Expert Review of Medical Devices

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ierd20

Feasibility of a wearable biosensor device to characterize exercise and sleep in neurology residents

Kellyann Niotis, Nabeel Saif, Marialaura Simonetto, Xian Wu, Peter Yan, Jessica P. Lakis, Ingrid Estrada Ariza, Adam P. Buckholz, Nivita Sharma, Matthew E. Fink & Richard S. Isaacson

To cite this article: Kellyann Niotis, Nabeel Saif, Marialaura Simonetto, Xian Wu, Peter Yan, Jessica P. Lakis, Ingrid Estrada Ariza, Adam P. Buckholz, Nivita Sharma, Matthew E. Fink & Richard S. Isaacson (2021) Feasibility of a wearable biosensor device to characterize exercise and sleep in neurology residents, Expert Review of Medical Devices, 18:11, 1123-1131, DOI: 10.1080/17434440.2021.1990038

To link to this article: https://doi.org/10.1080/17434440.2021.1990038

Published online: 22 Oct 2021.

EXPERT REVIEW OF MEDICAL DEVICES
2021, VOL. 18, NO. 11, 1123–1131
https://doi.org/10.1080/17434440.2021.1990038

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

Feasibility of a wearable biosensor device to characterize exercise and sleep in neurology residents

Kellyann Niotis[a], Nabeel Saif[b], Marialaura Simonetto[c], Xian Wu[d], Peter Yan[e], Jessica P. Lakis[f], Ingrid Estrada Ariza[b], Adam P. Buckholz[g], Nivita Sharma[h], Matthew E. Fink[b] and Richard S. Isaacson[b]

[a]2019–2020 McGraw Fellow in Neurology Research; Department of Neurology, Weill Cornell Medicine and New York-Presbyterian, New York, NY, USA; [b]Department of Neurology, Weill Cornell Medicine, New York, NY, USA; [c]Departments of Internal Medicine and Neurology, Weill Cornell Medicine and New York-Presbyterian, New York, NY, USA; [d]Division of Biostatistics and Epidemiology, Weill Cornell Medicine and Department of Healthcare Policy & Research, New York-Presbyterian, New York, NY, USA; [e]Department of Neurology, Beth Israel Deaconess Hospital-Milton Center for Specialty Care, Milton, MA, USA; [f]Office of Development, New York-Presbyterian, New York, NY, USA; [g]Department of Internal Medicine, Weill Cornell Medicine and New York-Presbyterian, New York, NY, USA; [h]Weill Cornell Medical College, New York, NY, USA

ABSTRACT

Background: Research suggests optimizing sleep, exercise and work-life balance may improve resident physician burnout. Wearable biosensors may allow residents to detect and correct poor sleep and exercise habits before burnout develops. Our objectives were to evaluate the feasibility of a wearable biosensor to characterize exercise/sleep in neurology residents and examine its relationship to self-reported, validated survey measures. We also assessed the device's impact on well-being and barriers to use.

Methods: This prospective cohort study evaluated the WHOOP Strap 2.0 in neurology residents. Participants completed regular online surveys, including self-reported hours of sleep/exercise, and validated sleep/exercise scales at 3-month intervals. Autonomic, exercise, and sleep measures were obtained from WHOOP. Changes were evaluated over time via linear regression. Survey and WHOOP metrics were compared using Pearson correlations.

Results: Sixteen (72.7%) of 22 eligible participants enrolled. Eleven (68.8%) met the minimum usage requirement (6+ months) and were classified as 'consecutive wearers.' Significant increases were found in sleep duration and exercise intensity. Moderate-to-low correlations were found between survey responses and WHOOP measures. Most (73%) participants reported a positive impact on well-being. Barriers to use included 'Forgetting to wear' (20%) and 'not motivational' (23.3%).

Conclusion: Wearable biosensors may be a feasible tool to evaluate sleep/exercise in residents.

ARTICLE HISTORY
Received 25 June 2020
Accepted 4 October 2021

KEYWORDS
Activity trackers; wearables; biosensors; graduate medical education; neurology; resident wellness; WHOOP

1. Introduction

Resident physician burnout is a complex issue exacerbated by long duty hours, demanding work environments and various psychosocial stressors. In a recent survey, 45.2% of trainees reported burnout and neurology residents ranked 4th among the most burned out [1]. Burnout has been associated with depression, anxiety, and adverse patient outcomes [2]. This has led to an increasing prevalence of wellness initiatives among medical training programs. However, their effectiveness has rarely been evaluated [3–5].

While the development of physician burnout is multifactorial, research suggests that optimizing sleep, exercise, and work-life balance improves well-being [4,5]. However, most studies on the effects of these factors on burnout rely on self-reported measures [6]. With the increasing proliferation of wearable biosensors, real-time collection of sleep and autonomic measures has become possible. Such devices may allow trainees to conveniently detect and correct poor sleep and autonomic patterns before burnout develops. Wearable devices also have the potential to replace self-reported data with more objective measures of sleep, cardiovascular exertion, and autonomic regulation. With the ongoing popularity of wearable devices, emerging technologies in advanced biosensing methods, such as skin-printed sensors and self-powered stretchable strain sensors, may only serve to increase the capability and accuracy of these devices in the future [7,8].

The WHOOP Strap 2.0 (WHOOP Inc., Boston, USA), a wearable biosensor initially developed for elite athletes, uses multiple validated sensors to continuously record biometrics data [9,10]. While most consumer wearable devices record sleep duration and heart rate (HR), WHOOP has several unique features that make it suitable for investigation in medical trainees. First, it is one of the few commercially available wearables that internally computes heart rate variability (HRV), a validated metric for autonomic resiliency that is influenced by stress [11–14]. Therefore, HRV may potentially serve as a quantitative surrogate for residents' stress. In addition, the evidence suggests that increased awareness of HRV with biofeedback training improves stress and anxiety [15]. Second, WHOOP uses algorithms to analyze biometrics and provide users with individualized assessments of cardiovascular exertion, sleep quality, and sleep

CONTACT Richard S. Isaacson rii9004@med.cornell.edu Department of Neurology, Weill Cornell Medicine/New York-Presbyterian, 428 East 72th Street, Suite 400, New York, NY 10021, USA

debt over time. Third, it may serve as an active wellness intervention by providing personalized sleep and exercise recommendations via a smartphone application. Further, our research team had experience using the device in a prior feasibility study in an outpatient clinical research environment that included some subjects in similar age ranges as the current study [16].

Several studies have used wearable devices to evaluate sleep, exercise, and well-being in medical trainees, but only two of these employed the WHOOP [3,17–21]. Both studies focused solely on sleep metrics in orthopedic surgery residents over a brief 4-week duration [22,23]. To our knowledge, this is the first study to explore the utility of WHOOP in evaluating sleep, exercise, autonomic measures, and well-being of medical trainees over a 12-month period.

The primary objective of this study was to examine the feasibility of WHOOP to longitudinally characterize autonomic, exercise, and sleep patterns in neurology resident physicians. Since the reliability of self-reported sleep and exercise is uncertain, we also examined the correlation between measurements recorded by WHOOP, surveys and validated sleep and exercise scales. To better understand the utility of WHOOP as a wellness intervention tool, we assessed the impact on well-being and barriers to continued use.

2. Methods

2.1. Study overview

This was a prospective cohort study involving both data collection from WHOOP and participant-completed questionnaires. Data collection occurred from March 2018 through June 2019. Blinded usernames for WHOOP accounts and e-mails for study correspondence and surveys were used. Participants were informed that study data could not be used to inappropriately influence employment or educational decisions made by the training program. Institutional Review Board approval was obtained, and each participant provided informed consent to participate in the study (Protocol # 1,707,018,393). Study procedures are outlined in Figure 1.

2.2. Study population

Participants consisted of postgraduate year (PGY) physicians from the New York-Presbyterian/Weill Cornell Neurology Residency Program in New York City. Eligibility criteria were as follows: (1) PGY-1 to PGY-3 categorical residents, (2) willingness to complete at least 80% of surveys, (3) willingness to wear WHOOP, and (4) owning a smart phone. Eligible residents were invited to participate via e-mail [24,25].

2.3. Description of WHOOP device

The WHOOP Strap 2.0 is a water resistant wrist-worn device that contains a 3-axis accelerometer, photo-plethysmography (PPG) heart rate monitor, capacitive touch sensor and ambient temperature sensor that continuously records data. WHOOP 2.0 has a battery life of approximately 30 hours and may be charged while wearing.

Like many newer wearables, WHOOP monitors sleep with an accelerometer, a validated tool that demonstrates sleep tracking performance comparable to traditional actinography [26–28]. Sleep stages are automatically detected using a PPG-monitor, which has been validated against traditional polysomnography [29]. Recent studies have validated the accuracy of WHOOP-measured total nightly sleep time (TST), rapid eye movement (REM) sleep and slow wave sleep (SWS) against polysomnography [7,8]. In addition, its PPG-HR monitor has been validated against electrocardiography [7,28]. In a comparison study evaluating the accuracy of nine popular wearables (including WHOOP 2.0, FitBit Ionic, Apple Watch Series 3, and Garmin Vivo Smart 4) to record TST, the Oura Smart Ring followed by WHOOP 2.0 were shown to be the most accurate. The other devices in this study were found to underestimate TST [30].

WHOOP calculates several other exercise and autonomic metrics. Using PPG, WHOOP internally computes HRV using a validated method (the root mean square of successive differences between heartbeats during SWS) [31]. Cardiovascular strain score (Strain), measured on a 21-point scale, is an individualized metric accounting for total cardiovascular exertion and time spent in one's maximum HR zone. WHOOP records a workout both manually (by the wearer in the smartphone application) and automatically (if HR is elevated for at least 15 minutes and earns a Strain of at least 8.0). Based on these metrics, WHOOP sends wearers notifications with individualized recommendations regarding sleep need, sleep consistency and cardiovascular exertion [32].

Throughout the study period, resting heart rate (RHR), HRV, Strain, total workouts per 24-h period (TotWO), workout strain (WO Strain) and TST in hours were electronically transmitted for each participant through the anonymized WHOOP cloud platform. Participants had access to real-time data via the WHOOP smartphone application.

2.4. Study procedures

Upon enrollment, participants completed a baseline survey collecting demographics, as well as self-reported average weekly hours of exercise and sleep. The baseline survey included two validated sleep and exercise scales, the Rapid Assessment of Physical Activity (RAPA), and the Epworth Sleepiness Scale (ESS) [31,33]. Participants were provided user instructions and a WHOOP Strap 2.0 to wear for a goal of 12 months Similar surveys were completed at 3 month intervals. At 12 months, a final survey was administered, which included additional questions evaluating attitudes toward WHOOP, impact on well-being and barriers to continued use. E-Mail reminders were sent to encourage compliance with surveys and usage of the device.

2.5. Outcome measures

The primary objective was to evaluate the feasibility of using WHOOP to longitudinally monitor residents' data. This outcome was measured by the mean length of time participants wore the device. Secondary outcome measures included (1)

change in WHOOP-measured TST, (2) change in objective exercise frequency (TotWOs) and intensity (Strain and WO

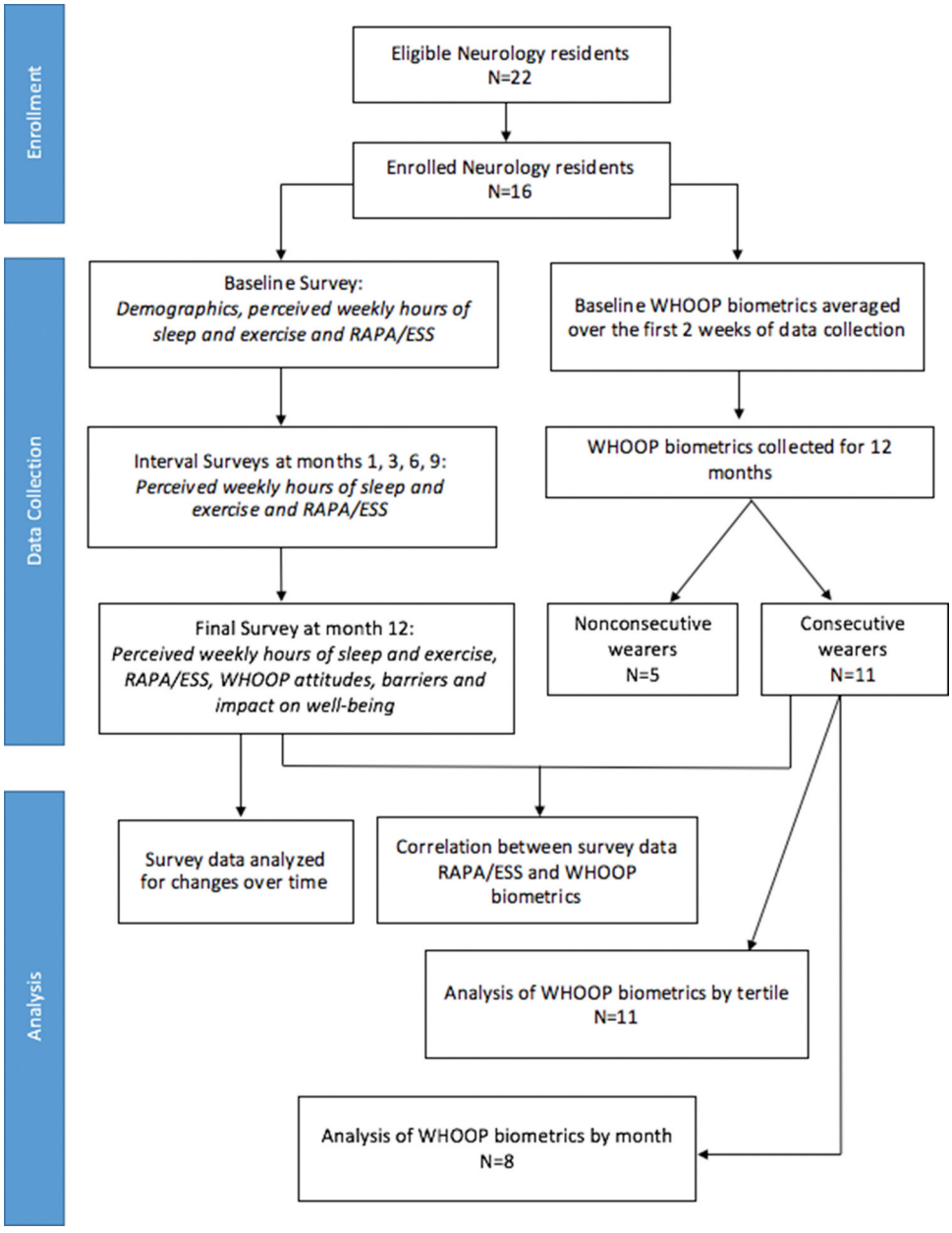

Figure 1. Study procedures outline.

Strain), (3) change in RHR recorded during SWS, (4) change in HRV, (5) change in self-reported TST and exercise (average weekly number of workouts and minutes of exercise), (6) change in ESS and RAPA scores, (7) correlation between WHOOP-measured sleep and exercise metrics, self-reported measures, and validated ESS and RAPA scales, (8) self-reported barriers to use, (9) impact on well-being, and (10) attitudes toward WHOOP measured via Likert Scales.

2.6. Statistical methods

Baseline WHOOP measures were calculated for each participant by averaging the first 2 weeks of data. For the purpose of comparing baseline characteristics, participants were divided *a priori* into two groups: consecutive wearers (wore WHOOP for ≥6 months) and nonconsecutive wearers (<6 months). Unpaired t-tests were used to analyze for differences in demographics and baseline biometrics between wearer groups.

In consecutive wearers, individual and cohort WHOOP-collected data were analyzed over time in two ways: (1) change by month (including participants who wore WHOOP for 12 months) and (2) change by tertile of time spent in the study (including all consecutive wearers).

For monthly analysis, WHOOP metrics were averaged every month. The qualifying month of participation was one in which there were no more than 5 days of missing data. To maintain homogeneity given the variable durations of participation, analyses were also performed by averaging WHOOP metrics within tertiles of time each participant wore the device. For example, if a participant wore WHOOP for 12 months, each tertile would consist of 4 months of data, whereas if a participant wore WHOOP for 9 months, each

tertile would consist of 3 months of data. Both month-to-month and tertile-to-tertile changes within the cohort were evaluated using linear regression.

Survey data were compared at months 0, 1, 3, 6, 9 and 12 using one-way analysis of variance (ANOVA). Pearson correlation coefficients were used to compare survey data to corresponding WHOOP metrics.

Given the small sample size of this feasibility study, no power calculation was conducted. P-values <0.05 were considered to be statistically significant. Analyses were conducted in SAS 9.4 (SAS Institute, NC, USA) and GraphPad Prism 8.0.0 (GraphPad Software, CA, USA).

Data can be made available via request to the corresponding author.

3. Results

Of 22 eligible participants, 16 (72.7%) enrolled. The mean time of wearing WHOOP was 7.9 months, with eight (50%) participants wearing WHOOP for 12 months. There were no significant differences in baseline demographics between consecutive and nonconsecutive wearers. Compared to nonconsecutive wearers, consecutive wearers had significantly higher baseline HRV (57.2 ms vs. 45.3 ms, $p < 0.001$), Strain (8.9 pts. vs. 6.7 pts., $p < 0.001$), and WO Strain (7.7 pts. vs. 6.9 pts., $p = 0.002$). Nonconsecutive wearers had longer TST (7.4 h vs. 6.2 h, $p = 0.02$) (Table 1).

3.1. WHOOP measures

Table 2 summarizes changes in WHOOP-recorded metrics on an individual basis and across the cohort. At 12 months, the cohort had significantly increased mean TST by 29 minutes (slope = 0.04, 95% CI 0.01 to 0.08 hrs., $p = 0.03$), Strain by 1.9 points (slope = 0.16, 95% CI 0.06 to 0.26 points, $p = 0.004$), and WO Strain by 4.1 points (slope = 0.34, 95% CI 0.23 to 0.45 points, $p < 0.001$) (Figure 2(a, c and e)). RHR significantly decreased by a mean of 2.4 bpm (slope = −0.20, 95% CI. −0.38 to −0.02 bpm, $p = 0.04$) (Figure 2g). No significant changes were found in HRV (slope = 0.30, 95% CI −0.45 to 1.05, $p = 0.39$) and TotWOs (slope = 0.05, 95% CI −0.02 to 0.11, $p = 0.12$).

In the last tertile, the cohort significantly increased mean TST by 20 minutes (slope = 0.10, 95% CI 0.08 to 0.12 hrs., $p = 0.002$), Strain by 0.9 points (slope = 0.46, 95% CI 0.40 to 0.52 pts., $p = 0.001$) and WO Strain by 1.9 points (slope = 0.93, 95% CI 0.45 to 1.41 pts., $p = 0.01$) (Figure 2(b,d and f)). There was a trend toward decreased RHR (slope = −0.91, 95% CI −2.82 to 1.02 bpm, $p = 0.18$) (Figure 2(h)). No significant changes were found in HRV (slope = 0.26, 95% CI −6.44 to 6.96 ms, $p = 0.88$) or TotWO (slope = 0.04, 95% CI −0.45 to 0.57, $p = 0.39$).

Table 1. Cohort demographics and WHOOP baseline biometrics.

Demographics				
	Overall	Consecutive wearers	Nonconsecutive wearers	
Characteristics	(N = 16)	(N = 11)	(N = 5)	*P-value*
Age, y; mean (SD)	27.3(1.7)	27.6 (1.6)	26.6 (1.8)	0.31
Sex, n (%)				0.30
Women	10 (62.5)	8 (72.7)	2 (40.0)	
Men	6 (37.5)	3 (27.3)	3 (60.0)	
Ethnicity, n (%)				0.77
Caucasian	9 (56.2)	6 (54.5)	3 (60.0)	
African American	1 (6.2)	1 (9.1)	0 (0)	
Latino	1 (6.2)	0 (0)	1 (20.0)	
Asian	4 (25.0)	3 (27.3)	1 (20.0)	
Middle Eastern	1 (6.2)	1 (9.1)	0 (0)	
PGY, n (%)				0.77
PGY1	4 (25.0)	2 (18.2)	2 (40.0)	
PGY2	9 (56.2)	7 (63.6)	2 (40.0)	
PGY3	3 (18.8)	2 (18.2)	1 (20.0)	
Months worn, mean (SD)	7.9 (4.9)	10.9 (2.0)	1.4 (0.5)	

WHOOP Baseline Biometrics

Biometrics, mean (SD)	**Consecutive wearers** (N = 11)	**Nonconsecutive wearers** (N = 5)	*P-value*
RHR (bpm)	63.9 (7.1)	63.8 (6.5)	0.96
HRV (ms)	57.2 (38.5)	45.3 (4.1)	<0.001
Strain	8.9 (1.9)	6.7 (1.7)	<0.001
TotWO (no.)	0.3 (0.3)	0.1 (0.1)	0.17
WO Strain	7.7 (3.8)	6.9 (2.7)	0.002
TST (hours)	6.2 (1.1)	7.4 (1.2)	0.02

Table 2. Changes in WHOOP-recorded biometrics by month and by tertile.

PARTICIPANTS WITH SIGNIFICANT CHANGES IN BIOMETRICS (No., % of total)		
	By 12 months (N = 8)	By Tertile (N = 11)
Increased Strain	4 (50%)	6 (54.5%)
Increased WO Strain	4 (50%)	6 (54.5%)
Increased TotWO	4 (50%)	5 (45.5%)
Increased TST	3 (37.5%)	7 (63.6%)
Decreased RHR	5 (62.5%)	5 (45.5%)
Increased HRV	4 (50%)	4 (36.4%)

COHORT CHANGES IN BIOMETRICS

	By 12 months (n = 8)			By Tertile (n = 11)		
	Slope	95% CI	*P-value*	Slope	95% CI	*P-value*
Strain	0.16	0.06 to 0.26 pts.	0.004	0.46	0.40 to 0.52 pts.	0.001
WO Strain	0.34	0.23 to 0.45 pts.	<0.001	0.93	0.45 to 1.41 pts.	0.01
TotWO	0.05	−0.02 to 0.11	0.12	0.04	−0.45 to 0.57	0.39
TST	0.04	0.01 to 0.08 hrs.	0.03	0.10	0.08 to 0.12 hrs.	0.002
RHR	−0.20	−0.38 to −0.02 bpm	0.04	−0.91	−2.82 to 1.02 bpm	0.18
HRV	0.30	−0.45 to 1.05 ms	0.39	0.26	−6.44 to 6.96 ms	0.88

3.2. Survey responses

Interval survey response rate was 87.3%. There were no differences in survey measures from baseline to any follow-up periods (Table 3). No strong correlations were found between WHOOP-measured and self-reported survey data (Table 3).

Final survey response rate was 93.8%. The majority (72.7%) of consecutive wearers reported WHOOP had a positive impact on their wellness, most notably by increasing exercise frequency and/or intensity. Most consecutive wearers (63.6%) also reported that they would continue to wear WHOOP after the study period. Despite increased WHOOP-measured sleep,

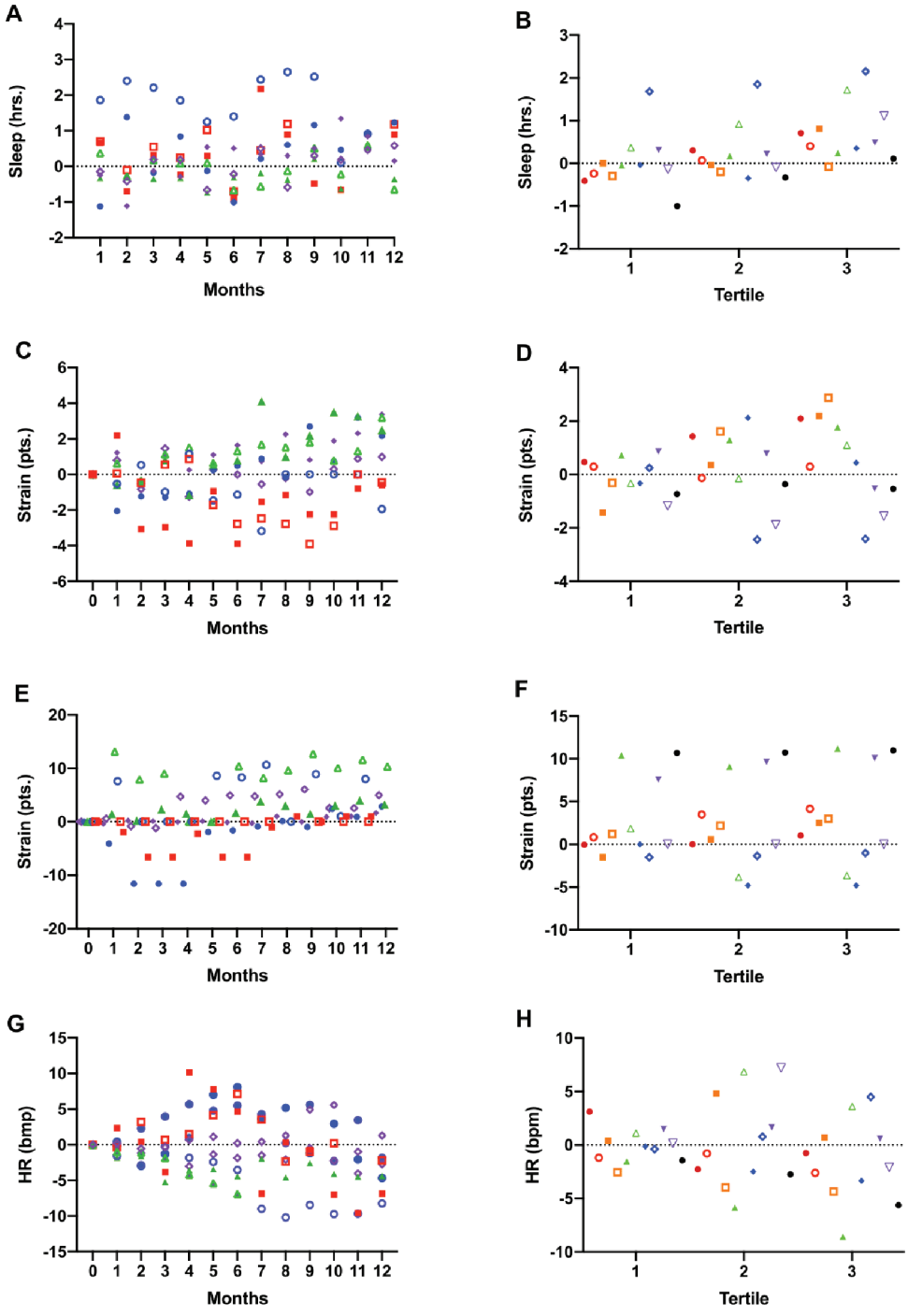

Figure 2. Biometric Changes from Baseline Over Time Panels A, C, E and G: Each colored symbol represents an individual's monthly change from baseline for the biometric of interest over 12 months. N = 8. Panels B, D, F and H: Each colored symbol represents an individual's change from baseline for the biometric of interest over tertiles. N = 11. (a) TST over 12 months‡ (b) TST over tertiles‡ (c) Strain over 12 months‡ (d) Strain over tertiles‡ (e) WO Strain over 12 months‡ (f) WO Strain over tertiles‡ (g) RHR over 12 months‡ (h) RHR over tertiles ‡Denotes statistically significant change in the cohort's mean metrics.

participants reported WHOOP had neutral (47%) or little-to-no effect (33%) on sleep duration or bedtime (Figure 3(a)). Forgot to wear" (23.3%), 'not motivational' (20%), and having a 'short battery life' (16.7%) were the most cited limitations of WHOOP (Figure 3(b)). Two participants dropped out of the study due to adverse events, including the development of a self-diagnosed 'ulnar neuropathy' (n = 1) and an 'allergic reaction' (n = 1).

4. Discussion

Our study suggests that a wearable device may be a feasible way to longitudinally characterize sleep, exercise, and autonomic regulation in resident physician trainees. While prior studies have used wearables to measure the potentially harmful impact of residency on trainee sleep and physical activity, our study is the first to use such a device in an attempt to optimize sleep, exercise, and autonomic regulation, which

Table 3. Self-reported survey data over time and correlation between WHOOP-measured and self-reported survey data using Pearson correlation coefficient.

SELF-REPORTED SURVEY MEASURES							
Survey measures, mean (SD)	Pre (n = 16)	1mo (n = 11)	3mo (n = 11)	6mo (n = 11)	9mo (n = 9)	12mo (n = 15)	*P-value*
RAPA	4 (2)	5 (2)	5 (2)	5 (2)	6 (1)	5 (2)	0.24
ESS	8 (4)	8 (5)	8 (6)	8 (5)	6 (4)	7 (4)	0.34
Average sleep (h)	6 (1)	6 (1)	6 (1)	6 (1)	6 (1)	7 (1)	0.86
Average exercise (h)	4 (2)	4 (3)	4 (2)	4 (1)	5 (1)	4 (2)	0.96

CORRELATION BETWEEN WHOOP-RECORDED AND SELF-REPORTED MEASURES (n = 11)				
	Strain	WO Strain	TotWO (no.)	TST (hrs.)
RAPA	0.35[b]	0.45[a]	0.42[a]	
ESS				−0.49[a]
Average sleep (h)				0.38[b]
Average exercise (h)	0.20[b]	0.52[a]	0.43[a]	

a = moderate correlation (0.40–0.59); b = weak correlation (0.20–0.39).

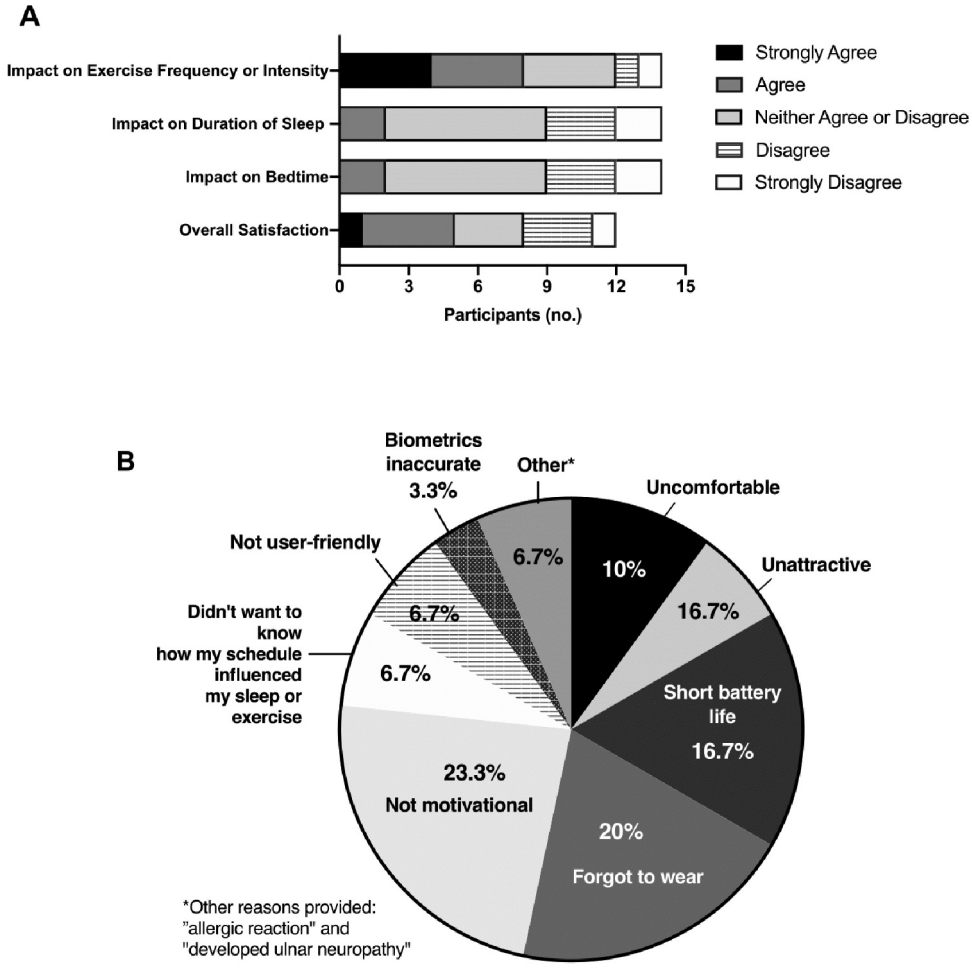

Figure 3. Quality measures.

research suggests can reduce burnout [4]. Further, our study assessed the WHOOP device over a 12 month period, considerably longer than prior studies with study periods of 14 days to 6 months [3,17–21].

The observed baseline biometric differences between consecutive and nonconsecutive wearers may provide insight into who may benefit most from wearables such as WHOOP. For example, we found that consecutive wearers slept fewer hours at baseline, which may imply that sleep-deprived residents were more likely to commit to wearing the device more regularly as a motivating force to correct poor sleep habits. Additionally, we found consecutive wearers to be more physically active at baseline, which may suggest an inherent value and prioritization of exercise in this group. Further investigation is needed to identify which subpopulations of residents are best served by this type of wellness intervention.

Prior studies have reported on the wide intra- and inter-individual variability of HRV and sleep duration in healthy

adults using wearable technology [11,13,14,33,34]. Additionally, Quer *et al.* reported that an individual's RHR is fairly consistent over time but differs from others by up to 70 bpm [35]. To remove a component of between-person variability, Figure 2 depicts participants' changes in biometrics from their individual baseline.

Due to the wide interindividual variability, examining changes on an individual basis rather than using cohort averages may provide additional insights. For example, our sample has a wide baseline range for HRV (range: 22.4 to 133.9 ms), which may explain the lack of significant cohort-wide changes for this metric. However, examination on an individual basis demonstrates that six (75%) participants at 12 months and seven (63.6%) participants at the final tertile increased their HRV, which suggests the development of more robust autonomic regulation.

Interestingly, we found a discrepancy in the change in RHR between monthly and tertile analysis. While we found that the cohort's RHR significantly decreased at 12 months, this change was not significant via tertile analysis. One possible explanation for this discrepancy is that the additional participants included in the tertile analysis had fewer TotWOs and lower Strain and WO Strain scores. Therefore, these participants were less likely to undergo cardiac conditioning as endurance training contributes to improvements in RHR [34]. Another consideration is that these participants may not have had sufficient time to significantly change the autonomic measures [36].

Due to the nature of this feasibility study, our study is limited in its ability to provide in-depth analysis of sleep and exercise patterns. Differences in the types of exercises participants chose to engage in may have contributed to discrepancies in RHR, HRV, Strain, and WO Strain since endurance training has unique cardiovascular effects over resistance and weight training, which are known to have more concentrated effects on skeletal muscle physiology [34]. Future studies should include a comprehensive evaluation of sleep stages (e.g. REM, deep) and exercise types (e.g. cardiovascular, strength training).

The moderate-to-low correlation between survey and WHOOP measures aligns with previous findings demonstrating poor correlation between subjective and objective sleep and exercise measures [37,38]. Recall bias may at least partially explain this finding. However, the validated RAPA/ESS scores and WHOOP measures also demonstrated moderate-to-weak correlations. While this implies caution should be taken when using these scales in medical trainees, it is important to consider the potential inaccuracy of some WHOOP measurements (e.g. Strain scores). These measurements are internally computed by WHOOP's 'proprietary algorithms' and are not available to the public or the investigators. As such, it is unclear exactly how these scores are calculated, and there are currently no published studies that demonstrate these measures are valid markers for cardiac exertion. Thus, the interpretation of these correlations is limited. Further study is needed to compare various wearable devices and their correlation to self-reported sleep and exercise measures in this population.

Our study has several limitations, including the aforementioned small sample size. Additionally, our study did not account for scheduling factors, such as night shifts and rotations with extended work hours, which have been showed to decrease HRV and disrupt circadian rhythm in health-care workers [14,17]. Since the life of a physician trainee has unique challenges and highly unpredictable stressors, future studies comparing metrics from age-matched controls in other professions would be useful.

There are further limitations related to aspects of the WHOOP device itself. In addition to the aforementioned uncertain validity of the cardiovascular exertion scores, TotWO may be underestimated because some activities (e.g. weightlifting, yoga) might not generate the sustained HR elevation/Strain to trigger auto-detection of a workout. Additionally, WHOOP requires users to remember to charge it, and participants' failure to do so led to inconsistencies in the number of days eligible for data analyzation per month between participants. This aligns with participants citing 'forgot to wear' and 'short battery life' as top barriers to use. One user reported the development of an 'ulnar neuropathy' (made by a self-diagnosis) which resolved after no longer using the device. One 'allergic reaction' was characterized as skin redness where the sensor touched the epidermis as another reason for study discontinuation. While these were self-reported adverse effects and not clinically diagnosed by a clinician, it is plausible that wearing a tight wrist strap imposes direct pressure on a peripheral nerve to cause a neuropathy, and on the skin resulting in adverse outcomes [39]. Additionally, studies have reported that prolonged use of wearable fitness trackers is associated with the development of a skin rash [40,41]. Newer trackers, including WHOOP Strap 3.0 which was released after completion of this study, have improved battery life. Addressing battery life, education on skin hygiene (e.g. cleaning the device and skin each day) and other reasons for discontinuation may better optimize the use of these devices in future studies.

Finally, as with all prospective studies using wearable devices, it is unclear whether the positive behavioral changes we observed were driven by WHOOP itself or the Hawthorne effect. Retrospective studies using data collected from wearables are limited, but one study may imply that these devices facilitate behavioral change that is independent of active engagement with the study. In a large cohort of more than 90,000 adult FitBit wearers, RHR decreased by an average of 2 bpm over the course of 1 year, a decrease that was similarly found in our cohort (2.4 bpm) [35]. While we observed improvements in measures that support WHOOP's utility as a wellness intervention for residents, further studies are needed to demonstrate long-term device-driven behavioral change.

5. Conclusion

Our study aimed to evaluate the feasibility of the WHOOP biosensors to longitudinally monitor sleep, exercise, and autonomic measures in neurology residents. While our study was limited by its sample size and variable participation, we reported significant improvements in our cohort's TST, RHR, Strain, and WO Strain. Given the importance of these metrics in maintaining a healthy lifestyle, our data support the feasibility of using these devices as a wellness intervention in select

resident groups. Nevertheless, future studies employing various devices across multiple academic institutions with a larger cohort are warranted to draw definitive conclusions.

Funding

This research was funded by the Department of Neurology at Weill Cornell Medicine and philanthropic support of grateful patients of the Alzheimer's Prevention Clinic, Weill Cornell Memory Disorders Program.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions statement

Name	Location	Role	Contribution
Kellyann Niotis MD	Weill Cornell Medicine and New York Presbyterian Hospital, New York	PGY-4 Resident, 2019-2020 McGraw Fellow in Neurology Research	Conceptualization, Data Curation, Investigation, Methodology, Statistical analysis, data interpretation, writing original draft and critical editing
Nabeel Saif MS	Weill Cornell Medicine, New York	Research coordinator	Investigation, Methodology, data interpretation, critical revision of the manuscript, project administrator
Marialaura Simonetto MD MS	Weill Cornell Medicine and New York Presbyterian Hospital, New York	PGY-1 Resident	Conceptualization, Investigation, Methodology, preformed statistical analysis, data interpretation, critical revision of manuscript
Xian Wu MPH	Weill Cornell Medicine, New York	Research Biostatistician	Investigation, Methodology, preformed statistical analysis
Peter Yan MD MS	Beth Israel Deaconess Hospital-Milton Center for Specialty Care, Milton, Ma	Associate Professor	Conceptualization, Investigation, Methodology, data interpretation, critical revision of manuscript
Jessica P. Lakis BS	New York Presbyterian Hospital, New York	Research Associate	Investigation, Methodology, data interpretation, writing original draft and critical editing
Ingrid Estrada Ariza MD	Weill Cornell Medicine	Research Associate	Investigation, Methodology, data interpretation, critical revision of manuscript
Adam P. Buckholz MD	Weill Cornell Medicine and New York Presbyterian Hospital, New York	PGY-5 Fellow	Investigation, Methodology, data interpretation, critical revision of manuscript
Matthew E. Fink MD	Weill Cornell Medicine and New York Presbyterian Hospital, New York	Department Chair	Conceptualization, Methodology, Investigation, Funding and resources, data interpretation, critical editing of manuscript
Richard S. Isaacson MD	Weill Cornell Medicine and New York Presbyterian Hospital, New York	Corresponding author; associate program director	Conceptualization, Methodology, Investigation, Funding and resources, data interpretation, critical editing of manuscript

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Dyrbye LN, Burke SE, Hardeman RR, et al. Association of clinical specialty with symptoms of burnout and career choice regret among US resident physicians. J Am Med Assoc. 2018;320(11):1114–1130.
2. Koutsimani P, Montgomery A, Georganta K. The relationship between burnout, depression, and anxiety: a systematic review and meta-analysis. Front Psychol. 2019. DOI:10.3389/fpsyg.2019.00284
3. Poonja Z, O'Brien P, Cross E, et al. Sleep and exercise in emergency medicine residents: an observational pilot study exploring the utility of wearable activity monitors for monitoring wellness. Cureus. 2018;10(7). DOI:10.7759/cureus.2973
4. Raj KS. Well-being in residency: a systematic review. J Grad Med Educ. 2016;8(5):674–684.
5. Patel RS, Sekhri S, Bhimanadham NN, et al. A review on strategies to manage physician burnout. Cureus. 2019. DOI:10.7759/cureus.4805
6. Trockel MT, Menon NK, Rowe SG, et al. Assessment of physician sleep and wellness, burnout, and clinically significant medical errors. JAMA Network Open. 2020. DOI:10.1001/jamanetworkopen.2020.28111.
7. Liang H, He Y, Chen M, et al. Self-powered stretchable mechanoluminescent optical fiber strain sensor. Adv Intell Syst. 2021;2100035. DOI:10.1002/aisy.202100035.
8. Kao H-LC, Bedri A, Lyons K. SkinWire: fabricating a self-contained on-skin PCB for the hand. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;2(3):Article 116.
9. Berryhill S, Morton CJ, Dean A, et al. Effect of wearables on sleep in healthy individuals: a randomized cross-over trial and validation study. J Clin Sleep Med. 2020. DOI:10.5664/jcsm.8356.

10. Miller DJ, Lastella M, Scanlan AT, et al. A validation study of the WHOOP strap against polysomnography to assess sleep. J Sports Sci. 2020. DOI:10.1080/02640414.2020.1797448.
11. Goffeng EM, Nordby KC, Tarvainen MP, et al. Fluctuations in heart rate variability of health care workers during four consecutive extended work shifts and recovery during rest and sleep. Ind Health. 2018. DOI:10.2486/indhealth.2017-0100.
12. Robinson C, Lawless R, Zarzaur BL, et al. Physiologic stress among surgeons who take in-house call. Am J Surg. 2019;218 (6):1181–1184.
13. Nunan D, Sandercock GRH, Brodie DA. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. Pacing Clin Electrophysiol. 2010. DOI:10.1111/j.1540-8159.2010.02841.x
14. Berntson GG, Thomas Bigger J, Eckberg DL, et al. Heart rate variability: origins methods, and interpretive caveats. Psychophysiology. 1997. DOI:10.1111/j.1469-8986.1997.tb02140.x.
15. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. Psychol Med. 2017. DOI:10.1017/S0033291717001003
16. Isaacson RS, Hristov H, Saif N, et al. Individualized clinical management of patients at risk for Alzheimer's dementia. Alzheimer's Dement. 2019. DOI:10.1016/j.jalz.2019.08.198.
17. Marek AP, Nygaard RM, Liang ET, et al. The association between objectively-measured activity, sleep, call responsibilities, and burnout in a resident cohort. BMC Med Educ. 2019. DOI:10.1186/s12909-019-1592-0.
18. Feng S, Yi JS, Deitz G, et al. Relationships between sleep, activity, and burnout in ophthalmology residents. J Surg Educ. 2020. DOI:10.1016/j.jsurg.2020.09.003
19. Mendelsohn D, Despot I, Gooderham PA, et al. Impact of work hours and sleep on well-being and burnout for physicians-in-training: the resident activity tracker evaluation study. Med Educ. 2019;53(3):306–315.
20. Fargen KM, Spiotta AM, Turner RD, et al. Operation la sierra: a novel wellness initiative for neurological surgery residents. J Grad Med Educ. 2016. DOI:10.4300/JGME-D-15-00641.1
21. Schrager JD, Shayne P, Wolf S, et al. Assessing the influence of a fitbit physical activity monitor on the exercise practices of emergency medicine residents: a pilot study. JMIR mHealth uHealth. 2017;5(1):e2.
22. Sochacki KR, Dong D, Peterson LE, et al. The measurement of orthopaedic surgeon quality and quantity of sleep using a validated wearable device. JAAOS Glob Res Rev. 2018. DOI: 10.5435/jaaosglobal-d-18-00065.
23. Sochacki KR, Dong D, Peterson L, et al. Using a validated wearable device. Arthrosc Sport Med Rehabil. 2019;1(2):e115–e121.
24. Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. Metabolism. 2018. DOI:10.1016/j.metabol.2017.10.008
25. Lee HA, Lee HJ, Moon JH, et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. Psychiatry Investig. 2017. DOI:10.4306/pi.2017.14.2.179.
26. Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep. 2013. DOI:10.5665/sleep.3142.
27. Fonseca P, Weysen T, Goelema MS, et al. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. Sleep. 2017. DOI:10.1093/sleep/zsx097.
28. Georgiou K, Larentzakis AV, Khamis NN, et al. Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv). 2018. DOI:10.2478/folmed-2018-0012
29. Stone JD, Rentz LE, Forsey J, et al. Evaluations of commercial sleep technologies for objective monitoring during routine sleeping conditions. Nat Sci Sleep. 2020. DOI:10.2147/NSS.S270705.
30. Stone JD, Rentz LE, Forsey J, et al. Evaluations of commercial sleep technologies for objective monitoring during routine sleeping conditions. Nat Sci Sleep. 2020. doi:10.2147/NSS.S270705
•• **A study evaluating the accuracy of various commercially available wearable devices to measure sleep during in-home conditions.**
31. Topolski TD, LoGerfo J, Patrick DL, et al. The Rapid Assessment of Physical Activity (RAPA) among older adults. Prev Chronic Dis. 2006;3(4):1–8.
32. Johns MW. Reliability and factor analysis of the Epworth sleepiness scale. Sleep. 1992;15(4):376–381.
33. Fagard RH. A population-based study on the determinants of heart rate and heart rate variability in the frequency domain. Verh K Acad Geneeskd Belg. 2001;63(1):57-89.
34. Jaiswal SJ, Quer G, Galarnyk M, et al. Association of sleep duration and variability with body mass index: sleep measurements in a large US population of wearable sensor users. JAMA Intern Med. 2020. DOI:10.1001/jamainternmed.2020.2834
35. Quer G, Gouda P, Galarnyk M, et al. Inter- And intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: retrospective, longitudinal cohort study of 92,457 adults. PLoS One. 2020. DOI:10.1371/journal.pone.0227709
36. Borresen J, Lambert MI. Autonomic control of heart rate during and after exercise: measurements and implications for monitoring training status. Sport Med. 2008;38(8):633–646.
37. Prince SA, Adamo KB, Hamel ME, et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008. DOI:10.1186/1479-5868-5-56
38. Girschik J, Fritschi L, Heyworth J, et al. Validation of self-reported sleep against actigraphy. J Epidemiol. 2012. DOI:10.2188/jea.JE20120012
39. Helfenstein Júnior M. Uncommon compressive neuropathies of upper limbs. Best Pract Res Clin Rheumatol. 2020. DOI:10.1016/j.berh.2020.101516
40. Ráthonyi G, Ráthonyi-Odor K, Bendíková E, et al. Wearable activity trackers usage among university students. Eur J Contemp Educ. 2019. DOI: 10.13187/ejced.2019.3.600.
41. Turakhia MP, Desai M, Hedlin H, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the Apple Heart Study. Am Heart J. 2019. DOI:10.1016/j.ahj.2018.09.002.