Temple

STUDY 01 / TRACKING HEART RATE AT THE TEMPLE / WHITE PAPER BASED ON INTERNAL STUDY

Temple vs. five flagship wearables: a heart rate accuracy study

COMPARING TEMPLE WITH APPLE WATCH SERIES 12, FITBIT AIR, WHOOP MG, GARMIN FORERUNNER 970 & OURA RING 5



STUDY OVERVIEW

Abstract

Background: Most athletes rely on real-time measurement of heart rate through wearables to guide their training and recovery. While electrocardiogram (ECG)-based heart rate monitors have higher accuracy, optical sensors are more widely used in wearables. These sensors are prone to movement-induced errors, which render them less accurate during exercise such as running. We compared the accuracy of heart rate measured by six optical sensors, including Temple, a wearable attached to the temple region, and five flagship consumer wearables, against an ECG chest strap heart rate monitor.

Methods: Forty-three healthy adults (nine women and thirty-four men; aged 20 to 35 years) walked and ran a graded treadmill protocol through heart rate Zones 1 to 5 while wearing Temple, Apple Watch Series 12, WHOOP MG, Garmin Forerunner 970, Fitbit Air, and Oura Ring 5 at the same time. The reference was a Polar H10 ECG chest strap with its electrodes taped to the skin. In a substudy, four location-agnostic optical sensors (Polar Verity Sense) were worn simultaneously, one each on the temple, chest, thumb, and wrist. Heart rate was compared second-by-second over 20.5 hours of recording.

Results: Temple had the lowest mean absolute error (MAE) at 1.18 bpm (95% CI 1.10 to 1.26), followed by Apple Watch Series 12 (1.94), Fitbit Air (2.96), WHOOP MG (3.14), Garmin Forerunner 970 (3.18), and Oura Ring 5 (5.21), with failure episodes (readings more than 20 bpm off for 10 seconds or more) excluded. With those episodes retained among otherwise valid readings, Temple's MAE remained 1.18 bpm. Temple’s MAE was significantly lower than that of each of the five flagship wearables, over the whole session and in every HR zone (Holm-adjusted paired t-tests, $p < 0.05$). In high-intensity Zones 4 and 5, Temple's MAE was 1.16 bpm, and no participant's MAE with Temple exceeded 5 bpm. Across the exercise protocol, only 1.8% of Temple's readings were large errors (more than 5 bpm off), compared with 9.0% to 36.1% for the other wearables.

Conclusion: Temple measured heart rate with lower error and fewer large errors than five flagship wearables, across rest to near-maximal running, and held its accuracy in high-intensity zones.

Key takeaways:

- Temple had the lowest heart rate error in this controlled indoor treadmill study, where six optical wearables, including Temple, were tested against a taped Polar H10 reference. Temple showed a mean absolute error of 1.18 bpm (95% CI 1.10 to 1.26) with failure episodes (readings more than 20 bpm off for 10 seconds or more) excluded, compared with 1.94 to 5.21 bpm for Apple Watch Series 12, Fitbit Air, WHOOP MG, Garmin Forerunner 970, and Oura Ring 5.
- Temple’s mean absolute error was significantly lower than that of each of the five flagship wearables, over the whole session and in every heart rate zone (Holm-adjusted paired t-tests).
- Accuracy held up in high-intensity Zones 4 and 5. Temple's error was 1.16 bpm, and no participant's mean absolute error exceeded 5 bpm (0 of 42); in Zone 5 alone, it was 0.95 bpm.
- Temple made the fewest large errors (readings more than 5 bpm off). Over the exercise protocol, only 1.8% of its readings were more than 5 bpm off (even lower in Zones 4 and 5, at 1.4%), against 9.0% to 36.1% for the other wearables.
- The temple stood out as a sensor site. When four body locations were compared simultaneously, Polar Verity Sense, a location-agnostic optical sensor, had the lowest observed mean absolute error at the temple (3.50 bpm), compared with 5.10 to 6.03 bpm at the chest, thumb, and wrist.

01 / CONTEXT

1. Introduction

Heart rate is one of the most important signals tracked by a fitness wearable. Athletes use it to pace a run, hold a training zone, and judge recovery. The training load and health features of most wearables are built on it. It has to be accurate at rest and, above all, in high-intensity Zones 4 and 5 of the common five-zone scheme, at 80% of maximum heart rate and above. At these intensities, lactate-threshold and interval sessions are run, and a few beats decide whether an athlete is in the right zone. Optical sensors are less accurate during activity than at rest [1], and their error depends on the exercise, with vigorous whole-body movement being the hardest case [2].

Most consumer wearables measure heart rate with photoplethysmography (PPG) at the wrist or finger. At these locations, arm swing and tendon movement add motion artifacts, the signal can lock onto a cadence, and skin contact can loosen [1, 3, 5]. Some studies evaluating sensor accuracy across skin tones report larger errors during exercise among participants at the higher end of the skin tone scale [4], while others find no clear difference [1].

A sensor's position influences how well it works. The same optical sensor is more accurate higher on the arm than at the wrist [5], and head-worn sensors have outperformed the wrist during swimming [6, 7]. The anterior temple sits over branches of the superficial temporal artery [8] and away from hand and wrist movement.

Chest-strap electrocardiography (ECG) is the usual reference outside the laboratory [9]. The Polar H10 agrees closely with clinical ECG during exercise [10, 11], but chest straps can lose electrode contact during movement [9, 12]. To further raise the benchmark, we taped the reference strap to the skin and wore a second strap, buckled without tape, to serve as a real-world reference. We compared Temple with five flagship wearables, second-by-second, from rest to near-maximal running, with particular attention given to high-intensity Zones 4 and 5. In a substudy, identical optical sensors were worn at four sites simultaneously for direct comparison.

2. Methods

2.1 Participants

Healthy adults aged 18 to 55 years, who were habitually active and able to run on a treadmill, were eligible. Exclusion criteria included any known or suspected cardiovascular, respiratory, or metabolic conditions; medications that alter heart rate or its response to exercise; musculoskeletal injuries that made running unsafe; pregnancy; skin conditions or tattoos at a sensor site; and any other condition that, in the investigator's judgement, made vigorous exercise inadvisable.

Forty-three participants were analyzed (Table 1). Skin tone was rated visually on the Monk Skin Tone Scale [13].

Table 1. Participants

Characteristic	Mean (SD) or n
Participants	43 (9 women, 34 men)
Age, years	23.5 (2.9); range 20 to 35
Monk skin tone 1 / 2 / 3 / 4 / 5 / 6	1 / 2 / 8 / 17 / 11 / 4

SD, standard deviation.

2.2 Devices

Temple is a small wearable worn on the anterior temple, above the zygomatic arch and lateral to the lateral orbital rim, and held by a medical-grade Grip. It measures reflectance PPG with green, red, and infrared light, together with accelerometry, and reports heart rate once per second. Its heart rate data was retrieved through the Temple app's internal export function.

Apple Watch Series 12 and Oura Ring 5 ran in workout mode, and their heart rates were read from the Apple Health export, with about one reading every 5 seconds. WHOOP MG and Fitbit Air broadcast heart rate over Bluetooth once per second, and Garmin Forerunner 970 about twice per second (Table 2). The reference was a Polar H10 ECG chest sensor mounted on a Polar Pro Strap (size XS-S, M-XXL, or XXXL, chosen to fit each participant's chest circumference). It reported heart rate over Bluetooth. A second Polar H10, buckled but untaped, was used only for the reliability analysis. Bluetooth streams were logged on a study computer. Recordings took place in October 2026, with every device on the latest firmware available at the time.

02 / METHODS

Table 2. Devices, sites, and data streams

Device	Maker	Site	Data stream	Interval	Analyzed
Temple	**Temple**	**temple**	**Temple app internal export**	**1 second**	**43**
Polar H10, taped (reference)	Polar Electro	chest, taped	Bluetooth broadcast	1 second	43
Apple Watch Series 12	Apple	forearm (upper 22, bottom 21)	Apple Health export, workout mode	about 5 seconds	43
Fitbit Air	Google	forearm (upper 17, bottom 25)	Bluetooth broadcast	1 second	42
WHOOP MG	WHOOP	forearm (upper 24, bottom 19)	Bluetooth broadcast	1 second	43
Garmin FR970	Garmin	forearm (upper 22, bottom 21)	Bluetooth broadcast	about 0.5 seconds	43
Oura Ring 5	Oura Health	index, middle or ring finger	Apple Health export, workout mode	about 5 seconds	43
Polar H10, untaped	Polar Electro	chest, about 1 cm below the reference	Bluetooth broadcast	1 second	43
Polar Verity Sense (substudy)	Polar Electro	temple, chest, thumb, wrist	Bluetooth broadcast	1 second	29

Analyzed: participants with at least 2 minutes of scored data; for Polar Verity Sense, participants with usable recordings at all four sites. Fitbit Air was not recorded for one participant.

2.3 Placement and fitting

Participants fitted both chest straps themselves, following an illustrated checklist (Fig. 1A and B). Electrodes and skin were dampened with water. The reference strap was placed directly below the chest muscles, with the sensor sitting comfortably on the sternum. The skin above and below the strap was patted dry, and the strap was fixed with two strips of commercial kinesiology tape: an uncut strip along the upper edge, and a strip cut out around the sensor along the lower edge. The untaped strap sat about 1 cm below the reference strap at the same tension, with its sensor directly below the reference sensor and the two straps not touching each other.

Four wrist devices were worn at once, facing outward on the back of the forearm and fastened snugly, as each manufacturer advises for heart rate tracking. Two were worn on each forearm: one at the usual position just above the wrist bone (bottom), and one 2 to 3 cm higher up the forearm (upper) (Fig. 1C). Positions were rotated across participants. The upper position is consistent with WHOOP's advice to wear its sensor about 1 inch above the wrist bone, or slightly higher, rather than close to the bone [14]; the bottom position is where a watch is commonly worn. Oura Ring 5 was worn on the index, middle, or ring finger, as Oura advises, and sized for a snug fit with Oura's sizing kit [15]. In a substudy, Polar Verity Sense optical sensors were placed at the temple, chest, thumb, and wrist, held with custom-made double-sided tape and covered with kinesiology (KT) tape. Devices were sanitized between participants.

02 / METHODS

2.4 Protocol

Sessions took place in an indoor laboratory in Gurugram, India, at 23-25°C. Participants first sat at rest, then walked and ran on a treadmill through heart rate Zones 1 to 5 in order, with a target of 3.5 minutes at rest and in each zone, and finished with a walking cool-down (Fig. 1D). The operator adjusted speed and incline to hold each zone on the live reference heart rate. Zones were calculated as a percentage of age-predicted maximum heart rate, 208 − 0.7 × age [16]: rest below 50%; Zone 1, 50% to 60%; Zone 2, 60% to 70%; Zone 3, 70% to 80%; Zone 4, 80% to 90%; and Zone 5, 90% or more. Zones 4 and 5 are high-intensity zones. For analysis, each second was assigned to a zone from the reference heart rate.

Counting seconds with a reference reading, participants spent a mean of 4.3 minutes at rest, and 4.6, 4.6, 5.5, 4.9, and 3.9 minutes in Zones 1 to 5 respectively. Sessions lasted 28.6 minutes on average (between 16.3 to 42.0 minutes), for a total of 20.5 hours. Peak heart rate averaged 185.4 bpm, and 40 of 43 participants spent at least 60 seconds in Zone 5.

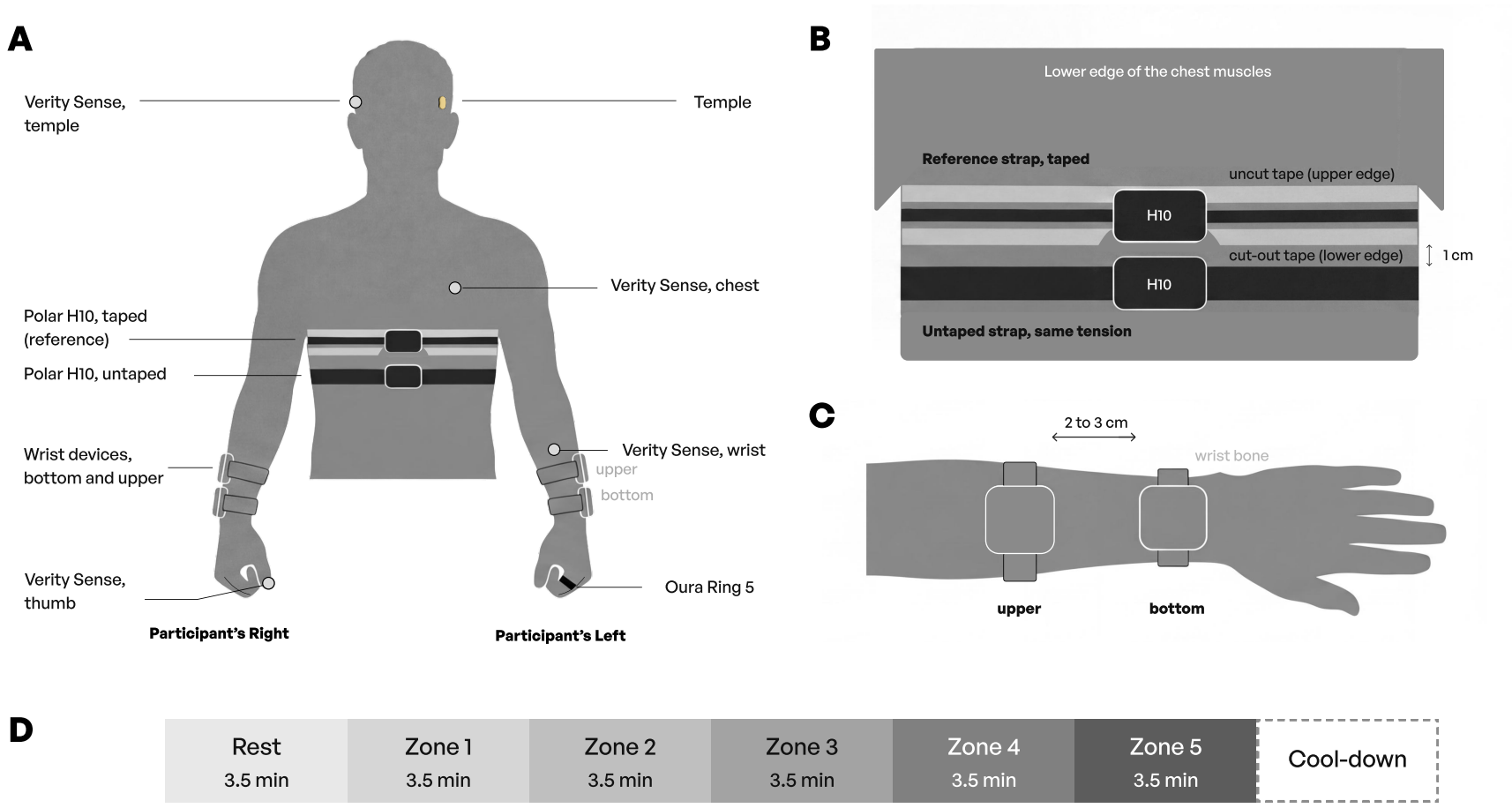


Figure 1. Set-up and protocol. (A) Device sites: Temple at the temple; two Polar H10 chest straps, the upper one taped (reference) and the lower one untaped; four wrist devices, two on each forearm at the bottom and upper positions; Oura Ring 5 on a finger; Polar Verity Sense at the temple, chest, thumb, and wrist in the substudy. (B) Chest straps: the taped reference strap and the untaped strap below it. (C) Wrist positions: bottom, just above the wrist bone, and upper, 2 to 3 cm higher. (D) Protocol: seated rest and heart rate Zones 1 to 5, 3.5 minutes each, then a walking cool-down.

2.5 Data processing

Reference. The reference for each second was the heart rate broadcast by the taped Polar H10, averaged over its readings in that second. It was processed by the same rules as every device (below), with no further artifact rejection.

02 / METHODS

Scoring. Each device reading was compared with the reference heart rate in the same second. When a device sent more than one reading in one second, as the Garmin Forerunner 970 does, the readings were averaged. Values were not interpolated or carried forward; readings outside 36–240 bpm were treated as missing data, and gaps were not counted as errors. Apple Watch Series 12 and Oura Ring 5 were scored on their own readings, about once every 5 seconds. Timestamps were used as recorded, with no delay correction, and each session ran from the start to the end of the reference recording. A device counted for a participant with at least 2 minutes of scored data (120 seconds, or 24 readings for Apple Watch Series 12 and Oura Ring 5), and in a given zone with at least 5 data points in that zone. Data coverage was the share of session time in which a device was reporting. A gap between valid readings longer than 2.5 seconds (10.5 seconds for Apple Watch Series 12 and Oura Ring 5) counted as lost time, less the device's normal interval of 1 second (5 seconds), and gaps at the start and end of the session counted the same way.

Failure episodes. We flagged failure episodes the same way for every device, using in-house thresholds. A failure episode was a run of consecutive data points, each more than 20 bpm from the reference and no more than 10 seconds apart, that lasted 10 seconds or more, counting the first and last second. An episode was a drop if the device read below the reference, and a spike if it read above it. Temple had one such episode (a 12-second drop), and the other wearables had 4 to 62 episodes. Episodes were removed from the accuracy analysis and counted under reliability. Polar Verity Sense recordings were discarded as soon as a sensor fell or shifted, and the site comparison included the 29 participants with usable recordings at all four sites.

SCORING PATHWAY

Reference broadcast

Same-second Bluetooth average

Same-second scoring

No interpolation or delay correction

Two reporting views

Accuracy and reliability

2.6 Statistics

For each participant and device, we computed the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) over every scored second [9]. Device values are means of participant values, with 95% confidence intervals (CI) from 4,000 bootstrap resamples of participants. Bias (device minus reference) and 95% limits of agreement were calculated using the Bland-Altman method for repeated measures [17, 18]. To compare Temple with each comparator, we kept only the seconds that both devices reported (requiring at least 20 such seconds per participant, or 5 within a zone). For each participant, we computed the comparator's mean absolute error minus Temple's. The mean difference is given with a 95% CI from the same bootstrap of participants and was tested using a paired t-test, with p-values adjusted by the Holm method across the five comparators within each analysis: whole session, rest, and each zone [19]. Tests were two-sided with alpha 0.05. Analyses were performed in Python 3 using pandas, NumPy, and SciPy.

02 / METHODS

2.7 Reliability

Large errors and dropouts are recognized problems for consumer heart rate devices during exercise, and reporting the share of readings within set error bands is recommended [9]. We therefore defined reliability as how often a device was far from the reference: the percentage of data points (one scored second, or one reading for Apple Watch Series 12 and Oura Ring 5) more than 5 bpm from the reference (a large error), split into over-reads and under-reads, and the percentage of data points inside a drop or spike episode. Reliability was assessed only for the exercise protocol, from the start of Zone 1 to the end of Zone 5; seated rest and cool-down were excluded. The window began once the reference stayed in Zone 1 or above for 30 seconds (at least 15 reference readings, no more than 5 seconds apart). It ended at the last second in Zone 5, or at peak heart rate for the two participants who never reached Zone 5 (41 of 43 did). It averaged 20.7 minutes.

3. Results

3.1 Overall accuracy

Temple had a lower error than each of the five flagship wearables (Table 3; Fig. 2A-C). Its MAE was 1.18 bpm (95% CI: 1.10 to 1.26), its MAPE was 0.98%, and its RMSE was 1.97 bpm. Apple Watch Series 12 followed at 1.94 bpm, then Fitbit Air (2.96), WHOOP MG (3.14), Garmin Forerunner 970 (3.18), and Oura Ring 5 (5.21). Temple's data coverage was 99.6%, highest among the six optical wearables. With failure episodes retained, Temple's MAE was 1.18 bpm, and the ranking was unchanged. Temple's MAE also remained the lowest, with the wrist devices counted only in their upper position (section 3.7).

Table 3. Accuracy over the whole session (failure episodes removed)

Device	n	MAE, bpm (95% CI)	MAPE, %	RMSE, bpm	Bias, bpm	95% limits of agreement, bpm
Temple	**43**	**1.18 (1.10 to 1.26)**	**0.98**	**1.97**	**−0.31**	**−4.30 to 3.68**
Apple Watch Series 12	43	1.94 (1.65 to 2.28)	1.59	3.08	−0.61	−7.18 to 5.96
Fitbit Air	42	2.96 (2.61 to 3.35)	2.41	4.43	−0.70	−9.70 to 8.30
WHOOP MG	43	3.14 (2.61 to 3.69)	2.52	4.74	−0.84	−11.09 to 9.40
Garmin FR970	43	3.18 (2.89 to 3.54)	2.72	4.77	−0.38	−9.78 to 9.02
Oura Ring 5	43	5.21 (4.42 to 6.12)	4.24	8.18	−2.70	−18.88 to 13.47

MAE, MAPE, and RMSE are means of participant values; bias and limits of agreement from the Bland-Altman method for repeated measures. Failure episodes were removed. Fitbit Air was not recorded for one participant.

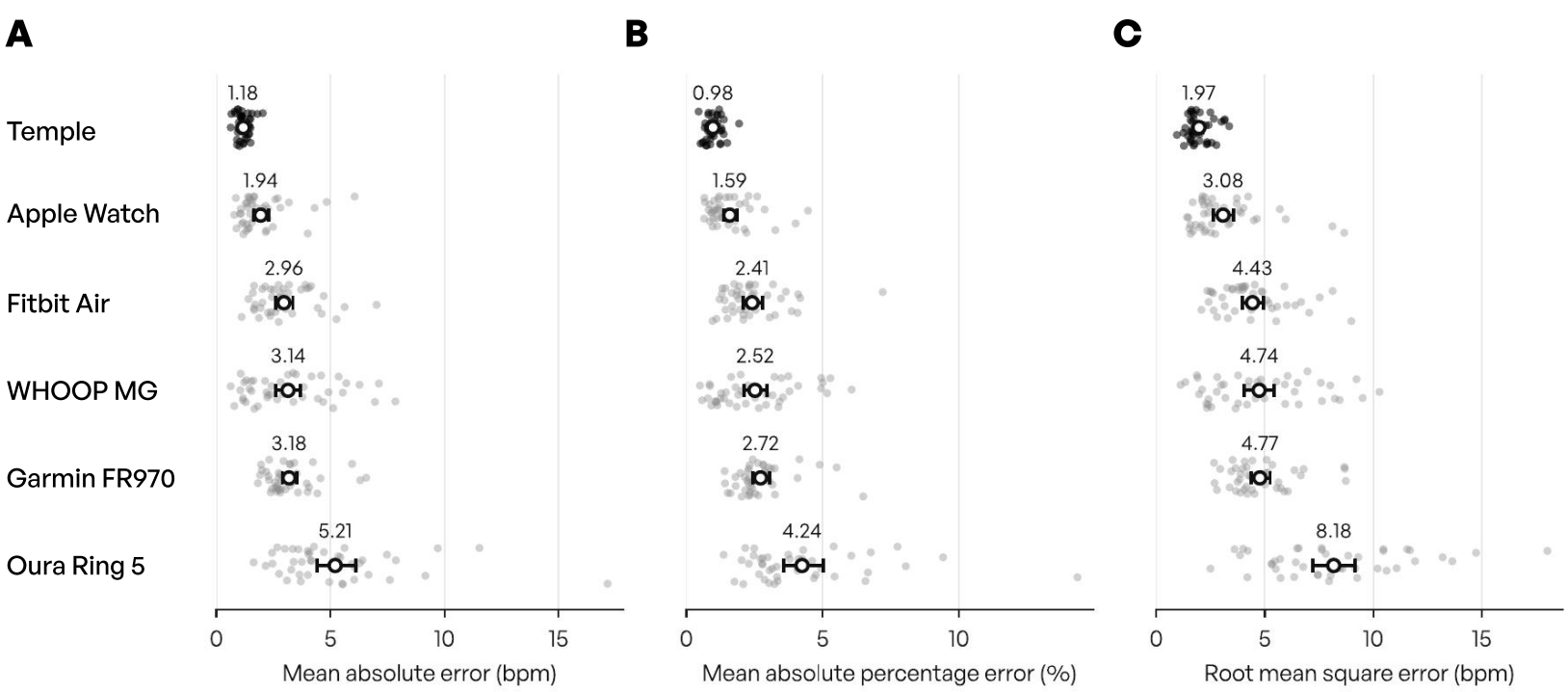


Figure 2. Overall accuracy. (A) Mean absolute error (MAE), (B) mean absolute percentage error (MAPE), and (C) root mean square error (RMSE) against the taped Polar H10. Dots represent participants; open circles represent means with 95% CIs; the printed value is the mean. Failure episodes are removed.

03 / RESULTS

3.2 Head-to-head and by zone

On the seconds both devices reported, Temple's MAE was 0.85 bpm lower than that of Apple Watch Series 12 (95% CI 0.55 to 1.19; Holm $p < 0.001$). Its margin was 1.80 bpm over Fitbit Air, 1.98 bpm over WHOOP MG, 2.03 bpm over Garmin Forerunner 970, and 4.01 bpm over Oura Ring 5 (all $p < 0.001$). At rest and across all zones, Temple had the lowest error, ranging from 0.95 bpm in Zone 5 to 1.52 bpm at rest (Fig. 3), and its MAE was significantly lower than that of each of the five flagship wearables.

A Mean absolute error at rest, in each zone and overall (bpm)

DEVICE	REST	ZONE 1	ZONE 2	ZONE 3	ZONE 4	ZONE 5	OVERALL
Temple	1.52	1.30	1.21	1.04	1.31	0.95	1.18
Apple Watch	1.98	2.27	2.41	2.09	1.73	1.84	1.94
Fitbit Air	2.84	3.90	4.06	2.81	2.72	2.10	2.96
WHOOP MG	2.52	3.49	4.38	3.23	2.85	3.54	3.14
Garmin FR970	3.74	4.24	4.24	2.63	3.11	1.95	3.18
Oura Ring 5	4.62	5.91	5.44	5.26	5.23	5.43	5.21

B

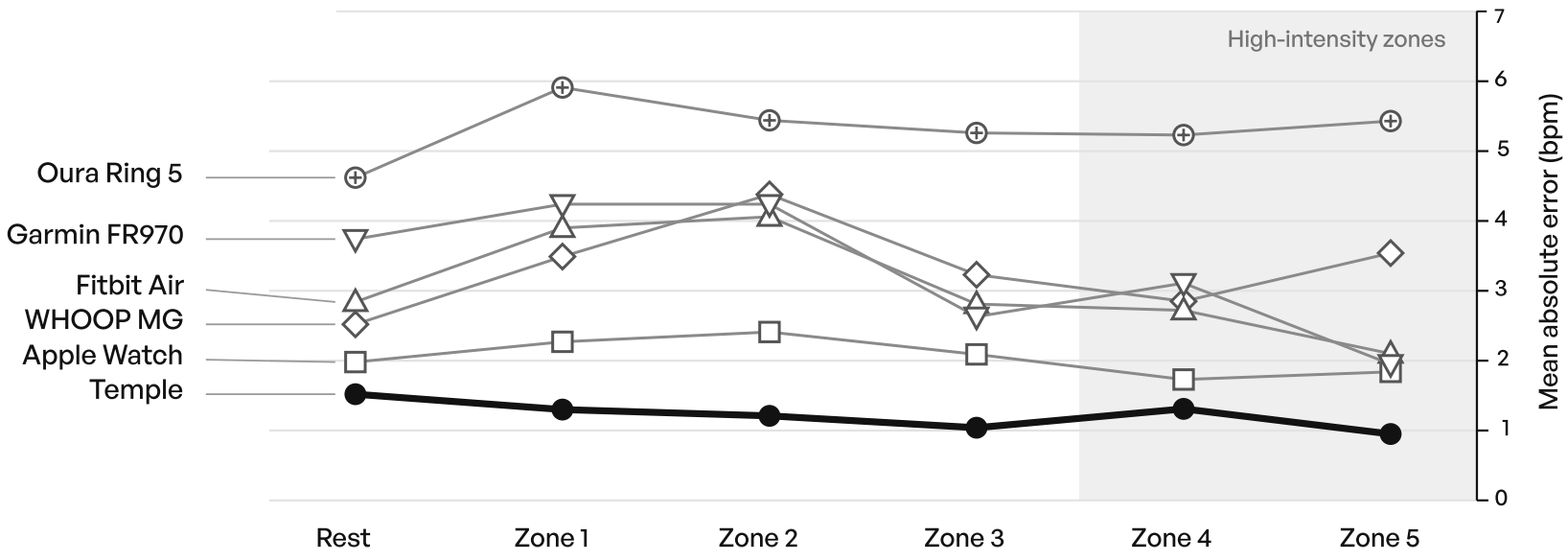


Figure 3. Temple's error was the lowest at rest and in every zone. (A) Mean absolute error (MAE) at rest, in each heart rate zone, and over the whole session (Overall), mean of participants; darker cells mark larger error. (B) The same values across the zones, one line per device; the shaded band marks the high-intensity Zones 4 and 5. Failure episodes were removed.

03 / RESULTS

3.3 Agreement with the reference

Temple's readings stayed closest to the reference. On average, 97.8% of a participant’s data points were within 5 bpm of the reference for Temple, against 92.3% for Apple Watch Series 12, 84.7% for Fitbit Air, 82.4% for WHOOP MG, 83.9% for Garmin Forerunner 970, and 69.8% for Oura Ring 5.

3.4 Bland-Altman analysis

Temple's bias was −0.31 bpm, and its 95% limits of agreement ran from −4.30 to 3.68 bpm, the narrowest among the six optical wearables (Fig. 4). The limits were −7.18 to 5.96 bpm for Apple Watch Series 12 and widened to −18.88 to 13.47 bpm for Oura Ring 5 (Table 3).

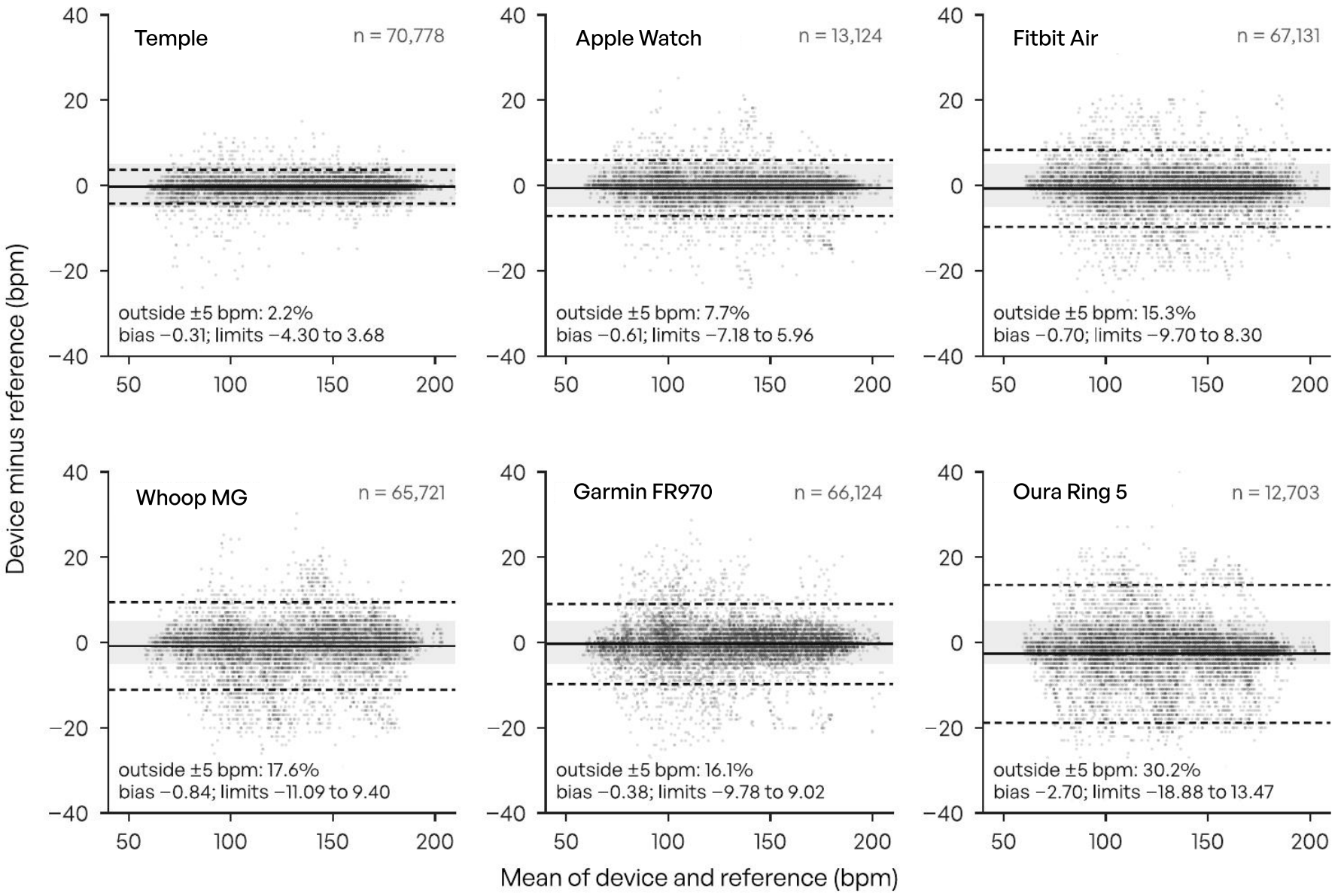


Figure 4. Temple had the narrowest limits of agreement among the six optical wearables tested. Bland-Altman plots on a common scale of ±40 bpm: device minus reference against the mean of the two. Every panel shows 11,000 data points drawn at random from that device's data, so all devices are shown at the same density (Apple Watch Series 12 and Oura Ring 5 report about every 5 seconds); n is the device's total number of data points. Shaded band: ±5 bpm, with the mean share of a participant's data points outside it. Solid line: bias; dashed lines: 95% limits of agreement for repeated measures. Shares, bias, and limits use all data points. Failure episodes removed.

03 / RESULTS

3.5 High-intensity zones

Temple's accuracy held in the high-intensity zones. With all valid readings retained, over every second of the session in Zones 4 and 5, its MAE in Zones 4 and 5 was 1.16 bpm (95% CI 1.06 to 1.26), and no participant's mean absolute error exceeded 5 bpm (0 of 42; worst 2.29 bpm), against 1 of 41 for Apple Watch Series 12, 6 of 41 for Fitbit Air, 11 of 41 for WHOOP MG, 7 of 42 for Garmin Forerunner 970, and 25 of 42 for Oura Ring 5. In Zone 5 alone, Temple's MAE was 0.95 bpm, and again, no participant's mean absolute error exceeded 5 bpm (0 of 41; worst 2.52 bpm), compared with 3 of 39 (Apple Watch Series 12) to 22 of 41 (Oura Ring 5) for the other wearables. Participant numbers differ because some participants did not reach a zone, some devices recorded no readings in it, and a device is counted in a zone only if it has at least 5 data points there.

In Zones 4 and 5 of the exercise protocol, Temple under-read by more than 5 bpm in 1.2% of its data points and over-read in 0.22% of its data points, compared with 7.1% to 34.3% and 0.41% to 2.8% for the other wearables (Fig. 5).

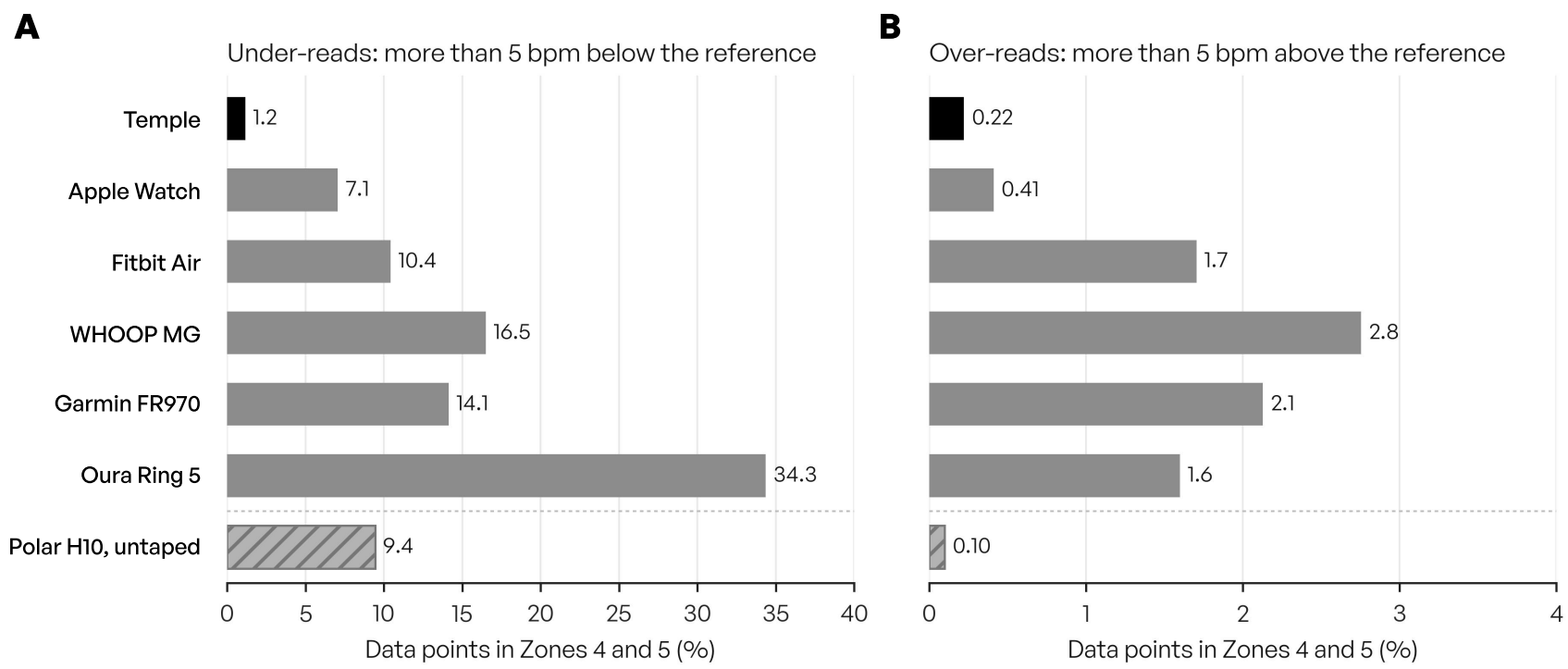


Figure 5. Temple stayed accurate in the high-intensity zones of the treadmill protocol. Data points in the high-intensity Zones 4 and 5 of the exercise protocol that were (A) more than 5 bpm below the reference (under-reads) and (B) more than 5 bpm above it (over-reads), as a percentage of each device's data points in those zones. All valid readings. The untaped Polar H10 (hatched, below the dashed line) is shown for reliability only. Percentages are pooled across all participants' data points.

03 / RESULTS

3.6 Reliability

Over the exercise protocol, Temple made the fewest large errors (Table 4). Only 1.8% of its data points were more than 5 bpm off, against 9.0% to 36.1% for the other wearables. Temple produced no drops or spikes during this period.

Table 4. Reliability over the exercise protocol, % of data points (all valid readings)

Device	Data points	>5 bpm off	Over-reads	Under-reads	Drops	Spikes	>5 bpm off, Zones 4 and 5
Temple	**51,444**	**1.8**	**0.61**	**1.2**	**0.00**	**0.00**	**1.4**
Apple Watch Series 12	9,421	9.0	2.3	6.7	0.15	0.00	7.5
Fitbit Air	49,907	17.3	4.4	13.0	1.2	0.03	12.1
WHOOP MG	49,472	23.5	6.0	17.5	4.1	0.20	19.3
Garmin FR970	51,690	21.9	6.6	15.3	6.9	0.98	16.3
Oura Ring 5	10,191	36.1	6.5	29.5	9.7	0.03	35.9
Polar H10, untaped (reliability only)	50,321	6.4	0.12	6.3	4.9	0.00	9.5

Exercise protocol: from the start of Zone 1 to the end of Zone 5; seated rest and cool-down left out. >5 bpm off: more than 5 bpm from the reference (a large error). Over-reads and under-reads: more than 5 bpm above or below the reference. Drops and spikes: failure episodes, more than 20 bpm below or above the reference for 10 seconds or more.

A second, untaped Polar H10 was more than 5 bpm off for 6.4% of data points, almost all of them under-reads, most of which were drops (4.9% of data points), possibly because of electrode-contact loss; in Zones 4 and 5, it under-read for 9.4% (Fig. 5A). Between its failures, it tracked the reference closely: its MAE was 0.59 bpm with failure episodes removed and 2.12 bpm with all readings kept (Temple 1.18 bpm in both).

03 / RESULTS

3.7 Sensor site

When one optical sensor model (Polar Verity Sense) was worn at four sites at once, its lowest observed MAE was at the temple: 3.50 bpm, against 5.10 bpm at the chest, 5.62 bpm at the thumb, and 6.03 bpm at the wrist (29 participants; all readings; Fig. 6A). The Temple device, in the same participants, had an MAE of 1.13 bpm.

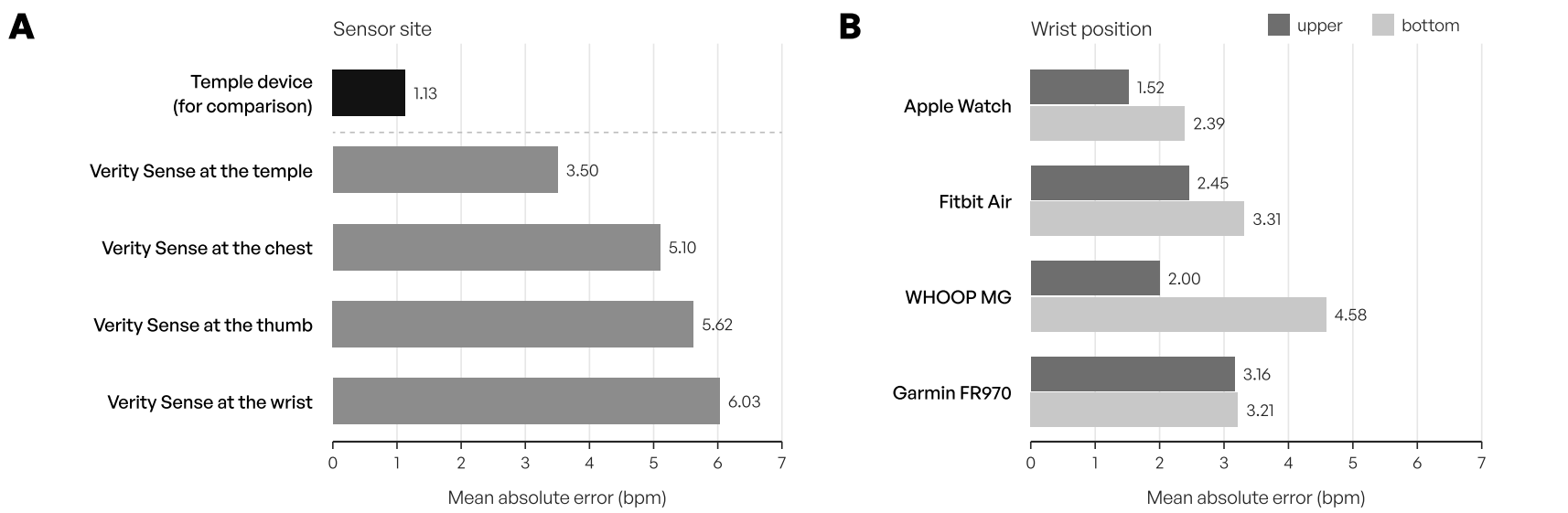


Figure 6. One optical sensor had its lowest error at the temple site. (A) Mean absolute error (MAE) of identical Polar Verity Sense sensors worn simultaneously at four sites, in the 29 participants with usable recordings at all four sites (all readings), with the Temple device in the same participants shown for comparison. (B) MAE of each wrist device in the upper and bottom positions, mean of participants (n = 17 to 25 per position); positions were rotated across participants. Failure episodes removed in B.

Every wrist device had lower error in the upper position than in the bottom position (Fig. 6B): Apple Watch Series 12: 1.52 vs 2.39 bpm; Fitbit Air: 2.45 vs 3.31 bpm; WHOOP MG: 2.00 vs 4.58 bpm; and Garmin Forerunner 970: 3.16 vs 3.21 bpm. With the wrist devices counted only in their better, upper position, Temple's error was still the lowest overall (Temple 1.18 bpm; wrist devices 1.52 to 3.16 bpm), at rest and in every zone (Table 5).

Table 5. Mean absolute error at rest, in each zone, and overall with the wrist devices in the upper position only (bpm, failure episodes removed)

DEVICE	N	REST	ZONE 1	ZONE 2	ZONE 3	ZONE 4	ZONE 5	OVERALL
Temple	43	1.52	1.30	1.21	1.04	1.31	0.95	1.18
Apple Watch Series 12, upper	22	1.93	1.74	1.79	1.40	1.39	1.39	1.52
WHOOP MG, upper	24	2.07	2.54	2.85	1.74	1.64	2.02	2.00
Fitbit Air, upper	17	2.31	2.95	3.08	2.11	2.34	2.09	2.45
Garmin FR970, upper	22	3.60	3.81	3.76	2.76	3.06	2.25	3.16
Oura Ring 5	43	4.62	5.91	5.44	5.26	5.23	5.43	5.21

Mean participant values. Wrist devices: participants who wore the device in the upper position. Temple and Oura Ring 5: all participants. Failure episodes were removed.

3.8 Subgroups

Temple’s MAE was 1.03 bpm in women (n = 9) and 1.22 bpm in men (n = 34), 1.11 bpm in participants with Monk skin tones 1 to 4 (n = 28) and 1.31 bpm in participants with tones 5 to 6 (n = 15). In each of these groups, Temple had the lowest MAE among the six optical wearables tested.

04 / INTERPRETATION

4. Discussion

Temple had lower error than five flagship wearables at rest and in every heart rate zone, and it was the most reliable device in the study. Over the exercise protocol, it had the fewest readings far from the reference, and fewer than a second, untaped Polar H10 ECG chest strap.

The Apple Watch Series 12 was the strongest comparator, as in earlier independent work [20] and a previous white paper that found it more accurate than other consumer wearables [21]. Temple's error was lower still. Wrist and finger errors were not confined to the hardest running, in line with reports that wrist error varies with the type of exercise [2], and is not limited to the highest intensities [22].

Of the four sites tested using a single optical sensor model, the temple had the lowest observed MAE. At the anterior temple, the sensor lies near the branches of the superficial temporal artery [8], away from hand and wrist movements that disturb wrist PPG during exercise [3]. Temple is always worn one way, and does not move once applied, so people wear it every day exactly as it was worn in this study. We did not test accuracy during everyday activities. People wear wrist devices higher or lower on the arm. In this study, all four were more accurate when worn higher up the arm (Fig. 6B).

Chest straps depend on electrode contact, which the manufacturer lists as a requirement for a correct reading [12]. The tape held our reference strap in place. A second, untaped Polar H10 had more large errors than Temple, most often in Zones 4 and 5, possibly due to electrode contact loss.

Limitations

This study took place at one site, with one session per participant and treadmill running only. Other forms of exercise, such as cycling, strength training, and outdoor running were not tested. The sample included 43 participants aged 20 to 35 years, 34 of whom were men. Their skin tones spanned Monk 1 to 6; tones 7 to 10 were not represented. The reference was a chest-strap ECG, not a 12-lead recording. Comparators ran the firmware of the time, which may change. Wrist positions were rotated, not randomized, so the upper-versus-bottom comparison is between participants, and the main wrist results pool the two positions; upper-position results are in Table 5. The untaped strap sat slightly lower than its usual position, directly below the chest muscles, which may have contributed to its errors. The Verity Sense substudy was smaller, and some recordings were cut when a sensor dropped off.

04 / INTERPRETATION

Conclusions

Temple measured heart rate with a mean absolute error of 1.18 bpm from rest to near-maximal running, lower than each of the five flagship wearables tested against a taped ECG chest strap. In high-intensity Zones 4 and 5, its error was 1.16 bpm, with no participant's mean absolute error exceeding 5 bpm, and it made the fewest large errors of any device in the study. In a separate analysis of one optical sensor worn at four sites, the temple had the lowest observed MAE. Temple is always worn one way, and does not move once applied, so people wear it every day exactly as it was worn in this study. How a wrist device performs depends on whether it’s worn higher or lower on the arm. These findings apply to the tested devices and placements during this controlled treadmill protocol.

Declarations

The study was approved by an independent ethics committee, and all participants gave written informed consent before taking part. This was an observational device-comparison study in healthy volunteers and not a clinical trial. Temple Private Limited, which makes Temple, funded, ran, and analyzed the study. De-identified raw and cleaned heart rate data for Temple, the five flagship wearables, and both chest straps are available here. Generative AI was used to assist with the analysis and drafting; all analyses and text were reviewed and verified before publication.

05 / SOURCES

5. References

1. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med. 2020;3:18. doi:10.1038/s41746-020-0226-6

2. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc. 2017;49(8):1697-1703. doi:10.1249/MSS.0000000000001284

3. Fine J, Branan KL, Rodriguez AJ, Boonya-Ananta T, Ajmal, Ramella-Roman JC, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel). 2021;11(4):126. doi:10.3390/bios11040126

4. Hung SH, Serwa K, Rosenthal G, Eng JJ. Validity of heart rate measurements in wrist-based monitors across skin tones during exercise. PLoS One. 2025;20(2):e0318724. doi:10.1371/journal.pone.0318724

5. Moghaddam M, Collins JP, Gardner CE, Rabel MC. Impact of anatomical placement on the accuracy of wearable heart rate monitors during rest and various exercise intensities. Sensors (Basel). 2026;26(1):176. doi:10.3390/s26010176

6. Olstad BH, Zinner C. Validation of the Polar OH1 and M600 optical heart rate sensors during front crawl swim training. PLoS One. 2020;15(4):e0231522. doi:10.1371/journal.pone.0231522

7. Bartolomeu RF, Silva V, Pereira A, Morais G, Sokołowski K, Strzała M, et al. Agreement between a wristwatch and a free optical sensor with a chest strap in measuring HR variations during front crawl swimming. Appl Sci (Basel). 2025;15(11):5848. doi:10.3390/app15115848

8. Pinar YA, Govsa F. Anatomy of the superficial temporal artery and its branches: its importance for surgery. Surg Radiol Anat. 2006;28(3):248-253. doi:10.1007/s00276-006-0094-z

9. Mühlen JM, Stang J, Lykke Skovgaard E, Judice PB, Molina-Garcia P, Johnston W, et al. Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network. Br J Sports Med. 2021;55(14):767-779. doi:10.1136/bjsports-2020-103148

10. Schaffarczyk M, Rogers B, Reer R, Gronwald T. Validity of the Polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. Sensors (Basel). 2022;22(17):6536. doi:10.3390/s22176536

11. 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(7):1525-1532. doi:10.1007/s00421-019-04142-5

12. Polar Research and Technology. Polar H10 heart rate sensor system [Internet]. White paper. Kempele (Finland): Polar Electro Oy; 2019 Nov 11 [cited 2026 Oct 30]. Available from: https://www.polar.com/img/static/whitepapers/pdf/polar-h10-heart-rate-sensor-white-paper.pdf

13. Monk E. The Monk Skin Tone Scale [Preprint]. SocArXiv; 2023 May 4. doi:10.31235/osf.io/pdf4c

14. WHOOP. What affects WHOOP accuracy? [Internet]. Boston (MA): WHOOP; 2026 Aug 13 [cited 2026 Oct 30]. Available from: https://support.whoop.com/s/article/What-Affects-WHOOP-Accuracy

15. Oura Health. How to choose the right Oura Ring size [Internet]. Oulu (Finland): Oura Health; [cited 2026 Oct 30]. Available from: https://support.ouraring.com/hc/en-us/articles/360025590653

05 / SOURCES

16. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol. 2001;37(1):153-156. doi:10.1016/S0735-1097(00)01054-8

17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327(8476):307-310. doi:10.1016/S0140-6736(86)90837-8

18. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat. 2007;17(4):571-582. doi:10.1080/10543400701329422

19. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6(2):65-70.

20. Hajj-Boutros G, Landry-Duval MA, Comtois AS, Gouspillou G, Karelis AD. 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(2):165-177. doi:10.1080/17461391.2021.2023656

21. Apple Inc. Apple Watch heart rate accuracy study [Internet]. Cupertino (CA): Apple Inc.; 2026 Sep [cited 2026 Oct 30]. Available from: https://www.apple.com/health/pdf/Heart_Rate_Accuracy_Study_2026.pdf

22. Jeong Y, Yoon S, Lee J, Jeong WM, Ha MS. Comparison of heart rate measurement accuracy among commercially available photoplethysmography-based wearable devices across exercise intensities. Phys Act Nutr. 2026;30(2):142-150. doi:10.20463/pan.2026.0032

© 2026 Temple Private Limited. All rights reserved. Bluetooth is a registered trademark of Bluetooth SIG, Inc. Apple Watch, Apple Health, Oura, WHOOP, Fitbit, Garmin, Polar and other product and company names mentioned here are trademarks of their respective owners. Temple is not affiliated with or endorsed by these companies.