European Journal of Applied Physiology (2022) 122:1975–1990
https://doi.org/10.1007/s00421-022-04951-1

INVITED REVIEW

Wearable activity trackers–advanced technology or advanced marketing?

Ren-Jay Shei[1] · Ian G. Holder[2] · Alicia S. Oumsang[2] · Brittni A. Paris[2] · Hunter L. Paris[2]

Received: 10 December 2021 / Accepted: 4 April 2022 / Published online: 21 April 2022

Abstract

Wearable devices represent one of the most popular trends in health and fitness. Rapid advances in wearable technology present a dizzying display of possible functions: from thermometers and barometers, magnetometers and accelerometers, to oximeters and calorimeters. Consumers and practitioners utilize wearable devices to track outcomes, such as energy expenditure, training load, step count, and heart rate. While some rely on these devices in tandem with more established tools, others lean on wearable technology for health-related outcomes, such as heart rhythm analysis, peripheral oxygen saturation, sleep quality, and caloric expenditure. Given the increasing popularity of wearable devices for both recreation and health initiatives, understanding the strengths and limitations of these technologies is increasingly relevant. Need exists for continued evaluation of the efficacy of wearable devices to accurately and reliably measure purported outcomes. The purposes of this review are (1) to assess the current state of wearable devices using recent research on validity and reliability, (2) to describe existing gaps between physiology and technology, and (3) to offer expert interpretation for the lay and professional audience on how best to approach wearable technology and employ it in the pursuit of health and fitness. Current literature demonstrates inconsistent validity and reliability for various metrics, with algorithms not publicly available or lacking high-quality validation studies. Advancements in wearable technology should consider standardizing validation metrics, providing transparency in used algorithms, and improving how technology can be tailored to individuals. Until then, it is prudent to exercise caution when interpreting metrics reported from consumer-wearable devices.

Keywords Physical activity · Wearable technology · Fitness · VO_2max · Energy expenditure · Health monitoring · Step count

Abbreviations

ECG	Electrocardiography
EE	Energy expenditure
GPS	Global positioning system
HR	Heart rate
HRV	Heart rate variability
PPG	Photoplethysmography
RMR	Resting metabolic rate
$\dot{V}O_2$	Oxygen consumption
$\dot{V}O_{2max}$	Maximal oxygen consumption

Communicated by Michael I Lindinger.

✉ Ren-Jay Shei
reshei@alumni.iu.edu

[1] Indiana University Alumni Association, Indiana University, 1000 E 17th Street, Bloomington, IN 47408, USA

[2] Department of Sports Medicine, Pepperdine University, Malibu, CA, USA

Introduction

Wearable technology has gained in popularity among a broad segment of the general population including elite, competitive, and recreational athletes, as well as both the physically active and sedentary general population. Even in 2015, approximately 1 in 8 respondents (12.5%) to an annual nationwide consumer mail panel survey in the United States indicated that they currently used a wearable activity monitor (Omura et al. 2017). Recent survey data of fitness trends in 2019 showed that wearable technology was the number one most popular trend (Thompson 2018), and the market for such devices continues to grow. These devices are broadly marketed to both the general population for physical fitness and health monitoring as well as specifically to elite athletes as a training tool (Wahl et al. 2017). As such, the public health relevance of such wearable devices is increasing and may impact areas, such as physical activity, well-being, cardiovascular health, mortality risk, dietary habits,

among others. For example, higher volumes of physical activity energy expenditure measured by wearable devices was recently shown to be associated with reduced mortality rates, and that higher-intensity activity reduced mortality rates to a larger extent than lower-intensity activity (Strain et al. 2020). Recent advances in technology have resulted in a myriad of wrist-based sensors being built into the current generation of fitness watches (Fig. 1). These include digital 3-axis accelerometers, pulse oximeters, optical heart rate sensors, thermometers, barometers, magnetometers, among others. These sensors, in combination with ever-improving algorithms—most of them proprietary—have led to fitness companies marketing these devices as being capable of estimating and monitoring such physiological parameters as step count, heart rate (HR), sleep quality, sleep rhythm, energy expenditure (EE), maximal oxygen consumption ($\dot{V}O_{2max}$), peripheral oxygen saturation, and the "training effect" of both individual and cumulative exercise bouts.

Despite these purported advances in physiological monitoring capabilities, few published data exist to support the

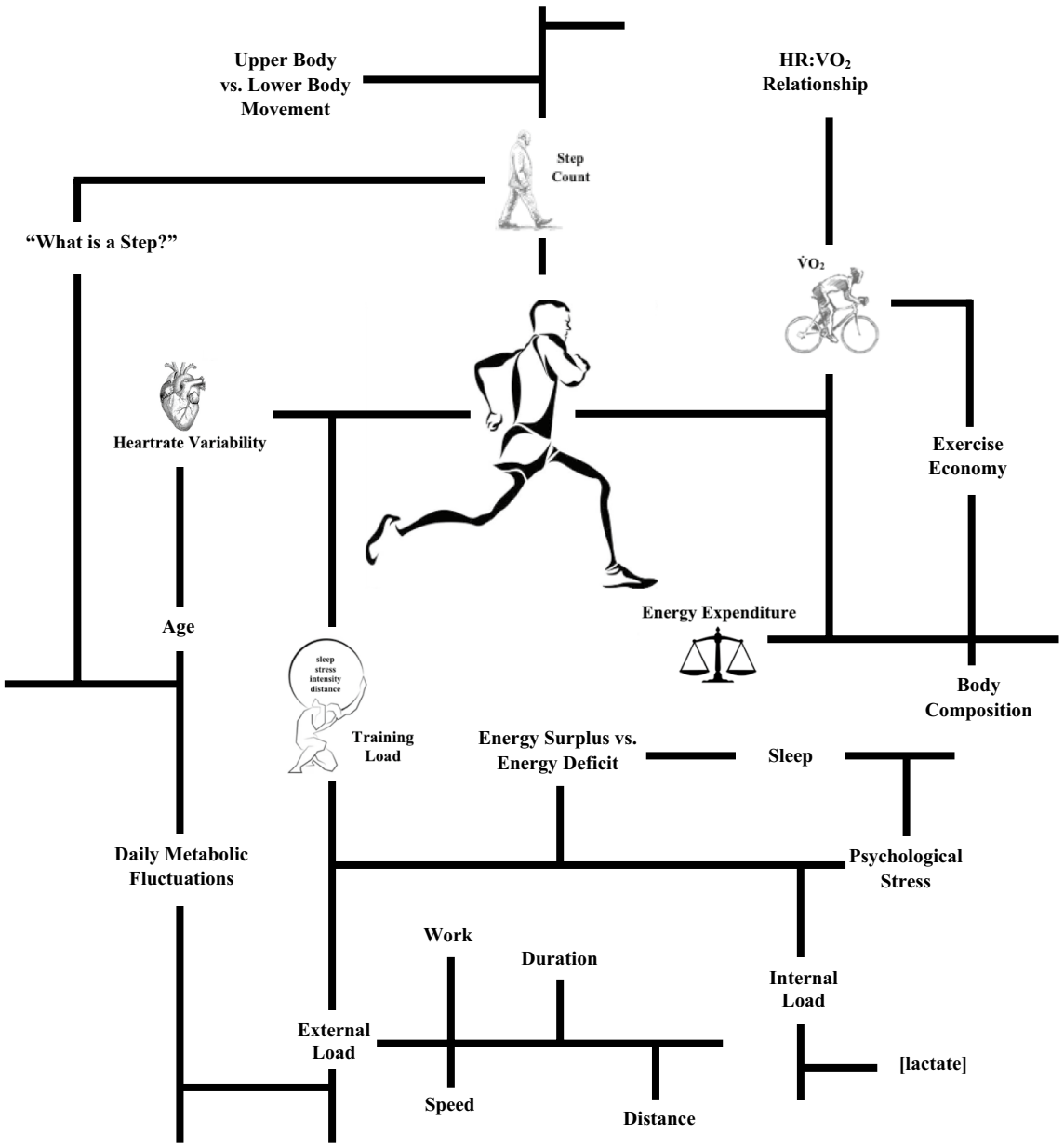

Fig. 1 Web of variables assessed by wearable devices and factors that must be considered in accurately reporting these variables. Variables directly connected to the athlete are those recorded by technologies. Outer variables are factors that influence the inner measure. For example, "training load" must consider both "external load" and "internal load," which themselves must consider "work," "distance," "duration," etc

validity and reliability of such tracking. Due to the widespread use of these wearable-fitness tracking devices to promote health, fitness, and an overall active lifestyle, rigorous and transparent reporting of validation studies should be encouraged to improve precision, accuracy, and reliability, and to engender trust in the consumers who purchase and utilize such devices. To date, studies investigating the validity and reliability are sparse, with wide disparity in findings, likely due to a variety of reasons including differences in devices tested, study population, and experimental design. This review summarizes the available studies testing the validity and reliability of wearable-fitness devices, discusses several publicly available algorithms to estimate select physiological parameters, and presents current knowledge gaps and future directions for the wearable technology field to address. Specific metrics discussed below include $\dot{V}O_2$ and $\dot{V}O_{2max}$ estimation, EE estimation, step count estimation, HR and HR variability (HRV), which were all selected on the basis of being common measures used in applied physiology research and are commonly tracked by end-users of consumer-wearable devices. These metrics also have the most published research data available on their validity compared to gold standard laboratory or research methods of data collection. Additionally, more applied metrics, such as training load, stress, and sleep, are discussed as well. These variables are also measured in physiology research, although they are less common than the aforementioned metrics, and they have been selected for discussion because of their special interest to the lay population of end-users.

$\dot{V}O_2$ and $\dot{V}O_{2max}$ estimation

Oxygen transport and utilization are the most important determinants of endurance exercise performance. The amount of oxygen that an individual utilizes per unit time ($\dot{V}O_2$) provides a comprehensive view of aerobic capacity and pathophysiology. Traditional $\dot{V}O_2$ measurement is undertaken in a laboratory setting using open-circuit spirometry and indirect calorimetry in which expired gases are collected and analyzed and $\dot{V}O_2$ is calculated using the Haldane transformation. Continuous measurement of $\dot{V}O_2$ during an incremental exercise test to exhaustion is considered the gold standard measurement of maximal aerobic capacity, or $\dot{V}O_{2max}$. The major drawback of such testing is the need for properly trained staff, expensive equipment, and the lack of a field-based or "real-world" setting. Therefore, attempts have been made to either adapt equipment to be more portable, or to estimate $\dot{V}O_2$ based on surrogate physiological parameters, such as heart rate, workload, or accelerometry data (Carrier et al. 2020; Helm et al. 2021; Snyder et al. 2021).

Though $\dot{V}O_2$ estimation has become more common in wearable devices, research concerning the accuracy of wearables in estimating $\dot{V}O_2$ and $\dot{V}O_{2max}$ has struggled to keep pace. Data from (Passler et al. 2019) compared five commercially available wrist-worn devices against respiratory gas analysis and found a mean absolute percentage error of > 10% [considered a high error (Henriksen et al. 2020)] for the majority of devices. Even within a given brand, $\dot{V}O_2$ was sometimes *over*estimated, but other times *under*estimated. Particularly prominent in discussing validity and reliability amongst wearable devices is the, Towards Intelligent Health and Well-Being Network of Physical Activity Assessment (INTERLIVE). This consortium of six European universities and one industrial partner develops best-practice guidelines for evaluating consumer wearables. A recent review from INTERLIVE (Molina-Garcia et al. 2022) regarding the validity of wearables in estimating $\dot{V}O_2$ and $\dot{V}O_{2max}$ concluded that population-level estimates of $\dot{V}O_{2max}$ demonstrate some degree of accuracy, but the margin of error is large for $\dot{V}O_{2max}$ estimates at the individual level. Within the realm of oxygen utilization, some devices also report arterial oxygen saturation (SaO_2 or SpO_2). Recently, Zhang and Khatami (Zhang and Khatami 2022) summarized existing data on the validity of this measure, especially in the Apple and Garmin smartwatches. They concluded that, despite millions of people utilizing smartwatches to monitor oxygen saturation (and possible sleep apnoea), and although the Apple Watch seems to be more accurate than the Garmin, none of the popular smartwatches had been well validated. Overall, findings suggest that for $\dot{V}O_2$ estimates, as well as oxygen saturation estimates, wrist-worn activity trackers lack the accuracy for sport and healthcare application.

Discrepancies between laboratory-based $\dot{V}O_2$ calculations and wearable technologies may be explained both by the algorithms companies utilize as well as individual variations in exercising $\dot{V}O_2$. Perhaps the most significant challenge in judging the methodology used by wearable technologies to estimate $\dot{V}O_2$ is the highly secretive, proprietary nature of the variables and algorithms used to estimate $\dot{V}O_2$. From the little public information available, such as from patent filings and white papers, it appears some devices utilize running speed and the inclination of the terrain (both presumably derived from the GPS signal of the device) to estimate theoretical $\dot{V}O_2$. Technologies then apply a "quality" filter to improve data quality. Ultimately, $\dot{V}O_{2max}$ is based on a constructed curve of HR vs estimated workload (Firstbeat Technologies Ltd.; US Patent Application Publication). Without full transparency in precisely how each wearable device estimates the parameters reported to the user, it is impossible to rigorously evaluate the validity and reliability of these metrics.

Considering that wearable technologies often estimate $\dot{V}O_2$ from HR, along with other variables, the accuracy of

HR measurement becomes relevant. Whereas heart rate monitors utilizing chest straps have a high level of validity and reliability (Gillinov et al. 2017), the accuracy of wrist-worn devices [which rely on photoplethysmography (PPG) to measure HR] remains questionable as exercise type and intensity, as well as other factors, such as skin tone, temperature, humidity, and proximity of the sensor to the skin, may influence application (Støve et al. 2019). Acknowledging the inherent flaws when relying solely on heart rate or accelerometry to estimate $\dot{V}O_2$ (Fudge et al. 2007), some companies attempt to overcome limitations by combining methodologies. For example, one recent method attempts to estimate $\dot{V}O_2$ from HR and body movement in daily life (Bonomi et al. 2020). This approach uses activity recognition techniques to identify contextual setting (e.g., cycling, sedentary, walking, etc.) and then apply prediction equations specific to a given setting. This method may guide future advancements. While merging accelerometry and heart rate improves accuracy (Haskell et al. 1993), $\dot{V}O_2$ predictions still differ from laboratory-based measurements, particularly at high and low exercise intensities (Acheson et al. 1980; Brage et al. 2003). Discrepancies also exist amongst the algorithms wearable technologies use, and though general information is public—such as how Firstbeat and Garmin utilize submaximal exercise data (US Patent Application Publication) whereas Polar V800 relies upon HRV at rest (Polar)—proprietary information makes accessing these equations difficult and limits testing for validation and reliability. Further discrepancies between the predictive equations used in wearable technologies and laboratory-based calculation of $\dot{V}O_2$ may be explained considering the individual response to exercise. Although the relationship between HR and $\dot{V}O_2$ is linear at sub-maximal intensities, HR can be influenced by age, sex, training status, efficiency, economy, emotional state, and many other factors, making the translation from heart rate to $\dot{V}O_2$ prone to errors (Fudge et al. 2007). Differences also exist in the biomechanical characteristic of locomotion which influence $\dot{V}O_2$. For example, during running exercise, differences in vertical displacement between individuals alters the metabolic cost of running. Even within an individual, biomechanical changes occur within an event (Winter et al. 2016), with form usually deteriorating throughout an exercise bout as an individual tires. Chronic changes also occur with training, and training generally improves running economy thereby altering the $\dot{V}O_2$ for a given pace and distance. Other individual responses to exercise, such as the presence of the $\dot{V}O_2$ slow component or cardiac drift, may also result in a disconnect between predictive equations and actual $\dot{V}O_2$.

Although current uncertainties in the validity of $\dot{V}O_2$ estimation in wearable technologies limit usefulness in health and medical practices, accuracy of wearable technology is optimized when $\dot{V}O_2$ prediction equations are based on a person's unique exercise characteristics (Fudge et al. 2007). Application of these technologies will therefore expand when technology permits a more individualized equation based on physiological and biomechanical responses to exercise, and when algorithms become more easily accessible and available for testing.

Energy expenditure estimation

Energy expenditure (EE) can be defined as the calories burned at rest or during physical activity. Direct calorimetry quantifies heat production, and while it represents the gold standard for measuring human metabolic rate, indirect calorimetry is the more common method utilized for assessing EE, even within the research laboratory (Kenny et al. 2017). While the doubly-labeled water technique represents the most reliable measure of indirect calorimetry (particularly for quantifying free-living EE), open-circuit spirometry is the customary measure, primarily because of its ease of use and comparatively lower price (Ainslie et al. 2003; O'Driscoll et al. 2020). Another method of quantifying caloric expenditure—and one used by wearable devices—is to calculate EE from anthropomorphic data (such as body mass) and exercise parameters recorded through GPS and accelerometry. Because EE data are often used to promote body mass regulation, the ability of wearable devices to accurately report EE has implications for health and disease.

Although insufficient validation studies exist for many of the features promoted by wearable technology, EE is one of the more frequently examined physiological outputs. In general, research demonstrates strong reliability for wearable technologies (Evenson et al. 2015), but poor validity (Düking et al. 2020; Fuller et al. 2020; O'Driscoll et al. 2020; Argent et al. 2022). When pitted against the gold standard doubly label water technique, accelerometers demonstrated varying degrees of validity, prompting researchers to call for further development and evaluation of wearable technology (Plasqui and Westerterp 2007; Plasqui et al. 2013; Murakami et al. 2019). As $\dot{V}O_2$ estimation improves when identifying contextual setting (Bonomi et al. 2020), evidence also suggests that when accelerometry algorithms recognize various types of physical activity, estimation of energy expenditure improves (Bonomi et al. 2009). Within this framework, data generally agree that the ability of wearable technology to accurately report EE depends on the type of exercise (Boudreaux et al. 2018), the intensity of exercise (Roos et al. 2017; Wahl et al. 2017), sex and skin tone of user (Shcherbina et al. 2017), and the specific device in question (Kendall et al. 2019; O'Driscoll et al. 2020).

Though calculations for EE are proprietary and prevent specific analysis, discrepancies between laboratory-based measurement of EE and wearable technologies likely arise from limitations in EE calculations as well as individual variation in EE that would not be captured using traditional equations. Common calculations for EE utilize body mass, age, activity status, accelerometry data, and heart rate. Assuming height, weight, age, and sex are recorded correctly, calculations could fall short if levels of activity status are unclear (e.g., "I was highly-active last week but moderately-active the week prior, so what is my activity status?") or if heart rate data are inaccurate (as mentioned in the previous section on $\dot{V}O_{2\,max}$). Individual variation in EE represents an alternative explanation for the observed discrepancies and a large hurdle for wearable technologies to overcome, particularly given contemporary limitations on technological capabilities. While discussion on the individual variations in energy expenditure is beyond the scope of this paper, we offer a few example scenarios of where energy expenditure could be altered and yet go unnoticed by wearable technologies: Resting metabolic rate (RMR) accounts for the largest contributor to total daily energy expenditure and is itself dependent on body size and body composition. While body size may be accurately assessed by inputting height and weight into wearable devices, many of these activity trackers are unable to assess body composition. Even if body fat percentages were known, many calculations on EE do not account for lean mass vs. fat mass or variations in EE that exist within lean tissue. Furthermore, a large variability in RMR—up to 250 kcal per day—exists outside of differences in body composition (Johnstone et al. 2005), and again would not be assessed using the traditional calculations for EE. Finally, energy balance also influences RMR where RMR may demonstrate an adaptive response to caloric restriction thereby predisposing individuals to weight regain (Fothergill et al. 2016). For example, following weight loss, RMR for a given fat-free mass may be reduced, thereby lowering the total daily energy expenditure and the calories that can be consumed if body mass is to be maintained. In addition to body composition then, an accurate estimation of EE considers both chronic and acute alterations to energy balance. It is unlikely that current wearable technologies permit such considerations. For the many people using wearable technologies for weight loss (or maintenance of lost weight), this last point becomes particularly relevant in that wearable devices may be most accurate for those in energy balance but may be used most often by those in energy *im*balance. Therefore, those most dependent upon wearable devices for EE may be the ones most likely to experience inaccuracies in EE estimation. Each of these examples represents not only the individual response to energy expenditure, but also the complexities of quantifying caloric expenditure and potential pitfalls that may explain the lack of validity in wearable technologies regarding EE.

In addition to limiting our understanding of the validity and reliability of these devices, the proprietary nature of the energy expenditure calculations also prevents proper use of these devices. For example, when using wearable devices to determine caloric expenditure for a given exercise session, are the number of calories burned for a given exercise session based on net or gross expenditure? Perhaps more importantly, is this clear to the user? If, for a given 30 min exercise session, a wearable device displays that 350 cal were burned, does this mean that an *additional* 350 cal were burned above resting metabolic rate, or does this mean that a *total* of 350 cal were burned which includes those attributable to resting metabolic rate? Overall, given the consistent finding that wearable devices lack validity in calculating EE, and given the inconsistencies in the nature of inaccuracies throughout devices, refinements are in order before these devices can be relied upon to report EE.

Step count/physical activity

Whereas physical activity promotes musculoskeletal health and disease prevention, a sedentary lifestyle is linked with muscle atrophy, decreased quality of life, and a less favorable cardiometabolic profile (Riel et al. 2016). While well known that physical activity is meaningful, less is known about what constitutes meaningful physical activity. Metrics toward this goal have been developed, including the recommendation to achieve 10,000 steps per day (Hatano 1993; Choi et al. 2007). In pursuing meaningful physical activity then, monitoring daily step count represents a valuable component of health promotion and one of the major functions endorsed by wearable technologies. In the research setting, the gold standard for quantifying steps is via hand tally, where physical activity is tracked by video recording and two reviewers independently assign manual step counts (Dijkstra et al. 2008; Riel et al. 2016). When video tracking is unavailable, the use of accelerometers is preferred to self-report (Riel et al. 2016), and one of the more common accelerometers used in the research setting is the ActiGraph—a triaxial accelerometer placed near the hip above the right anterior superior iliac spine (Riel et al. 2016).

As physical activity and step count are common features of wearable devices, a number of reviews have examined these functions in various technologies and under a variety of settings—laboratory, free-living, walking, running, etc. While some studies (Wahl et al. 2017; Montes et al. 2020) report promising outcomes for reliability and validity of step count, a 2020 systemic review (Fuller et al. 2020) examining 158 publications, 9 different wearable brands,

and 45 devices concluded that, according to the wider body of literature, wearable devices are accurate for measuring step count in the laboratory, but exhibit a wider range of inaccuracy in free-living environments. Even within brands, validity differed. For example, whereas the Fitbit Charge tended to underestimate steps, the Fitbit Classic overestimated step count. Variability also existed for intra-device reliability, where step count differed, not only within the same company, but also within the same device. Another 2020 systematic review (Henriksen et al. 2020) examined devices that used a triaxial accelerometer system and found that a large heterogeneity between study protocols (test duration, laboratory vs. free-living, reported metrics, statistical analyses, model investigated, etc.) limited conclusions on the overall accuracy of these technologies. One of the issues manufacturers and researchers face is determining, "What is a step?" Does it include the shuffle of the elderly, slide of the tennis player, leap of the dancer, march-in-place of the military person? For laboratory and validation purposes, as long as proprietary algorithms set the criteria for what constitutes a step, validation studies will prove difficult as they may define a step according to different criteria. Need exists for standardized performance benchmarks industries can meet to permit uniform testing (Bassett et al. 2017).

A variety of step counters exists, utilizing diverse internal mechanisms, and placed on different regions of the body. Recent trends, however, favor wrist- and arm-worn activity trackers that utilize triaxial accelerometers to detect movement. While some data support the relationship between wrist-worn devices and observer-counted steps (El-Amrawy and Nounou 2015; Chen et al. 2016), one of the main sources of error is when upper-body movement fails to reflect lower body locomotion. For example, wrist-worn devices may inappropriately record steps when folding laundry (Chen et al. 2016), brushing teeth, or when worn on the dominant vs. non-dominant wrist. Conversely, some steps occur without wrist movement, such as when walking while pushing a stroller (Chen et al. 2016). Under these circumstances, devices fail to appropriately log steps. Every arm movement of daily living is not accompanied by a step, nor vice versa—a discrepancy manufacturers have attempted to overcome. For example, some devices have altered sensitivity thresholds whereby a certain acceleration must be met before a step is counted. This appropriately limits the counting of steps due to subtle wrist movements but fails to count steps in slower-moving or clinical populations. Other devices utilize a filter where movement must be maintained for ≥ 4 s to be counted as a step (Bassett et al. 2017). This method, however, eliminates steps of daily, household movements, which is concerning because frequent, short-duration bouts occur in daily activity (Orendurff et al. 2008). Ultimately, step counters are more accurate when placed at the hip or ankle compared to the wrist [exceptions exist depending on the age of the user and intensity of the activity (Mandigout et al. 2019; Fuller et al. 2020)]. Although device placement in the laboratory often leans on proximity of the variable being measured—such as the heart rate strap worn on the torso or step count accelerometer worn on the waist—consumer preference dictates that functionality and fashion be united, shifting emphasis towards wrist-worn devices.

While some studies (Evenson et al. 2015; Wahl et al. 2017) report a general level of accuracy or reliability for certain devices and under set conditions, recent systematic reviews (Fuller et al. 2020; Henriksen et al. 2020) demonstrate less confidence in the ability of current technology to accurately report step count and raise questions about consistency, even within brands. A recent study (Montes et al. 2020) investigating five different wearable devices under both walking and jogging conditions found that when manual counters recorded ~ 800 steps, wearable devices were off by as much as 50. Another investigation (Wahl et al. 2017) estimated that over the course of a marathon, where 60,000 steps may be taken, some devices would be within 100 steps whereas others off by nearly 8,000. Therefore, while wearable devices appear useful for tracking physical activity and step count, one must consider that error waxes or wanes depending on the type of activity, the intensity, and the environment where exercise occurs. Regarding step count and physical activity tracking, INTERLIVE noted the lack of consistency in how validity is evaluated, and identified various domains that validation studies should consider (Johnston et al. 2021). To confidently fulfill step goals then, users, technology companies, and validation procedures must take additional steps.

Heart rate and heart rate variability

Heart rate (HR) increases with exertion and is used to indicate exercise intensity or as a correlate for VO_2 and physiological stress. Changes in heart rate represent useful signposts for training adaptations—where a lower heart rate for a given workload represents improved cardiorespiratory and muscular fitness. Within the clinical setting, HR is commonly measured via electrocardiogram (ECG), which detects depolarization and repolarization of the heart using electrodes placed on the chest, or by photoplethysmography such as in a pulse oximeter. In healthy populations, and in particular during exercise, HR is monitored via short-range telemetry using an electrical chest strap that transmits data via radiowave to a receiver or watch.

Heart Rate Variability (HRV) describes the time variation between consecutive heartbeats and reflects the health of the autonomic nervous system. A healthy heart displays some amount of time oscillation between beats, whereas a diseased heart exhibits either a metronome-like rhythm, or

abnormal variations between beats (Shaffer and Ginsberg 2017). Changes in HRV may indicate overtraining or illness, and are useful in evaluating training intensities when comparing HRVs between exercise bouts (Hinde et al. 2021). Similar to HR, HRV is quantified using ECG, where algorithms calculate variability between ventricular contractions (Shaffer et al. 2014; Hinde et al. 2021).

The ability of wrist-worn devices to accurately and reliably capture HR has been investigated over a wide range of devices and brands with results demonstrating both accuracy and reliability, inside and outside the laboratory (Düking et al. 2020) Accuracy of HR wavers, however, with alterations in exercise intensity (Spierer et al. 2015; Thiebaud et al. 2018; Müller et al. 2019; Thomson et al. 2019; Chow and Yang 2020). Regarding accuracy and reliability of HRV, a systematic review (Georgiou et al. 2018) examined eighteen studies and found high correlations for wearable HRV and classic ECG at rest. Similar to HR, HRV accuracy and reliability decreased during exercise—a finding supported by a 2021 investigation (Hinde et al. 2021) that examined thirty-two portable devices and found that validity and reliability decreased as HR and exercise intensity increased.

Previously, wearable technologies capable of detecting HR did so via chest strap and associated watch. This technique compared favorably with ECG (Leger and Thivierge 1988; Goodie et al. 2000). Recent devices, however, detect HR and HRV using PPG, although many retain compatibility with HR-measuring chest straps. PPG employs light-emitting diodes and detectors to monitor light absorption in the blood. Transmission of light through the tissue, or reflection of light from the tissue, alter light intensity and are associated with changes in tissue perfusion (Tamura et al. 2014). Changes in perfusion then indicate contraction or relaxation of the heart. Validity of devices that use PPG may be compromised due to a number of scenarios including motion of the extremities, wrist positioning, disturbances in sensor–blood interface (such as sweat), and skin pigmentation. (Spierer et al. 2015) also found that when the devices were evaluated over specific exercise modalities, such as stair climbing and resistance training, reliability waned—a finding confirmed in later research (Shcherbina et al. 2017). A study on trail running similarly demonstrated poor validity with PPG-based HR sensors (Navalta et al. 2020). To optimize the performance of PPG monitoring, most manufacturers recommend routine cleaning of the PPG sensor—wiping excess dirt, sweat, or other debris from the skin—prior to placement, and then securing the sensor tightly to minimize motion. In real-world settings, it can be difficult to achieve these recommendations given that some users sweat heavily during an exercise bout, some run in inclement weather or conditions that impair PPG sensor performance (e.g., trail running and open-water swimming, which may increase the debris and sediment that contact the sensor), and some may experience discomfort when a wrist-worn unit is tightened sufficiently to ensure optimal PPG performance.

HRV can also be detected using PPG (Giles et al. 2016), and although ECG has been considered superior due to clear detection of ventricular contraction, updated algorithms in PPG devices have improved pulse detection. Losses in accuracy associated with the exercising state may be due to a number of physiological and technological factors, such as sympathetic stiffening of the blood vessels and sampling rate of the devices themselves (Shaffer et al. 2014). Finally, because HRV exhibits variation between people, baseline values for HRV must be established using resting, personal data prior to the influence of other stimuli. Otherwise, HRV values by themselves lack meaningful interpretation.

Daily fluctuations in body temperature, circulating hormones, sleep cycle, and metabolism contribute to HRV, and gold standard techniques in HRV assessment account for these by taking a 24 h HRV recording. Many wearable devices lack the battery capacity for 24 h measurement and rely instead upon short-term (5 min) data acquisition. When short-term measurements are used, recommendations call for resting measurements, which influences the use of short-term HRV when exercising (Shaffer and Ginsberg 2017). HRV can also change based on age, sex, health, and HR itself (Shaffer and Ginsberg 2017). Therefore, algorithms for HRV and especially the diagnostic outcome of HRV measurement would necessarily take these variables into account. When using short-term recordings (compared with 24 h recordings), recommendations suggest utilizing frequency domains instead of time interval domains (Malik et al. 1996). Therefore, if wearable algorithms utilize short-term recordings (which is likely) they will be most effective if transforming the data into domains based on frequency. Though short-term recordings refer to HRV measurement over a 5 min span, even shorter recordings ($\leq$1 min) have become more prevalent. The usefulness of these snapshot views remains to be assessed and is not endorsed by leading scientific societies (Hinde et al. 2021). Once again, the INTERLIVE collaboration between universities and industry offers a best-practice protocol for validating PPG devices, which includes considering instrument placement, target population, and testing conditions amongst other considerations (Mühlen et al. 2021). Ultimately, while wearables may provide a general overview of fitness and health, future devices will permit a 24 h view of HRV and demonstrate more consistency with ECG technology, especially during heavy exercise sessions.

Training load, sleep, and stress

Health and performance hang from a tenuous line separating undertraining from overtraining. Individuals—coaches, athletes, or those pursuing health and wellbeing—seek to

promote stress-induced adaptation while avoiding any injury or chronic fatigue elicited by the over-trained or overstressed state. While appropriate physical stress from exercise promotes health, additional sources of stress (be it physical, social, environmental, etc.) can compile and create an overburdened state with an elevated risk of injury or illness (Hamlin et al. 2019). Monitoring training stress—the physiological strain resulting directly from training sessions (Paquette et al. 2020)—represents a useful diagnostic technique to prevent harm and optimize adaptation. The quantification of training load considers both external loads (e.g., speed, distance, duration, work) and internal loads (e.g., HR, HRV, blood lactate) during a given time. Alongside exercise-specific stressors, the monitoring of training load may also consider sleep quantity and quality as a potential diagnostic for assessing physiological strain. Finally, daily stress resulting from non-training factors, such as work, relationships, financial stress, etc., may also contribute to overall training load and can be assessed through visual analog scales and questionnaires. In the laboratory, the gold standard for sleep tracking is polysomnography, which records cortical and electro-ocular activity via electrodes on the scalp (De Zambotti et al. 2019). Given the vast number of contributors to physiological status and overall stress, however, no gold standard for training load quantification currently exists.

For training load to be meaningful it must relate to an outcome, such as injury occurrence or performance, and investigations must validate the strength of this relationship. Limited data assess the dose–response relationship between training load measurements and training outcomes. One investigation (Sanders et al. 2017) assessed the relationships between various measures of training load and aerobic fitness in trained cyclists and found the strongest dose–response relationship when measures considered individual characteristics in the calculation of training load rather than assuming a one-size-fits-all approach with pre-established exercise intensity levels. Similar to many of the previously-discussed variables, the rapid rate of industry development and deployment, combined with a paucity of performance standards, means that assessment of accuracy and reliability struggles to keep pace with emerging techniques for quantifying training load (Passfield et al. 2022). A 2018 review of consumer wearables for monitoring stress and sleep found that 5% of technologies had been formally validated (Peake et al. 2018). Need exists for validation of training load algorithms that are sport-specific with thresholds that technology companies must meet to promote their devices as accurately assessing training load. Recently, a group investigating the use of sleep-tracking devices (Menghini et al. 2021) suggested one such evaluation technique. Further efforts such as this will promote validity and reliability of future technology.

Technological advancement in wearable devices enables the acquisition of massive datasets which opens new opportunities for insight into health and disease but also creates new challenges in determining how best to use data in assessing these outcomes. Regarding training load, wrist-worn devices reliably track days, distance, duration of exercise, as well as speed, cadence, and even surrogates for ground reaction forces (Moore and Willy 2019). These data are then combined and weighted to assess training load. Several quantification methods have been proposed, each relying upon a unique combination of variables, and each prescribing different weight to those variables. For example, "training impulse" uses exercise duration and mean heart rate during exercise to assess the intensity and overall load of the exercise session. However, mean work rate or total accumulated work may not be truly reflective of the stress of an individual exercise bout. Time spent in different intensity domains, particularly at high intensity, may have little effect on overall average heart rate, but may induce significant training stress (Black et al. 2017; Clark et al. 2018). A different method calculates exercise intensity based on an individual's heart rate and blood lactate relationship. Still others utilize session rating of perceived exertion, where athletes rate their perceived level of difficulty for a given bout of exercise. Sleep patterns may also reflect the stress of training, and sleep is included in some quantifications of training load by integrating movement data from accelerometers with cardiac and environmental information (Menghini et al. 2021). Training load can also be sport-specific, and the development of power meters in cycling enables power output to be included in training load calculations for cyclists.

In assessing the utility of sleep-tracking devices, Menghini et al. (2021) state that "unstandardized, undisclosed, and often unvalidated data outcomes and algorithms are among the main challenges." This sentiment can be expanded to the current state of training load monitoring in general as we remain uncertain of which variables best identify the level of stress incurred on an individual. In addition to validation standards then, advancements in training load quantification will establish best practices for which variables to rely upon (and how heavily). In the future, additional variables may also emerge, and technology will take an individualized approach to training load, where a certain distance and pace for one person incurs a different training load compared to that same distance and pace in another person. Technology may also benefit from an ability to sieve through the more general term, "stress" and pinpoint anatomical, site-specific load (Moore and Willy 2019). Until a method exists in which both internal and external loads are reliably accounted for, the calculation of training load by wearable devices remains deficient.

Qualitative synthesis and expert opinion

It should be clear that wearable activity trackers have advanced rapidly over the past decade and continue to feature new capabilities and metrics of interest to the end users. Improvements in sensors, engineering, algorithms, and incorporating some level of individual data (such as height, weight, activity/training level, etc.) have all expanded and enhanced the capabilities of these devices (Düking et al. 2018). Yet, for all the advances that have been made, significant gaps still exist in validity and reliability, particularly in real-world performance circumstances. Because of the variety in manufacturer models, algorithms, and even the advertised purpose of the wearable device, this topic is quite broad and factors, such as the relatively small number of studies and lack of transparency from the manufacturers make it difficult to draw rigorous scientific conclusions. Nevertheless, recent meta-analyses on specific metrics along with available validation studies provide some understanding of the validity and reliability of these devices. In reviewing the available data on measures, such as $\dot{V}O_2$, EE, step count, and others, several themes emerge.

First, while some measurements appear to be valid and reliable in more controlled circumstances, such as at rest or low-intensity activity, under more dynamic conditions, such as high-intensity exercise, validity and reliability, are significantly lower, and likely undependable. This issue is particularly problematic given the application of these devices as training tools and aids to improve health and fitness. Low confidence in these metrics suggests that athletes and other users should not build their training and dietary plans around the $\dot{V}O_2$, EE, training load, and other measures that are given by these wearable devices. Doing so may present significant risk of over- or underestimation and could result in imbalanced training load, energy imbalance, and other factors ultimately impairing performance, health, and fitness.

Second, the proprietary nature of algorithms used to estimate these metrics impedes validation testing and likely slows innovation. This presents a challenge because independent validation of algorithms would significantly improve consumer and scientific confidence in how these calculations are made, and whether assumptions used in these measures are indeed valid and applicable to the population or individual using these devices. Moreover, open-source, or otherwise publicly available data could help drive innovation and improvement in these algorithms and enhance transparency and consumer confidence. Systematic reviews of individual metrics have shown that although some algorithms perform well at the population level, the estimation error at the individual level is still large and thus, do not provide useful information for the end-user of most consumer-wearable devices. One possibility is that a "one-size-fits-all" algorithm that is universally applied to all users is not nuanced enough to accurately estimate these metrics at the individual level. Future research should focus on not only validating algorithms at the population level, but also reducing estimation error at the individual level.

Third, accounting for individual differences and environmental factors is both complex and challenging, yet essential for the best performance of wearable devices. Adapting different algorithms to specific sub-populations based on age, sex, training history, presence or absence of cardiometabolic and/or respiratory disease, race, ethnicity, body composition, etc. may yield results more useful to a broader audience. Moreover, estimates are almost certainly impacted by inter-individual variability in factors, such as running economy (or cycling efficiency, or economy/efficiency in other exercise modalities), actual vs. predicted maximal HR, muscle fiber type distribution, $\dot{V}O_2$ slow component and cardiac drift during prolonged exercise, and biomechanical, musculoskeletal, and neurophysiological factors that influence stress, recovery and injury, many of which are not properly accounted for in algorithms used to estimate common metrics.

Acknowledging the complexity of these challenges, future development should encourage more robust testing and transparency. Particularly in light of the recent COVID-19 pandemic, which necessitated many people to monitor their own exercise regimens to a greater extent due to loss of organized exercise classes, gym access, face to face coaching, etc., it seems prudent to hold these wearable activity tracking devices to a higher standard. It would seem logical to develop a standard analogous to a "Phase 3" clinical trial validating the efficacy and safety of medical therapies and devices and indeed the International Federation of Sports Medicine recently published a perspective proposing the establishment of a global standard for wearable devices in sport and fitness (Ash et al. 2020). The proposed quality assurance standard would commission testing of marketing claims and endorse the claims that are validated. However, practical constraints, such as budget, facilities, and human resources, necessary for testing and validation may make widespread adoption of such a standard difficult. Ultimately, the nature of this field of wearable activity trackers is dynamic, one of constant change, updates, redesign, and new models, and will necessitate continual review and research. In light of this, it is likely that the science will always be lagging to supporting new devices emerging into the market. If so, this begs the question: Is it the role of scientists to validate the claims of manufacturers? Rather than place the burden on the scientific community, manufacturers should be encouraged to provide full transparency for the algorithms

they use in their wearable devices with rigorous, transparent, and complete reporting of algorithm development, validation, and real-world testing.

Interpretation for lay public/practical application

Highlighting the validity and reliability challenges in wearable technology then begs the questions, "How are people utilizing this technology?" and "What are the general levels of acceptance for metrics reported by these devices?" According to a 2019 survey, the majority of wearable technology users turn to their devices for tracking step count (60%). Heart rate (44%), calories burned (42%), and sleep monitoring (40%) represented the next highest metrics (Global Web Index 2020). Figure 2 provides a normalized representation of average acceptance for various metrics. Users mentioned managing their fitness levels (47%) and feeling in control of their health (45%) as the primary reasons they use this technology (Global Web Index 2020). Indeed, this ability to collect activity data supports one's adoption of wearable technology (Canhoto and Arp 2017). Additionally, users expressed a desire to track more health parameters in the future, with blood pressure (53%), heart rate (51%), blood sugar (45%), cardiac issues (42%), and stress issues (42%) at the top of the list (Global Web Index 2020).

Wearable technology is here to stay (Market Research Future; Pew Research Center; SurveyMonkey; Thompson), which likely means that users will continue collecting health- and fitness-related data, but will also desire assistance in interpreting these data. In fact, over half of survey respondents in 2014 were willing to share their data with a physician (Accenture Interactive). In another survey, two thirds of respondents expressed a desire for a physician or health coach to guide them in understanding their wearable technology data in making lifestyle changes. Three fourths of respondents were willing to pay for this service (The Harris Poll). Thus, as the use of wearable technology grows more widespread, and as companies producing these devices add features that generate even more data, health professionals will likely find their patients or clients asking them for recommendations based on self-collected health and fitness data. But if data are unreliable and lack validity, health professionals may struggle in guiding patients or clients to the next best steps for their health and fitness goals. Figure 3 summarizes the Mean Absolute Percentage Error for various physiological variables recorded by consumer-wearable devices reported in recent studies.

Metric	General Acceptance
VO_2	✓✓
VO_2max	✓
Energy Expenditure	✓✓
Step Count	✓✓✓
Heart Rate	✓✓✓✓
Heart Rate Variability	✓✓
Training Load	✓
Sleep & Stress	✓

Fig. 2 Normalized representation of average metric acceptance. Consensus is taken from expert opinion and other literature. ✓ = not generally accepted. ✓✓✓✓ = widely accepted

So, what can be done? Perhaps this is the push the wearable technology industry needs (as the previous section mentioned) to create standards for new devices and capabilities (Cardinale and Varley 2017), as well as marketing what those true capabilities are. Additionally, the health field will need to devise best practices so providers know how to handle the data their patients or clients present to them. This could include comparing the wearable technology data with already accepted measures of health and fitness (e.g., comparing time and distance walked using the wearable device with the 6-Minute Walk Test norms chart) (Chiauzzi et al. 2015).

Furthermore, validity and reliability issues may be addressed if wearable technology users and their health providers look at data trends over time, instead of simply as vital signs taken in the moment. For instance, if a wearable technology user starts experiencing a higher-than-normal resting HR, this could signal overtraining and/or possible sickness, whereas receiving a resting heart rate measurement of 80 bpm in the doctor's office only indicates that the patient is not currently experiencing tachycardia or bradycardia. Indeed, studies show that deviations in one's wearable technology metrics over time can signal disease, or risk for disease (Li et al. 2017; Rose et al. 2019; Lown et al. 2020). Currently, although wearable devices can track changes over time, wearable devices cannot detect what is a meaningful change. Rather, various algorithms have been created and must be applied to find anomalies in the user's generated data to predict future illness (Lown et al. 2020; Alavi et al. 2021; Sunny et al. 2022). These algorithms require multiple steps to be useful, including first downloading the data to an application programming interface (API), processing it for uniformity, and filling in missing data points (Sunny et al. 2022). For healthcare providers to accurately predict disease risk based on deviations in user's wearable technology data, current restrictions require them to employ the use of

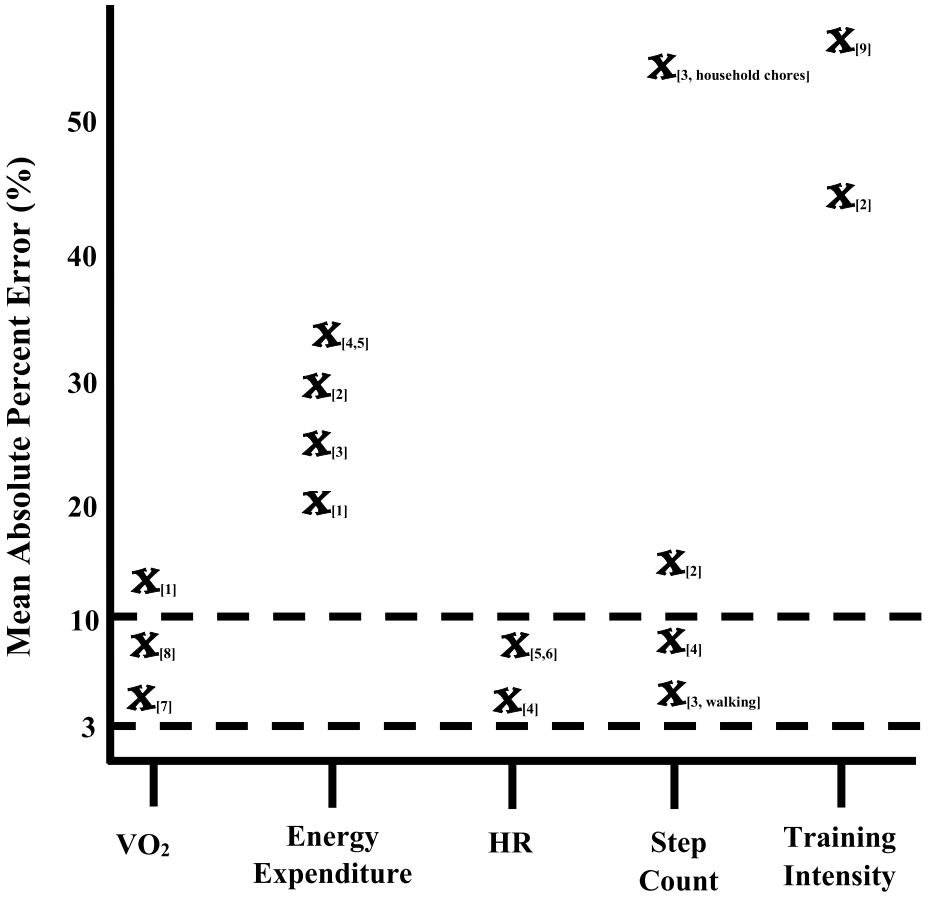

Fig. 3 Mean absolute percentage error (MAPE) for various physiological variables recorded by wearable devices from recent investigations. MAPE indicates the predictive accuracy of devices. Although no standardized thresholds exist for high or low error, MAPE > 3% has been considered high for laboratory-based studies and > 10% has been considered high for studies in free-living conditions. Note: studies normally found a *range* for MAPE. Therefore, "X" indicates the approximate average of various devices and scenarios tested. 1, Passler (2019); 2, Henriksen (2020); 3, Nelson (2016); 4, Montoye (2017); 5, Wallen (2016); 6, Navalta (2020); 7, LeBoeuf (2014); 8, Carrier (2020); 9, Henriksen (2021)

these relevant algorithms, which are external to the wearable device itself (Guk et al. 2019; Sunny et al. 2022).

If consumers and their health providers are monitoring changes over time, this is a path to greater personalized health and fitness training (Li et al. 2017; Rose et al. 2019). As stated earlier, $\dot{V}O_{2_{max}}$ estimation, HRV, and training load become more accurate when an individualized, and not a one-size-fits-all, approach is used. Designers of wearable technology could create a more robust, yet user-friendly, set-up process for new purchases, making it more individualized to the wearer. This could include showing users pictures of various types of exercise, allowing them to swipe through options, thereby obtaining a closer look at their true physical activity levels. For example, apps such as the "Diet ID" have recently been developed using this method, giving health providers a more accurate view of patient or client nutrition. Figure 4 provides a comparison between how users are relying upon consumer-wearable devices along with an expert recommendation on how best to utilize these devices.

If wearable technology companies provide consumers with more individualized capabilities, this would improve the generated outputs and represent an additional marketing tool to boost sales. It would also enable the identification of a specific baseline of health for each user so deviations that signal disease could be found more easily (Guk et al. 2019). Furthermore, creating wearables with the ability to track data changes themselves would simplify data processing and make these devices more valuable to the health field. While consumers and health professionals should continue to push the wearable technology industry toward greater standards for reliability and validity, improving data tracking and implementing greater individualization can provide a win for all stakeholders.

Summary and conclusion

Wearable technologies are powerful tools for health and fitness and have become indispensable training tools for athletes of all levels. Yet, for all their merits, significant limitations exist, primarily related to the validity and reliability of the metrics these devices purport to measure. The rapid rate of development and deployment of new technologies, sensors, algorithms, and other components of these devices may lead athletes and other users to believe that these sophistications are highly accurate and

	Current Practice	Expert Advice	Expert Rating
$\dot{V}O_2$ and $\dot{V}O_2$max	• Higher use among athletes/amateur athletes • Superfluous feature for lay person	• More individualization needed for better accuracy • Chance of over/underestimation • Explanation for lay public needed on what this measure is and why it's important to them	★★☆☆☆
Energy Expenditure	• 47% use wearables to manage their fitness [2] • 42% use wearables as a calorie burner [2] • Ability to collect activity data is a major factor for intention to adopt a wearable [3]	• Chance of over/underestimation • With options for goal setting, reminders to move, and social support, clinicians may find wearables useful to help patients initially monitor calories consumed and physical activity completed to determine behavioral changes needed [6]	★★★☆☆
Step Count	• 60% use wearables to track steps [2] • Ability to collect activity data is a major factor in intention to adopt a wearable [3]	• Differences in accuracy between and within brands depending on activity • Helpful to track steps over time to attempt to exceed user's personal best • Could be demotivating if not reaching the preset number of steps [7]	★★★☆☆
Heart Rate (HR)	• 44% use wearables to track HR [2]	• Accuracy decreases with increasing exercise intensity • Accuracy best with chest strap • Helpful to track resting HR over time to monitor for sickness or overtraining • Helpful to monitor HR over time doing similar activities to see if any deviations from the norm [4,5]	★★★★☆

Fig. 4 Comparison between how users are relying upon wearable devices with expert recommendation and 5-star rating on how best to rely upon wearable devices. 5-star, excellent reliability; 4-star, good reliability; 3-star, moderate reliability; 2-star, poor reliability; 1-star, very poor reliability. 1, Evenson (2015); 2, Global Web Index (2020); 3, Canhoto (2017); 4, Rose (2019); 5, Li (2017); 6, McDonough (2021); 7, Kerner (2017); 8, Montgomery-Downs (2012)

Heart Rate Variability	• Superfluous feature for lay person • Desire to track cardiac data, but want BP and HR rather than HRV [2]	• More individualization needed for better accuracy • Accuracy best with chest strap • Users' HRV should be determined over 24 hours, but currently many wearables lack the battery power • Explanation for lay public needed on what this measure is and why it's important to them	★★☆☆☆
Training Load	• Higher use among athletes/amateur athletes • 47% use wearables to manage their fitness [2]	• More individualization needed for better accuracy • Chance of over/underestimation • Could be more accurate if training load estimations were sport-specific • Unless working with a trainer or coach, the lay person doesn't know how much training they should be doing for their goals/lifestyle; risk of injury greater if relying on wearables alone	★☆☆☆☆
Sleep	• 40% use wearables to track sleep [2]	• Currently lack validating standards • Overestimate hours slept and underestimate wakeful moments [1,8] • Helpful to compare data from several nights of sleep and record activities performed and food/drink consumed prior to sleep, as well as next day energy levels then make changes as necessary	★★★☆☆

Fig. 4 (continued)

valid, yet published data suggest this may be the case in limited circumstances, primarily at rest and during low-intensity activity. These limitations make it difficult to be confident in metrics the end-user is seeing from these devices and challenging for the practitioner to interpret the meaning of the data generated by these devices. Improving transparency in development and validation of these metrics, along with better tailoring to individuals should increase the validity and reliability of these devices. While it is acknowledged that no technology or device performs perfectly under all circumstances, the breakdown in performance of many wearable devices in certain real-world settings begs the question, are they more advanced technology or advanced marketing to the end user? Scientists and practitioners alike would do well to remember that good science is often not good marketing; and conversely

(and perhaps more importantly), good marketing is not always good science.

Author contributions RJS and HLP conceived of the idea for the work; RJS, IGH, ASO, BAP, HLP drafted the manuscript and critically revised the work; all authors reviewed and approved the final manuscript.

Funding No funding was received for this work.

Declarations

Conflict of interest The authors declare no conflicts of interest. Ren-Jay Shei is an employee of Coherus BioSciences. The work described herein is solely reflective of the author's (R-JS) personal views and is unrelated to his job duties with Coherus BioScience. These views do not constitute an endorsement by Coherus BioSciences, do not represent the views of Coherus BioSciences, and Coherus BioSciences had no role in the conception, writing, revision, or final approval of the manuscript. Brittni Paris is co-owner of Smart Fit Womxn LLC, a wellness coaching company. The work described here is BAP's personal view and does not reflect the view of Smart Fit Womxn.

References

Accenture Interactive Consumer Privacy Concerns for Wearable Tech

Acheson K, Campbell I, Edholm O et al (1980) The measurement of daily energy expenditure—an evaluation of some techniques. Am J Clin Nutr 33:1155–1164

Ainslie PN, Reilly T, Westerterp KR (2003) Estimating human energy expenditure. Sports Med 33:683–698

Alavi A, Bogu GK, Wang M et al (2022) Real-time alerting system for COVID-19 and other stress events using wearable data. Nat Med 28:175–184

Argent R, Hetherington-Rauth M, Stang J et al (2022) Recommendations for determining the validity of consumer wearables and smartphones for the estimation of energy expenditure: expert statement and checklist of the INTERLIVE network. Sports Med. https://doi.org/10.1007/s40279-022-01665-4

Ash GI, Stults-Kolehmainen M, Busa MA et al (2020) Establishing a global standard for wearable devices in sport and fitness: perspectives from the New England chapter of the American college of sports medicine members. Curr Sports Med Rep 19:45–49

Bassett DR, Toth LP, LaMunion SR, Crouter SE (2017) Step counting: a review of measurement considerations and health-related applications. Sports Med 47:1303–1315

Black MI, Jones AM, Blackwell JR et al (2017) Muscle metabolic and neuromuscular determinants of fatigue during cycling in different exercise intensity domains. J Appl Physiol 122:446–459

Bonomi AG, Plasqui G, Goris AH, Westerterp KR (2009) Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. J Appl Physiol 107:655–661

Bonomi AG, Ten Hoor GA, De Morree HM et al (2020) Cardiorespiratory fitness estimation from heart rate and body movement in daily life. J Appl Physiol 128:493–500

Boudreaux BD, Hebert EP, Hollander DB et al (2018) Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 50:624–633

Brage S, Wedderkopp N, Franks PW et al (2003) Reexamination of validity and reliability of the CSA monitor in walking and running. Med Sci Sports Exerc 35:1447–1454

Canhoto AI, Arp S (2017) Exploring the factors that support adoption and sustained use of health and fitness wearables. J Mark Manag 33:32–60

Cardinale M, Varley MC (2017) Wearable training-monitoring technology: applications, challenges, and opportunities. Int J Sports Physiol Perform 12:S2-55

Carrier B, Creer A, Williams LR et al (2020) Validation of garmin fenix 3 HR fitness tracker biomechanics and metabolics (VO2max). J Meas Phys Behav 3:331–337

Chen M-D, Kuo C-C, Pellegrini CA, Hsu M-J (2016) Accuracy of wristband activity monitors during ambulation and activities. Med Sci Sports Exerc 48:1942–1949

Chiauzzi E, Rodarte C, DasMahapatra P (2015) Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 13:1–6

Choi BC, Pak AW, Choi JC (2007) Daily step goal of 10,000 steps: a literature review. Clin Invest Med 30:E146–E151

Chow H-W, Yang C-C (2020) Accuracy of optical heart rate sensing technology in wearable fitness trackers for young and older adults: validation and comparison study. JMIR MHealth UHealth 8:e14707

Clark IE, Vanhatalo A, Bailey SJ et al (2018) Effects of two hours of heavy-intensity exercise on the power-duration relationship. Med Sci Sports Exerc 50:1658–1668

De Zambotti M, Cellini N, Goldstone A et al (2019) Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 51:1538

Dijkstra B, Zijlstra W, Scherder E, Kamsma Y (2008) Detection of walking periods and number of steps in older adults and patients with Parkinson's disease: accuracy of a pedometer and an accelerometry-based method. Age Ageing 37:436–441

Düking P, Fuss FK, Holmberg H-C et al (2018) Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR MHealth UHealth 6:e9341

Düking P, Giessing L, Frenkel MO et al (2020) Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: validation study. JMIR MHealth UHealth 8:e16716

El-Amrawy F, Nounou MI (2015) Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 21:315–320

Evenson KR, Goto MM, Furberg RD (2015) Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 12:1–22

Firstbeat Technologies Ltd. Firstbeat Technologies White Paper: Automated Fitness Estimation

Firstbeat Technologies Ltd. Firstbeat Technologies White Paper: VO2 estimation

Fothergill E, Guo J, Howard L et al (2016) Persistent metabolic adaptation 6 years after "the biggest loser" competition. Obesity 24:1612–1619

Fudge BW, Wilson J, Easton C et al (2007) Estimation of oxygen uptake during fast running using accelerometry and heart rate. Med Sci Sports Exerc 39:192–198

Fuller D, Colwell E, Low J et al (2020) Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR MHealth UHealth 8:e18694

Georgiou K, Larentzakis AV, Khamis NN et al (2018) Can wearable devices accurately measure heart rate variability? a systematic review. Folia Med (plovdiv) 60:7–20

Giles D, Draper N, Neil W (2016) Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol 116:563–571
Gillinov S, Etiwy M, Wang R et al (2017) Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 49:1697–1703
Global web index (2020) Digital healthcare report: understanding the evolution and digitization of healthcare
Goodie JL, Larkin KT, Schauss S (2000) Validation of polar heart rate monitor for assessing heart rate during physical and mental stress. J Psychophysiol 14:159
Guk K, Han G, Lim J et al (2019) Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials 9:813
Hamlin MJ, Wilkes D, Elliot CA et al (2019) Monitoring training loads and perceived stress in young elite university athletes. Front Physiol 10:34
Haskell WL, Yee MC, Evans A, Irby PJ (1993) Simultaneous measurement of heart rate and body motion to quantitate physical activity. Med Sci Sports Exerc 25:109–115
Hatano Y (1993) Use of the pedometer for promoting daily walking exercise. ICHPER 29:4–8
Helm MM, Carrier B, Davis DW et al (2021) Validation of the garmin fenix 6S maximal oxygen consumption (VO2max) estimate. Int J Exerc Sci: Conf Proceed 14:29
Henriksen A, Johansson J, Hartvigsen G et al (2020) Measuring physical activity using triaxial wrist worn polar activity trackers: a systematic review. Int J Exerc Sci 13:438
Hinde K, White G, Armstrong N (2021) Wearable Devices suitable for monitoring twenty four hour heart rate variability in military populations. Sensors 21:1061
Johnston W, Judice PB, García PM et al (2021) Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. Br J Sports Med 55:780–793
Johnstone AM, Murison SD, Duncan JS et al (2005) Factors influencing variation in basal metabolic rate include fat-free mass, fat mass, age, and circulating thyroxine but not sex, circulating leptin, or triiodothyronine. Am J Clin Nutr 82:941–948
Kendall B, Bellovary B, Gothe NP (2019) Validity of wearable activity monitors for tracking steps and estimating energy expenditure during a graded maximal treadmill test. J Sports Sci 37:42–49
Kenny GP, Notley SR, Gagnon D (2017) Direct calorimetry: a brief historical review of its use in the study of human metabolism and thermoregulation. Eur J Appl Physiol 117:1765–1785
Leger L, Thivierge M (1988) Heart rate monitors: validity, stability, and functionality. Phys Sportsmed 16:143–151
Li X, Dunn J, Salins D et al (2017) Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 15:e2001402
Lown M, Brown M, Brown C et al (2020) Machine learning detection of Atrial Fibrillation using wearable technology. PLoS One 15:e0227401
Malik M, Bigger JT, Camm AJ et al (1996) Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Eur Heart J 17:354–381
Mandigout S, Lacroix J, Perrochon A et al (2019) Comparison of step count assessed using wrist-and hip-worn Actigraph GT3X in free-living conditions in young and older adults. Front Med. https://doi.org/10.3389/fmed.2019.00252
Market research future heart rate monitor market overview
Menghini L, Cellini N, Goldstone A et al (2021) A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep 44:zsaa170
Molina-Garcia P, Notbohm HL, Schumann M et al (2022) Validity of estimating the maximal oxygen consumption by consumer wearables: a systematic review with meta-analysis and expert statement of the INTERLIVE network. Sports Med. https://doi.org/10.1007/s40279-021-01639-y
Montes J, Tandy R, Young J et al (2020) Step count reliability and validity of five wearable technology devices while walking and jogging in both a free motion setting and on a treadmill. Int J Exerc Sci 13:410
Moore IS, Willy RW (2019) Use of wearables: tracking and retraining in endurance runners. Curr Sports Med Rep 18:437–444
Mühlen JM, Stang J, Skovgaard EL et al (2021) Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE network. Br J Sports Med 55:767–779
Müller AM, Wang NX, Yao J et al (2019) Heart rate measures from wrist-worn activity trackers in a laboratory and free-living setting: validation study. JMIR MHealth UHealth 7:e14120
Murakami H, Kawakami R, Nakae S et al (2019) Accuracy of 12 wearable devices for estimating physical activity energy expenditure using a metabolic chamber and the doubly labeled water method: validation study. JMIR MHealth UHealth 7:e13938
Navalta JW, Montes J, Bodell NG et al (2020) Concurrent heart rate validity of wearable technology devices during trail running. Plos One 15:e0238569
O'Driscoll R, Turicchi J, Beaulieu K et al (2020) How well do activity monitors estimate energy expenditure? a systematic review and meta-analysis of the validity of current technologies. Br J Sports Med 54:332–340
Omura JD, Carlson SA, Paul P et al (2017) National physical activity surveillance: users of wearable activity monitors as a potential data source. Prev Med Rep 5:124–126
Orendurff MS, Schoen JA, Bernatz GC et al (2008) How humans walk: bout duration, steps per bout, and rest duration. J Rehabil Res Dev 45:1077–1090
Paquette MR, Napier C, Willy RW, Stellingwerff T (2020) Moving beyond weekly "distance": optimizing quantification of training load in runners. J Orthop Sports Phys Ther 50:564–569
Passfield L, Murias JM, Sacchetti M, Nicolò A (2022) Validity of the training-load concept. Int J Sports Physiol Perform 17:507–514
Passler S, Bohrer J, Blöchinger L, Senner V (2019) Validity of wrist-worn activity trackers for estimating VO2max and energy expenditure. Int J Env Res Public Health 16:3037
Peake JM, Kerr G, Sullivan JP (2018) A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 9:743
Pew research center about one-in-five Americans use a smart watch or fitness tracker
Plasqui G, Westerterp KR (2007) Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity 15:2371–2379
Plasqui G, Bonomi AG, Westerterp KR (2013) Daily physical activity assessment with accelerometers: new insights and validation studies. Obes Rev 14:451–462
Polar tech support
Riel H, Rathleff CR, Kalstrup PM et al (2016) Comparison between Mother, ACTIGRAPH wGT3X-BT, and a hand tally for measuring steps at various walking speeds under controlled conditions. PeerJ 4:e2799
Roos L, Taube W, Beeler N, Wyss T (2017) Validity of sports watches when estimating energy expenditure during running. BMC Sports Sci Med Rehabil 9:1–8
Rose SMS-F, Contrepois K, Moneghetti KJ et al (2019) A longitudinal big data approach for precision health. Nat Med 25:792–804
Sanders D, Abt G, Hesselink MK et al (2017) Methods of monitoring training load and their relationships to changes in fitness and

performance in competitive road cyclists. Int J Sports Physiol Perform 12:668–675
Shaffer F, Ginsberg JP (2017) An overview of heart rate variability metrics and norms. Front Public Health 5:258
Shaffer F, McCraty R, Zerr CL (2014) A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 5:1040
Shcherbina A, Mattsson CM, Waggott D et al (2017) Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 7:3
Snyder NC, Willoughby CA, Smith BK (2021) Comparison of the polar V800 and the garmin forerunner 230 to predict VO2max. J Strength Cond Res 35:1403–1409
Spierer DK, Rosen Z, Litman LL, Fujii K (2015) Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 39:264–271
Støve MP, Haucke E, Nymann ML et al (2019) Accuracy of the wearable activity tracker garmin forerunner 235 for the assessment of heart rate during rest and activity. J Sports Sci 37:895–901
Strain T, Wijndaele K, Dempsey PC et al (2020) Wearable-device-measured physical activity and future health risk. Nat Med 26:1385–1391
SurveyMonkey consumer health and wellness: What it means to brands in 2021
Sunny JS, Patro CPK, Karnani K et al (2022) Anomaly Detection framework for wearables data: a perspective review on data concepts. Data Anal Algorithms Prospects Sens 22:756
Tamura T, Maeda Y, Sekine M, Yoshida M (2014) Wearable photoplethysmographic sensors—past and present. Electronics 3:282–302
The Harris poll trends in wearable fitness
Thiebaud RS, Funk MD, Patton JC et al (2018) Validity of wrist-worn consumer products to measure heart rate and energy expenditure. Digit Health 4:2055207618770322
Thompson WR (2018) Worldwide survey of fitness trends for 2019. Acsms Health Fit J 22:10–17
Thompson WR worldwide survey of fitness trends for 2021. ACSM's health and fitness journal. (2021)
Thomson EA, Nuss K, Comstock A et al (2019) Heart rate measures from the apple watch, fitbit charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 37:1411–1419
US patent application publication firstbeat technologies US. Patent
Wahl Y, Düking P, Droszez A et al (2017) Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 8:725
Winter S, Gordon S, Watt K (2016) Effects of fatigue on kinematics and kinetics during overground running: a systematic review. J Sports Med Phys Fitness 57:887–899
Zhang Z, Khatami R (2022) Can we trust the oxygen saturation measured by consumer smartwatches? Lancet Respir Med. https://doi.org/10.1016/S2213-2600(22)00103-5