Minor Research Project

Exploring Health Insights from Polar Vantage V3, Polar Verity Sense and Polar H10 Integration within the RADAR-base Platform

Famke Schulting
MSc. Bioinformatics & Biocomplexity, Utrecht University
Examiner: Dr. Anke Huss[1]
Supervisors: MSc. Wouter Franke[2] and Dr. Bastiaan de Graaf[2]
[1] Fac. of Veterinary Medicine, Dept. for Risk Assessment Sciences (IRAS), Utrecht University
[2] The Hyve B.V., Utrecht
August 13, 2024

Table of Contents

Abstract

The emergence of wearable technology has led to the emergence of the integrated data collection platform RADAR-base. The objective of this study was to assess the added value of three newly integrated Polar devices – the Polar Verity Sense, the Polar Vantage V3, and the Polar H10 - to the RADAR-base platform.

These wearables were evaluated in terms of their ease of use, robustness, and the quality of their data collected. To this end, heart rate measurements of 13 participants were taken during different activity phases, using the newly integrated devices and a Fitbit Charge 2 for comparison. Measurements were compared to the Polar H10, as this chest strap uses ECG, which is reported to be a standard in measuring heart rate (1).

The Polar Verity Sense was excluded due to connectivity issues. The Bland-Altman plots revealed a mean bias of 1,21 bpm for the Polar Vantage V3, of which 95% of values fell within –8.57 and 10.98 bpm. For the Fitbit Charge 2, a mean bias of 5,64 bpm was found, with 95% of values within -14,42 and 25,69 bpm. High agreement was found as the Lin's concordance correlation coefficients (r_c) were 0,98 (substantial) and 0,91 (moderate) for the Polar Vantage V3 and Fitbit Charge 2, respectively (2).

Based on these findings, we conclude that while each of the analyzed devices is suitable for use in research, the Vantage V3 appears to be the most suitable in terms of consistency, ease of use and accuracy. Nevertheless, it is crucial for researchers to consider the differences in usage and data processing methodologies of these devices when designing their patient monitoring studies, and that they limit HR data collection throughout larger studies to a single device as much as possible.

Layman's summary

Wearables, which are devices like fitness trackers and smartwatches, are becoming more popular. These devices can measure various biomarkers, such as skin temperature, step count, sleep performance and heart rate. Heart rate is an important biomarker in research, as many conditions can cause changes within a person's heart rate, which means we can potentially use heart rate to detect such diseases. Traditionally, long term studies gather heart rate data by using specialized devices or during doctor visits. Nowadays, heart rate can be gathered using wearables, which is not only more convenient for patients but also gives researchers and clinicians a much larger body of data, as heart rate is continuously monitored. To store these measurements, platforms such as RADAR-base were developed, which support simultaneous data collection of different biomarkers from various sources. Recently, Polar wearables were integrated into the RADAR-base platform, To evaluate the performance of these wearables and the quality of their heart rate measurements, an experiment was set up. For this, 13 participants were asked to wear the Polar devices and a Fitbit Charge 2 simultaneously during a 15-minute experiment in which their heart rate was measured. These participants were first asked to stay seated for 5 minutes, so that their heart rate could be measured during rest. Next, participants were asked to cycle on a home trainer, to measure their heart rate during exercise. And finally, participants were asked to remain seated again, to measure their heart rate during recovery. The heart rate measurements of the different devices were compared to the measurements of the Polar H10 device. The Polar H10 uses ECG, which is considered to be very accurate for measuring HR. One of the devices, the Polar Verity Sense had connectivity issues and was therefore excluded from our study. Different statistical tests showed that the heart rate measurements of the remaining devices, the Fitbit Charge 2 and Polar Vantage V3, were highly similar.

Therefore, we believe that each of the tested devices can be used in research. Nevertheless, in our opinion, the Polar Vantage V3 would be most suitable for patient monitoring studies, as this device is very easy to use and its heart rate measurements were found to be close to the standard.

That being said, as all these devices are used differently and measure heart rate using different methods, researchers should think critically about which device would best suit their research needs and select a single device.

Introduction

In recent years, the use of wearable technology has revolutionized the way fitness and health information is monitored. Due to the development of new wearable sensors, an opportunity for the continuous monitoring of patients is emerging in healthcare. Leveraging such data enables a deeper understanding of disease ethology, improved diagnosis and prognosis, and potentially even detecting disease relapse at early stages, suitable for intervention (3). The ease of use of these novel devices also significantly reduces the burden of clinical research on patients, particularly if they reduce the number of necessary doctor visits (4).

Alongside the development of wearables, platforms for (integrated) data collection using wearables have been developed such as Vivalink (5) and ICON (6). Other platforms, such as the Empatica Health Monitoring Platform, are designed for data collection through a specific device or brand (7). Another platform that has been developed to leverage wearables is RADAR-base (Remote Assessment of Disease And Relapses): an open source platform that is used to collect data from wearables and mobile applications (8). RADAR-base enables researchers to collect a wide range of data streams in real-time without fear of data-loss and converges all of them into a single location. This significantly facilitates subsequent data analysis, as managing data from different sources can impose a significant challenge. RADAR-base can be used for passive monitoring of participants via their phone and wearable sensors such as the Empatica E4 (9), Pebble 2 (10), Fitbit Charge 2 (11), Biovotion (12) and Faros (13), to record a variety of biomarkers. For instance, RADAR-base was recently used in combination with the Fitbit Charge 2 to evaluate heart rate control in a randomized trial where patients were treated with either digoxin or beta-blockers (14).

Many studies using wearables focus on monitoring patient heart rate. This illustrates the particular interest of monitoring patient heart rate. Worldwide, cardiovascular diseases are the leading cause of death, taking an estimated 18,6 million lives per year (15). Consequently, heart rate measurements are highly sought after as a biomarker in research. Variations in heart rate over time can reveal important clinical insights into the health of the physiological system that produces these changes. For instance, elevated resting heart rate can be associated with the development of cardiovascular disease and increased mortality with conditions such as COPD, hypertension, diabetes and cardiovascular disease (16–18). Heart rate measurements are commonly obtained through either PPG or ECG sensors. In an electrocardiogram (ECG), the heart's electrical activity is recorded through repeated cardiac cycles. Each heartbeat is associated with corresponding signal phase and characteristics, which can be seen in the ECG-plot. When determining heart rate, the so-called QRS complex, as shown in Figure 1 marks the ventricular depolarization. The heart rate is determined by the RR-interval, which is the time between two R waves. Each R wave is produced by depolarization of the main mass of the ventricles (19). In PPG, Optical Heart Rate (OHR) sensors use photoplethysmography (PPG) to measure changes in blood volume under the skin. PPG makes use of LEDs which emit light into your skin, and a photodiode that detects the intensity of the light reflecting from the skin. As changes in the volume of blood flow in the wrist affect the amount of light reflected, the blood flow through the capillaries in the skin can be measured and heart rate can be determined. This PPG technology has been widely used in commercially available medical devices, as it can simultaneously measure blood pressure, oxygen saturation, cardiac output, and respiration (20). Such PPG-based devices are more commonly used due to their smaller size, lower cost, greater comfort, and ease of use compared to ECG devices (21). However, as ECG is still considered to be the gold standard for measuring HR, it has recently been adopted into commercial wearables (22).

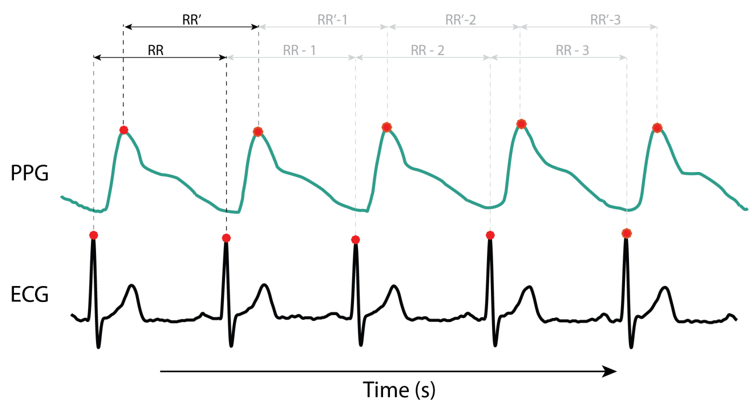

Figure 1: The RR and RR' intervals of PPG and ECG signals are used to calculate HR
Instantaneous heart rate (bpm) can be calculated by dividing 60 s by either the RR' or RR value (5). Wearables make use of algorithms that all look for these heartbeat signatures. To eliminate noise in these HR signals, these algorithms, such as the Pan-Tompkins algorithm employ sophisticated filtering steps (23). This figure is reprinted from Vandenberk et al. (24).

The rising number of devices also poses a challenge for data collection platforms: Deciding what devices to integrate and maintain. While integration of many devices into a platform is obviously beneficial, the costs involved in such efforts can be significant.
One newer brand is Polar Electro Oy, which is commonly known as Polar. Polar markets itself as a global leader in the health and fitness industry due to their expertise in

development and production of heart rate monitors. As such, we received several requests from researchers to include Polar devices (such as The Polar Vantage V3; a wrist-worn device designed for continuous physiological monitoring using a PPG sensor. And the Polar Verity Sense; another device using PPG, which can be worn around your arm or temple). into the RADAR-base platform.
We integrated these devices into the platform to solve our research question: How do these Polar devices perform in terms of consistency, robustness and quality of their data collected within the RADAR-base environment, compared to other, previously integrated PPG devices. As a test case we selected the Polar devices and a Fitbit Charge 2, a device that was easily accessible and has already been widely used in combination with the RADAR-base platform (14).

To this end, we conducted an experiment where participants wore the newly integrated devices simultaneously during various phases of activity, along with the previously integrated Fitbit Charge 2 (Figure 2). We then assessed the alignment of the HR measurements obtained from the Polar Vantage V3, Polar Verity Sense, and Fitbit Charge 2 with those of the Polar H10, to characterize the quality of the obtained data.

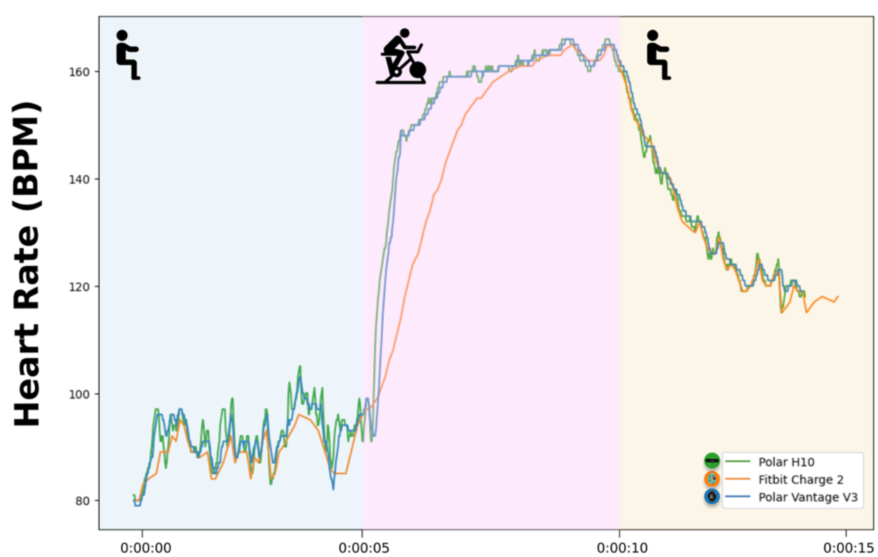

Figure 2: HR will be measured using different devices during a 15-min testing session
HR measurements will be collected during a resting phase (highlighted in blue), exercise phase (highlighted in pink) and recovery phase (highlighted in yellow) of one participant. The Polar H10 (green line), Fitbit Charge 2 (orange line) and Polar Vantage V3 (blue line) were worn simultaneously to compare the different HR measurements.

Materials and methods

Materials

Sensors and recorded signals

Within this study, heart rate measurements from the Polar Vantage V3 *(SN: C4095P0622722)*, Polar Verity Sense *(SN: M4094J1178250)*, Polar H10 *(SN: M4041W0558072)* and Fitbit Charge 2 *(SN: 2af235a15a16)* were compared (Table 1). The Polar Vantage V3, Polar Verity Sense and Polar H10 each provide the HR measurements with a frequency of 1.0 Hz, while the Fitbit Charge 2 has a frequency of 0.1 Hz.

During the experiment, participants were asked to exercise on a home trainer (VirtuFit FB1.0i, (25)). A home trainer was selected to minimize the impact of movement on the PPG sensors, which are sensitive to motion.

Fitbit Charge 2	**Polar Vantage V3**	**Polar Verity Sense**	**Polar H10**
Heart Rate in bpm (PPG)	Heart Rate in bpm (PPG)	Heart Rate in bpm (PPG)	Heart Rate in bpm (ECG)
Heart Rate Variability	PP-interval in ms	PP-interval in ms	RR-interval in ms
Breathing Rate (during sleep)	Accelerometer data	Accelerometer data	Accelerometer data
Calories	Wrist ECG *	Photoplethysmography (PPG) values	Electrocardiography (ECG) in V
SpO_2	SpO_2 *		
Skin Temperature	Skin Temperature *		
Sleep stage	Sleep stage *		
Step Counts	Step Counts *		

Table 1: Overview of the physiological parameters measured by the different devices
In blue, the parameters that are included in the comparison are shown. The starred Polar Vantage V3 features are not supported by Polar's SDK and therefore cannot be integrated into the RADAR-base platform using the pRMT app at this time.

RADAR-base

The RADAR-base platform can integrate monitoring data of various data sources. In Management Portal, a project was registered to which each participant was added and assigned with the correct sources. Next the experiment was conducted, and the collected heart rate data was either sent via Bluetooth to the pRMT application or, in the case of Fitbit Charge 2, was gathered using a third-party RESTful API connection to the S3 storage of the RADAR-base instance (Figure 7).

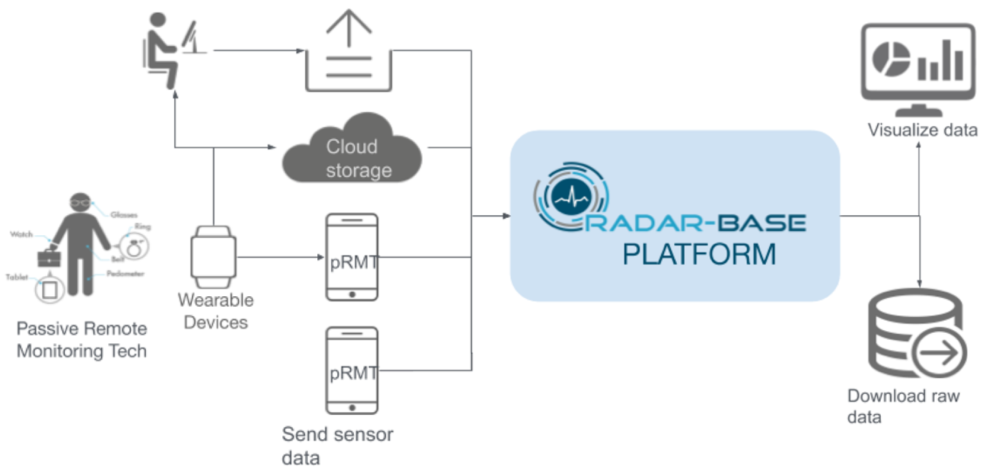

Figure 7: The RADAR-base platform is used for passive data collection
The RADAR-base platform enables the collection of passive monitoring data via various data sources. These include either phone sensor data or data from wearables that are connected via Bluetooth to the pRMT app, manually uploaded data or data that is gathered via a third-party RESTful API connection. The collected data can be directly visualized or the raw materials can be downloaded. Adapted from Ranjan Y et al (26).

Methods

Software development process

For this study, updates were made to the RADAR-base platform to enable data collection using Polar devices to the RADAR-base platform. For this, the RADAR-commons-android library was used to create a new plugin for the RADAR-base pRMT app (28,29). For the development of this plugin, the publicly available SDK of Polar was utilized (30). This SDK was used to establish the connection of the Polar devices to the pRMT app via Bluetooth. Subsequently, HR data measured by the devices, was sent to the storage center of the RADAR-base instance, using a AVRO file format (27,31).

Study population

13 participants (10 male and 5 female, age 23-52), that did not have any known cardiovascular condition, were recruited on a voluntary basis, and gave informed consent to participate in the experiment. A privacy scan was conducted to assess whether the study adhered to the key principles of the GDPR, for which proper and fitting technical and organizational measures were implemented.

Experimental study design

The experimental study was set up as follows:
Each participant wore the devices simultaneously during a single setting. The Polar H10 was worn around the chest, while the Polar Vantage V3 and Fitbit Charge 2 were worn randomly on either wrist and the Polar Verity Sense was worn on a randomly chosen upper arm. All Polar devices were connected to the pRMT app via Bluetooth. The Fitbit Charge 2 was connected to the Fitbit application.
Participants were instructed to keep their arms stable on the handlebars during all activity phases to minimize movement interference with the PPG sensors. After a 1-minute normalization period to ensure stable heart rate signals, participants were instructed to follow this protocol (Figure 2):

1. Remain seated for 5 minutes (*resting phase*), to get a baseline recording of the heart rate.
2. Cycle on a home trainer for 5 minutes (*exercise phase*). Participants were asked to give a moderate to heavy effort.
3. Remain seated again for 5 minutes (*recovery phase*), to measure their heart rate while it recovers again after the exercise.

Analysis

Excluded data

For 3 participants, we observed in the Bland-Altman plots that the collected HR data for one or both PPG devices fell outside of the 95% CI (Sup. Figure 2). We excluded these patients from subsequent analyses.
Furthermore, the Polar Verity Sense showed connectivity issues throughout the entire study, resulting in an extremely low (N = 247) number of recorded data points that could be aligned to the Polar H10, making this data unsuitable for meaningful statistical

interpretation without introducing strong biases. Consequently, due to this small number of measurements compared to the other devices, we opted to exclude the Polar Verity Sense from all statistical analyses (Sup. Table 5).
To match the lower frequency of 0.1 Hz of the Fitbit Charge 2, the measurements of all the other devices, which have a frequency of 1.0 Hz, were adjusted to this frequency. To achieve this, the mean of every 10 HR measurements was calculated. Subsequently, only every tenth HR measurement was set to this mean and included in the analysis.
The datapoints of each device were matched to corresponding Polar H10 datapoints for all statistical testing. If no matching datapoint was recorded, the datapoint was excluded.

Statistical analysis

To assess the distribution of HR measurements from the tested devices, Shapiro-Wilk tests were performed on each dataset. As the results indicated non-normal distributions for each dataset, further analysis consisted of non-parametric tests. As discussed by Sartor et al., performing a Student *t* test in device validation studies is not appropriate, as a *t* test assesses difference, meaning when the null hypothesis that the two means are equal is rejected, that does not yet prove that these two means are equal (32). Therefore, the Lin's concordance coefficient (r_c), which is the concordance between a new test or measurement and a gold standard test or measurement, was reported instead (33).
In this study, the Lin's concordance correlation coefficient (r_c) was used to assess the concordance between the HR measurements of each device and the Polar H10. These r_c values were interpreted according to Mahon et al, who reported values below 0,90 to be poor, from 0,90 to 0,95 to be moderate, 0,95 to 0,99 to be substantial and values higher than 0,99 to be almost perfect (2).The root mean square error (RMSE) and normalized root mean square error (NRMSE) were performed to evaluate the range of error between the HR measurements of the Polar H10 and the other devices. Lastly, Bland-Altman plots were generated to visualize the differences between the two measurements against their averages (34).

Data analyses were carried out using Python 3.10.9 (35) (with libraries: Pandas 1.5.3 (36), NumPy 1.23.5 (37), SciPy 1.10.0 (38), Matplotlib 3.7.0 (39) and Sklearn 1.2.1 (40)).

Results

In total, 13 participants were part of this study. Of each participant, the HR measurements (in bpm) acquired by the Polar Vantage V3, Polar Verity Sense and Fitbit Charge 2 were each compared to the acquired measurements of the Polar H10, to assess their accuracy during different phases of activity. In total 26505 datapoints were recorded: 11398 from the Polar H10, 11327 from the Vantage V3, 2098 from the Polar Verity Sense and 1682 from the Fitbit Charge 2. These time points were aligned, resulting in 1218 time points and 3654 measurements, ranging from 52 to 166 bpm.

High quality of measurements in 3 out of 4 devices

During each experiment, the Polar H10, Polar Vantage V3 and Fitbit Charge 2 performed without meaningful issues or failures. However, for 12 out of the 13 participants, the Polar Verity Sense sensor lost connection during the *exercise* phase. Due to this, the Polar Verity Senses' data was not considered of high enough quality for meaningful comparison and was therefore not included in further analysis (Sup. Table 5, Sup. Figure 1).

Moreover, for all 13 participants Bland-Altman plots were generated. These showed that there were 3 participants of which most of their HR measurements fell outside the 95% CI, which is why these participants were excluded from further analysis (Sup. Figure 2). After inspecting their individual plots, we saw that these outliers either occurred due to deviating HR measurements of the Fitbit Charge 2 (for subject-5) or both PPG devices (for subject-3 and subject-6) (Sup. Figure 3).

High similarity between Polar devices

The distribution of HR measurements of each device was similar across all analyzed devices (Figure 3). Notably, the two Polar devices (Polar H10 and Polar Vantage V3) show a particularly high similarity in their density curves. This indicates that HR measurements recorded by the two Polar devices are closely aligned, reflecting the consistency within the Polar brand.

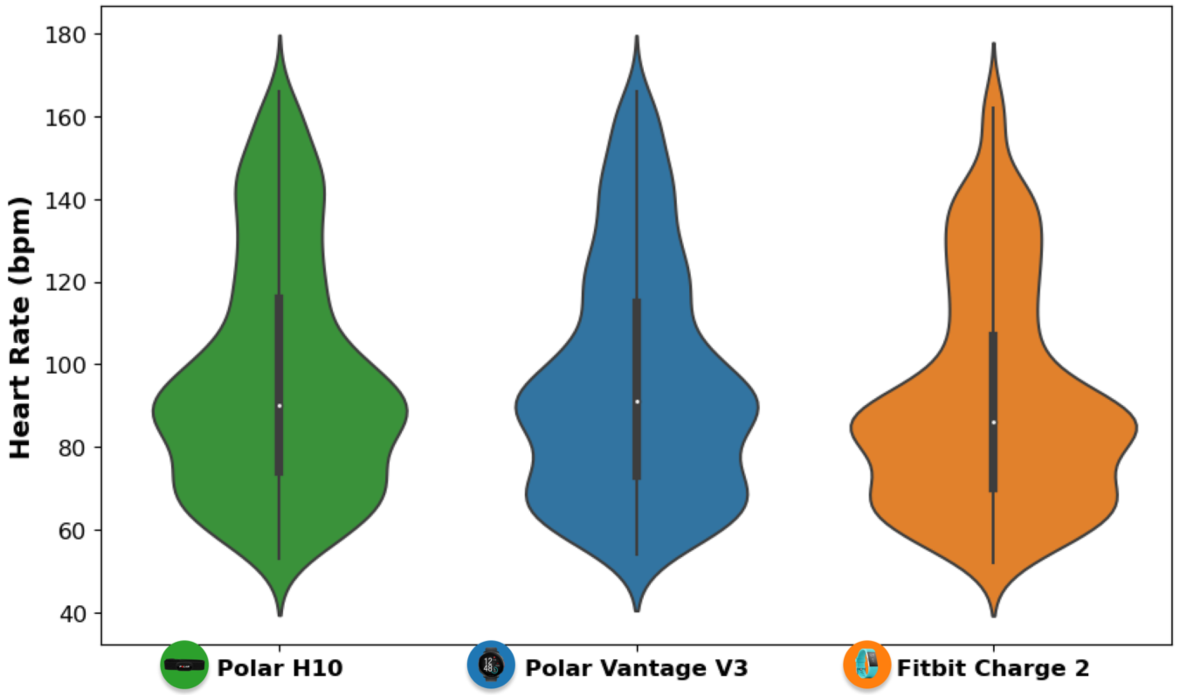

Figure 3: High similarity between distribution of HR measurements for Polar devices
Violin plot showing the density of all recorded HR measurements per device for the Polar H10 (green), Polar Vantage V3 (blue) and the Fitbit Charge 2 (orange). High similarity in distribution of HR measurements can particularly been seen for the Polar devices, reflecting the consistency within the Polar brand.

Moreover, the RMSE values of 5.03 and 11.62, along with NRMSE values of 4% and 10% for the Polar Vantage V3 and Fitbit Charge 2, respectively, suggest a high degree of consistency and similarity in HR measurements from these devices (Sup. Table 6).

Furthermore, the Lin's concordance correlation coefficients (r_c) showed there was a substantial agreement between the Polar H10 and Polar Vantage V3 HR measurements ($r_c = 0.98$), and a moderate agreement to the Fitbit Charge 2 HR measurements ($r_c = 0.91$) (Figure 4) (2).

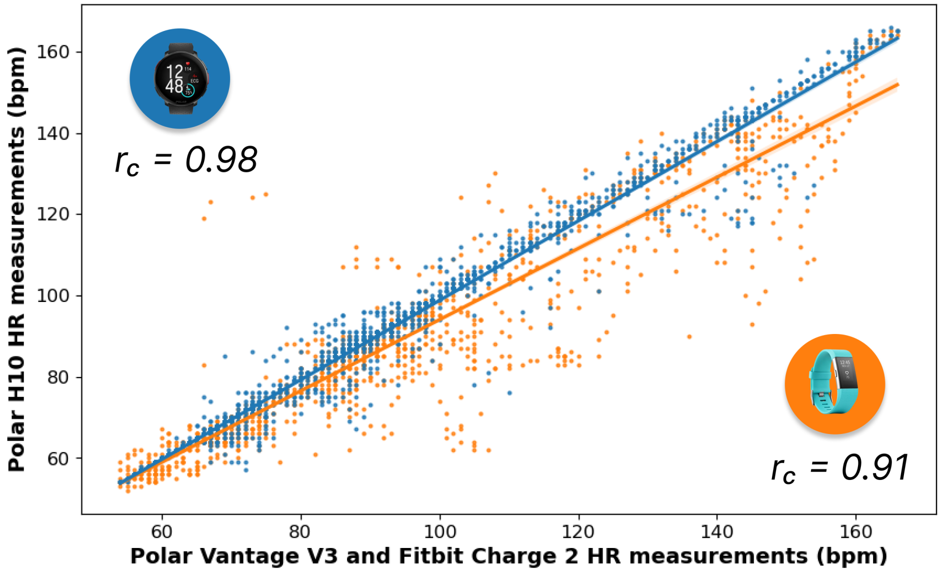

Figure 4: Polar Vantage V3 has the highest correlation with Polar H10
HR measurements of Polar Vantage V3 (blue) correlate more closely to the Polar H10, compared to HR measurements of Fitbit Charge 2 (orange). Moreover, the HR measurements of the Fitbit Charge 2 are more widely scattered, indicating greater discrepancies in HR measurements when compared to the Polar H10.

Moreover, Bland-Altman plots were created to visualize the agreement between each device-pair (Figure 5). These also show that the similarity between the Polar devices was the highest. The mean bias between Polar H10 and Polar Vantage V3 was 1.21, while the mean bias between Polar H10 and Fitbit Charge 2 was 5.64 bpm.
For both the Fitbit Charge 2 and the Polar Vantage V3 (although to a lesser extent), some measurements fell outside the lower limit of agreement (LLA) and the upper limit of agreement (ULA). Most measurements that lay outside these limits were HR measurements recorded during the *exercise* phase.

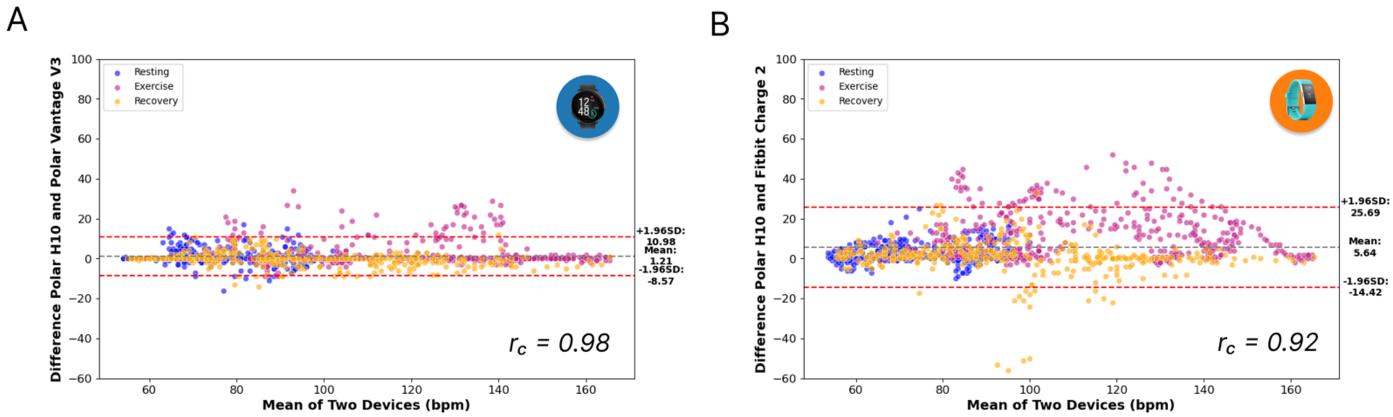

Figure 5: Polar Vantage V3 shows highest agreement with Polar H10
Bland-Altman plots show more agreement between Polar H10 and Polar Vantage V3 (panel A) than Polar H10 and Fitbit Charge 2 (panel B). This is evidenced by the more centered mean bias and narrowed spread of differences, as well as the higher correlation coefficient (r_c = 0.98). *Especially for the Fitbit Charge 2, most outliers occurred during the exercise phase.*

Similarity per subject between devices

We found that even after exclusion of the three participants of which most HR measurements fell outside the 95% CI of the BA plots, there were still subjects for which the r_c was considered poor (< 0,90) if we determined the r_c for each subject individually. While for the Polar Vantage V3, most r_c values that were calculated based on the HR measurements of each subject individually were high, between 0.92 and 0.99, for two subjects it was much lower (r_c = 0.74 and r_c = 0.87 for subject-10 and 2, respectively). However, for the Fitbit Charge 2, more dispersion between the different subjects was found, as the r_c ranged between 0.71 – 0.96 for all subjects, apart from two subjects (r_c = 0.41 and r_c = 0.33 for subject-7 and 2, respectively).
When observing the HR measurements of these subjects, dispersions could indeed be seen. While HR measurements of the Polar H10 showed consisted patterns and expected behavior based on the type of activity, the other devices either measured a too low HR, or had a time syncing error at the start of the transition from the *rest* to *exercise phase* (Sup. Figure 4). These observations seemingly indicate that the performance of the device is linked to individual subject characteristics.

Time is needed for equilibration after large changes in heart rate

Another key observation we made was that each device reacts differently to rapid, large changes in heart rate, as the reported HRs diverge and subsequently converge after they reach a plateau during the *exercise* phase of the experiment. From the HR measurements plots, we saw that the Fitbit Charge 2 especially showed a lag when transitioning from the *rest* to *exercise* phase when compared to the Polar devices (Figure 6). This effect was also observed in the Bland-Altman plots: Datapoints with large deviations predominantly occurred during the *exercise* phase. Consequently, we analyzed the effect of the experiment phase on the correlations.

For the Polar Vantage V3 there was no clear difference in correlation across the different phases, as for all phases the r_c was found to be between 0.96-0.99.

In contrast, the Fitbit Charge 2 had the lowest correlation during the exercise phase (r_c = 0.82), while the *rest* and *recovery* phase were in line with the Polar Vantage V3 (r_c = 0.91 and r_c = 0.94 respectively). We attribute this lower correlation to the lag observed after transitioning to the *exercise* phase, which might be a result of the lower recording frequency of the Fitbit Charge 2, or due to differences in the HR detection- and smoothing-algorithms of these devices.

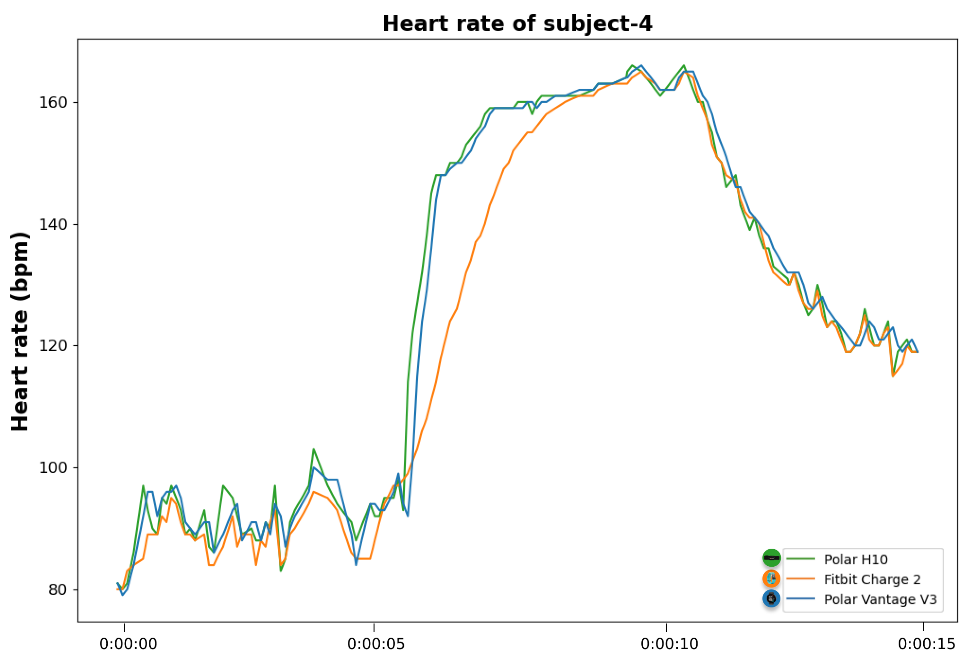

Figure 6: **Fitbit Charge 2 lags after transitioning from *rest* to *exercise* phase**

The Fitbit Charge 2 (orange) shows a bigger delay when transitioning from rest to exercise. The Polar H10 (green) and Polar Vantage V3 (blue) HR measurements are very similar, which might be explained by the potentially similar proprietary algorithms of Polar for measuring HR.

Discussion & Conclusion

In this study, we evaluated the integration of three different Polar devices and compared them to the previously integrated Fitbit Charge 2 into the RADAR-base platform. These devices, apart from the Polar Verity Sense which suffered from too many connectivity issues to be properly evaluated, all showed high correlation with the ECG-based measurements of Polar H10, our selected standard. Based on these results we believe each of these devices would be suitable for use in research.

Although we were able to successfully assess the quality of HR measurements of the Polar Vantage V3, Polar H10 and Fitbit Charge 2, this evaluation was not based on all the recorded data. As mentioned, 3 participants for which most of their HR measurements fell outside the 95% CI, were excluded from the analysis. While this exclusion of 23% of the participants is high, we believe this exclusion was necessary as we suspect that it may have been related to improper wearing of the devices.
This relates to another limitation, which is the experimental length. Since we recorded the HR of our subjects in a single setting of 15 minutes, one could argue that there was not enough time for subjects to really get acquainted with these devices. An improvement would therefore be to increase our experiment length, with the hope that such outlying HR measurements do not largely affect our results.
Another way to ensure that there are no outlying measurements that may be the result of user error, is to test these devices without the need of human subjects. Instead, ECG electrical signals and PPG signals could be simulated using a simulator such as the AECG100 (41), a device that can simulate ECG electrical signals and PPG signals simultaneously. By using simulated signals, environmental noise is eliminated, allowing for precise and accurate testing of wearable devices with less potentially confounding factors.

Another point of discussion is the use of the Polar H10 as our standard. Currently an ECG Holter instrument is reported to be the gold standard for recording HR in research, as it was found to be more accurate in calculating HR than PPG based measurements (1). Nevertheless, Gilgen-Ammann et al. demonstrated that the Polar H10 was just as accurate as the ECG Holter instrument during resting and low-intensity activities. Moreover, the Polar H10 provided a higher quality RR interval signal during more intense activities, leading to its recommendation as a gold standard for HR measurements (42). In this study, we therefore used the Polar H10 to record a baseline to compare the PPG-based Polar devices to, and indeed found that the data it produced was the most consistent of all tested devices.

In addition to the Polar H10, the PPG-based devices included in the analysis were also found to produce sufficient quality data to make meaningful conclusions. However, special care must be taken when making direct comparisons between devices, as the devices each have a different lag time, as shown in Figure 6. We believe this is the result of the devices each using a different algorithm to calculate HR. HR algorithms typically identify a specific 'signature' of a heartbeat, such as the R-peak of the QRS complex or the P-peaks in PPG signals. To eliminate noise in these HR signals, sophisticated filtering steps are employed, such as those in the Pan-Tompkins algorithm (20). As both Fitbit's and Polar's HR algorithms are proprietary, we can only speculate that differences between them are the cause of the observed differences, illustrating the need for open science. One known factor that could have contributed is the recording frequency. The Fitbit Charge 2 has a recording frequency 10 times lower than the evaluated Polar devices. Thus, the device likely needs more time to collect the number of measurements required to confidently smooth the signal, leading to longer lag times.

As mentioned, the Lin’s concordance correlation coefficients indicated a higher correlation between the Polar H10 and Polar Vantage V3 HR measurements (r_c = 0.98) compared to the Fitbit Charge 2 HR measurements (r_c = 0.91). Moreover, the correlation between Polar Vantage V3 and Fitbit Charge 2, both of which measure HR using a PPG sensor, was found to be 0.82 (Sup. Figure 5). This indicates that despite the Polar devices using different sensors to determine HR (ECG vs. PPG), the comparability between the two Polar devices is high.
It is possible that the proprietary algorithms of Polar use a similar approach in filtering HR data, which could explain the high similarity between their readings.
Despite this observation, it is important to realize that the correlation between all three devices is quite high. This demonstrates that all devices are individually good at measuring HR and could be suitable for use in a patient monitoring study. However, to minimize confounding factors arising from inherent differences between devices, we recommend comparing heart rate (HR) data collected using the same device.

When comparing our HR measurements obtained by the different wearable devices with previous research, many similarities can be found. For instance, Helmer et al. compared multiple consumer wearables, including the Fitbit Sense, equipped with PPG sensors, and reported high correlations ($r \geq 0.95$) between these devices and the clinical gold standard ECG (1). They concluded that these devices are promising for patient-monitoring of HR.

Nevertheless, other studies have indicated that wearable devices, such as the Fitbit Charge HR and the Apple Watch, tend to underestimate HR measurements (46,47). Furthermore, our observation that the wearable devices tend to disagree more during the *exercise* phase is something that has also been previously reported (44,45).

However, as the data quality of wearables grows, we are likely to reap the rewards of continuous remote monitoring in the form of digital biomarkers. Integrative analysis of multiple data streams will therefore play an increasingly important role, making platforms like RADAR-base invaluable parts of the research ecosystem. As observed by Sartor et al. we will need to carefully assess the quality of these devices in a structural manner to guide researchers in making the right choice for their study and avoid spending valuable resources integrating devices which do not add (enough) value (32). As differences between devices exist, such as their ease of use or the way in which the data is pre-processed, researchers should really consider which device will best suit their study design and needs. For instance, for large-scale population research, the accuracy of a device might be less critical compared to factors like cost and ease of use, especially if the objective is to collect as much data as possible. However, for studies focused on specific health conditions, the quality of the device, and consequently its price, becomes more significant.

To conclude, our results showed that both the Polar Vantage V3 and Fitbit Charge 2 highly correlated with our standard, the Polar H10, which showed very consistent patterns. Therefore, we consider all these devices to be suitable for research purposes. Nevertheless, we believe the Polar Vantage V3 to have an edge over the Fitbit Charge 2. Firstly, the Polar devices send data directly to the RADAR-base pRMT app, whereas the Fitbit Charge 2 stores data in the Fitbit Cloud, which can be a privacy concern. Moreover, the Polar Vantage V3 also showed the highest correlation with ECG measurements from the Polar H10 and can measure HR at a higher frequency than the Fitbit Charge 2. Subjectively, we also found the Polar Vantage V3 to be more comfortable for the participants and easy to use for the researchers, which is why we see its benefit over the Fitbit Charge 2.

Conflict of interest

The primary author of this research was employed by the Hyve B.V. for the duration of this research project. The implementation of the Polar devices on the RADAR-base platform, developed as part of this research, is intended to be used in services sold by the Hyve.

Privacy statement

The privacy scan, which confirmed adherence to GDPR principles, as well as the participant information letter and consent forms, are available upon request to the first author.

Data and code availability

The repository of RADAR-base can be found at: https://github.com/RADAR-base
The raw data generated during this study are not available as they have been destroyed to comply with data privacy and confidentiality regulations.
However, sample data, the scripts used in the analysis, the tested outcomes, and all generated plots are available as a Jupyter notebook in the following repository: https://github.com/fschulting/wearables_analysis

Acknowledgements

I would like to thank Bastiaan de Graaf for his guidance and supervision on this project, and the Hyve for providing the necessary resources to complete this project. I would also like to acknowledge the Hyve's collaboration with King's College London (KCL) in developing the RADAR-base platform. And lastly, I extend my gratitude to Polar for providing a discount on the devices used in this research.

References

1. Etiwy M, Akhrass Z, Gillinov L, Alashi A, Wang R, Blackburn G, et al. Accuracy of wearable heart rate monitors in cardiac rehabilitation. Cardiovasc Diagn Ther. 2019 Jun;9(3):262–71.

2. Mahon G. A Proposal for Strength-of-Agreement Criteria for Lin's Concordance Correlation Coefficient. In 2005 [cited 2024 Aug 6]. Available from: https://www.semanticscholar.org/paper/A-Proposal-for-Strength-of-Agreement-Criteria-for-Mahon/0bed3bdb75247d09d4c5f1398cd309272fd26ad5

3. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The Impact of Wearable Technologies in Health Research: Scoping Review. JMIR Mhealth Uhealth. 2022 Jan 25;10(1):e34384.

4. Rosa C, Marsch LA, Winstanley EL, Brunner M, Campbell ANC. Using digital technologies in clinical trials: current and future applications. Contemp Clin Trials. 2021 Jan;100:106219.

5. Vivalink. Vivalink. [cited 2024 Jul 10]. Digital Health Technologies for Remote Patient Monitoring. Available from: https://www.vivalink.com

6. ICON Digital Platform | ICON plc [Internet]. [cited 2024 Aug 12]. Available from: https://www.iconplc.com/solutions/technologies/icon-digital-platform

7. Empatica [Internet]. [cited 2024 Jul 10]. Empatica Health Monitoring Platform. Available from: https://www.empatica.com/platform/home

8. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, et al. RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices. JMIR Mhealth Uhealth. 2019 Aug 1;7(8):e11734.

9. Empatica [Internet]. [cited 2024 Jul 11]. Empatica Store | E4 wristband. Available from: https://www.empatica.com/store/e4-wristband

10. Kickstarter [Internet]. 2017 [cited 2024 Jul 11]. Pebble 2, Time 2 + All-New Pebble Core. Available from: https://www.kickstarter.com/projects/getpebble/pebble-2-time-2-and-core-an-entirely-new-3g-ultra

11. 101 Guide for Fitbit Charge 2 [Internet]. [cited 2024 Jul 11]. Available from: https://www.fitbit.com/global/nl/charge2/charge2-101

12. Biofourmis | Revolutionize the delivery of care [Internet]. [cited 2024 Jul 11]. Available from: https://www.biofourmis.com/

13. Bittium [Internet]. [cited 2024 Jul 11]. Available from: https://www.bittium.com/medical/bittium-faros

14. Gill SK, Barsky A, Guan X, Bunting KV, Karwath A, Tica O, et al. Consumer wearable devices for evaluation of heart rate control using digoxin versus beta-blockers: the RATE-AF randomized trial. Nat Med. 2024 Jul 15;1–7.

15. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol. 2020 Dec 22;76(25):2982–3021.

16. Omlor AJ, Trudzinski FC, Alqudrah M, Seiler F, Biertz F, Vogelmeier CF, et al. Time-updated resting heart rate predicts mortality in patients with COPD. Clin Res Cardiol. 2020 Jun 1;109(6):776–86.

17. Palatini P. Resting Heart Rate as a Cardiovascular Risk Factor in Hypertensive Patients: An Update. Am J Hypertens. 2021 Apr 20;34(4):307–17.

18. Böhm M, Schumacher H, Teo KK, Lonn EM, Mahfoud F, Ukena C, et al. Resting heart rate and cardiovascular outcomes in diabetic and non-diabetic individuals at high cardiovascular risk analysis from the ONTARGET/TRANSCEND trials. European Heart Journal. 2020 Jan 7;41(2):231–8.

19. Noble RJ, Hillis JS, Rothbaum DA. Electrocardiography. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations [Internet]. 3rd ed. Boston: Butterworths; 1990 [cited 2024 Jun 10]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK354/

20. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007 Mar;28(3):R1-39.

21. Tang Q, Chen Z, Ward R, Menon C, Elgendi M. PPG2ECGps: An End-to-End Subject-Specific Deep Neural Network Model for Electrocardiogram Reconstruction from Photoplethysmography Signals without Pulse Arrival Time Adjustments. Bioengineering (Basel). 2023 May 23;10(6):630.

22. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol. 2021 Aug;18(8):581–99.

23. Pan J, Tompkins WJ. A Real-Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering. 1985 Mar;BME-32(3):230–6.

24. Vandenberk T, Stans J, Mortelmans C, Van Haelst R, Van Schelvergem G, Pelckmans C, et al. Clinical Validation of Heart Rate Apps: Mixed-Methods Evaluation Study. JMIR Mhealth Uhealth. 2017 Aug 25;5(8):e129.

25. VirtuFit FB1.0i foldable exercise bike | GladiatorFit | Sport and Fitness Equipment [Internet]. GladiatorFit CH. 2024 [cited 2024 Aug 1]. Available from: https://www.gladiatorfit.ch/en/product/virtufit-fb10i-foldable-exercise-bike/

26. Ranjan Y, Chang J, Sankesara H, Conde P, Rashid Z, Dobson RJB, et al. RADAR-IoT: An Open-Source, Interoperable, and Extensible IoT Gateway Framework for Health Research. Sensors. 2024 Jan;24(14):4614.

27. RADAR-base/RADAR-Schemas [Internet]. RADAR-base; 2024 [cited 2024 Jul 5]. Available from: https://github.com/RADAR-base/RADAR-Schemas

28. RADAR-base/radar-commons-android [Internet]. RADAR-base; 2024 [cited 2024 Jul 4]. Available from: https://github.com/RADAR-base/radar-commons-android

29. RADAR-base/radar-prmt-android: Application to be run on an Android device to interact with the wearable devices & phone sensors for passive data streaming [Internet]. [cited 2024 Jul 4]. Available from: https://github.com/RADAR-base/radar-prmt-android

30. polarofficial/polar-ble-sdk [Internet]. Polar Electro; 2024 [cited 2024 Jul 4]. Available from: https://github.com/polarofficial/polar-ble-sdk

31. Apache Avro [Internet]. [cited 2024 Aug 12]. Apache Avro™ 1.11.1 Documentation. Available from: https://avro.apache.org/docs/1.11.1/

32. Sartor F, Papini G, Cox LGE, Cleland J. Methodological Shortcomings of Wrist-Worn Heart Rate Monitors Validations. J Med Internet Res. 2018 Jul 2;20(7):e10108.

33. Lin LIK. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. 1989;45(1):255–68.

34. Martin Bland J, Altman Douglas G. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. The Lancet. 1986 Feb 8;327(8476):307–10.

35. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009. 242 p.

36. McKinney W. pandas: a Foundational Python Library for Data Analysis and Statistics. In 2011 [cited 2024 Jul 24]. Available from: https://www.semanticscholar.org/paper/pandas%3A-a-Foundational-Python-Library-for-Data-and-McKinney/1a62eb61b2663f8135347171e30cb9dc0a8931b5

37. Van Der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Comput Sci Eng. 2011 Mar;13(2):22–30.

38. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020 Mar;17(3):261–72.

39. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9(3):90–5.

40. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12(85):2825–30.

41. WhaleTeq [Internet]. [cited 2024 Jul 9]. AECG100_Health Wearables Testing_Products. Available from: https://www.whaleteq.com/en/product.php?act=view&gad_source=5&gclid=EAIaIQobChMIoaXsouSYhwMVMiN7Bx2vIAyrEAAYASAAEgJ9pvD_BwE&id=21

42. 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 Jul 1;119(7):1525–32.

43. 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 Aug 30;22(17):6536.

44. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. Ann Intern Med. 2017 Apr 18;166(8):610–2.

45. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. J Sports Sci Med. 2016 Sep;15(3):540–7.

46. Kroll RR, Boyd JG, Maslove DM. Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study. J Med Internet Res. 2016 Sep 20;18(9):e253.

47. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of Wrist-Worn Heart Rate Monitors. JAMA Cardiology. 2017 Jan 1;2(1):104–6.

Supplementary materials

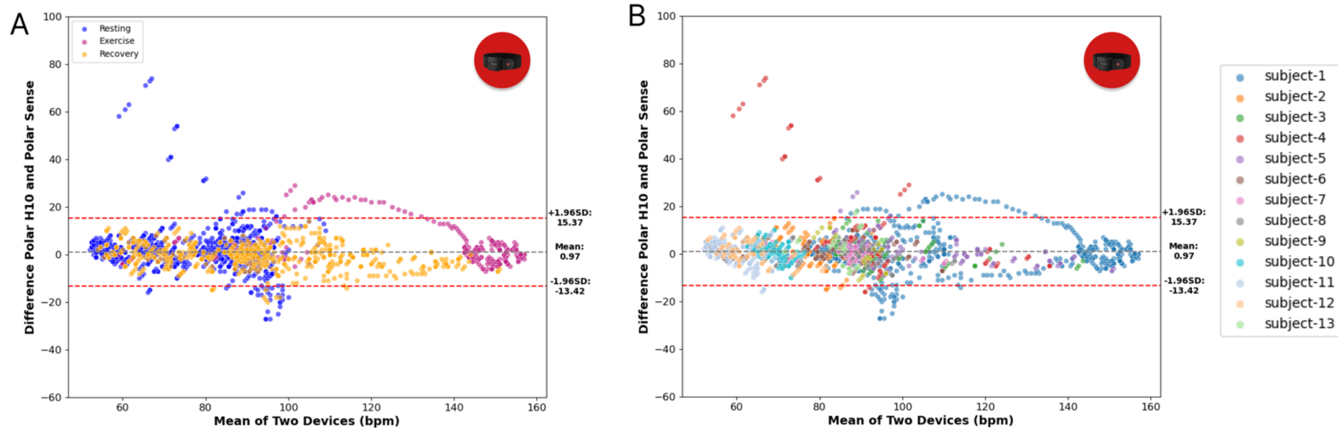

Figure S1: Connectivity issues of Polar Verity Sense led to large gap in HR measurements during *exercise* phase

Due to the connectivity issues of the Polar Verity Sense in 12 of the 13 participants, only from subject-1 data was recorded during the exercise phase. In total, only 238 measurements were recorded during the exercise phase, and 1176 and 708 during the rest- and recovery phase respectively. The BA-plot in panel A reveals this sparse number of datapoints during the exercise phase. From the BA-plot in panel B it became clear that all these measurements during exercise came from one subject (subject-1) only. We considered these measurements during rest and recovery to potentially be unreliable and therefore opted to exclude all measurements from further analysis.

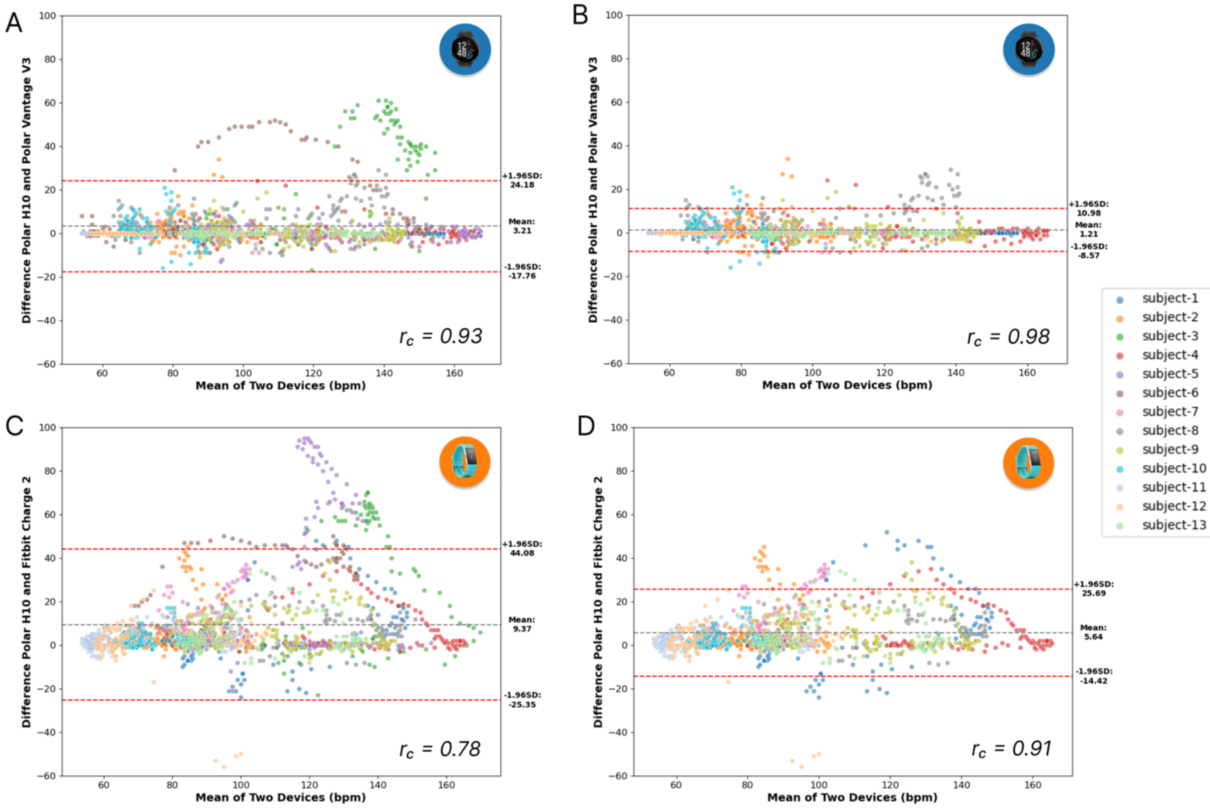

Figure S2: Bland-Altman plots showing the three excluded participants

The HR measurements of subject-3 (in green) and subject-6 (in brown) were outliers for Polar Vantage V3 (panel A), and participant subject-5 (in purple) for Fitbit Charge 2 (panel C) as most of their HR measurements fell outside the 95% CI. After exclusion of these participants, the mean bias and SD were more centered (panel B and D).

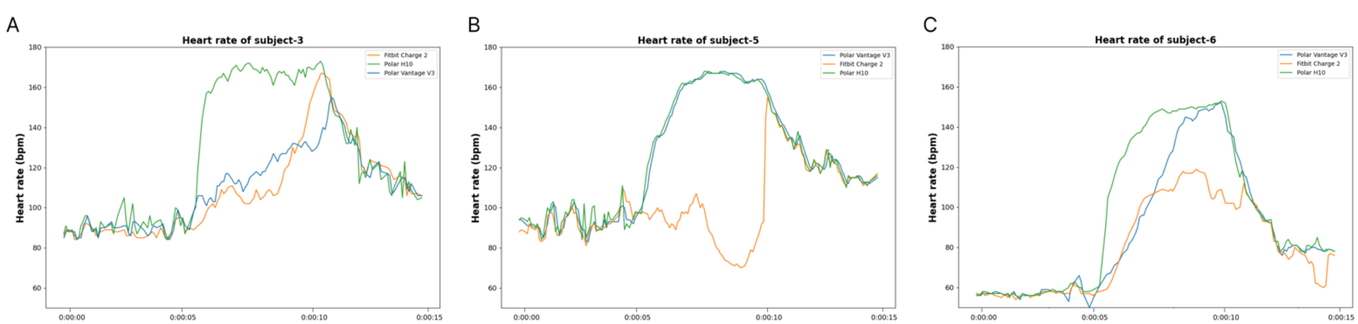

Figure S3: HR measurements of three excluded participants

Three participants were excluded based on the Bland Altman plots (Sup. Figure 2). The participants had either deviant HR measurements of the Fitbit Charge 2 (for subject-5 in panel B) or deviant HR measurements of both PPG devices (for subject-3 in panel A and subject-6 in panel C).

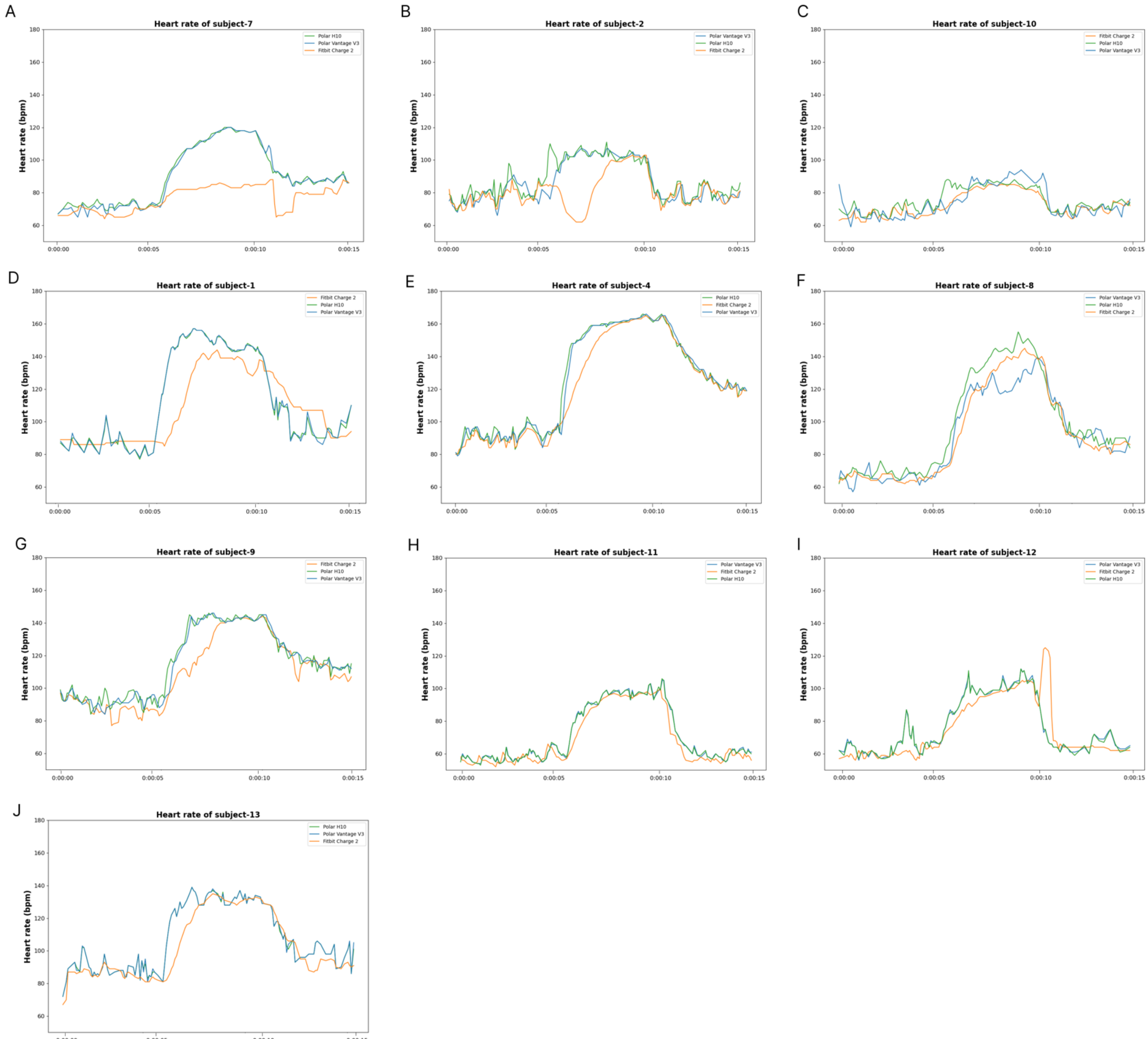

Figure S4: The HR measurements of all participants

For subject-7 (panel A) and subject-2 (panel B), the Fitbit Charge 2 was not properly measuring during the exercise phase. For subject-10 (panel C), both the Fitbit Charge 2 and Polar Vantage V3 had a lag when transitioning to exercise phase, in comparison to the Polar H10. For all other subjects (panels D to J), no large dispersions were observed in the individual HR plots.

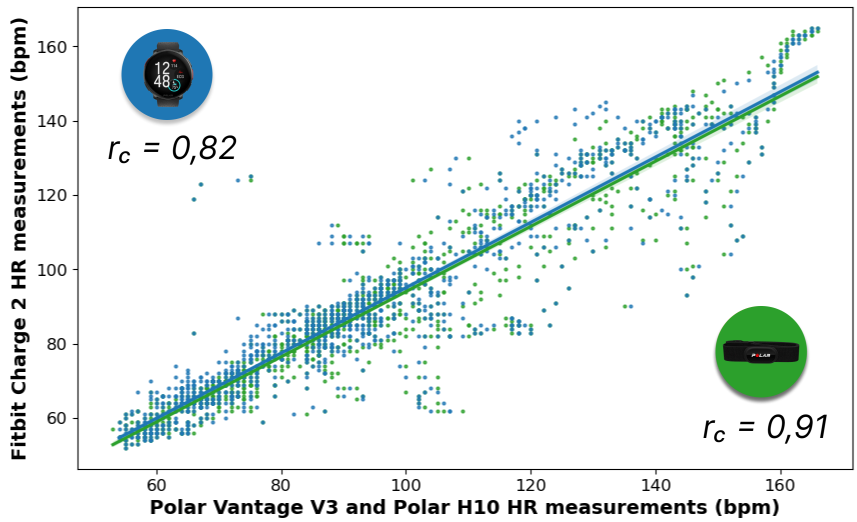

Figure S5: Correlations of Polar Vantage V3 and Polar H10 to Fitbit Charge 2

HR measurements of Polar H10 (green) correlate more to the Fitbit Charge 2 compared to HR measurements of Polar Vantage V3 (blue), even though the latter two both use a PPG sensor for their HR measurements.

Device	***Number of measurements***					***Freq. (Hz)***	***Measurement fidelity (%)***	***Total % excluded***
		Total	**Rest**	**Exercise**	**Recovery**			
Polar Verity Sense	Before	2122	1176	238	708	1.0	0.18	
	After	-	-	-	-	-		100
Polar Vantage V3	Before	11339	3880	3870	3589	1.0		
	After	1218	415	410	393	0.1	0.97	89.26
Polar H10	Before	11398	3900	3896	3602	1.0		
	After	1218	415	410	393	0.1	0.97	89.28
Fitbit Charge 2	Before	1648	553	564	567	0.1		
	After	1218	415	410	393	0.1	1.0	26.07

Table S1: Overview of total number of HR measurements

Polar Verity Sense recorded relatively fewer measurements compared to the other devices and was therefore excluded. Additionally, the data of three subjects were excluded since most of their measurements fell outside the 95% CI (Figure S2).
Moreover, the frequency of measurements was adjusted to match the frequency of the Fitbit Charge 2 device.

Device	***Activity***	***Lin's concordance correlation coefficient (r_c)***	***RMSE (bpm)***	***NRMSE***
Polar Vantage V3	*Full*	0.98	5.13	0.05
	Rest	0.96	3.78	0.08
	Exercise	0.96	7.26	0.07
	Recovery	0.99	3.38	0.03
Fitbit Charge 2	*Full*	0.91	11.68	0.10
	Rest	0.91	5.35	0.11
	Exercise	0.81	17.33	0.16
	Recovery	0.94	8.90	0.08

Table S2: Polar Vantage V3 correlates more to Polar H10 than the Fitbit Charge 2
The Lin's concordance correlation coefficient of the Polar Vantage V3 are for each activity phase higher than the ones of the Fitbit Charge 2. Furthermore, the RMSE and NRMSE were lower for Polar Vantage V3 HR measurements. Most disagreements in HR measurements were found during the exercise period.