applied sciences

MDPI

Article

Are We Really Training at the Desired Intensity? Concurrent Validity of 16 Commercial Photoplethysmography-Based Heart Rate Monitors

Pablo Oropesa [1], Alejandro Sánchez-Pay [2,*], Elena Conesa-Ros [1], Antonino Bianco [3], Jesús J. Ruiz-Navarro [1,4,*] and Alejandro Martínez-Cava [1]

[1] Human Performance and Sports Science Laboratory, Faculty of Sport Sciences, University of Murcia, 30700 Murcia, Spain; poropesam@um.es (P.O.); econesaros@um.es (E.C.-R.); alejandro.martinez12@um.es (A.M.-C.)
[2] Performance Analysis and Biomechanics in Sport, Department of Physical Activity and Sport, Faculty of Sport Sciences, University of Murcia, 30700 Murcia, Spain
[3] Sport and Exercise Sciences Research Unit, Department of Psychology, Educational Science and Human Movement, University of Palermo, 90133 Palermo, Italy; antonino.bianco@unipa.it
[4] Aquatics Lab, Department of Physical Education and Sports, Faculty of Sport Sciences, University of Granada, 18071 Granada, Spain
* Correspondence: aspay@um.es (A.S.-P.); jjruiznavarro@um.es (J.J.R.-N.)

Abstract

The validity and accuracy of photoplethysmography (PPG)-based wearable heart rate (HR) monitors remain debatable. This study aimed to determine the concurrent validity of HR records from a wide range of contemporary PPG monitors across the full spectrum of exercise intensities and running conditions. Ten well-trained male endurance athletes performed several incremental maximal aerobic (IMA) tests and high-intensity interval training (HIIT) sessions while wearing a Polar H9 chest strap, as the reference criterion; 16 PPG-based heart rate monitors were tested (Amazfit Fit5; Apple Ultra and SE; Garmin 35, 45, 235, and 935; Polar M200, M430, OH1, Vantage M, Vantage V, Vantage V3, and Verity Sense; Galaxy Watch 5, Suunto 3 Fitness). The results showed excellent ICC (>0.90) versus the reference device across IMA and HIIT tests. Overall, the ICC decreased, and magnitudes of error increased (BIAS, SEM, and CV) as the intensity increased. Moreover, lower ICC values and greater BIAS, SEM, and CV were observed during the HIIT compared to the IMA test. Nevertheless, notable differences between devices were observed in magnitude of errors, accuracy, data loss, and read failures. In conclusion, PPG-based HR monitor validity is device-dependent and therefore exercise data from running training and competitions should be interpreted with caution.

Keywords: activity monitors; physical activity; PPG; validation; wearables

Check for updates

Academic Editor: Argyris Toubekis

Received: 6 November 2025
Revised: 16 December 2025
Accepted: 17 December 2025
Published: 22 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Training load monitoring is a crucial part of the athletes' development process, aiming to understand and attribute responses to their training. In this context, heart rate (HR) monitoring is one of the most used methods to quantify the internal training load in sport [1], as it facilitates the identification and monitoring of the intensity zones throughout the training process [2]. Chest straps have been the most validated and widely used to detect and monitor HR [3]. However, these bands often present some inconveniences

during its use, such as discomfort. As an alternative, a new HR monitoring method known as photoplethysmography (PPG) was developed [4].

Based on the detection of blood volume variations, PPG is an optical measurement of the peripheral pulse at the skin surface [5]. This method revolutionized sport and exercise science, being extensively adopted in the sports industry, with brands producing different PPG devices [6]. However, the range of brands and models available has increased considerably during recent years, offering a wide spectrum of options, from affordable to high-end devices, that still require proper validation [7]. Without such validation, the use of these devices may not only fail to benefit the users but could even be counterproductive, as they might provide misleading or inaccurate information about the actual training intensity, thereby negatively impacting their monitoring or training interpretation.

In this context, different PPGs have been validated during walking, running, cycling, or swimming [8–11]. Due to the controlled conditions offered by treadmill running protocols, they represent the most traditional approach for validation [12]. Previous studies have assessed HR agreement against electrocardiogram (ECG) measurements [13]. However, because of the high cost, limited accessibility, discomfort, and complexity associated with electrode placement, chest straps have been validated against ECG [14], and are now commonly used as criterion standards in most PPG validation studies [3,15,16].

A key factor when validating a PPG sensor is exercise intensity [17]. Overall, PPG devices show low error rates at rest or during light exercise; however, the accuracy might decrease as exercise intensity increases considerably [18]. This decline is attributed to the greater amount of movement at higher intensities, which can introduce artifacts and interfere with signal quality [19]. Because of that, it is important to assess PPG devices across the full spectrum of exercise intensities, including interval-based activities that involve abrupt changes in HR. Otherwise, a device validated at low-intensity exercise might not be accurate during high-intensity training. Given the varying performance of PPG devices across different intensities and the large number of devices on the market that have not yet been validated under these conditions, it is essential to systematically assess their accuracy across a range of exercise intensities. Hence, the aim of this study was to determine HR validity of 16 wearable monitors during running at different intensities and protocols.

2. Materials and Methods

2.1. Participants

Ten well-trained male runners and triathletes spanning performance tiers 2 to 4 [20] volunteered to participate in this study (age: 27.0 ± 6.1 years; height: 175 ± 6.3 cm; body mass: 73.9 ± 9.1 kg; endurance training experience: 11.5 ± 5.1 years; maximum oxygen uptake: 60.2 ± 7.8 $mL \cdot kg^{-1} \cdot min^{-1}$) [21]. All the participants were familiar with the use of HR monitors and used them on a daily basis. To be included in the study, participants were required to meet the following inclusion criteria: (1) free of cardiovascular pathologies that would prevent them from performing the tests; (2) abstention from intense exercise for 48 h prior to each test. All the participants were informed about the purpose of the study and signed a written consent form before enrolling in the project. The study was conducted according to the Declaration of Helsinki and approved by the Ethics Commission of the University of Murcia (ID: M10/2024/022, approval date: 28 February 2024).

2.2. Study Design

To comply with the manufacturers' recommendations and avoid attaching 2 devices to the same part of the body, HR was simultaneously registered by different commercial devices during 14 testing sessions: 7 for the incremental maximal aerobic (IMA) test and

7 for the high-intensity interval training (HIIT) test. In each testing session, the subjects always wore an H9 chest strap (Polar Electro Oy Inc, Kempele, Finland) as the reference criterion device [22] and two or three additional devices (alternatives) to assess concurrent validity. The characteristics of the devices are described in Table 1.

Table 1. Characteristics of the heart rate measurement devices used.

Company	Model	Acronym	Position	Data Processing
Amazfit	Bip 5	AMAZ-BIP5	Wrist	APP Zepp
Apple	Watch SE	APPLE-SE	Wrist	APP Apple Fitness+
Apple	Watch ultra	APPLE-ULTRA	Wrist	APP Apple Fitness+
Garming	Forerunner 235	GAR-235	Wrist	Garmin Connect
Garming	Forerunner 35	GAR-35	Wrist	Garmin Connect
Garming	Forerunner 45	GAR-45	Wrist	Garmin Connect
Garming	Forerunner 935	GAR-935	Wrist	Garmin Connect
Polar	M200	POL-M200	Wrist	Polar FlowSync
Polar	M430	POL-M430	Wrist	Polar FlowSync
Polar	OH1	POL-OH1	Arm	Polar FlowSync
Polar	Vantage M	POL-VANTM	Wrist	Polar FlowSync
Polar	Vantage V	POL-VANTV	Wrist	Polar FlowSync
Polar	Vantage V3	POL-VANTV3	Wrist	Polar FlowSync
Polar	Verity Sense	POL-VS	Arm	Polar FlowSync
Samsung	Galaxy watch 5	SAM-GAL5	Wrist	APP Samsung Health
Suunto	3 Fitness	SUN-3	Wrist	APP Suunto

2.3. Procedure

2.3.1. Testing Procedures

All testing sessions were conducted in the same facilities at the same time of day (11:00–13.00 h) and the same controlled environmental conditions (21–23 °C and 50–50.4% of humidity; Kestrel 4000 Weather Metre, Nielsen-Kellerman Co., Boothwyn, PA, USA). The participants performed the tests in a randomized order. To ensure adequate recovery and to avoid the effects of residual fatigue, a rest period of 48–72 h was provided between trials.

2.3.2. Incremental Maximal Aerobic Test (IMA)

An incremental test was performed on a treadmill HP Cosmos Pulsar (H Cosmos Sports & Medical GMBH, Nussdorf Traunstein, Germany) with a 1% incline [23]. To address the entire HR spectrum, a modification of the test proposed by Cerezuela-Espejo et al. [24] was carried out. The test was conducted according to the following phases (Figure 1): (1) participants started the test seated for 1 min; (2) participants began running at 4, 6, 8, 9, 10, 11, ..., km· h^{-1} (1 min each stage) and increased their speed by 1 km· h^{-1} until they reached exhaustion; (3) subjects maintained a constant intensity of 8 km· h^{-1} (5 min); (4) participants returned to a sitting position (1 min) (Figure 1, top). The tests were performed until exhaustion to determine each participant's maximum HR (HR_{max}) and maximal aerobic speed (MAS) [21], which were subsequently used in the HIIT testing sessions.

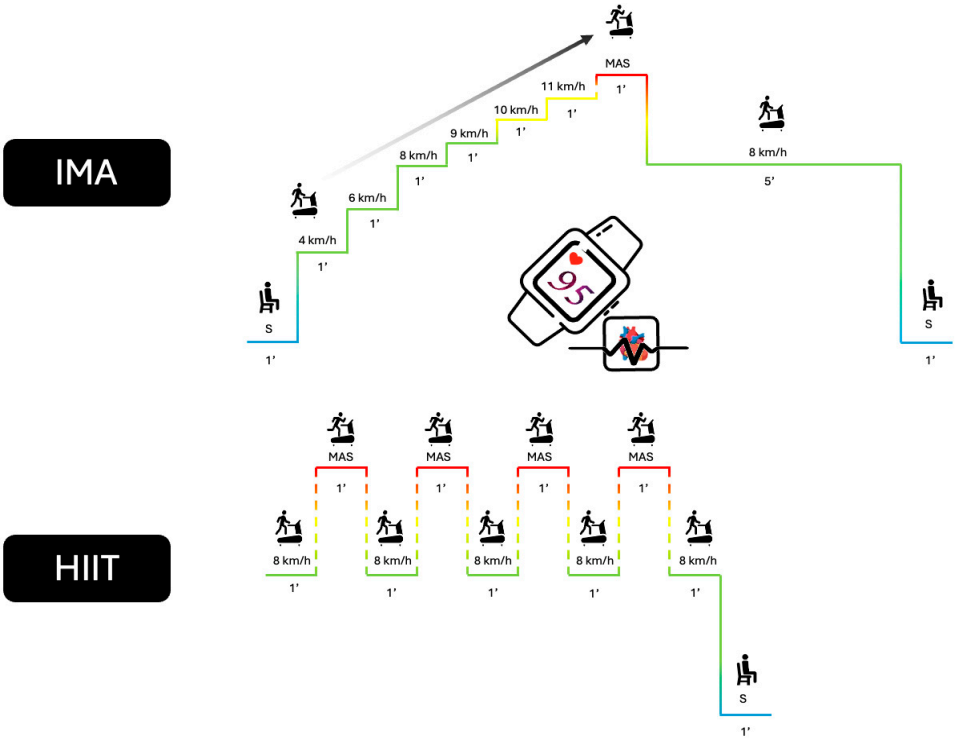

Figure 1. Incremental maximal aerobic (IMA) and high-intensity interval training (HIIT) tests.

2.3.3. High-Intensity Interval Training (HIIT) Test

A standard HIIT training protocol on the same treadmill was developed (Figure 1, bottom) [25]. The test consisted of four 1 min low-intensity intervals (running speed: 8 km· h^{-1}) alternated with four 1 min high-intensity intervals at each participant's 100% of their previously determined MAS, followed by an extra minute of a low-intensity interval at 8 km· h^{-1}. Then, the participants had to remain seated for 1 min.

2.3.4. Measurement Equipment and Data Acquisition

Before the beginning of each trial, the treadmill was calibrated according to the manufacturer's guidelines. HR per second was simultaneously recorded by different devices during the entire test. The Polar H9 chest strap was used as the reference criterion for the concurrent validity analysis given its excellent agreement and low magnitude of error compared to electrocardiogram device [22]. The batteries of all devices were fully charged before each session and checked along the data acquisition. A single researcher was responsible for placing the devices on each subject, strictly following the guidelines provided by each manufacturer to avoid the effects of misplacement. All the devices were positioned and turned on three mins before beginning each test. The synchronization was performed by up to four researchers simultaneously pressing the record button on each device, following the "three, two, one, press" signals. The same procedure was followed at the end of each test to terminate data recording.

2.3.5. Data Processing

The data were downloaded to a computer using the proprietary software corresponding to each device. Subsequently, the data were exported to an excel spreadsheet for refining and processing, which consisted of the following steps: (a) data pairing of the devices used in the trial; (b) count and remove records with zero or blank value (no record); (c) count and remove records of the alternative device with values of ±30% compared to the H9 chest strap; (d) average values every 20 s; and (e) create intensity ranges (<120 beats·min^{-1};

120–160 beats·min^{-1}; >160 beats·min^{-1}). Blank values and measurements deviating by more than 30% from the H9 chest strap were excluded to reflect the data that end users would realistically consider, as such values would be disregarded in practical use. Average values were obtained every 20 s to ensure an identical number of paired observations for each device–criterion vs. device–under-validation analysis. This last step was conducted to perform statistical analysis according to commonly used training zones [26].

2.3.6. Statistical Analyses

For the analysis of concurrent validity, a set of statistics that provided information on the level of agreement (absolute and relative terms) and magnitude of error was made:

- The intraclass correlation coefficients (ICCs) (1,k, one-way random effects, absolute agreement, multiple raters/measurements model) were calculated according to the guidelines presented by Koo and Li [27]. Moreover, ICCs were interpreted as <0.50, poor; 0.50–0.75, moderate; 0.75–0.90, good; >0.90, excellent [27].
- The standard error of measurement (SEM) was calculated from the square root of the mean square error term in a repeated-measures analysis of variance [28]. The results were presented both in absolute (beat·min^{-1}) and relative terms as a coefficient of variation (CV = 100 SEM/mean). For most sporting events, exercise performance tests, and sport technologies, CV should be lower than 5% [29,30].

Statistical analyses were performed using the SPSS software version 25.0 (IBM Corp., Armonk, NY, USA) and figures were designed using the GraphPad Prism software version 6.0 (GraphPad Software Inc., San Diego, CA, USA).

3. Results

During the IMA test, the participants reached an MAS of 19.2 ± 1.8 km· h^{-1} and reached a HR_{max} of 187 ± 8 beats·min^{-1}. The results of the concurrent validity of HR during IMA testing in different intensities are shown in Table 2. In general, the agreement (ICC value) is higher in the analysis of the total data compared to the different ranges of intensities in all devices. Similarly, at low intensity (<120 beat·min^{-1}), the ICC is higher compared to that at high intensities (>160 beat·min^{-1}) in all devices, except AMAZ-BIP5, POL-VANTV3, and POL-VS, which are almost identical. APPLE-ULTRA is the HR measurement device with the highest agreement and lowest magnitude of error (BIAS and SEM) across all intensity ranges. In addition, a low magnitude of error in relative terms were found in APPLE-ULTRA device (total = 0.8%; <120 beat·min^{-1} = 0.8%; 120–160 beat·min^{-1} = 0.9%; and >160 beat·min^{-1} = 0.4%; Figure 2). The analysis of missing data showed the GAR-35 to be the device with the highest percentage of lost data (9.5%), while the APPLE-ULTRA device showed no lost data, and did not require data refinement.

Table 3 shows the results of the concurrent validity of the HR data during the HIIT tests in different phases of the training. In general, the agreement (ICC value) is higher in resting time than it is during high-intensity training across all devices. APPLE-ULTRA is the HR measurement device with the highest agreement and lowest magnitude of error (BIAS and SEM) across all intensity ranges. In addition, a low magnitude of error in relative terms was found for the APPLE-ULTRA device (total = 1.0%; low intensity = 1.0%; high intensity = 0.9%; and cool down = 1.0%; Figure 3). The analyses of missing data and the data refinement process showed the GAR-35 to be the device with the highest percentage of lost data (6.8 and 12.7%, respectively), while AMAZ-BIP5, APPLE-SE, APPLE-ULTRA, POL-OH1, POL-VANTV, and POL-VS did not lose or require refined data.

Table 2. Concurrent validity of heart rate monitor devices during the incremental maximal aerobic test.

Device	Intensity	Agreement	Magnitude of Error		Data Lost and Refined	
		ICC	BIAS(SD)	SEM	Lost %	Refined %
AMAZ-BIP5	Total	0.997	−1 (3)	2	2.5	4.5
	<120 lat·min^{-1}	0.979	−2 (5)	4		
	120–160 lat·min^{-1}	0.985	0 (3)	2		
	>160 lat·min^{-1}	0.983	0 (2)	1		
APPLE-SE	Total	0.999	0 (1)	2	0.0	0.1
	<120 lat·min^{-1}	0.998	0 (2)	1		
	120–160 lat·min^{-1}	0.996	0 (1)	1		
	>160 lat·min^{-1}	0.899	−1 (5)	3		
APPLE- ULTRA	Total	1.000	0 (1)	1	0.0	0.0
	<120 lat·min^{-1}	0.998	0 (1)	1		
	120–160 lat·min^{-1}	0.994	0 (2)	1		
	>160 lat·min^{-1}	0.997	0 (1)	1		
GAR-235	Total	0.982	0 (9)	7	0.0	6.4
	<120 lat·min^{-1}	0.947	−1 (9)	6		
	120–160 lat·min^{-1}	0.828	3 (10)	7		
	>160 lat·min^{-1}	0.787	−2 (7)	5		
GAR-35	Total	0.944	−3 (14)	10	9.5	16.3
	<120 lat·min^{-1}	0.852	0 (11)	8		
	120–160 lat·min^{-1}	0.744	2 (13)	9		
	>160 lat·min^{-1}	0.161	−11 (13)	12		
GAR-45	Total	0.995	0 (4)	3	0.0	0.1
	<120 lat·min^{-1}	0.931	−3 (8)	6		
	120–160 lat·min^{-1}	0.978	0 (3)	2		
	>160 lat·min^{-1}	0.990	0 (2)	1		
GAR-935	Total	0.994	0 (5)	4	0.4	1.3
	<120 lat·min^{-1}	0.965	−1 (7)	5		
	120–160 lat·min^{-1}	0.958	1 (5)	3		
	>160 lat·min^{-1}	0.940	0 (4)	3		
POL-M200	Total	0.984	−2 (8)	6	0.0	2.2
	<120 lat·min^{-1}	0.974	−1 (6)	4		
	120–160 lat·min^{-1}	0.919	−1 (6)	5		
	>160 lat·min^{-1}	0.603	−4 (10)	7		
POL-M430	Total	0.992	−1 (6)	4	0.2	1.1
	<120 lat·min^{-1}	0.975	−1 (6)	4		
	120–160 lat·min^{-1}	0.918	−1 (7)	5		
	>160 lat·min^{-1}	0.916	−2 (4)	3		
POL-OH1	Total	0.998	−1 (3)	2	0.0	0.4
	<120 lat·min^{-1}	0.985	−1 (5)	4		
	120–160 lat·min^{-1}	0.991	0 (2)	2		
	>160 lat·min^{-1}	0.972	−1 (3)	2		
POL-VANTM	Total	0.989	−1 (7)	5	0.0	1.6
	<120 lat·min^{-1}	0.979	−1 (7)	5		
	120–160 lat·min^{-1}	0.898	0 (8)	5		
	>160 lat·min^{-1}	0.867	−2 (6)	4		
POL-VANTV	Total	0.989	−2 (7)	5	0.0	0.8
	<120 lat·min^{-1}	0.973	−1 (7)	5		
	120–160 lat·min^{-1}	0.976	−1 (4)	3		
	>160 lat·min^{-1}	0.673	−5 (9)	7		
POL-VANTV3	Total	0.996	−1 (4)	3	0.0	1.4
	<120 lat·min^{-1}	0.969	−2 (6)	4		
	120–160 lat·min^{-1}	0.983	0 (3)	2		
	>160 lat·min^{-1}	0.973	−1 (3)	2		
POL-VS	Total	0.997	0 (3)	2	0.0	0.4
	<120 lat·min^{-1}	0.976	−1 (5)	4		
	120–160 lat·min^{-1}	0.988	0 (2)	2		
	>160 lat·min^{-1}	0.986	0 (2)	1		
SAM-GAL5	Total	0.994	2 (4)	3	1.3	2.5
	<120 lat·min^{-1}	0.969	0 (6)	4		
	120–160 lat·min^{-1}	0.950	3 (4)	3		
	>160 lat·min^{-1}	0.951	1 (3)	2		

Table 2. *Cont.*

Device	Intensity	Agreement	Magnitude of Error		Data Lost and Refined	
		ICC	BIAS(SD)	SEM	Lost %	Refined %
SUN-3	Total	0.993	2 (6)	4	0.0	1.4
	<120 lat·min^{-1}	0.979	1 (6)	4		
	120–160 lat·min^{-1}	0.921	2 (7)	5		
	>160 lat·min^{-1}	0.932	2 (4)	3		

ICC: intraclass correlation coefficient; SEM: standard error of measurement; Lost: missing data (not registered or registered with zero value); SD: standard deviation.

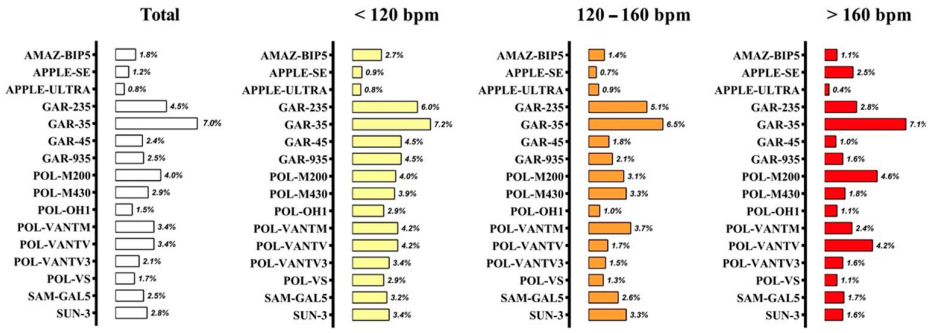

Figure 2. Relative standard error of measurement (CV) of heart rate monitor devices during the incremental maximal aerobic test.

Table 3. Concurrent validity of heart rate monitor devices during the high-intensity interval training test.

Device	Intensity	Agreement	Magnitude of Error		Data Lost and Refined	
		ICC	BIAS(SD)	SEM	Lost %	Refine %
AMAZ-BIP5	Total	0.977	−1 (6)	4	0.0	0.0
	Resting time	0.987	1 (5)	3		
	High intensity	0.925	−4 (6)	5		
	Cool down	0.978	4 (3)	3		
APPLE-SE	Total	0.997	0 (3)	2	0.0	0.0
	Resting time	0.997	0 (3)	2		
	High intensity	0.990	−1 (3)	2		
	Cool down	0.997	0 (2)	1		
APPLE- ULTRA	Total	0.999	0 (2)	1	0.0	0.0
	Resting time	0.999	0 (2)	1		
	High intensity	0.996	−1(2)	1		
	Cool down	0.998	1 (2)	1		
GAR-235	Total	0.889	−5 (12)	9	0.1	1.8
	Resting time	0.918	−7 (12)	10		
	High intensity	0.719	−6 (12)	9		
	Cool down	0.880	6 (10)	8		
GAR-35	Total	0.841	−9 (14)	12	6.8	12.7
	Resting time	0.879	−10 (13)	12		
	High intensity	0.702	−10 (12)	11		
	Cool down	0.834	8 (13)	11		
GAR-45	Total	0.984	0 (5)	4	2.4	0.0
	Resting time	0.991	1 (4)	3		
	High intensity	0.953	−2 (6)	4		
	Cool down	0.974	5 (4)	4		
GAR-935	Total	0.967	0 (7)	5	0.5	0.0
	Resting time	0.982	0 (6)	4		
	High intensity	0.938	−2 (6)	5		
	Cool down	0.836	9 (10)	10		
POL-M200	Total	0.875	−8 (14)	12	0.0	1.6
	Resting time	0.898	−8 (15)	12		
	High intensity	0.747	−9 (14)	12		
	Cool down	0.936	1 (10)	7		

Table 3. *Cont.*

Device	Intensity	Agreement	Magnitude of Error		Data Lost and Refined	
		ICC	BIAS(SD)	SEM	Lost %	Refine %
POL-M430	Total	0.971	−2 (7)	5	0.0	0.8
	Resting time	0.976	−1 (7)	5		
	High intensity	0.947	−4 (5)	5		
	Cool down	0.958	7 (5)	6		
POL-OH1	Total	0.988	−1 (5)	4	0.0	0.0
	Resting time	0.990	−1 (5)	4		
	High intensity	0.982	−2 (4)	3		
	Cool down	0.973	5 (4)	5		
POL-VANTM	Total	0.967	−1 (7)	5	0.0	0.1
	Resting time	0.975	0 (7)	5		
	High intensity	0.943	−4 (5)	5		
	Cool down	0.907	7 (7)	7		
POL-VANTV	Total	0.935	−6 (10)	8	0.0	0.0
	Resting time	0.942	−6 (11)	9		
	High intensity	0.885	−7 (8)	8		
	Cool down	0.960	3 (7)	5		
POL-VANTV3	Total	0.979	−1 (6)	4	0.0	0.2
	Resting time	0.983	0 (6)	5		
	High intensity	0.955	−4 (4)	4		
	Cool down	0.972	5 (4)	5		
POL-VS	Total	0.987	0 (5)	3	0.0	0.0
	Resting time	0.992	1 (4)	3		
	High intensity	0.971	−3 (4)	3		
	Cool down	0.973	5 (4)	5		
SAM-GAL5	Total	0.975	0 (6)	4	1.7	0.0
	Resting time	0.978	1 (6)	5		
	High intensity	0.944	−1 (6)	4		
	Cool down	0.974	4 (3)	4		
SUN-3	Total	0.961	−1 (8)	6	0.3	2.4
	Resting time	0.963	−2 (9)	6		
	High intensity	0.941	0 (7)	6		
	Cool down	0.946	2 (7)	6		

ICC: intraclass correlation coefficient; SEM: standard error of measurement; Lost: missing data (not registered or registered with zero value); SD: standard deviation.

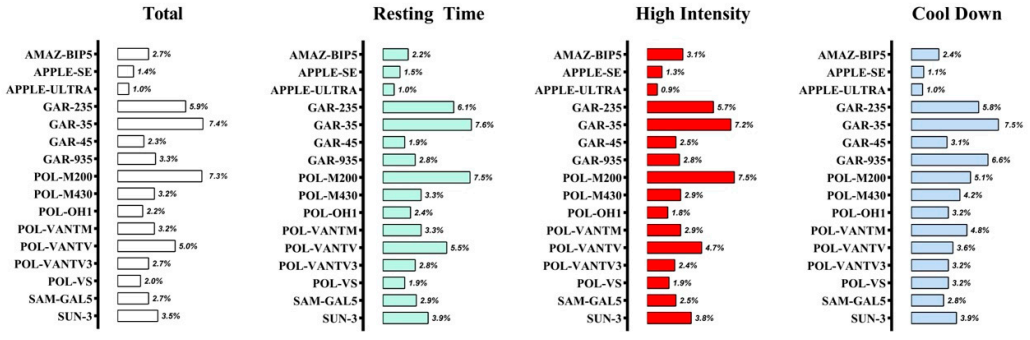

Figure 3. Relative standard error of measurement (CV) of heart rate monitor devices during the high-intensity interval training test.

4. Discussion

This study examined the concurrent validity of 16 commercial PPG-based heart rate monitors during two exercise protocols: an IMA test that spans the full HR spectrum and a HIIT test, where sudden changes in intensity are produced. Overall, the results showed excellent ICC for total intensity during the IMA and HIIT tests with the reference criterion (H9 chest strap); however, the magnitude of error varies depending on the device (BIAS, SEM, CV; Tables 2 and 3; Figures 2 and 3). In addition, it showed an overall tendency to

lower ICC and greater magnitudes of error (BIAS, SEM, CV; Tables 2 and 3; Figures 2 and 3) as the intensity increased. Nevertheless, there was certain difference between devices; for instance, while GAR-35 showed poor ICCs at high intensity during IMA test and a considerable magnitude of error (BIAS: −11, SEM: 12, CV: 7.2%), APPLE-ULTRA showed an excellent ICC and a tiny magnitude of error (BIAS: 0, SEM: 1, CV: 0.4%). These findings suggest that PPG-based HR monitors validity varies based on the device, and caution needs to be taken when interpreting the data of some devices, especially at high intensities and when performing interval activities.

Previous validation studies have established validity criteria for HR measurements as an ICC $\geq$ 0.90, a mean BIAS < 3 beats·min^{-1}, and an SEM < 5 beats·min^{-1}, with the satisfaction of at least the first and second criteria together [11,13]. During the IMA test, when checking the total intensity, all the devices met at the least the first and second criteria; however, GAR-235, GAR-35, POL-M200, POL-VANTM, and POL-VANTV did not meet the third criteria (Table 2). As the intensity increased, the validity of the devices decreased, more specifically, at high intensities (i.e., >160 beats·min^{-1}), GAR-235, GAR-35, POL-M200, POL-VANTM, and POL-VANTV did not satisfy the validation criteria (Table 2). These results are in line with those of previous studies that explored some of these models, which presented different results depending on the exercise and intensity assessed [16,31,32]. However, the results contrast those of previous studies that found better accuracy at greater intensities [3,16,33]. It was suggested that, with increased intensity, there is an improved perfusion that could decrease the error [3,16,33]. Nevertheless, Støve et al. [16] also suggested that, beyond 12.1 km· h^{-1}, the accuracy of these devices might decrease. Indeed, when comparing intensities, the aforementioned studies reached speeds of ~10 km· h^{-1}, which are considerably lower in comparison with those achieved in the current experiments (19.2 ± 1.8 km· h^{-1}). Therefore, at high speeds, greater arm movement combined with excessive sweating may impar measurement accuracy [16,34]. The formation of a thin layer of sweat between the skin and the sensor can alter the optical signal, in a manner similar (albeit less pronounced) to what has been discussed during swimming [10,35].

To date, most validation studies have been conducted under controlled exercise intensities, with participants walking, running, or swimming at steady speeds, or cycling at a constant power output [8,10,11,16,32]. However, previous research also indicated rapid intensity changes as a possible source of error [31,36]. To test these sudden changes in intensity, the HIIT protocol was conducted. Our results are indeed in agreement with previous research [31,36] and the majority of devices show worse results than during the IMA test, especially greater BIAS, SEM, and CV were observed (Tables 2 and 3; Figures 2 and 3). The reason behind this worse utility for HIIT might be due to the lag in PPG signals compared to ECG, which have been attributed to several physiological causes, such as the delay between changing HR and changes in blood volume at the wrist [36]. Despite this factor, APPLE-SE and, especially, APPLE-ULTRA were able to measure with a very small error (high intensity; BIAS, −1 and −1; SEM, 2 and 1; CV, 1.3 and 0.9%), suggesting that these devices might be useful in all conditions.

Data loss occurred in several devices (AMAZ-BIP5, GAR-235, GAR-35, GAR-45, GAR-935, POL-M430, SAM-GAL5, and SUN-3) during the HIIT and/or IMA protocols (Tables 2 and 3). For some devices, the missing data were negligible (<0.5%), whereas for others, the magnitude of signal loss was substantial (>2%), which could compromise HR validity during these measurements. Moreover, except for APPLE-SE and APPLE-ULTRA, all the devices required some data refinement (i.e., we removed records with values of ±30% compared to the H9 chest strap), indicating some measurement errors. These data should be interpreted with caution when assessing device validity. Blank values

and measurements deviating by more than 30% from the H9 chest strap were excluded to reflect data that end users would realistically consider; however, this exclusion may positively influence apparent device validity. For instance, see the amount of data lost and refined for GAR-35. This should therefore be taken into account when selecting a device. More specifically, this refinement might be negligible from a mean value perspective, but it may distort instantaneous readings and should be carefully considered when participating in exercises involving rapid and frequent HR fluctuations, as seen in real-time monitoring during HIIT. These findings align with previous research, showing the greater validity of Apple devices, as data were neither lost nor refined compared with other commercially available wearables [32,37].

This study provides a validation of a wide range of PPG-based HR monitors across two different running protocols, covering the complete HR spectrum during incremental exercise (i.e., IMA protocol) and during a typical HIIT protocol. This allows the results to be transferred to the most commonly performed exercises on running. However, it also presents some limitations. The fact that is performed only on running might limit its transferability to other disciplines such as cycling; nevertheless, it is important to note that running is the discipline with greater arm movements and that might induce more error. In addition, all measurements were performed indoors, which may not fully reflect the environmental variability and movement patterns encountered during outdoor running. Moreover, the sample comprised only healthy, young Caucasian males who were aerobically fit and within the normal range of body mass. Therefore, the findings may not be generalizable to females, individuals with different body sizes or skin tones, or to younger or older populations [38]. Finally, there was a likely lack of statistical independence, as repeated heart rate measures were nested within individuals. Treating these observations as independent may inflate precision and should be considered when interpreting the validity estimates.

5. Conclusions

The present study showed that all devices provided valid HR measurements when the full recording was considered; however, when data were analysed by intensity clusters, most devices struggled to maintain accuracy, with greater error (BIAS, SEM, and CV) at high intensities. Except for the Apple devices, the accuracy of the assessed devices decreased substantially during the HIIT protocol (Lower ICC, greater BIAS, SEM, and CV). Moreover, the amount of data lost and refined varies considerably across devices and should also be considered when evaluating their real-world validity. Altogether, these findings suggest that PPG-based HR monitors validity varies based on the device, and caution need to be taken when interpreting the data of some of the devices evaluated, especially at high intensities and when performing interval activities.

Author Contributions: Conceptualization, P.O., A.S.-P., A.B., J.J.R.-N., and A.M.-C.; Methodology, P.O., A.S.-P., J.J.R.-N., and A.M.-C.; Software, P.O., A.S.-P., A.B., J.J.R.-N., and A.M.-C.; Formal Analysis, P.O., A.S.-P., E.C.-R., J.J.R.-N., and A.M.-C.; Investigation, P.O., A.S.-P., E.C.-R., A.B., J.J.R.-N., and A.M.-C.; Data Curation, P.O., A.S.-P., E.C.-R., A.B., J.J.R.-N., and A.M.-C.; Writing—Original Draft Preparation, P.O., A.S.-P., E.C.-R., A.B., J.J.R.-N., and A.M.-C.; Writing—Review and Editing, P.O., A.S.-P., E.C.-R., A.B., J.J.R.-N., and A.M.-C.; Supervision, A.S.-P., J.J.R.-N., and A.M.-C. All authors have read and agreed to the published version of the manuscript.

Funding: AMC is partially funded by CAR Region de Murcia (ID: 39782/2024).

Institutional Review Board Statement: The study was conducted according to the Declaration of Helsinki and approved by the Ethics Commission of the University of Murcia (ID: M10/2024/022, approval date: 28 February 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available upon request to the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Achten, J.; Jeukendrup, A.E. Heart Rate Monitoring Applications and Limitations. *Sports Med.* **2003**, *33*, 517–538. [CrossRef]
2. Seiler, K.S.; Kjerland, G.Ø. Quantifying training intensity distribution in elite endurance athletes: Is there evidence for an "optimal" distribution? *Scand. J. Med. Sci. Sports* **2006**, *16*, 49–56. [CrossRef] [PubMed]
3. Stahl, S.E.; An, H.S.; Dinkel, D.M.; Noble, J.M.; Lee, J.M. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exerc. Med.* **2016**, *2*, e000106. [CrossRef] [PubMed]
4. Spierer, D.K.; Rosen, Z.; Litman, L.L.; Fujii, K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *J. Med. Eng. Technol.* **2015**, *39*, 264–271. [CrossRef] [PubMed]
5. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1. [CrossRef]
6. Montalvo, S.; Martinez, A.; Arias, S.; Lozano, A.; Gonzalez, M.P.; Dietze-Hermosa, M.S.; Boyea, B.L.; Dorgo, S. Commercial Smart Watches and Heart Rate Monitors: A Concurrent Validity Analysis. *J. Strength Cond. Res.* **2023**, *37*, 1802–1808. [CrossRef]
7. Alugubelli, N.; Abuissa, H.; Roka, A. Wearable Devices for Remote Monitoring of Heart Rate and Heart Rate Variability—What We Know and What Is Coming. *Sensors* **2022**, *22*, 8903. [CrossRef]
8. Ruiz-Malagón, E.; Ruiz-Alias, S.; García-Pinillos, F.; Delgado-García, G.; Soto-Hermoso, V.M. Comparison between photoplethysmographic heart rate monitor from Polar Vantage M and Polar V800 with H10 chest strap while running on a treadmill: Validation of the Polar Precision Prime™ photoplestimographic system. *Proc. Inst. Mech. Eng. Part P J. Sports Eng. Technol.* **2021**, *235*, 212–218. [CrossRef]
9. Schubert, M.; Clark, A.; De La Rosa, A. The Polar® OH1 Optical Heart Rate Sensor is Valid during Moderate-Vigorous Exercise. *Sports Med. Int. Open* **2018**, *2*, E67–E70. [CrossRef]
10. López-Belmonte, Ó.; Febles-Castro, A.; Gay, A.; Cuenca-Fernandez, F.; Arellano, R.; Ruiz-Navarro, J.J. Validity of the polar verity sense during swimming at different locations and intensities. *Scand. J. Med. Sci. Sports* **2023**, *33*, 2623–2625. [CrossRef]
11. Hettiarachchi, I.T.; Hanoun, S.; Nahavandi, D.; Nahavandi, S. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. *PLoS ONE* **2019**, *14*, e0217288. [CrossRef] [PubMed]
12. Navalta, J.W.; Montes, J.; Bodell, N.G.; Salatto, R.W.; Manning, J.W.; DeBeliso, M. Concurrent heart rate validity of wearable technology devices during trail running. *PLoS ONE* **2020**, *15*, e0238569. [CrossRef] [PubMed]
13. Jo, E.; Lewis, K.; Directo, D.; Kim, M.J.Y.; Dolezal, B.A. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. *J. Sports Sci. Med.* **2016**, *15*, 540–547. [PubMed]
14. Caminal, P.; Sola, F.; Gomis, P.; Guasch, E.; Perera, A.; Soriano, N.; Mont, L. Validity of the Polar V800 monitor for measuring heart rate variability in mountain running route conditions. *Eur. J. Appl. Physiol.* **2018**, *118*, 669–677. [CrossRef]
15. Bunn, J.A.; Wells, E.K.; Manor, J.P.; Webster, M.J. Evaluation of Earbud and Wristwatch Heart Rate Monitors during Aerobic and Resistance Training. *Int. J. Exerc. Sci.* **2019**, *12*, 374–384. [CrossRef]
16. Støve, M.P.; Haucke, E.; Nymann, M.L.; Sigurdsson, T.; Larsen, B.T. Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity. *J. Sports Sci.* **2019**, *37*, 895–901. [CrossRef]
17. Dooley, E.E.; Golaszewski, N.M.; Bartholomew, J.B. Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR mHealth uHealth* **2017**, *5*, e34. [CrossRef]
18. Wang, R.; Blackburn, G.; Desai, M.; Phelan, D.; Gillinov, L.; Houghtaling, P.; Gillinov, M. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol.* **2017**, *2*, 104–106. [CrossRef]
19. Schäfer, A.; Vagedes, J. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int. J. Cardiol.* **2013**, *166*, 15–29. [CrossRef]
20. Mckay, A.K.A.; Stellingwerff, T.; Smith, E.S.; Martin, D.T.; Mujika, I.; Goosey-Tolfrey, V.L.; Sheppard, J.; Burke, L.M. Defining training and performance caliber: A participant classification framework. *Int. J. Sports Physiol. Perform.* **2022**, *17*, 317–331. [CrossRef]
21. Pallarés, J.G.; Cerezuela-Espejo, V.; Morán-Navarro, R.; Martínez-Cava, A.; Conesa, E.; Courel-Ibáñez, J. A New Short Track Test to Estimate the VO_2max and Maximal Aerobic Speed in Well-Trained Runners. *J. Strength Cond. Res.* **2019**, *33*, 1216–1221. [CrossRef] [PubMed]
22. Gilgen-Ammann, R.; Schweizer, T.; Wyss, T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *Eur. J. Appl. Physiol.* **2019**, *119*, 1525–1532. [CrossRef] [PubMed]
23. Jones, A.M.; Doust, J.H. A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *J. Sports Sci.* **1996**, *14*, 321–327. [CrossRef] [PubMed]

24. Cerezuela-Espejo, V.; Courel-Ibáñez, J.; Morán-Navarro, R.; Martínez-Cava, A.; Pallarés, J.G. The relationship between lactate and ventilatory thresholds in runners: Validity and reliability of exercise test performance parameters. *Front. Physiol.* **2018**, *9*, 1320. [CrossRef]
25. Buchheit, M.; Laursen, P.B. High-intensity interval training, solutions to the programming puzzle: Part I: Cardiopulmonary emphasis. *Sports Med.* **2013**, *43*, 313–338. [CrossRef]
26. Pallarés, J.G.; Morán-Navarro, R.; Ortega, J.F.; Fernández-Elías, V.E.; Mora-Rodriguez, R. Validity and reliability of ventilatory and blood lactate thresholds in well-trained cyclists. *PLoS ONE* **2016**, *11*, e0163389. [CrossRef]
27. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163, Erratum in *J. Chiropr. Med.* **2017**, *16*, 346. [CrossRef]
28. Atkison, R.R.; Nevill, A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* **1998**, *26*, 217–238. [CrossRef]
29. Hopkins, W.G. Measures of Reliability in Sports Medicine and Science. *Sports Med.* **2000**, *30*, 375–381. [CrossRef]
30. Courel-Ibáñez, J.; Martínez-Cava, A.; Morán-Navarro, R.; Escribano-Peñas, P.; Chavarren-Cabrero, J.; González-Badillo, J.J.; Pallarés, J.G. Reproducibility and Repeatability of Five Different Technologies for Bar Velocity Measurement in Resistance Training. *Ann. Biomed. Eng.* **2019**, *47*, 1523–1538. [CrossRef]
31. Icenhower, A.; Murphy, C.; Brooks, A.K.; Irby, M.; N'dah, K.; Robison, J.; Fanning, J. Investigating the accuracy of Garmin PPG sensors on differing skin types based on the Fitzpatrick scale: Cross-sectional comparison study. *Front. Digit. Health* **2025**, *7*, 1553565. [CrossRef]
32. Hajj-Boutros, G.; Landry-Duval, M.A.; Comtois, A.S.; Gouspillou, G.; Karelis, A.D. 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. *Eur. J. Sport Sci.* **2023**, *23*, 165–177. [CrossRef]
33. Shcherbina, A.; Mikael Mattsson, C.; Waggott, D.; Salisbury, H.; Christle, J.W.; Hastie, T.; Wheeler, M.T.; Ashley, E.A. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J. Pers. Med.* **2017**, *7*, 3. [CrossRef] [PubMed]
34. Gillinov, S.; Etiwy, M.; Wang, R.; Blackburn, G.; Phelan, D.; Gillinov, A.M.; Houghtaling, P.; Javadikasgari, H.; Desai, M.Y. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med. Sci. Sports Exerc.* **2017**, *49*, 1697–1703. [CrossRef] [PubMed]
35. Olstad, B.H.; Zinner, C. Validation of the Polar OH1 and M600 optical heart rate sensors during front crawl swim training. *PLoS ONE* **2020**, *15*, e0231522. [CrossRef] [PubMed]
36. Bent, B.; Goldstein, B.A.; Kibbe, W.A.; Dunn, J.P. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit. Med.* **2020**, *3*, 18. [CrossRef]
37. Düking, P.; Giessing, L.; Frenkel, M.O.; Koehler, K.; Holmberg, H.C.; Sperlich, B. Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: Validation study. *JMIR mHealth uHealth* **2020**, *8*, e16716. [CrossRef]
38. Castaneda, D.; Esparza, A.; Ghamari, M.; Soltanpur, C.; Nazeran, H. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int. J. Biosens. Bioelectron.* **2018**, *4*, 195–202. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.