



Article

Assessing Heart Rate Using Consumer Technology Association Standards

Joel D. Reece ^{1,*}, Jennifer A. Bunn ², Minsoo Choi ¹ and James W. Navalta ³

¹ Exercise & Sport Science, Laie, HI, Brigham Young University, Oahu, HI 96762, USA; choiminsoo8138@gmail.com

² College of Health Sciences, Sam Houston State University, Huntsville, TX 77340, USA; jab229@shsu.edu

³ Department of Kinesiology and Nutrition Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA; james.navalta@unlv.edu

* Correspondence: joel.reece@byuh.edu

Abstract: It is difficult for developers, researchers, and consumers to compare results among emerging wearable technology without using a uniform set of standards. This study evaluated the accuracy of commercially available wearable technology heart rate (HR) monitors using the Consumer Technology Association (CTA) standards. Participants (N = 23) simultaneously wore a Polar chest strap (criterion measure), Jabra Elite earbuds, Scosche Rhythm 24 armband, Apple Watch 4, and Garmin Forerunner 735 XT during sitting, activities of daily living, walking, jogging, running, and cycling, totaling 57 min of monitored activity. The Apple Watch mean bias was within ± 1 bpm, and mean absolute percent error (MAPE) was $< 3\%$ in all six conditions. Garmin underestimated HR in all conditions, except cycling and MAPE was $> 10\%$ during sedentary, lifestyle, walk-jog, and running. The Jabra mean bias was within ± 5 bpm for each condition, and MAPE exceeded 10% for walk-jog. The Scosche mean bias was within ± 1 bpm and MAPE was $< 5\%$ for all conditions. In conclusion, only the Apple Watch Series 4 and the Scosche Rhythm 24 displayed acceptable agreement across all conditions. By employing CTA standards, future developers, researchers, and consumers will be able to make true comparisons of accuracy among wearable devices.

Keywords: wearable; tracking; devices



Citation: Reece, J.D.; Bunn, J.A.; Choi, M.; Navalta, J.W. Assessing Heart Rate Using Consumer Technology Association Standards. *Technologies* **2021**, *9*, 46. <https://doi.org/10.3390/technologies9030046>

Academic Editors: Manoj Gupta and Alessandro Tognetti

Received: 27 April 2021
Accepted: 23 June 2021
Published: 30 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The International Data Corporation estimates a 22.4% compound annual growth rate of wearable technology, equating to approximately 489.1 million units shipped globally in 2023, including smartwatches (22.3%), wristbands (14.3%), earwear (56.0%), and other wearable technology (7.5%) [1]. These devices are popular for day-to-day life as well as potential use in the health care industry [2], highlighting the importance of a standard of accuracy testing for these devices.

Standards that outline protocols and validation criteria for wearables have recently been presented by the Consumer Technology Association (CTA) for step counting [3], sleep [4], and heart rate [5]. These standards provide guidance for researchers to evaluate research-grade devices as well as commercial devices used by the lay public. Since the CTA released their standards for protocol and validation, researchers have begun to acknowledge them in their designs for step count [6–9], and HR [10–12]. However, only limited studies have actually implemented these standards [6].

Recent reviews of wearable technology have referred to or recommended the use of CTA's standards in validation research [13–17]. These CTA standards attempt to reconcile potential inaccuracies in current wearable heart rate monitor studies (e.g., a lack of investigation of diverse skin types [18]) and give common ground to compare the accuracy of devices.

Despite standardized assessment protocols developed by CTA [3–5] for the ever-growing wearable technology industry [1], implementation of these standards is still novel. Without implementation of these standards, it is difficult for developers, researchers, and consumers to compare the accuracy of these devices. Being able to compare accuracy will allow researchers to provide comparable information, developers to reconcile inaccuracies, and consumers to make better-educated decisions for their wearable tech purchases.

Therefore, the purpose of this study was to evaluate the accuracy of commercially available wearable technology heart rate monitors using the CTA standards. This study is among the first to evaluate the accuracy of commercially available wearable technology heart rate (HR) monitors using CTA standards. It was hypothesized that commercially available wearable technology devices would return valid heart rate measures compared to a criterion measure (Polar chest strap) during rest, lifestyle activity, walking, jogging, and cycling.

2. Materials and Methods

This was a prospective cross-sectional study design. This study was conducted with the approval of the university's institutional review board and in conjunction with the Declaration of Helsinki. All participants completed an informed consent and were given an opportunity to ask questions prior to study participation.

In accordance with the CTA protocol [5], 23 participants ($n = 13$ males, $n = 10$ females) completed testing. To be eligible for the study, participants had to be capable of completing the exercise testing, 18 years or older, and answer no to all the questions to the Physical Activity Readiness Questionnaire (PAR-Q). Participants were excluded if they had any skin wounds or tattoos where the heart rate sensors were placed. The CTA protocol dictates that at least 25% of the participants should be of lighter skin and (1–3 on Fitzpatrick Scale) and at least 25% of participants darker skin range (4–6 on Fitzpatrick Scale). Furthermore, at least 10% of the participants ($n > 2$, $n = 3$, 13%) with a body mass index (BMI) below $20 \text{ kg}\cdot\text{m}^{-2}$, and at least 10% of participants ($n > 2$, $n = 10$, 44%) above $25 \text{ kg}\cdot\text{m}^{-2}$. These CTA mandates were followed accordingly; participants were 23.57 ± 3.22 years old, 73.40 ± 16.17 kg, 171.53 ± 10.52 cm, and $24.87 \pm 4.45 \text{ kg}\cdot\text{m}^{-2}$. The median score was 3 on the Fitzpatrick Scale, with 57% of participants with lighter skin and 43% of participants with darker skin.

Participants wore five HR devices simultaneously, and all devices were worn in accordance to each manufacturer's guidelines and set to the appropriate mode for collection of HR data (See Figure 1). Devices were chosen based on availability, the ability to be placed at different locations, and because they were regarded as commonly used brands. The Polar H7 (Kempele, Finland) chest strap heart rate monitor (CHRM) was previously validated [19] and used as the benchmark heart rate assessment for comparison in accordance with the CTA protocol. It was placed directly on the skin at the height of the xiphoid process, and electrical conductance gel was used to ensure appropriate conductance. The devices tested included the Jabra Elite earbuds (Copenhagen, Denmark), the Scosche Rhythm 24 (Ventura County, CA, USA), the Apple Watch Series 4 (Cupertino, CA, USA), and the Garmin Forerunner 735 XT (Olathe, KS, USA). The Jabra earbuds were placed in both ears, the Scosche band was placed on the forearm just distal of the humeroulnar joint, the Apple Watch was placed on one wrist, and the Garmin watch was placed on the other wrist. The Scosche, Apple, and Garmin devices were randomly placed on dominant and non-dominant arms with each participant. The PerformTek app (Valencell, Inc, Raleigh, NC, USA) collected data from the Polar chest strap, Jabra earbuds, and Rhythm 24 devices. Data from the Apple and Garmin watches were collected through their respective apps. App profiles do not include participant identification.

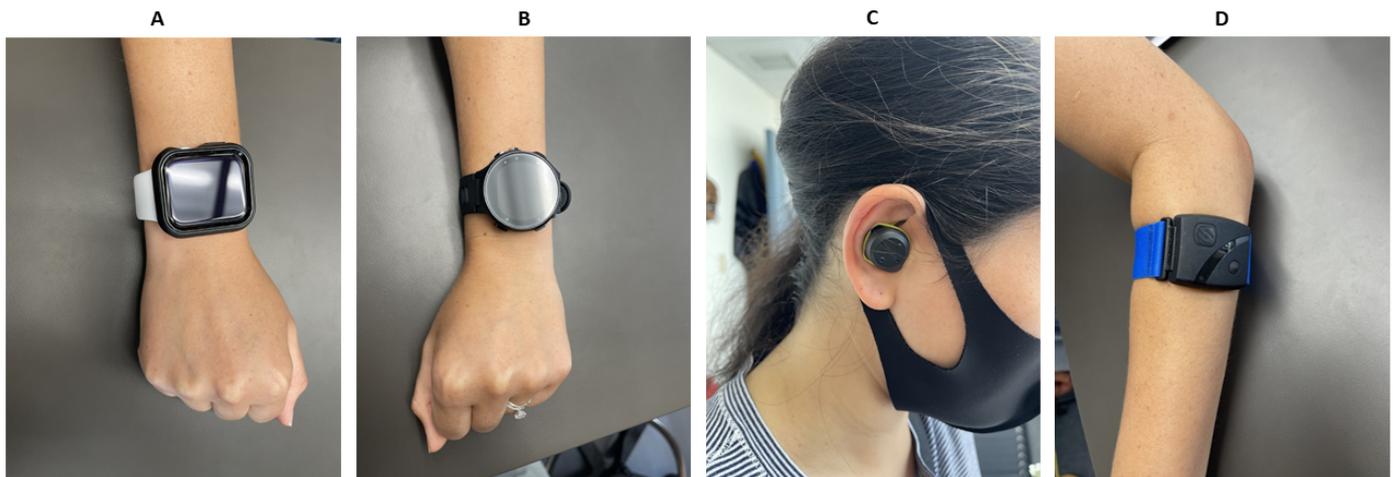


Figure 1. Placement of wearable heart rate monitors. (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices.

Technicians informed participants on the duration and characteristics of the protocol, as well as when each section of the test began and ended. Per the CTA protocol [5], the participant began seated quietly, refraining from engaging with external stimuli. Participants completed each section of the CTA protocol during one session (Figure 2). Technicians adhered to the timing and intensities indicated in the CTA protocol for the sedentary, lifestyle activities, walking, jogging, running, and cycling. For the lifestyle activities section, participants completed the two minutes of full body activities of daily living (9:30–11:30) folding laundry and simulated grocery shopping. Jogging/running was completed on a treadmill (Life Fitness, 95 Ti, Franklin Park, IL, USA). When cycling, participants kept their hands on the bicycle ergometer (Monark, Ergomedic 874E, Varberg, Sweden) handlebars in a natural position. Exercise intensities were defined by metabolic equivalent task (MET) level, ratings of perceived exertion (RPE), and/or HR max. Moderate intensity was defined as 3.0–6.0 METs, RPE 12–13 (Borg scale 6–20), or HR max of 64–76%. Vigorous intensity was defined as >6.0 METs, RPE of 14–17, HR max 77–95%.

| Sedentary 6 Minutes | Lifestyle 13 Minutes | Walking 7 Minutes | Dynamic Walk/Jog 12 Minutes | Running 7 Minutes | Cycling 12 Minutes |
|--|---|------------------------------------|--------------------------------|---------------------------------|---|
| 1 min. sit quietly | 1 min. sit quietly | 5 min. walking treadmill 2.5–4 mph | 1 min. stand quietly | 5 min. run treadmill 5.0–12 mph | 1 min. sit quietly on bike |
| 5 min. Seated, no talking, no moving, quiet room | 2 min. upper-body daily living activities | 1 min. stand quietly | 1 min. stand quietly | 1 min. stand quietly | 2 min. easy to moderate cycling |
| | 1 min. sit quietly | 1 min. stand quietly | 1 min. walk comfortable pace | 1 min. stand quietly | 2 min. vigorous cycling (85 rpm) |
| | 2 min. walk | 3 min. self-selected running | 1 min. walk comfortable pace | 1 min. stand quietly | 2 min. easy to moderate cycling |
| | 2 min. seated typing | 1 min. walk comfortable pace | 2 min. vigorous run | 1 min. stand quietly | 2 min. vigorous cycling (high resistance) |
| | 2 min. full body daily living activities | 2 min. more vigorous run | 1 min. walk comfortable pace | 1 min. stand quietly | 2 min. easy to moderate cycling |
| | 1 min. sit quietly | 1 min. walk comfortable pace | 2 min. vigorous run | 1 min. stand quietly | 2 min. easy to moderate cycling |
| | 1 min. stand quietly | 5 min. run treadmill 5.0–12 mph | 1 min. stand quietly | 1 min. stand quietly | 1 min. sit quietly on bike |

Figure 2. Summary of Consumer Technology Association standards for heart rate testing (5).

The Jabra, Scosche, and Polar data were obtained from the PerformTek app in one-second increments. The Apple and Garmin data were obtained from their respective apps, but these data were not provided in one-second increments. Rather, data were provided at variable time points, sometimes every second, sometimes every three seconds, and sometimes with long breaks of more than a minute in between data points. Data from each device was aligned according to the timestamp with the appropriate start time as the Polar chest strap in conjunction with the guidelines provided by the CTA protocol [5].

The data were then graphed to show any inherent delays with a device and aligned appropriately and analyzed. Additionally, almost all commercially available wearables have data dropout that may be due to damaged devices, failure of the tester to apply the testing devices properly, failure of the tester and/or the participant to follow the protocol, as well as other reasons that are inexplicable but are apparent when visualizing the data. It is common practice in research that data analysts make an honest effort to include all valid data acquired regardless of agreement between the reference device and test device, and report the percentage of data removed and retained per device [10]. Data were removed for all devices if the Polar chest strap registered a zero for HR. In addition, each device was also examined, and data were removed from analysis if there was a registered zero as the HR for a device or a locked HR. The total number and percentage of data points removed for each device were calculated. The following analyses were conducted after all zeroes were removed from the data sets.

Each device under test was compared only to the Polar chest strap. Accuracy was assessed for each device within each activity via mean bias, mean absolute percent error (MAPE), Bland-Altman analysis, an equivalence test, and intraclass correlation coefficient (ICC) using the two-way mixed model and absolute agreement in IBM SPSS (IBM Statistics version 26.0, Armonk, NY, USA). This analysis is consistent with previous literature employing multiple testing methods when aiming to prove that data are the same [6,14,20,21]. Equivalence tests were conducted using the two-one-sided-tests method with Jamovi statistical software [22]. Equivalence was determined with upper and lower boundaries set at 5% of the mean heart rate from the Polar chest strap for each activity. MAPE and Bland-Altman analyses were conducted using Microsoft Excel (Redmond, WA, USA). Accuracy of a device was considered if the MAPE was <10%, ICC > 0.90, and if it was equivalent to the Polar chest strap.

3. Results

Figures 3–8 demonstrate Bland-Altman plots of heart rate for each device during the six conditions of activity compared to the Polar H7. Tables 1–4 present the results for the Apple Watch series 4, Garmin 735 XT watch, Jabra Elite earbuds, and Scosche Rhythm 24 armband, respectively. There were several instances in which the Polar chest strap registered a heart rate of zero bpm. Data points were removed during the sedentary condition (875, 10.6%), lifestyle condition (499, 2.8%), walking (8, 0.1%), walk/jog (3, 0.02%), running (24, 0.2%), and cycling (132, 0.8%). The Apple and Scosche devices did not have any zeroes registered as HR, so no data points were removed from analysis for these devices, as indicated in Table 1. Data were removed from analysis for the Garmin watch for the lifestyle condition (88, 4.5%), and data from one participant malfunctioned during the lifestyle condition. Additionally, data from two participants malfunctioned during the walking, walk/jog, running, and cycling conditions. Thus, analysis of the Garmin watch was conducted with 22 participants for the lifestyle condition and 21 participants for walking, walk/jog, running, and cycling. The number and percentage of data points removed from the Garmin data are shown in Table 2. The Jabra earbuds registered zero as a HR as shown in Table 3 during the sedentary condition (48, 0.6%), lifestyle condition (225, 1.3%), walking (178, 1.8%), walk/jog (739, 4.5%), running (333, 3.4%), and cycling (295, 1.8%).

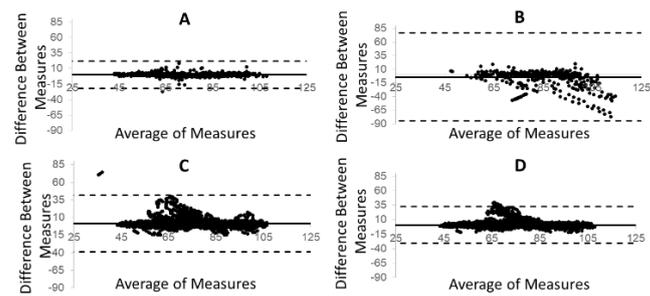


Figure 3. Bland–Altman plots of heart rate (bpm) during sedentary time for the (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices compared to the Polar H7. Dashed lines represent 95% limits of agreement.

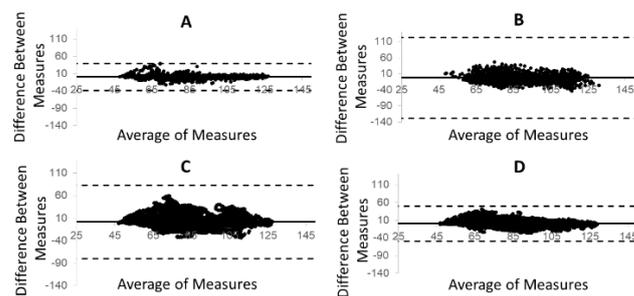


Figure 4. Bland–Altman plots of heart rate (bpm) during lifestyle activity for the (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices compared to the Polar H7. Dashed lines represent 95% limits of agreement.

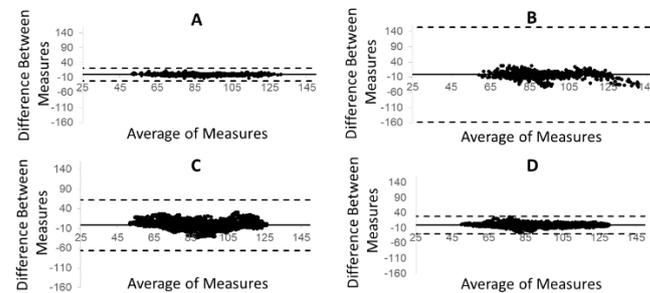


Figure 5. Bland–Altman plots of heart rate (bpm) during walking activity for the (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices compared to the Polar H7. Dashed lines represent 95% limits of agreement.

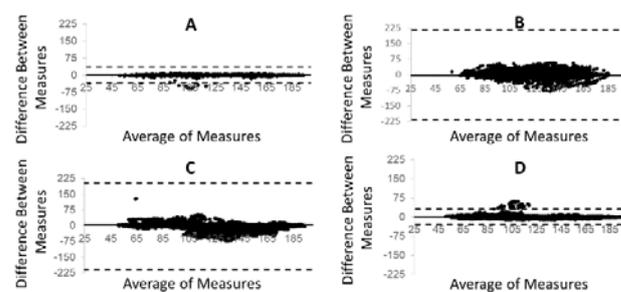


Figure 6. Bland–Altman plots of heart rate (bpm) during dynamic walk-jog activity for the (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices compared to the Polar H7. Dashed lines represent 95% limits of agreement.

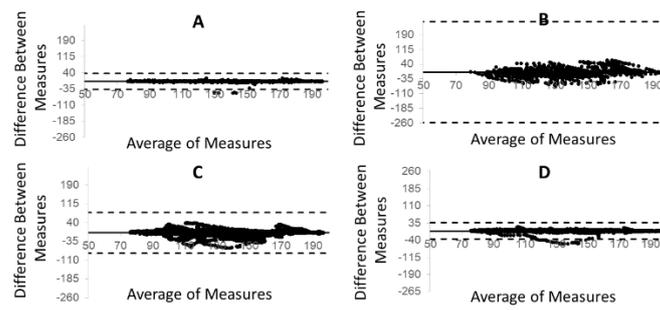


Figure 7. Bland–Altman plots of heart rate (bpm) during running activity for the (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices compared to the Polar H7. Dashed lines represent 95% limits of agreement.

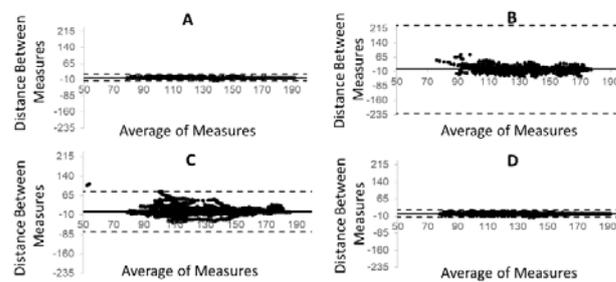


Figure 8. Bland–Altman plots of heart rate (bpm) during cycling activity for the (A) Apple Watch series 4, (B) Garmin Forerunner 735XT, (C) Jabra Elite earbuds, and (D) Scosche Rhythm 24 devices compared to the Polar H7. Dashed lines represent 95% limits of agreement.

Table 1. Results of all tests for the Apple Watch Series 4.

| | Sedentary | Lifestyle | Walking | Dynamic Walk-Jog | Running | Cycling |
|---|---|---|---|---|---|---|
| Total n data points (n removed, % removed) | 1416 (0, 0%) | 3455 (0, 0%) | 1868 (0, 0%) | 3342 (0, 0%) | 1934 (0, 0%) | 3302 (0, 0%) |
| Mean bias (bpm) | 0.19 (0.30) | 0.09 (3.32) | 0.02 (0.29) | 0.37 (1.00) | −0.16 (0.52) | 0.09 (0.04) |
| MAPE (%) | 1.69 | 2.95 | 1.50 | 1.17% | 0.92 | 0.62 |
| ICC | 0.989 | 0.963 | 0.991 | 0.993 | 0.992 | 0.998 |
| | (0.988, 0.990) | (0.960, 0.965) | (0.990, 0.991) | (0.992, 0.994) | (0.991, 0.993) | (0.998, 0.998) |
| LOA | −21.63 (−22.43, −20.83) to 22.01 (21.21, 22.82) | −38.58 (−39.58, −37.59) to 38.77 (37.78, 39.76) | −21.16 (−22.12, −20.19) to 21.19 (20.23, 22.16) | −35.05 (−35.67, −34.42) to 35.78 (35.16, 36.41) | −36.53 (−37.50, −35.56) to 36.22 (35.24, 37.19) | −13.40 (−15.26, −11.53) to 13.57 (11.71, 15.43) |

Table 2. Results of all tests for the Garmin 735 XT watch. Bold values indicate MAPE > 10% and ICC < 0.90.

| | Sedentary | Lifestyle | Walking | Dynamic Walk-Jog | Running | Cycling |
|---|--|---|---|---|---|---|
| Total n data points (n removed, % removed) | 461 (88, 4.5%) | 1964 (0, 0%) | 752 (0, 0%) | 2087 (0, 0%) | 1034 (0, 0%) | 1281 (0, 0%) |
| Mean bias (bpm) | −4.30 (115.9) | −3.20 (18.53) | −1.56 (23.42) | −2.32 (37.07) | −0.93 (69.41) | 3.14 (119.68) |
| MAPE (%) | 15.33 | 12.09 | 8.75 | 12.63 | 11.22 | 7.56 |
| | 0.444 | 0.710 | 0.802 | 0.732 | 0.710 | 0.809 |
| ICC | (0.368, 0.515) | (0.687, 0.733) | (0.775, 0.826) | (0.711, 0.751) | (0.678, 0.739) | (0.789, 0.827) |
| | −84.95 | −129.89 | −157.57 | −219.11 | −257.47 | −226.68 |
| LOA | (−92.85, −77.05) to 76.35 (68.45, 84.25) | (−134.52, −125.25) to 123.49 (118.85, 128.12) | (−166.00, −149.14) to 154.45 (146.02, 162.88) | (−226.28, −211.95) to 214.48 (207.32, 221.65) | (−269.61, −245.34) to 255.61 (243.47, 267.74) | (−237.05, −216.30) to 232.95 (222.57, 243.33) |

Table 3. Results of all tests for the Jabra Elite earbuds. Bold values indicate MAPE > 10% and ICC < 0.90.

| | Sedentary | Lifestyle | Walking | Dynamic Walk-Jog | Running | Cycling |
|---|-------------------------|-------------------------|-------------------------|----------------------------|-------------------------|-------------------------|
| Total n data points (n removed, % removed) | 7357 (48, 0.6%) | 17,195 (225, 1.3%) | 9475 (178, 1.8%) | 15,819 (739, 4.5%) | 8602 (333, 3.4%) | 15,266 (295, 1.8%) |
| Mean bias (bpm) | 1.32 (5.09) | 1.60 (15.31) | −1.57 (29.49) | 2.13 (506.44) | −1.68 (40.61) | 2.27 (27.07) |
| MAPE (%) | 4.14 | 7.58 | 7.91 | 10.80 | 7.91 | 7.15 |
| ICC | 0.937 | 0.826 | 0.833 | 0.906 | 0.923 | 0.904 |
| | (0.934, 0.939) | (0.821, 0.830) | (0.827, 0.839) | (0.903, 0.909) | (0.920, 0.926) | (0.901, 0.907) |
| LOA | −38.91 (−39.86, −37.96) | −80.27 (−82.08, −78.46) | −65.74 (−68.17, −63.30) | −208.08 (−210.98, −191.19) | −82.89 (−85.80, −79.99) | −75.49 (−77.80, −73.18) |
| | to 41.57 (40.62, 42.52) | to 83.48 (81.67, 85.29) | to 62.59 (60.16, 65.03) | to 205.34 (195.44, 215.23) | to 79.53 (76.62, 82.43) | to 80.04 (77.73, 82.34) |

Table 4. Results of all tests for the Scosche Rhythm 24 armband.

| | Sedentary | Lifestyle | Walking | Dynamic Walk-Jog | Running | Cycling |
|---|--|---|---|---|---|---|
| Total n data points (n removed, % removed) | 7405 (0, 0%) | 17,420 (0, 0%) | 9653 (0, 0%) | 16,558 (0, 0%) | 9636 (0, 0%) | 16,547 (0, 0%) |
| Mean bias (bpm) | 0.83 (3.96) | 0.30 (2.36) | 0.12 (0.24) | 0.37 (0.94) | −0.01 (0.78) | 0.17 (0.02) |
| MAPE (%) | 2.71 | 4.12 | 2.33 | 1.54 | 1.24 | 0.90 |
| ICC | 0.959 (0.957, 0.961) | 0.940 (0.939, 0.942) | 0.978 (0.978, 0.979) | 0.991 (0.991, 0.992) | 0.990 (0.990, 0.990) | 0.997 (0.997, 0.997) |
| LOA | 30.55 (−31.09, −30.02) to 32.21 (31.67, 32.74) | −49.50 (−50.14, −48.85) to 50.09 (49.45, 50.73) | −29.02 (−29.77, −28.27) to 29.25 (28.50, 30.00) | −35.34 (−35.79, −34.89) to 36.05 (35.60, 36.50) | −35.31 (−35.69, −34.94) to 35.29 (34.92, 35.67) | −15.88 (−17.46, −14.30) to 16.22 (14.64, 17.80) |

3.1. Apple

Table 1 shows the results of the Apple Watch Series 4. This device had the second least number of data points of the four devices tested. The Apple device had a mean bias within 1 bpm of the Polar chest strap, MAPE of less than 3%, and an ICC of > 0.96 in all six conditions. The Apple Watch performed the worst in the lifestyle activity (2.95% MAPE and ICC = 0.963) and best during cycling (0.62% MAPE and ICC = 0.998). Apple was equivalent to the Polar chest strap for all six testing conditions ($p < 0.001$).

3.2. Garmin

Table 2 provides the results of the Garmin 735 XT. The Garmin device had the least amount of data points of all four devices. The device tended to underestimate heart rate in all conditions, except cycling. MAPE was greater than 10% during sedentary, lifestyle, dynamic walk-jog, and running. Additionally, the ICCs were below 0.90 for all conditions and the limits of agreement were wide. The Garmin tested as not equivalent to the Polar chest strap for the sedentary condition ($p = 1.00$), but equivalent for the remaining five conditions ($p = 0.006$ for lifestyle, $p < 0.001$ for walking, dynamic walk-jog, running, and cycling).

3.3. Jabra

The results of the Jabra Elite earbud are shown in Table 3. The Jabra had the second highest number of data points available for analysis. The mean bias was within ± 5 bpm for each condition, except cycling, which was higher. MAPE exceeded 10% for dynamic walk-jog, was 7–8% for lifestyle, walking, running, and cycling, and was below 5% for the sedentary condition. ICCs were below 0.90 for lifestyle and walking, equal to 0.90 for dynamic walk-jog and cycling, and greater than 0.92 for sedentary and running. The Jabra was equivalent to the Polar chest strap for all six conditions ($p < 0.001$).

3.4. Scosche Rhythm 24

Table 4 shows the results of the Scosche Rhythm 24 armband. The Scosche device provided the highest number of data points of the four devices tested. Mean bias was within ± 1 bpm for all six conditions, and MAPE was less than 2% for dynamic walk-jog,

running, and cycling. MAPE was highest for the lifestyle condition (4.12%). The ICCs were 0.94 or higher for all conditions. The Scosche was equivalent to the Polar chest strap for all six conditions ($p < 0.001$).

4. Discussion

This study aimed to evaluate the agreement of commercially available wearable technology heart rate monitors to a criterion measure using the established CTA standards [5]. It was hypothesized that each device would return valid heart rate measures during rest, lifestyle activity, walking, jogging, and cycling. Our results indicate that the Apple Watch Series 4 and the Scosche Rhythm 24 display acceptable agreement with a criterion chest strap heart rate monitor across all conditions tested. The Jabra Elite earbuds were found to have acceptable agreement in the sedentary, running, and cycling conditions, whereas the Garmin Forerunner 735 XT did not satisfy our definition of validity for any condition tested.

Our research group has identified that there is often considerable lag between the release of wearable technology devices and independent investigations that report validity [14]. Toward this end, we are aware of only one other investigation that has determined validity of heart rate measures in the Apple Watch Series 4 [23]. The investigation was conducted in post-operative cardiac surgery patients with no exercise component and reported that agreement was higher in patients in atrial fibrillation ($rc = 0.86$) compared to those who were not ($rc = 0.64$) [23]. To our knowledge the current investigation represents the first report of the Apple Watch Series 4 with respect to heart rate validity during any form of exercise. Our results indicate acceptable heart rate agreement of the Apple Watch Series 4 across conditions including sedentary, lifestyle activities, walking, jogging, running, and cycling. Several other reports have found acceptable heart rate agreement for activities such as steady state treadmill exercise of increasing speed up to 6 mph [24], exercise blocks progressing from walking (3.0 and 4.0 mph) to running (6.0 and 7.5 mph) to cycling (100 W and 175 W) [25], and initial stages of the Bruce Treadmill test and stages of the 25-watt cycle protocol [26]. It should be noted that investigations into original versions of the Apple Watch have determined decreasing levels of validity as the intensity of exercise increases [27,28]. Further investigations are necessary to reveal whether this phenomenon applies subsequent versions of the Apple Watch.

The Scosche Rhythm 24 is a device that is designed to be located on the forearm and investigations reporting validity during exercise are beginning to emerge. One investigation using various walking and running speeds on the treadmill determined acceptable agreement between a previous version of the Scosche (Rhythm) arm strap with a chest strap monitor (overall $r = 0.93$, MAPE range = 2.22% to 6.67%) [29]. Additionally, no difference was reported in heart rate compared to a chest strap when participants performed walking and running on a treadmill and in an unrestricted setting ($p > 0.05$) [21]. The current investigation is in line with these previous reports in that the Scosche Rhythm 24 displayed acceptable agreement across all conditions tested. However, other investigations have reported lower validity measures during specific exercise applications. Although acceptable heart rate agreement with an ECG criterion device was reported at rest ($rc = 0.93$), cycle ($rc = 0.84$), and treadmill exercise ($rc = 0.92$), the Scosche Rhythm+ was not deemed valid while completing elliptical training ($rc = 0.41$ with arms, 0.27 without arm movement) [19]. During variable intensity trail running, better agreement from the Scosche Rhythm+ was observed during downhill running (MAPE = 3.8%, bias = 1.9 bpm, $rc = 0.885$) than when running was at a generally positive incline (MAPE = 6.2%, bias = 3.9 bpm, $rc = 0.699$) [21]. Thus, it is possible that, similar to what has been noted in the Apple Watch above, activities performed at greater intensities may result in a decreased agreement in heart rate measures obtained from the Scosche armband compared a criterion measure.

Earbud-based heart rate investigations are beginning to emerge in the exercise validity literature. Acceptable agreement has been reported during treadmill-based graded exercise testing with a prototype earbud device ($R^2 = 0.98$) [30], certain resistance training exercises

with Bose SoundSport Headphones (leg curl MAPE range = 4.46% to 6.48%) [31], as well as during treadmill and high intensity training exercises with the Jabra Pulse earbuds (MAPE = 2.48%, $rc = 0.943$; MAPE = 3.53%, $rc = 0.861$ respectively) [6]. On the other hand, graded exercise testing on a cycle ergometer resulted in heart validity that decreased as exercise intensity increased (MAPE at 50 W = 6.4%, MAPE at 200 W = 15.42%) [31]. Additionally, the Jabra Pulse earbud device tended to have less agreement compared to ECG when heart rate was above 100 bpm in patients with cardiac diseases and in participants with atrial fibrillation ($r^2 = 0.434$) [32]. Finally, the Jabra Elite Sport earbud device was determined to have poor agreement compared to a criterion chest strap during trail running at variable intensities of exercise (MAPE = 21.3%, $rc = 0.384$, ICC = 0.395) [21]. In the current investigation, the Jabra Elite earbuds acceptable agreement was observed in the sedentary, running, and cycling conditions but not during lifestyle tasks, walking, or the dynamic walk-jog. As earbud devices have shown varied results in returning accurate heart rate values, further investigations should be directed toward determining which factors negatively impact recorded measures. Unlike the previous devices discussed in which greater intensity appears to impact validity (Apple Watch, Scosche Rhythm), it is possible that greater engagement in active movement may decrease agreement to a criterion measure.

For a device to be considered valid in the current investigation, it had to meet the predetermined thresholds in every statistical test employed. Toward this end, the Garmin Forerunner 735 XT was not considered valid for any condition tested. Several Garmin Forerunner models have been tested in the exercise validity literature specific to heart rate. The Garmin 225 displayed large limits of agreement (−32.53 to 29.40 bpm) when compared to ECG during a walking protocol that included grades up to 8% [33], and large MAPE values in a protocol that consisted of walking and running (MAPE range = 9.88% to 27.38%) [34]. The Garmin 235 was found to have acceptable heart rate agreement during cycling at 150 W ($Rho = 0.889$), but low agreement during lower intensity cycling (50 W $Rho = 0.269$, 100 W $Rho = 0.462$) [35]. Additionally, heart rate measures in the Garmin 235 were considered acceptable when participants performed treadmill exercise up to 6 mph (Absolute percent difference = 6.1%) and stationary cycle exercise up to 125 W (Absolute percent difference = 4.6%), but not while completing elliptical exercise (Absolute percent difference = 13.7%) [19]. To our knowledge, we are the first to report heart rate validity in the Garmin Forerunner 735 XT. It appears that improvements must be made to increase the heart rate sensing capability in these devices.

It is unknown why the Polar chest strap returned readings of zero. It was most likely due to an issue with contact between the sensor and the skin, but greater investigation into the specifics was outside the scope of this study. All protocols were followed, including the use of conducting gel with the Polar chest strap. All devices were monitoring HR prior to recording data.

5. Conclusions

The purpose of this investigation was to determine heart rate validity of commercially available wearable technology devices using the established CTA standards [5]. The main conclusion from the current study indicates that only the Apple Watch Series 4 and the Scosche Rhythm 24 display acceptable agreement across all conditions. Although we hypothesized that every device would return valid heart rate measures during rest, lifestyle activity, walking, jogging, and cycling, that was not the case. Although activity type was not able to be synchronized, it did not influence heart rate of the Polar, Scosche, or Jabra data and did not appear to influence Apple data. Future research may investigate if heart rate data of the Garmin device is influenced by activity type. As wearable technology continues to advance, future investigations should employ CTA standards for developers, researchers, and consumers to make true comparisons of accuracy among wearables.

Author Contributions: Conceptualization, J.D.R., J.A.B. and J.W.N.; methodology, J.D.R., J.A.B. and J.W.N.; software, J.A.B. and J.W.N.; validation, J.A.B. and J.W.N.; formal analysis, J.A.B. and J.W.N.; investigation, J.D.R. and M.C.; resources, J.D.R., M.C. and J.A.B.; data curation, J.D.R. and J.A.B.; writing—original draft preparation, J.D.R., J.A.B. and J.W.N.; writing—review and editing, J.D.R., J.A.B., M.C. and J.W.N.; visualization, J.D.R., J.A.B. and J.W.N.; supervision, J.D.R. and M.C.; project administration and funding acquisition, J.D.R. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Application Approval Notification, IRB Application Number: # (19:19).

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author upon Institutional Review Board approval. The data are not publicly available due to resections of privacy and ethics.

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

References

1. IDC Worldwide Quarterly Wearable Device Tracker. 16 December 2019. Available online: <https://www.idc.com/getdoc.jsp?containerId=prUS45737919> (accessed on 15 December 2020).
2. Vijayalakshmi, K.; Uma, S.; Bhuvanya, R.; Suresh, A. A demand for wearable devices in health care. *Int. J. Eng. Technol.* **2018**, *7*, 1–4. [[CrossRef](#)]
3. Consumer Technology Association. *Physical Activity Monitoring for Step Counting, ANSI/CTA-2056*; Consumer Technology Association: Hopewell, VA, USA, 2016.
4. Consumer Technology Association. *Methodology of Measurements for Features in Sleep Tracking Consumer Technology Devices and Applications, ANSI/CTA/NSF-2052*; Consumer Technology Association: Hopewell, VA, USA, 2017.
5. Consumer Technology Association. *Physical Activity Monitoring for Heart Rate, ANSI/CTA-2065*; Consumer Technology Association: Hopewell, VA, USA, 2018.
6. Bunn, J.A.; Jones, C.; Oliviera, A.; Webster, M.J. Assessment of step accuracy using the Consumer Technology Association standard. *J. Sports Sci.* **2019**, *37*, 244–248. [[CrossRef](#)] [[PubMed](#)]
7. Montes, J.; Tandy, R.; Young, J.; Lee, S.P.; Navalta, J.W. 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.* **2020**, *13*, 410–426.
8. Moore, C.C.; McCullough, A.K.; Aguiar, E.J.; Ducharme, S.W.; Tudor-Locke, C. Toward harmonized treadmill-based validation of step-counting wearable technologies: A scoping review. *J. Phys. Act. Health* **2020**, *17*, 840–852. [[CrossRef](#)]
9. Navalta, J.W.; Montes, J.; Bodell, N.G.; Aguilar, C.D.; Lujan, A.; Guzman, G.; Kam, B.K.; Manning, J.W.; DeBeliso, M. Wearable device validity in determining step count during hiking and trail running. *J. Meas. Phys. Behav.* **2018**, *1*, 86–93. [[CrossRef](#)]
10. Bunn, J.; Wells, E.; Manor, J.; Webster, M. Evaluation of earbud and wristwatch heart rate monitors during aerobic and resistance training. *Int. J. Exerc. Sci.* **2019**, *12*, 374–384.
11. Nelson, B.W.; Allen, N.B. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-h period: Intraindividual validation study. *JMIR mHealth uHealth* **2019**, *7*, e10828. [[CrossRef](#)]
12. Nuss, K.J.; Sanford, J.L.; Archambault, L.J.; Schlemmer, E.J.; Blake, S.; Courtney, J.B.; Hulett, N.A.; Li, K. Accuracy of heart rate and energy expenditure estimations of wrist-worn and arm-worn Apple watches. *J. Meas. Phys. Behav.* **2019**, *2*, 166–175. [[CrossRef](#)]
13. Bassett, D.R.; Toth, L.P.; LaMunion, S.R.; Crouter, S.E. Step counting: A review of measurement considerations and health-related applications. *Sports Med.* **2017**, *47*, 1303–1315. [[CrossRef](#)] [[PubMed](#)]
14. Bunn, J.A.; Navalta, J.W.; Fountaine, C.J.; Reece, J.D. Current state of commercial wearable technology in physical activity monitoring 2015–2017. *Int. J. Exerc. Sci.* **2018**, *11*, 503–515. [[PubMed](#)]
15. Collins, T.; Woolley, S.I.; Oniani, S.; Pires, I.M.; Garcia, N.M.; Ledger, S.J.; Pandyan, A. Version reporting and assessment approaches for new and updated activity and heart rate monitors. *Sensors* **2019**, *19*, 1705. [[CrossRef](#)] [[PubMed](#)]
16. Haghayegh, S.; Khoshnevis, S.; Smolensky, M.H.; Diller, K.R.; Castriotta, R.J. Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. *J. Med. Internet Res.* **2019**, *21*, e16273. [[CrossRef](#)]
17. Oniani, S.; Woolley, S.I.; Pires, I.M.; Garcia, N.M.; Collins, T.; Ledger, S.; Pandyan, A. Reliability assessment of new and updated consumer-grade activity and heart rate monitors. In Proceedings of the IARIA, The Ninth International Conference on Sensor Device Technologies and Applications, Venice, Italy, 16–20 September 2018.
18. 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](#)] [[PubMed](#)]
19. 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](#)]

20. Navalta, J.W.; Ramirez, G.G.; Maxell, C.; Radzak, K.N.; McGinnis, G.R. Validity and reliability of three commercially available smart sports bras during treadmill walking and running. *Sci. Rep.* **2020**, *10*, 7397. [[CrossRef](#)] [[PubMed](#)]
21. 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](#)]
22. The Jamovi Project. Jamovi (Version 1.2). Computer Software. Available online: <https://www.jamovi.org> (accessed on 1 June 2020).
23. Seshadri, D.R.; Bittel, B.; Brownsky, D.; Houghtaling, P.; Drummond, C.K.; Desai, M.; Gillinov, A.M. Accuracy of the Apple watch 4 to measure heart rate in patients with atrial fibrillation. *IEEE J. Transl. Eng. Health Med.* **2019**, *8*, 2700204. [[CrossRef](#)]
24. 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](#)]
25. Shcherbina, A.; Mattsson, C.M.; 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](#)]
26. Wallen, M.P.; Gomersall, S.R.; Keating, S.E.; Wisløff, U.; Coombes, J.S. Accuracy of heart rate watches: Implications for weight management. *PLoS ONE* **2016**, *11*, e0154420. [[CrossRef](#)]
27. Khushhal, A.; Nichols, S.; Evans, W.; Gleadall-Siddall, D.O.; Page, R.; O'Doherty, A.F.; Carroll, S.; Ingle, L.; Grant Abt, G. Validity and reliability of the apple watch for measuring heart rate during exercise. *Sports Med. Int. Open* **2017**, *1*, E206. [[CrossRef](#)] [[PubMed](#)]
28. Thomson, E.A.; Nuss, K.; Comstock, A.; Reinwald, S.; Blake, S.; Pimentel, R.E.; Brian, L.; Tracy, B.L.; Li, K. Heart rate measures from the Apple watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *J. Sports Sci.* **2019**, *37*, 1411–1419. [[CrossRef](#)] [[PubMed](#)]
29. 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](#)]
30. Leboeuf, S.F.; Aumer, M.E.; Kraus, W.E.; Johnson, J.L.; Duscha, B. Earbud-based sensor for the assessment of energy expenditure, HR, and VO₂max. *Med. Sci. Sports Exerc.* **2014**, *46*, 1046–1052. [[CrossRef](#)]
31. Boudreaux, B.D.; Hebert, E.P.; Hollander, D.B.; Williams, B.M.; Cormier, C.L.; Naquin, M.R.; Gillan, W.W.; Gusew, E.E.; Kraemer, R.R. Validity of wearable activity monitors during cycling and resistance exercise. *Med. Sci. Sports Exerc.* **2018**, *50*, 624–633. [[CrossRef](#)] [[PubMed](#)]
32. Leader, N.; Dorian, P.; Lam, J.; Lee, C.; Woo, A.; Chow, C. Evaluation of heart rate trackers in patients with atrial fibrillation. *Can. J. Cardiol.* **2018**, *34*, S152–S153. [[CrossRef](#)]
33. Claes, J.; Buys, R.; Avila, A.; Finlay, D.; Kennedy, A.; Guldenring, D.; Budts, W.; Cornelissen, V. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. *J. Med. Eng. Technol.* **2017**, *41*, 480–485. [[CrossRef](#)]
34. 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](#)]
35. 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](#)]