EUROPEAN JOURNAL OF SPORT SCIENCE
2023, VOL. 23, NO. 2, 165–177
https://doi.org/10.1080/17461391.2021.2023656

Routledge
Taylor & Francis Group

Check for updates

Wrist-worn devices for the measurement of heart rate and energy expenditure: A validation study for the Apple Watch 6, Polar Vantage V and Fitbit Sense

Guy Hajj-Boutros, Marie-Anne Landry-Duval*, Alain Steve Comtois, Gilles Gouspillou and Antony D. Karelis

Department of Exercise Science, Université du Québec à Montréal, Montreal Canada

ABSTRACT

The purpose of this study was to investigate the accuracy of 3 recently released wrist-worn devices (Apple Watch 6, Polar Vantage V and Fitbit Sense) for heart rate and energy expenditure during various activities. The study population consisted of 60 young healthy individuals (30 men and 30 women; age: 24.9 ± 3.0 years, BMI: 23.1 ± 2.7 kg/m2). Heart rate and energy expenditure were measured using the Polar H10 and Metamax 3B, respectively (reference measures) as well as with the 3 wrist-worn devices during 5 different activities (sitting, walking, running, resistance exercises and cycling). The Apple Watch 6 displayed the highest level of accuracy for heart rate measurement with a coefficient of variation (CV) (%) of less than 5% for all 5 activities, whereas the Polar Vantage V and the Fitbit Sense presented various degrees of accuracy (from high to poor accuracy) dependent on the activity (CVs between 2.44-8.80% and 4.14-10.76%, respectively). As for energy expenditure, all 3 devices displayed poor accuracy for all 5 physical activities (CVs between 14.68-24.85% for Apple Watch 6, 16.54-25.78% for Polar Vantage V and 13.44-29.66% for Fitbit Sense). Results of the present study indicate that the Apple Watch 6 was the most accurate for measuring heart rate across all 5 activities, whereas variable levels of accuracy for heart rate measurement for the Polar Vantage V and the Fitbit Sense were observed depending on the activity. As for energy expenditure, all 3 devices showed poor accuracy during all activities.

Highlights

- The Apple Watch 6 was the most accurate for measuring heart rate, whereas the Polar Vantage V and Fitbit Sense showed variable results dependent on the activity
- The Apple Watch 6, Polar Vantage V and Fitbit Sense showed poor accuracy for energy expenditure during 5 different physical activities
- Healthcare care professionals, athletes/coaches and the general population may want to proceed with caution on the clinical utility of energy expenditure of these devices during the implementation of an exercise training or nutritional programme.

KEYWORDS

Apple Watch 6; Polar Vantage V; Fitbit Sense; metamax 3B and physical activities

Introduction

It is important in clinical, exercise science, nutritional and epidemiological research to develop simple, practical and accurate devices for the measurement of energy expenditure and heart rate since they play a central role in the planning and implementation of physical activity and nutritional intervention programmes (Donnelly et al., 2009). It is essential to monitor heart rate accurately since this marker may be used to identify an individual's level of intensity during an exercise, and in turn, this could play an important role for exercise prescription and health outcomes. Furthermore, assessing energy expenditure using techniques that are reliable, non-invasive and inexpensive remains a major challenge in the field of exercise science. Currently, the exact measurement of energy expenditure requires laboratory methods that are not ideal for performing clinical/field research. For example, indirect calorimetry involves the wearing of a face mask, whereas the doubly labelled water technique is expensive and requires analyzing urine samples (Hills, Mokhtar, & Byrne, 2014).

Wrist-worn devices have been gaining popularity in the last few years from fitness coaches, athletes and the general population (Thompson, 2019). Several international companies such as Apple, Polar, and Fitbit have been competing in this rapidly growing market by regularly designing new devices using continuously advanced technologies. However, despite new technological advancements, a previous study using 3 popular wrist-worn devices (Apple Watch, Fitbit Charge HR, and

CONTACT Antony D. Karelis karelis.antony@uqam.ca Department of Exercise Science, Université du Québec à Montréal, C.P. 8888, Succursale centre-ville, Montreal, Quebec, H3C 3P8 Canada
*Equal first author

© 2022 European College of Sport Science

Garmin Forerunner 225) showed mean absolute percentage errors of approximately 1.1–24.4% for heart rate and 7.9–210.8% for energy expenditure during 4-minute stages of light, moderate, and vigorous intensity treadmill exercises in young adults (Dooley, Golaszewski, & Bartholomew, 2017). Two other studies using multiple wrist-worn models (Apple Watch, Basis Peak, TomTom Cardio, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, and Samsung Gear S2) also observed poor accuracy in energy expenditure during sitting, walking, running and cycling at different intensities in young adults (Shcherbina et al., 2017; Thiebaud et al., 2018). Furthermore, a recent systematic review reported that most older models of wrist-worn devices that are presently on the market from manufactures such as Apple, Fitbit, Garmin, Polar, and Samsung displayed variable results for heart rate and have poor accuracy for energy expenditure during a wide range of physical activities (e.g. walking, running and cycling) in young healthy adults (Fuller et al., 2020). Nevertheless, newer models are constantly being developed by manufacturers and sold in the marketplace usually without publicly providing any information on how their research studies were conducted and/or without presenting any of their results. Thus, conducting proper validation studies by independent research laboratories becomes of importance. Therefore, the purpose of the present study was to validate 3 recently released wrist-worn devices (Apple Watch 6, Polar Vantage V and Fitbit Sense) for heart rate and energy expenditure during sitting, walking, running, resistance exercises and cycling.

Methods

Participants

A total of 60 participants (30 men and 30 women) were recruited in this study between February 2021 and June 2021 using advertisement via emails and on social media websites. To be included in the study, participants had to meet the following criteria: (1) aged between 18–30 years old (2) non-obese (body mass index less than 30 kg/m^2), (3) physically active (participants self-reported their physical activity levels and had to meet the physical activity guidelines of at least 150 min of physical activity per week), (4) white skin colour (between 1 and 4 on the Fitzpatrick scale Fitzpatrick, 1988), (5) no physical limitations to practice physical activities, and (6) non-smoker. Exclusion criteria were: (1) chronic diseases such as cardiovascular disease, diabetes and cancer and (2) being pregnant. The study was conducted in accordance with the Declaration of Helsinki and all procedures were approved by the Ethics Committee of the Université du Québec à Montréal. All participants were fully informed about the nature, goal, procedures, and risks of the study, and gave their informed consent in writing.

Procedure

After screening, all participants were invited to the Department of Exercise Science at the Université du Québec à Montréal for one visit. Before the visit, participants were instructed to avoid eating at least 2 h before testing. Upon their arrival, anthropometric and body composition measurements were taken. Participants were then equipped with the Polar H10 chest strap, the MetaMax 3B and the 3 wrist-worn devices simultaneously (Apple Watch 6, Polar Vantage V, and Fitbit Sense) according to the manufacturer's instructions. Thereafter, participants performed 5 different free-living physical activities for 10 min each. Environmental conditions were kept constant with a mean room temperature of 19.9°C ± 0.6°C.

Body composition

Total body weight, body fat percentage and lean body mass were measured using dual energy X-ray absorptiometry (General Electric Lunar Prodigy; standard mode; software version 12.30.008, Madison, WI, USA). Calibration was executed daily with a standard phantom prior to each test. Also, standing height (± 0.1 cm) was measured using a wall stadiometer (Perspective Enterprises, MI, USA). Body mass index = Body weight/Height (m^2) was then calculated.

Criterion measures

During all physical activity tasks, participants were wearing a breath-by-breath portable indirect calorimetry system (Metamax 3B, Cortex Biophysik GmbH, Leipzig, Germany), which was used as the criterion measure for energy expenditure. Prior to each test, the gas analyzers and the air volume turbine (using a 3 L syringe) were calibrated. It should be noted that previous research has reported that the MetaMax 3B is a valid and reliable system (Vogler, Rice, & Gore, 2010) and has been previously used in wrist-worn wearable validation studies (Duking et al., 2020; Gilgen-Ammann, Schweizer, & Wyss, 2019a). Furthermore, the Polar H10 chest strap (Polar Electro Oy, Kempele, Finland) was used as the criterion measure for heart rate and was connected to the MetaMax 3B software. This method has also been shown to be reliable (Gilgen-Ammann, Schweizer, & Wyss, 2019b) and has been previously used in wrist-worn

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

wearable validation studies (Gilgen-Ammann et al., 2019a; Müller et al., 2019).

Wearable devices

Three wrist-worn devices that measure heart rate using photoplethysmography were studied: Apple Watch 6 Version 7.0 (Apple Inc, Cupertino, CA), Polar Vantage V Firmware 5.1.8 (Polar Electro Oy, Kempele, Finland), and Fitbit Sense Version 128.4.17 (Fitbit Inc, San Francisco, CA). To our knowledge, detailed information on how energy expenditure was calculated does not seem to be available by the manufacturers for the Apple Watch 6, the Polar Vantage V and the Fitbit Sense. However, the user manual of the Apple watch 6 states that height, weight, gender and age are included in the calculation of energy expenditure. As for the Polar Vantage V, the user manual indicates that the following variables are used to calculate energy expenditure: body weight, height, age, gender, the intensity of the activity, maximum heart rate and maximal oxygen uptake. Finally, heart rate data during exercise are used to calculate energy expenditure for the Fitbit Sense as indicated in the user manual.

Before starting the activities, each participant's sex, body weight, height, handedness, and date of birth were entered in each of the wearable's software. All wearables were worn as instructed by the manufacturers (firmly and comfortably) and were worn simultaneously and synchronised at the same time. Specifically, the Polar Vantage V was placed 1 finger behind the wrist bone of the nondominant wrist of the participant, whereas the Fitbit Sense was placed 2 fingers behind the wrist bone of the dominant wrist. No particular information was given by the manufacturer for the placement of the Apple Watch 6. Therefore, based on preliminary testing in our laboratory for comfort and firmness, the Apple watch 6 was placed distal to the Polar Vantage V on the nondominant wrist of the participant. Moreover, the following activity modes were chosen that best represented each task. For all devices, *yoga* mode was selected for sitting. *Indoor run* (Apple), *treadmill running* (Polar), and *treadmill* (Fitbit) modes were selected for walking and running. T*raditional strength training* (Apple), *strength training* (Polar), and *weights* (Fitbit) modes were selected for resistance exercises. *Indoor cycle* (Apple), *indoor cycling* (Polar), and *spinning* (Fitbit) modes were selected for cycling.

Physical activities

All participants performed the following 5 activities in the same order for 10 min each: (1) sitting on a chair, (2) indoor walking on the treadmill at the participant's normal pace, (3) indoor running on the treadmill. For the first 5 min, participants ran at a lower self-selected speed and then ran at a higher self-selected speed the last 5 min. Before the start of the next activity, a 5 min rest period was given. (4) Resistance exercises using three different weight machines from Atlantis Precision Series (Atlantis Inc. Laval, Canada), which included a chest press, leg press, and straight-back seated cable row. Three sets of 10 repetitions with 1 min rest between each set was performed for each resistance exercise (total duration 12 min). Participants were instructed to self-select a weight that provided a moderate to vigorous intensity for each exercise. A two-minute rest period was given before the start of the next activity. (5) Cycling on an ergocycle (Precor 615 Commercial Upright Cycle, Lynnwood, WA) at a lower self-selected pace for the first 5 min and then at a higher self-selected pace the last 5 min. The starting and stopping time of each activity was noted on paper by the researcher. At the end of each activity, heart rate and energy expenditure data were recorded in each of the wearable's software applications and documented on paper by the researcher. It should be noted that the above-mentioned physical activities were chosen because they are popular in training/fitness centres and have been commonly reported in other wrist-worn validation studies (Dooley et al., 2017; Gilgen-Ammann et al., 2019a; Thiebaud et al., 2018; Shcherbina et al., 2017; Duking et al., 2020).

Statistical analysis

Statistical analyses were based on previous recommendations for validation studies (Duking, Fuss, Holmberg, & Sperlich, 2018; Hopkins, 2015). Descriptive, Pearson correlation, standardised typical error of the estimate (sTEE), coefficient of variation (CV), and Bland–Altman statistics were computed using a Microsoft Excel spreadsheet that was created and provided for validation studies in exercise science (Hopkins, 2015). The data are expressed as the mean ± standard deviation. In addition, the 90% confidence limits are shown for all statistical analyses. Mean absolute percentage error was also calculated: [(device – criterion)/criterion)] x 100. Pearson's correlations were performed to examine the relationship between the wearable values and the criterion measure. Pearson's r were interpreted as follows: ≥ 0.995: excellent; 0.95-0.994: very good; 0.85-0.94: good; 0.70-0.84: poor; 0.45-0.69: very poor; < 0.45: impractical (Hopkins, 2015). The sTEE was performed to estimate the confidence limits between each device and the criterion, the results were interpreted using half the thresholds of the modified Cohen scale: >2.0: impractical;

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1.0-2.0: very large; 0.6-1.0: large; 0.3-0.6: moderate; 0.1-0.3: small; <0.1: trivial (Hopkins, 2015). The CV (%) was performed to indicate the accuracy of the estimate by examining the variability in relation to the mean, the results were interpreted as follows: > 10%: poor accuracy; 5-10%: acceptable accuracy; < 5%: high accuracy. This interpretation was based on previous studies that used similar cut points as indicators of accuracy (Evenson & Spade, 2020; Feito, Bassett, & Thompson, 2012; Nelson, Kaminsky, Dickin, & Montoye, 2016). Bland–Altman analyses were performed to evaluate the extent of agreement between the devices and the criterion method with corresponding 95% limits of agreement (SD 1.96).

Results

Physical characteristics of the participants are presented in Table 1. Mean heart rate and energy expenditure values as well as mean absolute percentage error (MAPE) during all activities for the two measured criteria and the 3 wrist-worn wearable devices are shown in Table 2. The Apple Watch 6 had the lowest MAPE for heart rate for all activities (range: 0.6–2.9%) when compared to the Polar Vantage V (range: 1.8–5.7%) and the Fitbit Sense (range: 3.8–5.3%). Furthermore, large MAPEs were observed for the Apple Watch 6 (range: 14.9–47.8%), Polar Vantage V (range: 15.6–34.6%) and Fitbit Sense (range: 17.8–45.1%) for energy expenditure during all activities.

Table 3 shows the analysis of the accuracy of heart rate during all activities, which included the Pearson's r coefficients, sTEE and CV (%) with 90% confidence limits as well as their interpretations. Interpretations for the Pearson's coefficients for heart rate during the 5 different activities for the 3 devices were as follows: the Apple Watch 6 had 1 Excellent, 2 Very Good, 1 Good and 1 Poor; the Polar Vantage V had 0 Excellent, 2 Very Good, 1 Good and 2 Poor; the Fitbit Sense had 0 Excellent, 0 Very Good, 4 Good and 1 Poor. In addition, interpretations for the sTEE for heart rate during the 5 different activities for the 3 devices were as follows: the Apple Watch 6 had 1 Trivial, 1 Small, 2 Moderate and 1 Large; the Polar Vantage V had 0 Trivial, 1 Small, 2 Moderate and 2 Large; the Fitbit Sense had 0 Trivial, 0 Small, 2 Moderate and 3 Large. We found that the Apple Watch 6 displayed the highest level of accuracy for heart rate measurement with a CV (%) of less than 5% for all five activities. The Polar Vantage V showed acceptable accuracy for heart rate for walking and cycling as well as high accuracy for sitting, running and resistance exercises (CVs between 2.44-8.80%). We also noted that the Fitbit Sense seems to have a higher accuracy for heart rate during activities with a higher intensity. That is, the Fitbit Sense had poor accuracy for sitting (CV: 10.76%), acceptable accuracy for walking and resistance exercises (CVs: 6.76% and 5.89%, respectively), and high accuracy for running and cycling (CVs: 4.58% and 4.14%, respectively). All devices had a high level of accuracy for the running task (CVs: 4.67%, 4.02%, and 4.58%, respectively for the Apple Watch 6, Polar Vantage V, and Fitbit Sense). Furthermore, Bland–Altman analyses generally showed a good level of agreement for heart rate for all 3 devices across all activities (see Figures 1–5 in supplemental file). In addition, mean differences in heart rate for all 5 activities were the lowest with the Apple Watch 6 compared to the other devices (see Table 3). Moreover, when compared to the other activities, cycling had the highest mean differences in heart rate across all devices (see Table 3).

Table 1. Physical characteristics of the participants.

Variables	(n = 60)
Age (years)	24.9 ± 3.0
Height (m)	1.73 ± 0.1
Total body weight (kg)	69.3 ± 11.5
Body mass index (kg/m^2)	23.1 ± 2.7
Total fat mass (%)	19.9 ± 7.3
Total lean body mass (kg)	52.8 ± 10.6

Values are mean ± SD

The analysis of the accuracy of energy expenditure during all activities are presented in Table 4, which also included the Pearson's r coefficients, sTEE and CV (%) with 90% confidence limits as well as their interpretations. Interpretations for the Pearson's coefficients for energy expenditure during the 5 different activities for the 3 devices were as follows: the Apple Watch 6 had 1 Good, 3 Poor and 1 Impractical; the Polar Vantage V had 2 Poor, 2 Very Poor and 1 Impractical; the Fitbit Sense had 1 Good, 1 Poor, 3 Very Poor and 0 Impractical. In addition, interpretations for the sTEE for energy expenditure during the 5 different activities for the 3 devices were as follows: the Apple Watch 6 had 1 Moderate, 3 Large and 1 Impractical; the Polar Vantage V had 0 Moderate, 2 Large, 2 Very Large and 1 Impractical; the Fitbit Sense had 1 Moderate, 1 Large, 3 Very Large and 0 Impractical. The Apple Watch 6, Polar Vantage V, and Fitbit Sense had poor accuracy for all five physical activities (CVs between 14.68-24.85%, 16.54-25.78%, and 13.44-29.66%, respectively). Moreover, Bland–Altman analyses demonstrated poor levels of agreement for energy expenditure as shown by large random errors across all devices and activities (see Figures 1–5 in supplemental file). We also observed high variability in mean differences for energy expenditure as evidenced by large SDs and wide ranges in the limits of agreement across all devices and activities (see Table 4).

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Mean heart rate and energy expenditure as well as mean absolute percentage error during various activities.

Activity		Apple Watch 6	MAPE (Apple)	Polar Vantage V	MAPE (Polar)	Fitbit Sense	MAPE (Fitbit)
Heart rate (bpm)	Polar H10						
Sitting	69 ± 10.7	69 ± 10.6	0.6 ± 0.7	68 ± 10.9	1.8 ± 1.9	70 ± 10.1	4.6 ± 14.0
Walking	92 ± 12.0	92 ± 10.7	2.3 ± 3.4	93 ± 9.2	5.5 ± 6.3	90 ± 8.7	4.4 ± 5.0
Running	151 ± 12.4	150 ± 12.0	2.9 ± 3.7	146 ± 10.5	4.0 ± 3.2	148 ± 9.2	3.8 ± 3.3
Resistance exercises	119 ± 13.9	118 ± 13.3	1.4 ± 1.9	115 ± 12.4	3.5 ± 3.2	113 ± 12.0	5.3 ± 4.9
Cycling	147 ± 12.8	144 ± 12.3	2.2 ± 2.9	138 ± 11.9	5.7 ± 5.1	139 ± 10.2	5.2 ± 3.7
Energy expenditure (kcal)	MetaMax 3B						
Sitting	14 ± 2.8	17 ± 8.1	47.8 ± 44.2	14 ± 6.4	25.8 ± 48.1	12 ± 3.9	22.4 ± 13.4
Walking	45 ± 10.2	35 ± 10.3	24.1 ± 13.1	49 ± 11.3	15.6 ± 19.0	64 ± 10.2	45.1 ± 22.5
Running	117 ± 31.4	100 ± 24.7	14.9 ± 9.8	106 ± 22.8	15.7 ± 13.4	95 ± 20.6	17.8 ± 9.2
Resistance exercises	61 ± 19.7	64 ± 21.3	20.0 ± 20.0	77 ± 22.2	34.6 ± 32.6	44 ± 20.1	34.1 ± 16.7
Cycling	94 ± 25.3	80 ± 18.4	17.7 ± 10.8	91 ± 21.2	16.4 ± 18.6	69 ± 20.0	26.6 ± 14.5

Values are mean ± SD; bpm: beats per minute; kcal: kilocalories; MAPE: Mean absolute percentage error

Discussion

The purpose of the present study was to determine the accuracy of 3 recent wrist-worn devices for heart rate and energy expenditure during 5 different activities of moderate to vigorous intensities. One of the main findings of the present study indicates that the Apple Watch 6 appears to have the highest level of accuracy

Table 3. Analysis of the accuracy of heart rate measurements during various activities.

Activity	Apple Watch 6	Polar Vantage V	Fitbit Sense
Sitting	n = 60	n = 60	n = 58
Pearson's r	0.998 (0.997-0.999)	0.99 (0.98-0.99)	0.73 (0.61-0.82)
Interpretation of Pearson's r	Excellent	Very good	Poor
sTEE	0.06 (0.05-0.07)	0.16 (0.13-0.20)	0.94 (0.70-1.30)
Interpretation of sTEE	Trivial	Small	Large
CV (%)	0.90 (0.78-1.06)	2.44 (2.12-2.89)	10.76 (9.27-12.89)
Interpretation of CV (%)	High accuracy	High accuracy	Poor accuracy
Mean difference with limits of agreement (bpm)	0.08 ± 0.60 (−1.10–1.25)	−0.57 ± 1.73 (−3.96–2.81)	1.14 ± 7.43 (−13.43–15.70)
Walking	n = 60	n = 60	n = 60
Pearson's r	0.95 (0.92-0.97)	0.75 (0.63-0.83)	0.86 (0.79-0.90)
Interpretation of Pearson's r	Very good	Poor	Good
sTEE	0.34 (0.27-0.43)	0.90 (0.68-1.23)	0.60 (0.47-0.79)
Interpretation of sTEE	Moderate	Large	Large
CV (%)	4.15 (3.59-4.92)	8.80 (7.61-10.49)	6.76 (5.85-8.04)
Interpretation of CV (%)	High accuracy	Acceptable accuracy	Acceptable accuracy
Mean difference with limits of agreement (bpm)	0.07 ± 3.85 (−7.48–7.62)	1.39 ± 7.91 (−14.11–16.88)	−1.26 ± 6.21 (−13.44–10.91)
Running	n = 60	n = 60	n = 60
Pearson's r	0.84 (0.76-0.89)	0.88 (0.83-0.92)	0.85 (0.77-0.90)
Interpretation of Pearson's r	Poor	Good	Good
sTEE	0.64 (0.50-0.85)	0.53 (0.42-0.68)	0.63 (0.49-0.82)
Interpretation of sTEE	Large	Moderate	Large
CV (%)	4.67 (4.05-5.55)	4.02 (3.48-4.77)	4.58 (3.97-5.44)
Interpretation of CV (%)	High accuracy	High accuracy	High accuracy
Mean difference with limits of agreement (bpm)	−1.22 ± 6.53 (−14.02–11.57)	−5.01 ± 5.62 (−16.02–6.01)	−3.76 ± 6.62 (−16.73–9.22)
Resistance exercises	n = 60	n = 60	n = 60
Pearson's r	0.98 (0.97-0.99)	0.96 (0.93-0.97)	0.88 (0.82-0.92)
Interpretation of Pearson's r	Very good	Very good	Good
sTEE	0.19 (0.15-0.24)	0.31 (0.24-0.39)	0.54 (0.42-0.70)
Interpretation of sTEE	Small	Moderate	Moderate
CV (%)	2.30 (2.00-2.73)	3.58 (3.11-4.25)	5.89 (5.10-7.00)
Interpretation of CV (%)	High accuracy	High accuracy	Acceptable accuracy
Mean difference with limits of agreement (bpm)	−1.30 ± 2.84 (−6.87–4.28)	−4.30 ± 4.48 (−13.07–4.47)	−6.51 ± 6.62 (−19.48–6.46)
Cycling	n = 59	n = 59	n = 59
Pearson's r	0.93 (0.90-0.96)	0.78 (0.68-0.85)	0.89 (0.83-0.93)
Interpretation of Pearson's r	Good	Poor	Good
sTEE	0.38 (0.30-0.49)	0.80 (0.61-1.08)	0.51 (0.40-0.66)
Interpretation of sTEE	Moderate	Large	Moderate
CV (%)	3.23 (2.80-3.84)	5.72 (4.95-6.81)	4.14 (3.59-4.93)
Interpretation of CV (%)	High accuracy	Acceptable accuracy	High accuracy
Mean difference with limits of agreement (bpm)	−3.15 ± 4.74 (−12.44–6.14)	−8.64 ± 8.26 (−24.82–7.55)	−7.86 ± 6.14 (−19.89–4.17)

sTEE: standardised typical error of the estimate; CV: coefficient of variation; bpm: beats per minute

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

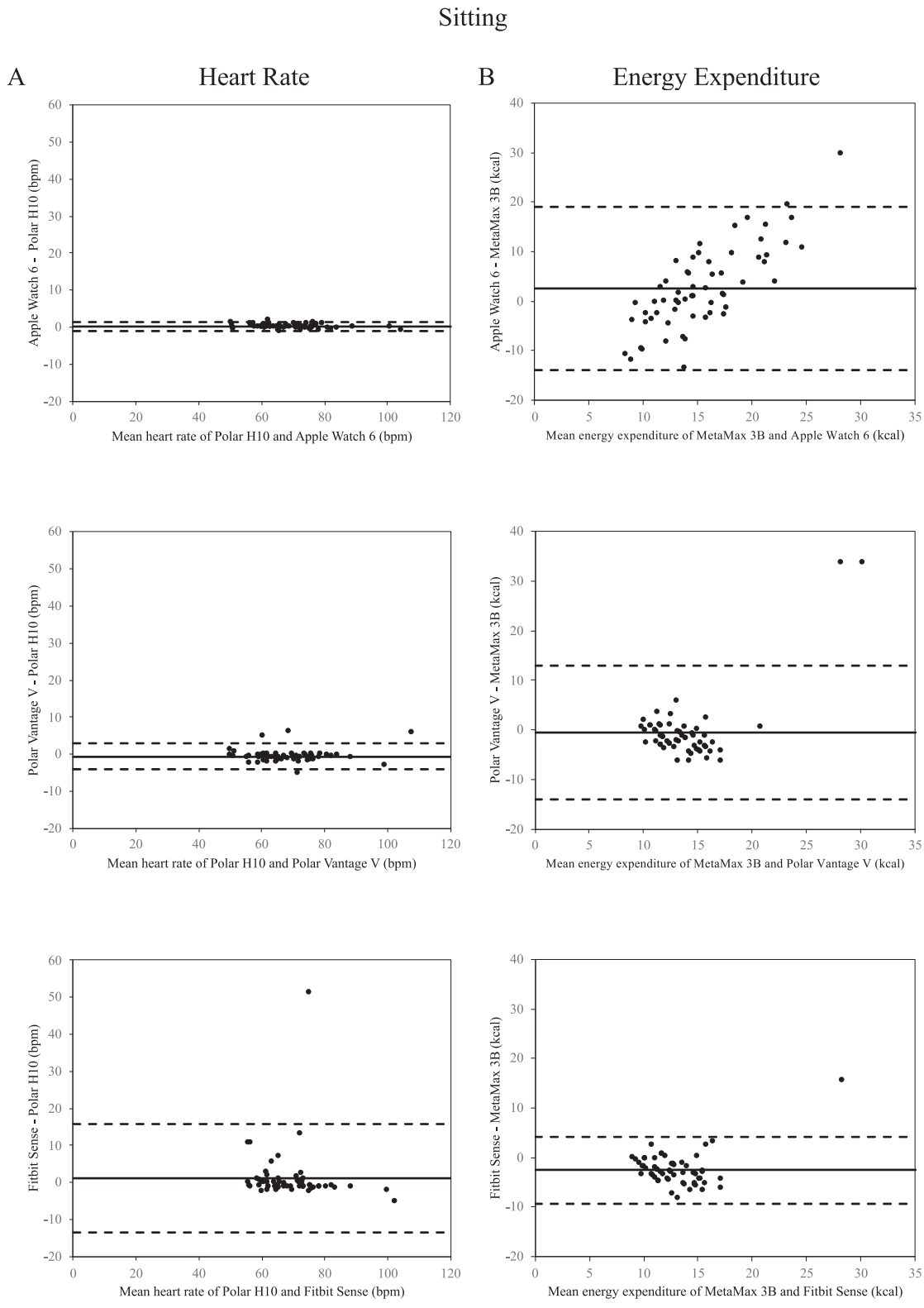

Figure 1. A (left panel): Bland-Altman analysis between the Polar H10 and the Apple Watch 6, Polar Vantage V and Fitbit Sense for heart rate during sitting. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI). B (right panel): Bland-Altman analysis between the MetaMax 3B and the Apple Watch 6, Polar Vantage V and Fitbit Sense for energy expenditure during sitting. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI).

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

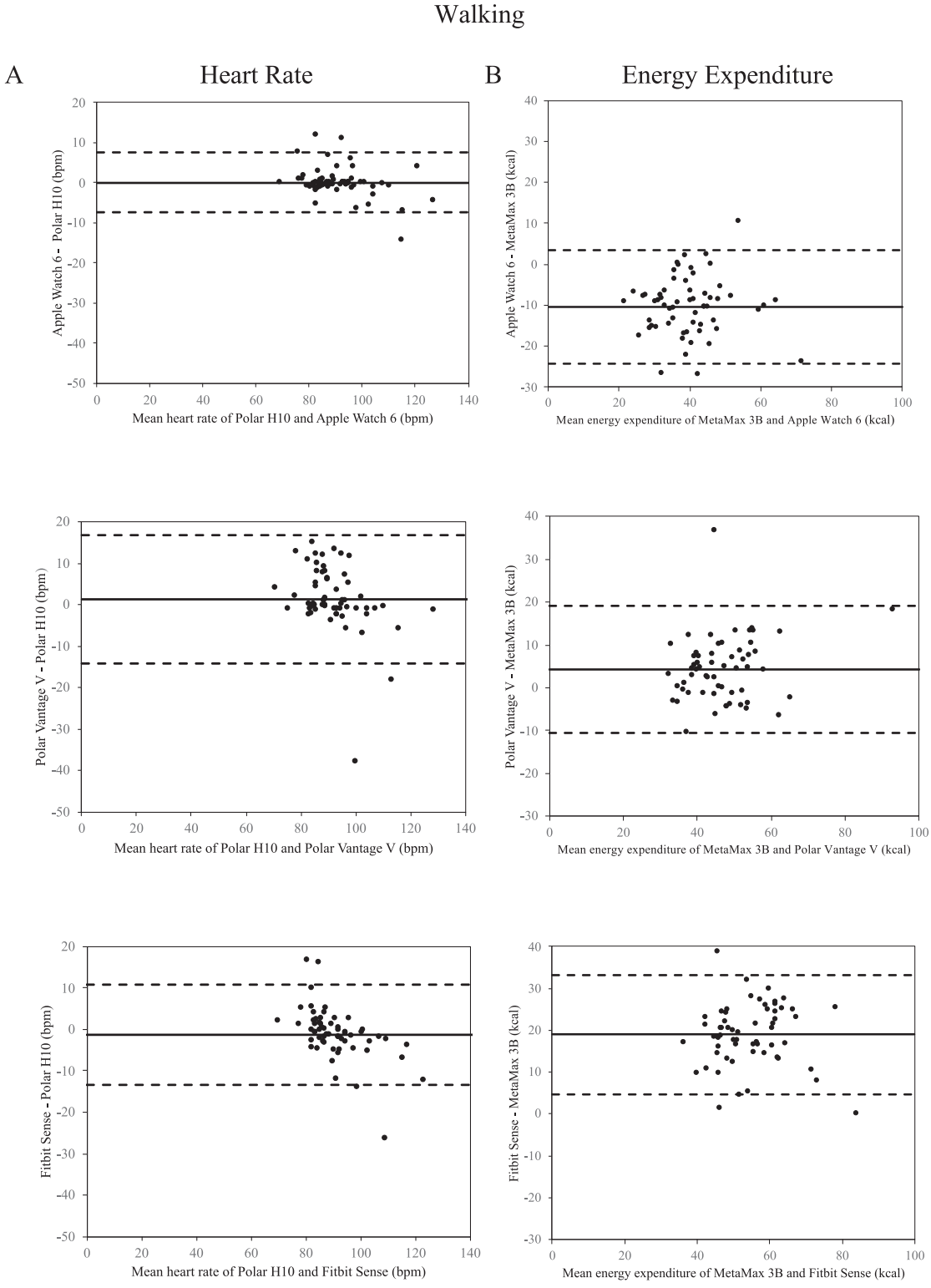

Figure 2. A (left panel): Bland-Altman analysis between the Polar H10 and the Apple Watch 6, Polar Vantage V and Fitbit Sense for heart rate during walking. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI). B (right panel): Bland-Altman analysis between the MetaMax 3B and the Apple Watch 6, Polar Vantage V and Fitbit Sense for energy expenditure during walking. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI).

for heart rate monitoring for all 5 activities (overall mean CV: 3.1% or 3.1 bpm discrepancy from the reference measure), followed by the Polar Vantage V (overall mean CV: 4.9% or 4.9 bpm discrepancy from the reference measure) with high accuracy for sitting, running and resistance exercises and then by the Fitbit Sense

17467391, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

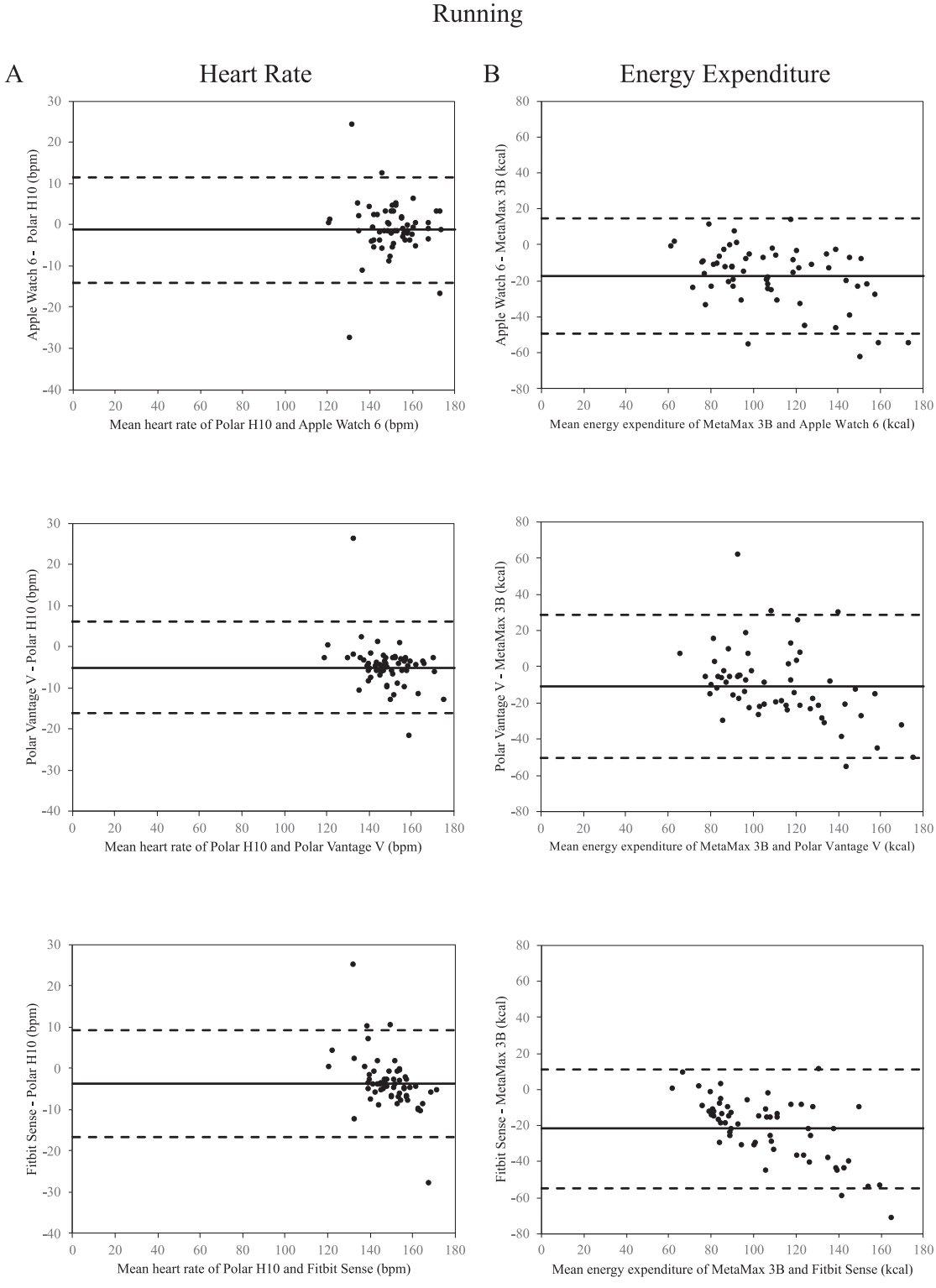

Figure 3. A (left panel): Bland-Altman analysis between the Polar H10 and the Apple Watch 6, Polar Vantage V and Fitbit Sense for heart rate during running. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI). B (right panel): Bland-Altman analysis between the MetaMax 3B and the Apple Watch 6, Polar Vantage V and Fitbit Sense for energy expenditure during running. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI).

(overall mean CV: 6.4% or 6.4 bpm discrepancy from the reference measure) with a high level of accuracy for running and cycling only. Interpretations of Pearson's r and sTEE as well as MAPEs also support the results of the CVs showing that the Apple Watch 6 was the most accurate for measuring heart rate. In addition, Bland–

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

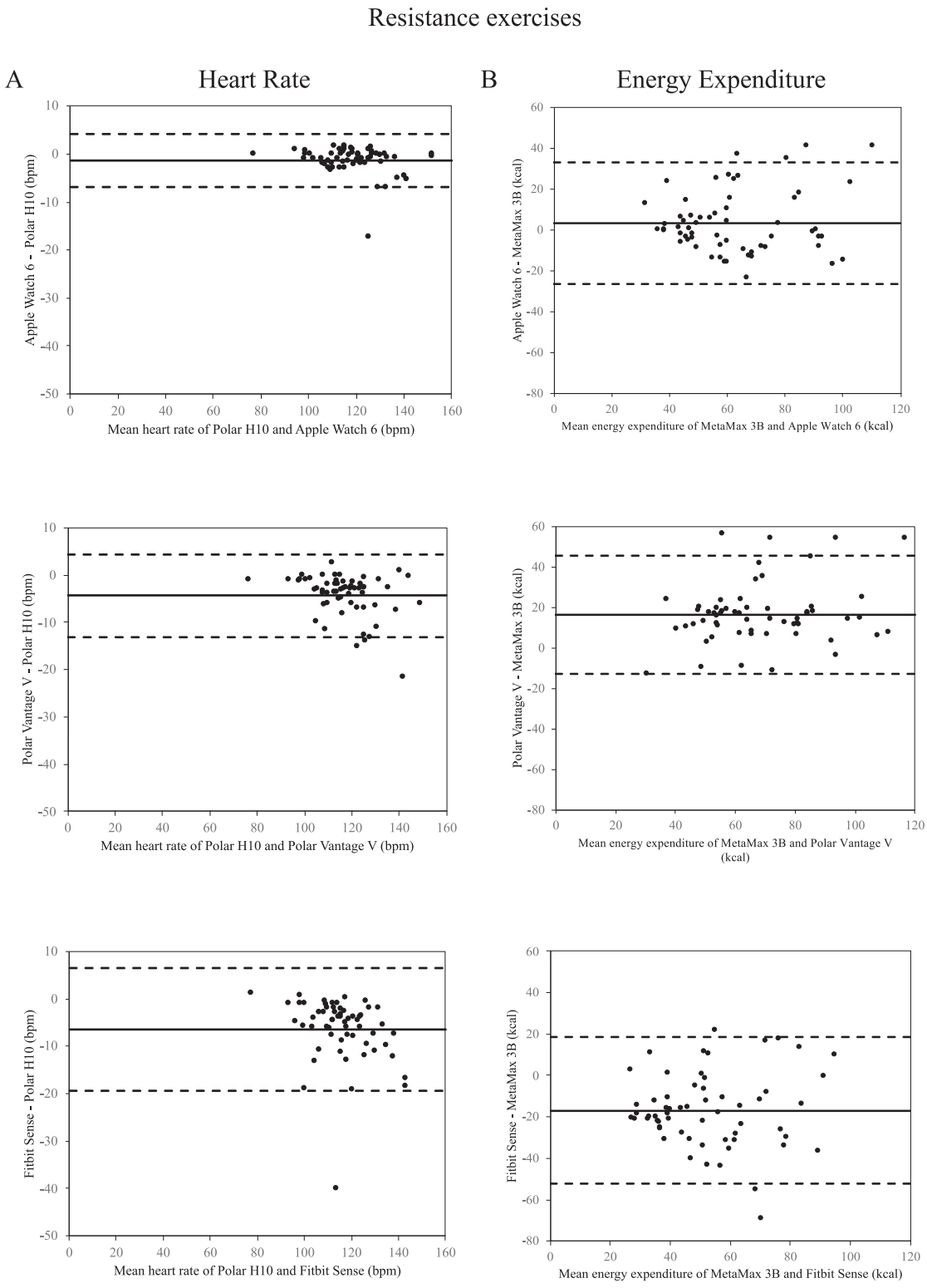

Figure 4. A (left panel): Bland-Altman analysis between the Polar H10 and the Apple Watch 6, Polar Vantage V and Fitbit Sense for heart rate during resistance exercises. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI). B (right panel): Bland-Altman analysis between the MetaMax 3B and the Apple Watch 6, Polar Vantage V and Fitbit Sense for energy expenditure during resistance exercises. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI).

17458391, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

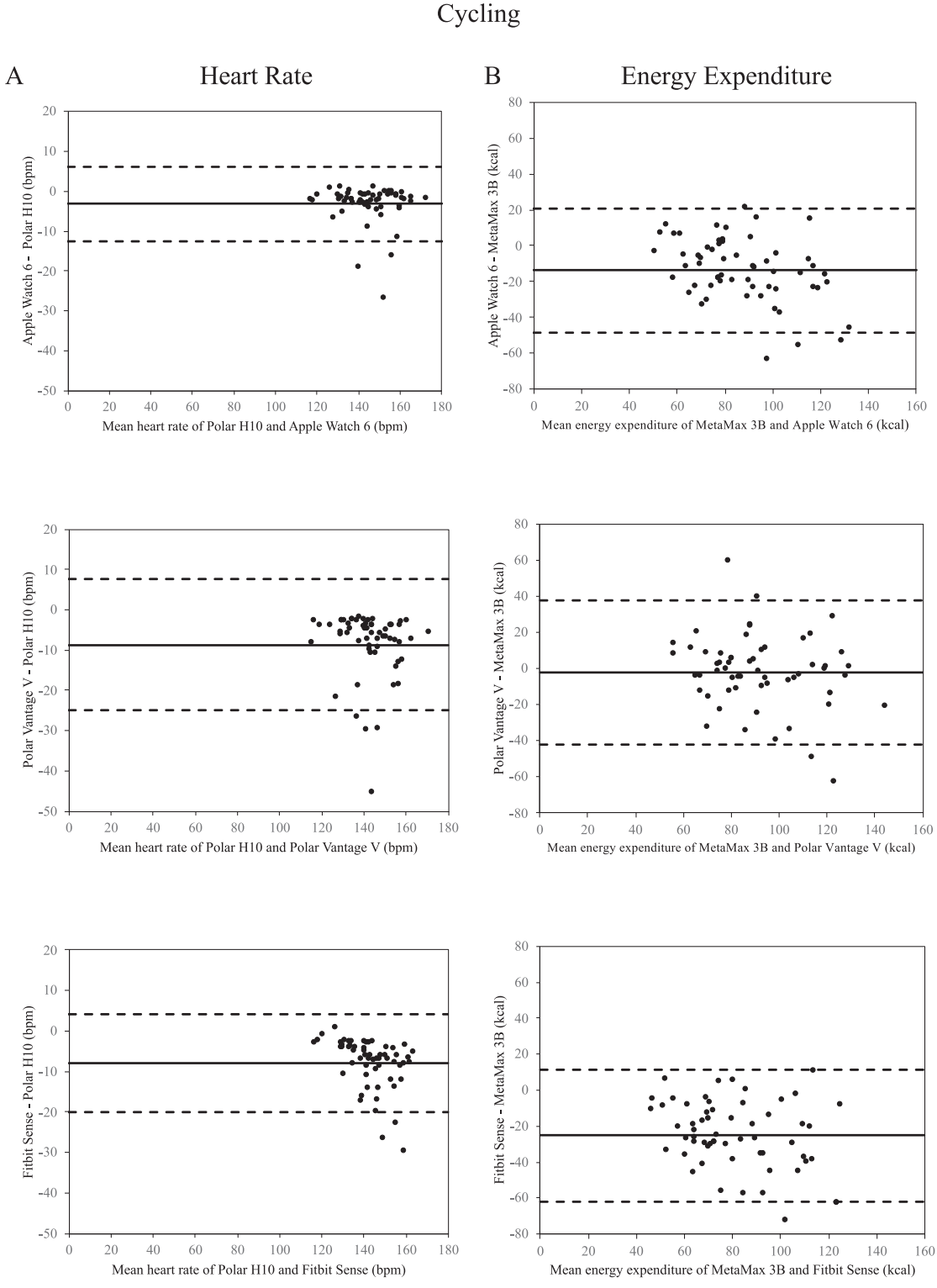

Figure 5. A (left panel): Bland-Altman analysis between the Polar H10 and the Apple Watch 6, Polar Vantage V and Fitbit Sense for heart rate during cycling. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI). B (right panel): Bland-Altman analysis between the MetaMax 3B and the Apple Watch 6, Polar Vantage V and Fitbit Sense for energy expenditure during cycling. Solid line represents the mean difference. Outer dotted lines represent limits of agreement (95% CI).

Altman analyses generally presented a good level of agreement for heart rate monitoring during all activities for all 3 devices. Similar results have been observed by previous studies who also reported that the Apple Watch 4 (Duking et al., 2020) and the Apple Watch 2 (Boudreaux et al., 2018) (albeit in older models) were

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4. Analysis of the accuracy of energy expenditure measurements during various activity.

Activity	Apple Watch 6	Polar Vantage V	Fitbit Sense
Sitting	n = 60	n = 59	n = 58
Pearson's r	0.05 (−0.16-0.27)	0.17 (−0.05-0.37)	0.55 (0.37-0.68)
Interpretation of Pearson's r	Impractical	Impractical	Very poor
sTEE	18.62 (3.64-6.06)	5.79 (2.49-20.82)	1.53 (1.07-2.48)
Interpretation of sTEE	Impractical	Impractical	Very large
CV (%)	22.61 (19.39-27.26)	22.25 (19.06-26.86)	18.55 (15.90-22.37)
Interpretation of CV (%)	Poor accuracy	Poor accuracy	Poor accuracy
Mean difference with limits of agreement (kcal)	2.57 ± 8.37 (−13.83–18.97)	−0.55 ± 6.92 (−14.10–13.01)	−2.58 ± 3.44 (−9.32–4.16)
Walking	n = 60	n = 59	n = 60
Pearson's r	0.75 (0.63-0.83)	0.69 (0.56-0.79)	0.72 (0.60-0.81)
Interpretation of Pearson's r	Poor	Very poor	Poor
sTEE	0.89 (0.68-1.22)	1.05 (0.78-1.49)	0.95 (0.72-1.32)
Interpretation of sTEE	Large	Very large	Large
CV (%)	15.61 (13.44-18.71)	16.54 (14.21-19.87)	16.23 (13.96-19.46)
Interpretation of CV (%)	Poor accuracy	Poor accuracy	Poor accuracy
Mean difference with limits of agreement (kcal)	− 10.40 ± 7.09 (−24.30–3.51)	4.29 ± 7.57 (−10.55–19.13)	18.94 ± 7.32 (−4.58–33.29)
Running	n = 60	n = 59	n = 60
Pearson's r	0.86 (0.79-0.91)	0.74 (0.63-0.83)	0.88 (0.82-0.92)
Interpretation of Pearson's r	Good	Poor	Good
sTEE	0.60 (0.47-0.78)	0.90 (0.68-1.24)	0.54 (0.42-0.69)
Interpretation of sTEE	Moderate	Large	Moderate
CV (%)	14.68 (12.64-17.58)	19.38 (16.62-23.34)	13.44 (11.58-16.07)
Interpretation of CV (%)	Poor accuracy	Poor accuracy	Poor accuracy
Mean difference with limits of agreement (kcal)	−17.31 ± 16.22 (−49.10–14.49)	−10.59 ± 20.16 (−50.10–28.91)	−21.84 ± 16.92 (−55.00–11.32)
Resistance exercises	n = 60	n = 59	n = 60
Pearson's r	0.74 (0.62-0.82)	0.71 (0.58-0.80)	0.61 (0.46-0.73)
Interpretation of Pearson's r	Poor	Poor	Very poor
sTEE	0.92 (0.69-1.26)	0.99 (0.74-1.40)	1.29 (0.93-1.94)
Interpretation of sTEE	Large	Large	Very large
CV (%)	24.85 (21.28-30.01)	25.78 (22.04-31.21)	29.66 (25.33-35.95)
Interpretation of CV (%)	Poor accuracy	Poor accuracy	Poor accuracy
Mean difference with limits of agreement (kcal)	3.14 ± 15.18 (−26.61–32.89)	16.33 ± 14.99 (−13.05–45.70)	−16.83 ± 17.95 (−52.01–18.36)
Cycling	n = 59	n = 58	n = 59
Pearson's r	0.72 (0.60-0.81)	0.60 (0.44-0.73)	0.69 (0.55-0.79)
Interpretation of Pearson's r	Poor	Very poor	Very poor
sTEE	0.96 (0.72-1.33)	1.32 (0.95-2.01)	1.06 (0.78-1.50)
Interpretation of sTEE	Large	Very large	Very large
CV (%)	20.94 (17.94-25.25)	24.25 (20.71-29.37)	22.10 (18.93-26.67)
Interpretation of CV (%)	Poor accuracy	Poor accuracy	Poor accuracy
Mean difference with limits of agreement (kcal)	−13.88 ± 17.59 (−48.35–20.59)	−2.29 ± 20.41 (−42.29–37.71)	−25.20 ± 18.74 (−61.93–11.52)

sTEE: standardised typical error of the estimate; CV: coefficient of variation; kcal: kilocalories

the most accurate for heart rate measurement during different activities followed by the Polar Vantage V (Duking et al., 2020). To our knowledge, no other study has examined the accuracy of heart rate in the Fitbit Sense. However, a recent systematic review found that different previous models from Fitbit tended to underestimate heart rate values by 3% (Fuller et al., 2020), whereas the Fitbit Sense had MAPEs between 3.8–5.3% in the present study. Our findings also showed that the level of accuracy for heart rate may be activity dependent for the Polar Vantage V and the Fitbit Sense. For example, only running had the highest level of accuracy for both devices when compared to the other activities. In addition, there seems to be a slight tendency to underestimate heart rate values in both the Polar Vantage V and the Fitbit Sense.

One of the major challenges in clinical/field research has been the difficulty to accurately measure energy expenditure using feasible methods. In the present study, all 3 wrist-worn devices presented poor accuracy across all 5 activities. That is, the Apple Watch 6, Polar Vantage V, and Fitbit Sense had overall mean CVs of 19.7% (range: 14.68–24.85%), 21.6% (range: 16.54–25.78%), and 20% (range: 13.44–29.66%), respectively. Moreover, the results of the CVs are supported by the findings of the interpretations of Pearson's r and sTEE as well as the MAPEs, which showed high variability for the Apple Watch 6 (range: 14.9–47.8%), Polar Vantage V (range: 15.6–34.6%) and Fitbit Sense (range: 17.8–45.1%) during all activities. Bland–Altman analyses also displayed poor levels of agreement as shown by large random errors across all devices and activities.

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Using running as an example and extrapolating the data to 1 h, overall mean differences for the Apple Watch 6, Polar Vantage V, and Fitbit Sense were −17.3 kcal/10 min or -103.8 kcal/hour, −10.6 kcal/10 min or -63.6 kcal/hour, and −21.8 kcal/10 min or -130.8 kcal/hour, respectively, suggesting that these 3 devices underestimated energy expenditure during running. These results are in line with a previous study who also showed that energy expenditure in the Apple Watch 4 (overall mean CV of 21%; range: 13.5–27.1%) and the Polar Vantage V (overall mean CV of 20%; range: 16.3–28.0%) lacked accuracy during different exercise intensities (Duking et al., 2020). Furthermore, another study investigated the accuracy of energy expenditure in the Polar Vantage M (comparable technology for the assessment of energy expenditure to the Polar Vantage V) during 7 different activities in young individuals (Gilgen-Ammann et al., 2019a). In that study, the authors concluded that the accuracy of energy expenditure was moderate to good with a MAPE of 21% (range: 9–31%) during the different activities. To our knowledge, no other study has validated energy expenditure for the Apple Watch 6 and Fitbit Sense. Therefore, it might be difficult to compare newer models to older ones since algorithms and technologies could be different. Indeed, 2 systemic reviews reported that multiple wearables devices (including previous models of Apple and Fitbit) had poor accuracy for the measurement of energy expenditure during various activities (Evenson, Goto, & Furberg, 2015; Fuller et al., 2020). Collectively, based on these findings, we would suggest that evaluating energy expenditure using these 3 wrist-worn devices does not provide an acceptable surrogate method for the estimation of energy expenditure in research studies. Manufacturers may need to re-evaluate their algorithms or technologies in order to increase the accuracy of their devices for the measurement of energy expenditure during different types of physical activities.

Our results may be useful for clinical and practical purposes. It is important to educate health care professionals (e.g. nutritionists and kinesiologists) and athletes/coaches regarding the accuracy of these devices. That is, healthcare care professionals, athletes/coaches and the general population may want to proceed with caution on the clinical utility of energy expenditure of these devices during the implementation of an exercise training or nutritional programme. It should also be noted that our study was only composed of young healthy physically active nonobese individuals. Therefore, our findings are limited to this population. Moreover, the present study was conducted in a laboratory setting that used 5 specific types of physical activities. Thus, our results are limited to these activities. Also, activities such as walking and running were performed on a treadmill and cycling on a stationary ergocycle. As a result, this may limit the practical implications of our study to an indoor setting only. Taken together, future studies may want to investigate the accuracy of these devices in different populations (e.g. obese, elderly) as well as in different settings such as the outdoors using different physical activities (e.g. running and cycling outdoors, climbing, swimming). Future validation studies may also want to consider measuring heart rate and energy expenditure for longer durations during various activities such as a 24-hour period.

In conclusion, results of the present study indicate that the Apple Watch 6 was the most accurate for measuring heart rate for all activities, whereas the accuracy of heart rate measurements varied between the different types of activities for the Polar Vantage V and Fitbit Sense (from high to poor accuracy). In addition, all 3 devices showed poor accuracy for assessing energy expenditure across all activities.

Acknowledgments

We would like to thank Jeanne Breault-Mallette, Laurent Jutras, Alexandra Lavoie-Lechasseur and Carole Roy for their technical assistance as well as the participants who accepted to be part of this study. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Université du Québec à Montréal: [Grant Number Non applicable].

Sources of support

This study was supported by start-up funds from the Université du Québec à Montréal. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Gilles Gouspillou is supported by a *Chercheur Boursier* Junior 2 salary award from the FRQS.

15367290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Authors' contributions

ADK, ASC, GG and GHB designed the research; MALD conducted the research; ADK, MALD and GHB analyzed the data; GHB and MALD wrote the first draft of the manuscript; ADK, ASC and GG contributed to the writing of the manuscript; ADK had primary responsibility for the design, writing and final content. All authors read and approved the final manuscript.

Reference

Boudreaux, B. D., Hebert, E. P., Hollander, D. B., et al. (2018). Validity of wearable Activity Monitors during cycling and resistance exercise. *Medicine & Science in Sports & Exercise*, *50*(3), 624–633.

Donnelly, J. E., Blair, S. N., Jakicic, J. M., et al. (2009). American college of Sports medicine position stand. Appropriate physical activity intervention strategies for weight loss and prevention of weight regain for adults. *Medicine & Science in Sports & Exercise*, *41*(2), 459–471.

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-Monitoring heart rate and physical activity wearable devices. *Jmir Mhealth and Uhealth*, *5*(3), e34.

Duking, P., Fuss, F. K., Holmberg, H. C., & Sperlich, B. (2018). Recommendations for Assessment of the reliability, sensitivity, and Validity of data provided by wearable sensors designed for Monitoring physical activity. *Jmir Mhealth and Uhealth*, *6*(4), e102.

Duking, P., Giessing, L., Frenkel, M. O., Koehler, K., Holmberg, H. C., & Sperlich, B. (2020). Wrist-Worn wearables for Monitoring Heart Rate and energy expenditure while sitting or performing light-to-vigorous physical activity: Validation study. *Jmir Mhealth and Uhealth*, *8*(5), e16716.

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *The international Journal of Behavioral Nutrition and Physical Activity*, *12*, 159.

Evenson, K. R., & Spade, C. L. (2020). Review of Validity and reliability of Garmin activity trackers. *J Meas Phys Behav*, *3*(2), 170–185.

Feito, Y., Bassett, D. R., & Thompson, D. L. (2012). Evaluation of activity monitors in controlled and free-living environments. *Medicine & Science in Sports & Exercise*, *44*(4), 733–741.

Fitzpatrick, T. B. (1988). The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, *124*(6), 869–871.

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 and Uhealth*, *8*(9), e18694.

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019a). Accuracy of the multisensory wristwatch Polar vantage's estimation of energy expenditure in various activities: Instrument validation study. *Jmir Mhealth and Uhealth*, *7*(10), e14534.

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019b). RR interval signal quality of a heart rate monitor and an ECG holter at rest and during exercise. *European Journal of Applied Physiology*, *119*(7), 1525–1532.

Hills, A. P., Mokhtar, N., & Byrne, N. M. (2014). Assessment of physical activity and energy expenditure: An overview of objective measures. *Frontiers in Nutrition*, *1*, 5.

Hopkins, W. (2015). Spreadsheets for analysis of validity and reliability. *Sportscience*, *19*, 36–42.

Müller, A. M., Wang, N. X., Yao, J., et al. (2019). Heart rate measures from wrist-worn Activity Trackers in a laboratory and free-living setting: Validation study. *Jmir Mhealth and Uhealth*, *7*(10), e14120.

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., & Montoye, A. H. (2016). Validity of consumer-based physical activity Monitors for Specific activity types. *Medicine & Science in Sports & Exercise*, *48*(8), 1619–1628.

Shcherbina, A., Mattsson, C. M., Waggott, D., et al. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, *7*(2).

Thiebaud, R. S., Funk, M. D., Patton, J. C., et al. (2018). Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *Digit Health*, *4*, 2055207618770322.

Thompson, W. R. (2019). Worldwide survey of fitness trends for 2020. *ACSM's Health & Fitness Journal*, *23*(6), 10–18.

Vogler, A. J., Rice, A. J., & Gore, C. J. (2010). Validity and reliability of the Cortex MetaMax3B portable metabolic system. *Journal of Sports Sciences*, *28*(7), 733–742.

17457290, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1080/17461391.2021.2023656, Wiley Online Library on [01/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License