medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Canadian Military Pain & Rehab Journal Dec 2023, Article 10

Accuracy, Utility and Applicability of the WHOOP Wearable Monitoring Device in Health, Wellness and Performance - a systematic review

Riham Khodr[1], Latifah Kamal[1], Amir Minerbi[2,3], Gaurav Gupta

Affiliations
1 University of Ottawa
2 Institute for Pain Medicine, Rambam Health Campus,
3 Faculty of Medicine, Technion Israel Institute of Technology

Abstract

Introduction: The WHOOP wearable device is gaining popularity in clinical and performance applications with the ability to track sleep and heart rate parameters and provide feedback on recovery and strain. With the claims on potential benefits, a critical evaluation of the underlying scientific literature and the accuracy of these devices is imperative.

Methods: Authors systematically reviewed studies examining the accuracy and clinical applications of the WHOOP device.

Results: The WHOOP appears to have acceptable accuracy for two-stage sleep and heart rate metrics, but depending on the study, room for improvement for four-stage sleep and heart rate variability identification. There are numerous preliminary studies looking at the WHOOP's ability to track and/or influence sleep and exercise behaviours at the cohort and/or population level. The impact of athletic performance and/or objective sleep is limited based on existing studies.

Discussion: The clinical application for the WHOOP, given the acceptable accuracy levels, continues to expand. Uses have included impact on sports performance, correlation with medical conditions (i.e. cognitive dysfunction), sleep and health behaviours in various populations. Limitations of existing accuracy trials include variable design and reporting metrics, while results from non-accuracy trials require further clinical validation for response rate and effect size.

Conclusion: The WHOOP wearable device has acceptable accuracy for sleep and cardiac variables to be used in clinical studies where a baseline can be established and, ideally, other clinical outcomes and gold standard tools can be employed.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

INTRODUCTION

Irrespective of the clinical or training setting, the optimal goal of monitoring physiological metrics is to optimize performance and wellness while preventing negative health outcomes (Kellmann, 2010), (Lundstrom, 2020). As technology evolves, so does the ability to provide non-invasive, portable and real-time feedback for sleep, heart variability (HRV) and activity levels (Plews et al., 2013, Prinsloo et al., 2014). The goal would be to use this data to inform decision-making to the benefit of the person tracking this information. For instance, high-performing groups (e.g., athletes, military, first responders) may benefit from automated feedback on physiological variables to direct activities and optimize preparedness and performance (Stone et al., 2020).

A variety of activity monitors are commercially available. Of these, the WHOOP wearable device (WHOOP Inc., Boston, MA) measures multiple variables and provides information on sleep, recovery and strain. Sleep quality (sleep disturbances and cycles), quantity (hours) and recovery (resting heart and rate heart rate variability) are used to provide the recovery score (i.e. ‘readiness to perform’). Strain uses continuous heart rate to measure energy expenditure and predict cardiovascular load during a given activity or throughout the day. (Lundstrom, 2020)

Although the WHOOP does not specifically measure energy metabolism, heart rate variability and sleep are sensitive to hemodynamic, endocrine, thermoregulatory, environmental, and psychological factors (A. Flatt et al., 2018; Lundstrom, 2020). In addition, training type, intensity and phase, baseline fitness, and body mass can also impact HRV and sleep metrics (Fortes et al., 2017); (Lundstrom, 2020) Despite more data being required, HRV and sleep appear to be correlated with reductions in athletic performance. (A. Flatt et al., 2018; A. A. Flatt & Howells, 2019; VanHeest et al., 2014; Woods et al., 2018, Woods et al., 2017, Reed et al., 2013, Sekiguchi et al., 2019, Bolin, 2019). Insufficient rest and sleep can impact cognitive and physical regeneration; therefore, identifying deficits could potentially impact outcomes (Dinges et al., 1997; Edwards & Waterhouse, 2009; Van Dongen et al., 2003; Vgontzas et al., 2004).

Therefore, using these tools to examine the autonomic system responsiveness to physiological stress could theoretically help examine the body’s ability to adapt to exercise stimulus and changes in health (C. R. Bellenger et al., 2016; Borresen & Lambert, 2008; Buchheit, 2014). While novel, the literature on the accuracy and clinical utility of these measures is still evolving. Studies on accuracy have looked at validation against gold standards, such as polysomnography (PSG) for sleep and electrocardiogram for heart rate variability (Berryhill et al., 2020). Other studies have looked at WHOOP use in performance, clinical prediction and treatment (Berryhill et al., 2020; Lundstrom, 2020).

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

While numerous limitations associated with "gold standard" measures of sleep and HRV (i.e. electrocardiograms, polysomnography, sleep surveys and diaries, actigraphy) exist, verification of commercially available activity trackers is still required (Depner et al., 2020; De Zambotti et al., 2019; Ibáñez et al., 2019; Kelly et al., 2012; Khosla et al., 2018; Stone et al., 2020). In this review, we look to systematically summarize the existing literature regarding the WHOOP wearable device concerning accuracy and applications.

METHODS

A systematic literature search was conducted on Google, Google Scholar and PubMed by 3 authors (RK, LK and GG) using keywords WHOOP, activity trackers, sleep tracking, heart rate variability, performance, clinical applications, validation and accuracy. Search strategy included "WHOOP" OR "activity trackers" OR "accuracy" OR "validation" OR "clinical applications" OR "sleep tracking" OR "heart rate variability" OR "performance" OR "sleep tracking".

Inclusion criteria were studies published in English, original research articles evaluating the accuracy of the WHOOP activity tracker, studies investigating the clinical applications or implications of using the WHOOP activity tracker in healthcare settings, studies conducted on human subjects of any age or health conditions and studies published from January 2000 to December 2023. Exclusion criteria were studies not written in English. reviews, meta-analyses, and opinion articles, studies not focusing on the WHOOP activity tracker, studies with insufficient data on accuracy or clinical applications and studies not conducted on human subjects.

We developed a standardized data extraction sheet to collect relevant information from each included study. This included fields for study characteristics (title, authors, publication year), study design, sample size, participant characteristics, WHOOP version/model, accuracy metrics, clinical outcomes, and any additional relevant information.

Given the number and heterogeneity of the studies, it was not possible to assess publication bias by means of any formal statistical tests.

A critical review was undertaken of these studies by the research team. Data was extracted from each article and prepared for a comparative analysis. Given the heterogeneity of the data, a meta-analysis of the data was not possible. Therefore, the analysis focused on a semi-quantitative and qualitative comparison looking at broad categories of accuracy and clinical/performance studies.

For accuracy studies, this review focused specifically on mean error measurements and not median error, limits of agreement, interquartile range or R^2. Data was extracted from the publication and supplementary data. For clarification or corrections, multiple primary authors were contacted with no response. Therefore, adjustments to

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

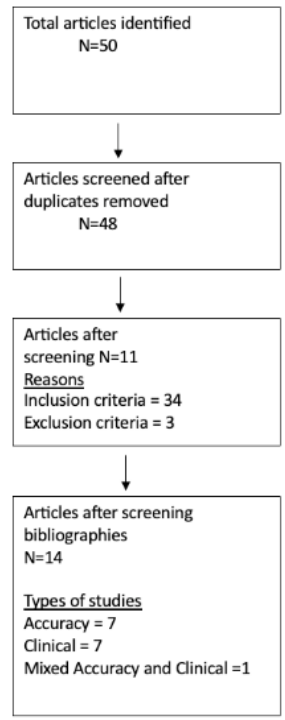
the presented data were made with consensus from the research team and, where possible, consistent with other data presented (Miller et al., 2020, Miller et al., 2021). Where manual and automatic detection were assessed separately, result ranges were displayed, and the absolute error was converted to percentage by dividing by the PSG measurements (Miller et al., 2021). Where only figures and no numbers were provided, calculations of mean absolute error were not possible (Grandner et al., 2023). The classification of bias for heart rate (HR) and HRV measures was done according to the following criteria proposed by Miller et al. 2022): 0.0–0.1, trivial; 0.1–0.3, small; 0.3–0.5, moderate; 0.5–0.7, large; 0.7–0.9, very large; 0.9–1.0, nearly perfect (C. Bellenger et al., 2021; Hopkins et al., 2009). Even with these guidelines, interpreting the sleep and HR/HRV outcomes was not easily possible from the available data in specific studies (Miller et al., 2022).

In addition, the values for bias did not correspond across all studies, nor did the reported analysis allow for standardization. Finally, commercial devices generally do not disclose their algorithms for detecting sleep states; therefore, light sleep was defined as the sum of N1 and N2, rapid eye movement (REM) and deep sleep (slow wave sleep) as N3 (Depner et al., 2020; Stone et al., 2020).

RESULTS

50 abstracts were identified, and 11 relevant studies were selected for further review. From these 11 articles, 4 additional studies were found in the references, for a total of 15 studies included in this review. See Figure 1 for flowchart.

Figure 1 Article Selection Flowchart



The results have been divided into studies looking at WHOOP accuracy validation and performance/clinical application.

Accuracy

Stone et al. showed a wide variation amongst the commercial devices tested, with error rates increasing for staging sleep (Stone et al., 2020). Miller et al. showed similar trends for sleep, wake and staging when comparing PSG to the WHOOP (Miller et al., 2020, 2022).

Bellenger et al. showed that WHOOP data required filtering to optimize ECG agreement, and correct interval

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

identification has a more significant impact on HRV than HR calculations (C. Bellenger et al., 2021). Logarithmic HRV transformation also appears to reduce percent bias and residuals, which in part explains its use in monitoring training status (C. Bellenger et al., 2021). Inaccuracies in WHOOP-derived slow wave sleep (SWS)/deep sleep identification may impact HR and HRV measures modulated at this sleep stage (Burgess et al., 2004; Cabiddu et al., 2012; Gronfier et al., 1999; Somers et al., 1993). In one study, WHOOP-derived SWS episodes limits of agreement (LOA) for heart rate and HRV approached and/or exceeded the smallest worthwhile change and coefficient of variation, which may, in turn, impact the day-to-day variability in WHOOP-derived HR and HRV (C. Bellenger et al., 2021).

Miller et al. concluded that “comparisons of reliability based on intraclass correlations should be made across devices within the same study as there is no clear threshold at which a device can be considered valid” (Miller et al., 2021). Furthermore, both WHOOP manual and automatic two and four sleep categorization could provide a practical alternative to PSG and perform well against other commercially available devices (Miller et al., 2021). Grandner et al. showed moderate and comparable sleep specificity with commercially available activity trackers and that personalized algorithms that adapt to user data over time can improve device performance (Grandner et al., 2023).

See Tables 1 and 2 for more information regarding accuracy studies.

Performance/Clinical Tracking

In studies looking at accuracy and clinical impact together, one study concluded that activity trackers can improve sleep quality and accurately measure sleep and cardiorespiratory variables (Berryhill et al., 2020). However, the impact of the WHOOP on athletic performance requires further investigation (Harms, 2018; Lundstrom, 2020).

There is emerging data suggesting HRV and time spent in slow wave sleep could correlate with cognitive function, thereby providing a non-invasive monitoring tool in pre-clinical cognitive impairment (Saif et al., 2019). Another study identified sleep patterns consistent with acute and chronic sleep deprivation amongst surgeons, declining post-call day 2 and recovering after post-call day 3 (Coleman et al., 2019). Additionally, poor sleep quality and quantity were seen in a group of orthopedic surgeons in another cohort study (Sochacki et al., 2018). Furthermore, one large-scale population monitoring showed a positive impact of physical distancing on sleep and exercise activity (Capodilupo & Miller, 2020), but a potential adverse association with mental health issues where sleep parameters are impaired (Czeisler et al., 2022).

See Table 3 for more information regarding the clinical and performance studies using the WHOOP.

Table 1: Accuracy Studies Looking at Sleep Metrics and WHOOP Versus a Standard

	N	Comparison	Generation	Total Sleep Time (TST)			Total Wake Time (TWT)			Light Sleep			Deep Sleep			REM		
				MAPE (%)	Bias	Sensitivity (%)	MAPE (%)	Bias	Sensitivity (%)	MAPE (%)	Bias	Sensitivity (%)	MAPE (%)	Bias	Sensitivity (%)	MAPE (%)	Bias	Sensitivity(%)
(Stone et al., 2020)	5	HBSMS & various CAT	2.0	8.8	0.27		60.3		-0.59	13.5	-0.17		46.2	0.15		57.8	0.10	
(Miller et al., 2020)	12	PSG	2.0	2.3	*22.5*	95	15.1	*22.5*	51	2.1	*34.6*	62	4.4	*20.1*	68	17.6	*34.1*	70
(Miller et al., 2021)	6	Actigraphy & PSG	2.0	6.4-10.1	-178.8 to 16.7	90-97	47.1-74.2	-17.8 to 16.7	45-60	22.2-23.8	-8.9 to 13.9	61-67	20.4-24.4	-15.5 to -6.6	61-63	35.0	6.5-8.8	66
(Grandner et al., 2023)	36	PSG, EEG headband, CAT	3.0	NC		90-91	NC			NC			NC			NC		
(Berryhill et al., 2020)	32	PSG	2.0	*4.2*						NC			NC			*0.01*		
(Miller et al., 2022)	53	PSG, ECG & CAT	3.0		-12.2	94		13.1	56		-15.6	58		-19.6	62		22.9	66

Mean Absolute Percentage Error (MAPE), Polysomnogram (PSG), Commercial Activity trackers (CAT), Home Based Sleep Monitoring Systems (HBSMS) , Electroencephalography (EEG), Not Calculable (NC). Electrocardiogram (ECG).

** NOTE: Data in red/italics could not be clarified*

Table 2: Accuracy Studies Looking at Cardiac Metrics and WHOOP Versus a Standard

	N	**Generation**	**Standard**	**Time Matched**		**Time Matched**		**Stage Matched**	
				Bias HR -	Bias HR - LOA	Bias HRV -	Bias HR - LOA	HR/HRV Bias-	HR/HRV Bias- LOA
(C. Bellenger et al., 2021)	6	2.0	ECG/PSG	trivial	trivial	trivial	small	trivial	Moderate to large
(Berryhill et al., 2020)	32	2.0	PSG	*trivial*			*small*		
(Miller et al., 2022)	53	3.0	PSG	-0.3	1.9	-4.5	7.6		

Polysomnogram (PSG), Electrocardiogram (ECG), Effect Size (ES), Commercial Activity trackers (CAT), Home Based Sleep Monitoring Systems (HBSMS)

** NOTE: Data in red/italics could not be clarified*

Table 3: Clinical and Performance Studies looking at the application of WHOOP

Trials	Design	Intervention	Measures	Outcome	N
(Berryhill et al., 2020)	Crossover RCT	WHOOP as a treatment device	sleep logs/PRO	Less sleep and sleep disturbance and increased naps Increased physical activity and HRV over study period	32
(Coleman et al., 2019)	Cohort	WHOOP as a monitoring device	Sleep metrics from WHOOP	Sleep patterns consistent with acute and chronic sleep deprivation amongst surgeons, declining post call day 2 and recovering after post call day 3	17
(Saif et al., 2019)	Cohort	WHOOP as a monitoring device	Whoop sleep metrics, user surveys, neuropsychological testing, biomarkers	Unsupervised machine learning techniques from biosensor device may be potentially useful for monitoring cognitive changes in preclinical AD.	33
(Sochacki et al., 2018)	Cohort	WHOOP as a monitoring device	Sleep metrics from WHOOP	Compared to standard population norms poor sleep quality and quantity worse with increased work hours	26

(Capodilupo & Miller, 2020)	Cohort	WHOOP as a monitoring device	Sleep and activity metrics from WHOOP	Improved exercise and sleep during pandemic related physical distancing restrictions	50,000
(Harms, 2018)	Controlled	WHOOP as a treatment device	Sleep and activity metrics from WHOOP, sports performance	No change for users in batting performance compare to non users	10
(Lundstrom, 2020)	Cohort	WHOOP as a monitoring device	Sleep and activity metrics from WHOOP, laboratory and resting metabolic rate, performance and patient survey	HRV correlated with measures of metabolism, but not performance	23
(Czeisler et al., 2022)	Cohort	WHOOP as a monitoring device	Sleep and activity metrics from WHOOP, and PRO anxiety and depression outcomes	Adverse mental health symptoms and pre- and mid-pandemic is associated with short sleep duration and inconsistent sleep timing	4912

Alzheimer's Dementia (AD) , Patient reported outcomes (PRO), Randomized Control Trial (RCT)

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DISCUSSION:

In the setting of health and performance, using wireless and portable devices like the WHOOP to measure autonomic nervous system adaptability (HRV) and sleep quality affords numerous advantages compared to electrical-based cardiac measures and polysomnography (Kellmann, 2010; Plews et al., 2013; Prinsloo et al., 2014, D'Souza et al., 2014, Georgiou et al., 2018; Lai & Kim, 2015; Luczak et al., 2020; Luedtke & Duoos, 2015, Kellmann, 2010; Plews et al., 2013; Prinsloo et al., 2014, Georgiou et al., 2018; Lai & Kim, 2015; Luczak et al., 2020; Luedtke & Duoos, 2015).

The clinical application for the WHOOP, given the acceptable accuracy levels, continues to expand. Uses have included impact on sports performance, correlation with medical conditions (i.e. cognitive dysfunction), sleep and health behaviours in various populations. Limitations of existing accuracy trials include variability in study design and reporting metrics, while results from non-accuracy trials require further clinical validation of response rate and effect size.

WHOOP Basics

The WHOOP uses plethysmography (PPG) to detect blood flow and heart rate changes between systole and diastole to calculate HRV (i.e. root mean square successive beat-to-beat interval differences) during slow wave sleep (Georgiou et al., 2018; Lai & Kim, 2015; Márquez & Molinero, 2013; Papageorgiou et al., 2018; Reed et al., 2013, D'Souza et al., 2014, Georgiou et al., 2018; Lai & Kim, 2015; Márquez & Molinero, 2013; Papageorgiou et al., 2018; Reed et al., 2013). Slow-wave sleep is theorized to play an integral role in physiological exercise recovery, therefore accurately measuring cardiac measures like HRV and heart rate could provide useful real-time feedback on the selection and monitoring of training activities (Gibbs et al., 2013; Papageorgiou et al., 2018; Scheid et al., 2009; Wade & Schneider, 1992, D'Souza et al., 2014, Georgiou et al., 2018; Lai & Kim, 2015; Márquez & Molinero, 2013; Papageorgiou et al., 2018; Reed et al., 2013, Gibbs et al., 2013; Papageorgiou et al., 2018; Scheid et al., 2009; Wade & Schneider, 1992). This hypothesis is supported by the correlation of certain sleep periods and growth hormone release, suggesting an optimization could improve performance and recovery (Shapiro et al., 1981).

Sleep Tracking Limitations

Previous studies of activity trackers capable of identifying four-stage sleep data have generally yielded high sensitivity when an individual is asleep, low to moderate sensitivity for someone being awake, and varying sensitivity for various stages of sleep (De Zambotti et al., 2019; Miller et al., 2020; Shambroom et al., 2012). PPG-based technologies have historically been prone to errors, in part through motion artifacts and variability in skin complexion (Allen, 2007; Bent et al., 2020; Butler et al., 2016; Sañudo et al., 2019). Extracting data from commercial platforms can be problematic since multiple measures can be related, and

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

standard deviations for each specific measure can influence the error measure (Stone et al., 2020).

In general, validation studies will also manually adjust activity tracker data in a research setting, meaning the autodetection accuracy may not be tested (de Zambotti et al., 2016, 2019; Kang et al., 2017; Maskevich et al., 2017; Meltzer et al., 2015; Miller et al., 2020). Even in the case of polysomnograms, technician-guided sleep staging can vary by up to 20%, which could influence the outcomes of studies on commercially available sleep and activity trackers (Collop, 2002). Therefore, the accuracy of sleep wearables in situations where manual adjustment of sleep times is performed by the user may vary. Finally, most accuracy validation studies have been done in healthy patients, with small sample sizes and variable nights of sleep, which may limit the generalizability of use (Miller et al., 2020).

Heart Rate and Heart Rate Variability Limitations

For HR and HRV measures, there is likely a natural day-to-day variability of 3-13% related in part to timing and body position, as well as pulse travel to the periphery (Al Haddad et al., 2011; C. Bellenger et al., 2021; Chen et al., 2020; A. A. Flatt & Howells, 2019; Hopkins et al., 2009; Nakamura et al., 2020; Plews et al., 2013; Selvaraj et al., 2008). Therefore a filter is likely appropriate where bias and level of agreement are less than the smallest worthwhile change and coefficient of variability. However, where the bias and LOA reach or exceed the smallest worthwhile change/coefficient of variation, these values may need to be interpreted against a device's own bias precision levels (Al Haddad et al., 2011; C. Bellenger et al., 2021; Chen et al., 2020; A. A. Flatt & Howells, 2019; Hopkins et al., 2009; Nakamura et al., 2020; Plews et al., 2013).

Relevance Future Direction

Therefore, given the inaccuracies and variability, context could guide use. For commercial applications for the average person, the error rates are not expected to impact decisions around sleep and activity selection significantly. For research including baseline measures, a raw data feed, a gold standard where possible, other clinical variables, and devices with accelerometers to account for body movement, those using higher wavelength PPG could potentially reduce bias as the technologies and algorithms continue to evolve (Stone et al., 2020). In order to address the challenges with current research, some authors have recommended standard reporting that clinicians can easily understand and apply to clinic settings and monitor for over-focusing on sleep metrics in a way that creates adverse health effects such as insomnia (Khosla & Wickwire, 2020).

In “treatment” trials, the clinical utility and plausibility of outcomes should be examined closely. Even where the effect size appears to be small, how reduced overall sleep and increased nap time biologically correspond to patients’ reports of less sleep disturbance should be explored further in larger,

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

participant-blinded trials (Berryhill et al., 2020). However, monitoring for sleep deprivation and/or cognitive decline in at-risk populations could be a strong value proposition for these tools, especially where outcomes are modifiable in a clinically relevant way (Coleman et al., 2019; Saif et al., 2019; Sochacki et al., 2018). Furthermore, based on current studies measuring population-level exercise and sleep data could be helpful from a public policy and planning standpoint (Capodilupo & Miller, 2020), but its impact on performance (e.g. sports, military, etc) needs to be clarified further (Harms, 2018; Lundstrom, 2020)

CONCLUSIONS

The WHOOP wearable device has acceptable accuracy for sleep and cardiac variables to be used in clinical studies where a baseline can be established and ideally other clinical outcomes and gold standard tools can be employed. Further study with easy-to-understand standardized measures is required to understand the accuracy of newer devices and algorithms, as well as their utility in clinical decision-making and outcomes.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

Al Haddad, H., Laursen, P. B., Chollet, D., Ahmaidi, S., & Buchheit, M. (2011). Reliability of resting and postexercise heart rate measures. *International Journal of Sports Medicine*, *32*(8), 598–605.

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, *28*(3), R1–R39.

Bellenger, C., Miller, D., Halson, S., Roach, G., & Sargent, C. (2021). Wrist-Based Photoplethysmography Assessment of Heart Rate and Heart Rate Variability: Validation of WHOOP. *Sensors* , *21*(10), 3571.

Bellenger, C. R., Fuller, J. T., Thomson, R. L., Davison, K., Robertson, E. Y., & Buckley, J. D. (2016). Monitoring Athletic Training Status Through Autonomic Heart Rate Regulation: A Systematic Review and Meta-Analysis. *Sports Medicine* , *46*(10), 1461–1486.

Bent, B., Goldstein, B. A., Kibbe, W. A., & Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digital Medicine*, *3*, 18.

Berryhill, S., Morton, C. J., Dean, A., Berryhill, A., Provencio-Dean, N., Patel, S. I., Estep, L., Combs, D., Mashaqi, S., Gerald, L. B., Krishnan, J. A., & Parthasarathy, S. (2020). Effect of wearables on sleep in healthy individuals: a randomized crossover trial and validation study. *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine*, *16*(5), 775–783.

Bolin, D. J. (2019). Sleep Deprivation and Its Contribution to Mood and Performance Deterioration in College Athletes. *Current Sports Medicine Reports*, *18*(8), 305–310.

Borresen, J., & Lambert, M. I. (2008). Autonomic control of heart rate during and after exercise : measurements and implications for monitoring training status. *Sports Medicine* , *38*(8), 633–646.

Buchheit, M. (2014). Monitoring training status with HR measures: do all roads lead to Rome? *Frontiers in Physiology*, *5*, 73.

Burgess, H. J., Penev, P. D., Schneider, R., & Van Cauter, E. (2004). Estimating cardiac autonomic activity during sleep: impedance cardiography, spectral analysis, and Poincaré plots. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *115*(1), 19–28.

Butler, M. J., Crowe, J. A., Hayes-Gill, B. R., & Rodmell, P. I. (2016). Motion limitations of non-contact photoplethysmography due to the optical and topological properties of skin. In *Physiological Measurement* (Vol. 37, Issue 5, pp. N27–N37). https://doi.org/10.1088/0967-3334/37/5/n27

Cabiddu, R., Cerutti, S., Viardot, G., Werner, S., & Bianchi, A. M. (2012). Modulation of the Sympatho-Vagal Balance during Sleep: Frequency Domain Study of Heart Rate Variability and Respiration. *Frontiers in Physiology*, *3*, 45.

Capodilupo, E. R., & Miller, D. J. (2020). *Changes in health promoting behavior during COVID-19 physical distancing: Utilizing Wearable Technology to Examine Trends in Sleep, Activity, and Cardiovascular Indicators of Health*. Public and

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Global Health. http://medrxiv.org/lookup/doi/10.1101/2020.06.07.20124685

Chen, Y.-S., Clemente, F. M., Bezerra, P., & Lu, Y.-X. (2020). Ultra-short-term and Short-term Heart Rate Variability Recording during Training Camps and an International Tournament in U-20 National Futsal Players. *International Journal of Environmental Research and Public Health*, *17*(3). https://doi.org/10.3390/ijerph17030775

Coleman, J. J., Robinson, C. K., Zarzaur, B. L., Timsina, L., Rozycki, G. S., & Feliciano, D. V. (2019). To Sleep, Perchance to Dream: Acute and Chronic Sleep Deprivation in Acute Care Surgeons. *Journal of the American College of Surgeons*, *229*(2), 166–174.

Collop, N. A. (2002). Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep Medicine*, *3*(1), 43–47.

Czeisler, M. É., Capodilupo, E. R., Weaver, M. D., Czeisler, C. A., Howard, M. E., & Rajaratnam, S. M. W. (2022). Prior sleep-wake behaviors are associated with mental health outcomes during the COVID-19 pandemic among adult users of a wearable device in the United States. *Sleep Health*, *8*(3), 311–321.

Depner, C. M., Cheng, P. C., Devine, J. K., Khosla, S., de Zambotti, M., Robillard, R., Vakulin, A., & Drummond, S. P. A. (2020). Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*, *43*(2). https://doi.org/10.1093/sleep/zsz254

de Zambotti, M., Baker, F. C., Willoughby, A. R., Godino, J. G., Wing, D., Patrick, K., & Colrain, I. M. (2016). Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiology & Behavior*, *158*, 143–149.

De Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M., & Baker, F. C. (2019). Wearable Sleep Technology in Clinical and Research Settings. *Medicine & Science in Sports & Exercise*, *51*(7), 1538–1557.

de Zambotti, M., Rosas, L., Colrain, I. M., & Baker, F. C. (2019). The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography. *Behavioral Sleep Medicine*, *17*(2), 124–136.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I. (1997). Cumulative Sleepiness, Mood Disturbance, and Psychomotor Vigilance Performance Decrements During a Week of Sleep Restricted to 4–5 Hours per Night. *Sleep*, *20*(4), 267–277.

D'Souza, A., Bucchi, A., Johnsen, A. B., Logantha, S. J. R. J., Monfredi, O., Yanni, J., Prehar, S., Hart, G., Cartwright, E., Wisloff, U., Dobryznski, H., DiFrancesco, D., Morris, G. M., & Boyett, M. R. (2014). Exercise training reduces resting heart rate via downregulation of the funny channel HCN4. *Nature Communications*, *5*, 3775.

Edwards, B. J., & Waterhouse, J. (2009). Effects of one night of partial sleep deprivation upon diurnal rhythms of accuracy and consistency in throwing darts. *Chronobiology International*, *26*(4), 756–768.

Flatt, A. A., & Howells, D. (2019). Effects of varying training load on heart rate variability and running performance among an

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Olympic rugby sevens team. *Journal of Science and Medicine in Sport / Sports Medicine Australia*, *22*(2), 222–226.

Flatt, A., Esco, M., & Nakamura, F. (2018). Association between Subjective Indicators of Recovery Status and Heart Rate Variability among Divison-1 Sprint-Swimmers. *Sportscience*, *6*(3), 93.

Fortes, L. S., da Costa, B. D. V., Paes, P. P., do Nascimento Júnior, J. R. A., Fiorese, L., & Ferreira, M. E. C. (2017). Influence of Competitive-Anxiety on Heart Rate Variability in Swimmers. *Journal of Sports Science & Medicine*, *16*(4), 498–504.

Georgiou, K., Larentzakis, A. V., Khamis, N. N., Alsuhaibani, G. I., Alaska, Y. A., & Giallafos, E. J. (2018). Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. *Folia Medica*, *60*(1), 7–20.

Gibbs, J. C., Williams, N. I., Mallinson, R. J., Reed, J. L., Rickard, A. D., & De Souza, M. J. (2013). Effect of high dietary restraint on energy availability and menstrual status. *Medicine and Science in Sports and Exercise*, *45*(9), 1790–1797.

Grandner, M. A., Bromberg, Z., Hadley, A., Morrell, Z., Graf, A., Hutchison, S., & Freckleton, D. (2023). Performance of a multisensor smart ring to evaluate sleep: in-lab and home-based evaluation of generalized and personalized algorithms. *Sleep*, *46*(1), zsac152.

Gronfier, C., Simon, C., Piquard, F., Ehrhart, J., & Brandenberger, G. (1999). Neuroendocrine processes underlying ultradian sleep regulation in man. *The Journal of Clinical Endocrinology and Metabolism*, *84*(8), 2686–2690.

Harms, N. (2018). *The Impact of WHOOP Technology on Sleep, Recovery, and Performance in NAIA Baseball Players*.

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*(1), 3–13.

Ibáñez, V., Silva, J., Navarro, E., & Cauli, O. (2019). Sleep assessment devices: types, market analysis, and a critical view on accuracy and validation. *Expert Review of Medical Devices*, *16*(12), 1041–1052.

Kang, S.-G., Kang, J. M., Ko, K.-P., Park, S.-C., Mariani, S., & Weng, J. (2017). Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. *Journal of Psychosomatic Research*, *97*, 38–44.

Kellmann, M. (2010). Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring: Preventing overtraining. *Scandinavian Journal of Medicine & Science in Sports*, *20*, 95–102.

Kelly, J. M., Strecker, R. E., & Bianchi, M. T. (2012). Recent developments in home sleep-monitoring devices. *ISRN Neurology*, *2012*, 768794.

Khosla, S., Deak, M. C., Gault, D., Goldstein, C. A., Hwang, D., Kwon, Y., O'Hearn, D., Schutte-Rodin, S., Yurcheshen, M., Rosen, I. M., Kirsch, D. B., Chervin, R. D., Carden, K. A., Ramar, K., Aurora, R. N., Kristo, D. A., Malhotra, R. K., Martin, J. L., Olson, E. J., … American Academy of Sleep Medicine Board of Directors. (2018). Consumer Sleep Technology: An American Academy of Sleep Medicine Position Statement. *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine*, *14*(5), 877–880.

Khosla, S., & Wickwire, E. M. (2020). Consumer sleep technology: accuracy and impact on behavior among healthy individuals

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

[Review of *Consumer sleep technology: accuracy and impact on behavior among healthy individuals*]. *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine*, *16*(5), 665–666.

Lai, P.-H., & Kim, I. (2015). Lightweight wrist photoplethysmography for heavy exercise: motion robust heart rate monitoring algorithm. *Healthcare Technology Letters*, *2*(1), 6–11.

Luczak, T., Burch, R., Lewis, E., Chander, H., & Ball, J. (2020). State-of-the-art review of athletic wearable technology: What 113 strength and conditioning coaches and athletic trainers from the USA said about technology in sports. *International Journal of Sports Science & Coaching*, *15*(1), 26–40.

Luedtke, D., & Duoos, B. (2015). Comparison of Four Feedback Methods Used to Help Improve Swimming Relay Exchanges - A Pilot Study. *International Journal of Aquatic Research and Education*, *9*(2), 8.

Lundstrom, E. (2020). *Effectiveness of Wearable Technology for Predicting Measures of Metabolism and Performance in Collegiate Division 1 Swimmers*.

Márquez, S., & Molinero, O. (2013). Energy availability, menstrual dysfunction and bone health in sports; an overview of the female athlete triad. *Nutricion Hospitalaria: Organo Oficial de La Sociedad Espanola de Nutricion Parenteral Y Enteral*, *28*(4), 1010–1017.

Maskevich, S., Jumabhoy, R., Dao, P. D. M., Stout, J. C., & Drummond, S. P. A. (2017). Pilot Validation of Ambulatory Activity Monitors for Sleep Measurement in Huntington's Disease Gene Carriers. *Journal of Huntington's Disease*, *6*(3), 249–253.

Meltzer, L. J., Hiruma, L. S., Avis, K., Montgomery-Downs, H., & Valentin, J. (2015). Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents. *Sleep*, *38*(8), 1323–1330.

Miller, D. J., Lastella, M., Scanlan, A. T., Bellenger, C., Halson, S. L., Roach, G. D., & Sargent, C. (2020). A validation study of the WHOOP strap against polysomnography to assess sleep. *Journal of Sports Sciences*, *38*(22), 2631–2636.

Miller, D. J., Roach, G. D., Lastella, M., Scanlan, A. T., Bellenger, C. R., Halson, S. L., & Sargent, C. (2021). A Validation Study of a Commercial Wearable Device to Automatically Detect and Estimate Sleep. *Biosensors*, *11*(6), 185.

Miller, D. J., Sargent, C., & Roach, G. D. (2022). A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults. *Sensors* , *22*(16), 6317.

Nakamura, F. Y., Antunes, P., Nunes, C., Costa, J. A., Esco, M. R., & Travassos, B. (2020). Heart Rate Variability Changes From Traditional vs. Ultra-Short-Term Recordings in Relation to Preseason Training Load and Performance in Futsal Players. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, *34*(10), 2974–2981.

Papageorgiou, M., Dolan, E., Elliott-Sale, K. J., & Sale, C. (2018). Reduced energy availability: implications for bone health in physically active populations. *European Journal of Nutrition*, *57*(3), 847–859.

Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. (2013). Training Adaptation and Heart Rate Variability in Elite Endurance Athletes: Opening the Door to Effective Monitoring. *Sports Medicine* , *43*(9), 773–781.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Prinsloo, G. E., Rauch, H. G. L., & Derman, W. E. (2014). A Brief Review and Clinical Application of Heart Rate Variability Biofeedback in Sports, Exercise, and Rehabilitation Medicine. *The Physician and Sportsmedicine*, *42*(2), 88–99.

Reed, J. L., De Souza, M. J., & Williams, N. I. (2013). Changes in energy availability across the season in Division I female soccer players. *Journal of Sports Sciences*, *31*(3), 314–324.

Saif, N., Yan, P., Niotis, K., Scheyer, O., Rahman, A., Berkowitz, M., Krikorian, R., Hristov, H., Sadek, G., Bellara, S., & Isaacson, R. S. (2019). FEASIBILITY OF USING A WEARABLE BIOSENSOR DEVICE IN PATIENTS AT RISK FOR ALZHEIMER'S DISEASE DEMENTIA. *The Journal of Prevention of Alzheimer's Disease*, 1–8.

Sañudo, B., De Hoyo, M., Muñoz-López, A., Perry, J., & Abt, G. (2019). Pilot Study Assessing the Influence of Skin Type on the Heart Rate Measurements Obtained by Photoplethysmography with the Apple Watch. *Journal of Medical Systems*, *43*(7), 195.

Scheid, J. L., Williams, N. I., West, S. L., VanHeest, J. L., & De Souza, M. J. (2009). Elevated PYY is associated with energy deficiency and indices of subclinical disordered eating in exercising women with hypothalamic amenorrhea. *Appetite*, *52*(1), 184–192.

Sekiguchi, Y., Adams, W. M., Benjamin, C. L., Curtis, R. M., Giersch, G. E. W., & Casa, D. J. (2019). Relationships between resting heart rate, heart rate variability and sleep characteristics among female collegiate cross-country athletes. *Journal of Sleep Research*, *28*(6), e12836.

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering & Technology*, *32*(6), 479–484.

Shambroom, J. R., Fábregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research*, *21*(2), 221–230.

Shapiro, C. M., Bortz, R., Mitchell, D., Bartel, P., & Jooste, P. (1981). Slow-wave sleep: a recovery period after exercise. *Science*, *214*(4526), 1253–1254.

Sochacki, K. R., Dong, D., Peterson, L. E., McCulloch, P. C., & Harris, J. D. (2018). The Measurement of Orthopaedic Surgeon Quality and Quantity of Sleep Using a Validated Wearable Device. *JAAOS: Global Research and Reviews*, *2*(10), e065.

Somers, V. K., Dyken, M. E., Mark, A. L., & Abboud, F. M. (1993). Sympathetic-nerve activity during sleep in normal subjects. *The New England Journal of Medicine*, *328*(5), 303–307.

Stone, J. D., Rentz, L. E., Forsey, J., Ramadan, J., Markwald, R. R., Finomore, V. S., Galster, S. M., Rezai, A., & Hagen, J. A. (2020). Evaluations of Commercial Sleep Technologies for Objective Monitoring During Routine Sleeping Conditions. *NSS News*, *12*, 821–842.

Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness:

medRxiv preprint doi: https://doi.org/10.1101/2024.01.04.24300784; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, *26*(2), 117–126.

VanHeest, J. L., Rodgers, C. D., Mahoney, C. E., & De Souza, M. J. (2014). Ovarian Suppression Impairs Sport Performance in Junior Elite Female Swimmers. *Medicine & Science in Sports & Exercise*, *46*(1), 156–166.

Vgontzas, A. N., Zoumakis, E., Bixler, E. O., Lin, H.-M., Follett, H., Kales, A., & Chrousos, G. P. (2004). Adverse effects of modest sleep restriction on sleepiness, performance, and inflammatory cytokines. *The Journal of Clinical Endocrinology and Metabolism*, *89*(5), 2119–2126.

Wade, G. N., & Schneider, J. E. (1992). Metabolic fuels and reproduction in female mammals. *Neuroscience and Biobehavioral Reviews*, *16*(2), 235–272.

Woods, A. L., Garvican-Lewis, L. A., Lundy, B., Rice, A. J., & Thompson, K. G. (2017). New approaches to determine fatigue in elite athletes during intensified training: Resting metabolic rate and pacing profile. *PloS One*, *12*(3), e0173807.

Woods, A. L., Rice, A. J., Garvican-Lewis, L. A., Wallett, A. M., Lundy, B., Rogers, M. A., Welvaert, M., Halson, S., McKune, A., & Thompson, K. G. (2018). The effects of intensified training on resting metabolic rate (RMR), body composition and performance in trained cyclists. *PloS One*, *13*(2), e0191644.