 min) and adolescents (by 20.6 min)	Sensitivity: 0.93 in children and 0.93 in adolescents Specificity: 0.68 in children and 0.58 in adolescents	Sensitivity: 0.93 in children and 0.91 in adolescents Specificity: 0.77 in children and 0.83 in adolescents
2018*At-home*	Liang and Martell (53)[d]	25 healthy young adults (10 women)	24.8 ± 4.4 y	-	Fitbit Charge 2 (Fitbit, Inc.)	> 85 %	-	Underestimated PSG TST (by 12.3 min), SOL (by 11.1 min), N1 + N2 sleep (by 42.4 min), REM sleep (by 11.6 min), and % N3 sleep (by 10.2 %). Overestimated PSG WASO (by 24.5 min) and % of WASO (by 6,5 %), % of N1 + N2 sleep (by 13.8 %), % of REM sleep (by 4.6 %), N3 sleep (by 39.8 min)	-	-
2018*In-lab*	Cook et al. (58)	43 clinical adult patients (29 females): 3 with a diagnosis of narcolepsy, 13 with idiopathic hypersomnia, 17 with idiopathic hypersomnia not otherwise specified, 6 with mild obstructive sleep apnea, and 4 with hypersomnolence related to another condition	33.3 ± 1.0 y	Actiwatch 2 (Phillips Respironics)	Jawbone UP3 (Jawbone Inc.)	> 85 %	Overestimated PSG TST (by 43.9 min) and SE (by 7.5%). Underestimated PSG SOL (by 12.9 min) and WASO (by 33.9 min).	Overestimated PSG TST (by 39.6 min) and SE (by 6.8%). Underestimated PSG SOL (by 5.1 min) and WASO (by 34.3 min)	Sensitivity: 0.97 Specificity: 0.31	Sensitivity: 0.97 Specificity: 0.39 Agreements for N1+N2 of 0.60, for N3 of 0.49, and for REM of 0.30
2018*In-lab*	Cook et al. (54)	49 adult patients (46 females) with suspected central disorders of hypersomnolence: 14 with idiopathic hypersomnia, 19 with idiopathic hypersomnia not otherwise specified/ unspecified, 6 with narcolepsy and 10 with mixed diagnoses	30.3 ± 9.8 y	-	Fitbit Alta HR (Fitbit, Inc.)	> 85 %	-	Overestimation of PSG TST (by 11.6 min), SE (by 2%) and N3 sleep (by 18.2 min)	-	Sensitivity: 0.96 Specificity: 0.58 Agreements for N1+N2 of 0.73, for N3 of 0.67, and for REM of 0.74 *No significant differences in sensitivity, specificity and agreements for sleep stages among sub-groups*

[a], unclear is how the time in bed (**TIB**) for the at-home PSG assessment was determined

[b], the experimental design of the study included 3 fixed night-time (10pm-7am; 11pm-7am; 12am-7am) and 2 fixed day-time (2pm-4pm; 3pm-4pm) "sleep opportunities" over three days. Analyses were based on averaged periods for night-time and day-time

[c], authors reported using a pre-product Polar fitness tracker corresponding to the commercially available Polar A370; authors calculated SE as TST/time between sleep onset and offset*100. Thus, the percentage does not account for the wake time between lights-off and sleep onset

[d], the comparison between Fitbit Charge 2 and PSG is based on a PSG portable clinical 1-channel EEG device (sleep stages were automatically analyzed in 30-s epochs and visually checked); **AHI**, apnea-hypopnea index (average number of apnea and hypopnea episodes per hour of sleep); **PLMI**, periodic limb movement index; **RDI**, respiratory disturbance index (average number of apnea and hypopnea episodes, and respiratory event-related arousals per hour of sleep); **REM**, rapid-eye-movement; **SOL**, sleep onset latency; **SE**, sleep efficiency; **TST**, total sleep time; **WASO**, wake after sleep onset

Table 2

An example of a qualitative approach to evaluate device performance.

An alternative way to evaluate the performance of a device is whether it adequately (compared to PSG) captures a significant literature effect (e.g., a group difference in sleep architecture between healthy individuals and those with a sleep disorder, sleep recovery after cognitive-behavioral treatments in insomnia sufferers, sleep alterations following acute stress-inducing experimental manipulation). For example, similarly to PSG, we previously showed that a multisensory sleep tracker (the ŌURA ring) was able to significantly detect the age-related decline in N3 sleep in an adolescence sample. This finding is encouraging given that the device showed its greatest limitation in PSG N3 classification (51% agreement in detecting PSG N3 sleep) (43).

Table 3

Assuming that the proprietary algorithms used by consumer wearables will remain proprietary, what else can the wearable industry do to facilitate the use of consumer wearable sleep-trackers in clinical and sleep research settings?

Open access to raw data	Allows application of publicly available algorithms to wearable raw accelerometer data (and/or plethysmography derived IBIs) obtaining a standardized sleep stage classification
Allow the choice of a specific version of the proprietary algorithm used for sleep classification when exporting/extracting sleep data	Allows consistency for data collection within a study period, by avoiding uncontrollable algorithm updates that may affect sleep parameter calculations Also allows researchers to choose a specific wearable device model using a specific algorithm with proven validation
Have a separate line of products more aligned with research and clinical needs	Would remove many concerns of using an uncontrolled consumer product for research and clinical sleep assessment
Increase partnership with sleep research and clinical centers	Allows access to domain expertise in basic sleep science and clinical sleep disorders, which can lead to consistent use of accepted terminology, and insight into the meaning and value of Big Data