Heart-to-Wear: Assessing the Accuracy of Heart Rate Sensor Measurements of Wearable Devices in Uncontrolled Environments

Simon Maximilian Wolf
University of Cologne
wolf@wim.uni-koeln.de

Patrick Seidel
University of Cologne
seidel@wim.uni-koeln.de

Tim Ockenga
University of Cologne
ockenga@wim.uni-koeln.de

Detlef Schoder
University of Cologne
schoder@wim.uni-koeln.de

Abstract

The growing popularity of wearable devices has enabled individuals to monitor their health and offers potential benefits for remote patient monitoring. However, the reliability of diagnoses provided by these non-approved medical devices remains uncertain. This study addresses the problem of assessing the measurement accuracy of heart rate recordings from wearable devices in both controlled and uncontrolled environments. Previous research has focused on evaluating accuracy in controlled settings, neglecting the impact of external factors on device performance. We conducted a comparative study with ten healthy individuals, recording heart rates during indoor cycling and outdoor activities. Participants wore two out of three tested smartwatches (Apple Watch Ultra, Garmin Enduro 2, Polar Pacer Pro) alongside a Polar H10 chest strap as a reference device. Our findings provide evidence that the Apple Watch Ultra and the Garmin Enduro 2 are particularly resistant to external factors that can occur during regular cycling activities.

Keywords: Wearable Devices, Measurement Accuracy, Heart Rate, Comparative Study

1. Introduction

In recent years, the number of connected wearable devices has increased rapidly and surpassed the one-billion-device mark for the first time in 2022 (Research and Markets, 2023). So far, there are numerous manufacturers on the market whose devices are designed to record and monitor a wide range of vital signs (Dunn et al., 2018; Khan et al., 2016). These devices not only provide users with valuable insights into their health status, but also enable medical staff to monitor their patients more closely and more frequently from a distance (Hall et al., 2014; Weenk et al., 2017). Applications range from simple biosensors to detect alcohol use disorder to highly complex smartwatches capable of accurately detecting complicated diseases such as atrial fibrillation or COVID-19 through their multitude of measurement metrics (Ates et al., 2021; Davis-Martin et al., 2021; Mishra et al., 2020; Perez et al., 2019). Nevertheless, since most of these devices are not approved medical devices, the reliability of their diagnoses is unclear so far.

In recent years, research has thoroughly investigated the measurement accuracy of various wearable devices, especially smartwatches. In addition to examining the accuracy of the step counter as well as the energy expenditure, the main focus is on evaluating the measurement accuracy of the heart rate recording of these devices (Germini et al., 2022; Henriksen et al., 2018). The experimental setup is always structured similarly, such that the devices are worn and subsequently evaluated in a variety of situations and at different physical stress levels for the test person (e.g. sitting, walking, jogging, cycling, etc.). In particular, various versions of the Apple Watch stand out in many studies due to their high measurement accuracy (Germini et al., 2022; Gillinov et al., 2017; Hajj-Boutros et al., 2023; Montalvo et al., 2022).

While these studies provide valuable insights into the accuracy of the devices under controlled conditions, they fail to account for the real-world variability and complexities of everyday life situations. They were all conducted under controlled conditions where participants performed steady-state aerobic exercises on machines like treadmills, stationary bicycles, or elliptical trainers (Martín-Escudero et al., 2023). Therefore, they may not accurately reflect the reliability of the devices in real-life scenarios and fail to generalize. To the best of the authors' knowledge, no study has specifically investigated how external factors encountered in everyday life can impact the measurement accuracy of wearable devices. For example, when wearing a wrist-worn device during outdoor biking, factors like bumps in the road or hand signals in traffic can disrupt the optimal position of the device on the skin. As these devices are used during various types of exercise, motion artifacts can introduce

URI: https://hdl.handle.net/10125/106768
978-0-9981331-7-1

unwanted interference or noise into the captured signal (Fine et al., 2021; Martín-Escudero et al., 2023). It is important to consider these factors to obtain a comprehensive understanding of the performance of wearable devices.

In this study, we tackle this research problem by evaluating the measurement quality of heart rate recordings from different wearable devices in a controlled environment against recordings in an uncontrolled environment while trying to answer the research question:

To what extent do external factors that occur in everyday situations influence the accuracy of heart rate measurement by wearable devices?

We recorded and evaluated the heart rate of ten healthy individuals (four female, six male) while cycling indoor on a stationary bike, as well as outdoor in public traffic. For each recording, the test persons were equipped with two out of three tested smartwatches (Apple Watch Ultra, Garmin Enduro 2, Polar Pacer Pro), as well as a Polar H10 chest strap, which was used as a reference device.

We found that the Apple Watch Ultra has the highest overall measurement accuracy, with an agreement of over 0.998, followed by the Garmin Enduro 2 and the Polar Pacer Pro. Our results show evidence that especially the Apple Watch Ultra and the Garmin Enduro 2 are very robust to external factors that occur during daily cycling activities. Combining findings from both controlled and real-world studies, allows for a more generalizable evaluation of the reliability and accuracy of these devices. This multidimensional approach helps to bridge the gap between laboratory research and real-life application.

2. Related Work

The investigation of the measurement accuracy of wearable devices has occupied the research community for several years. Germini et al. (2022) list various publications in their literature review that deal with the evaluation of the step counter, the heart rate sensor and the energy expenditure of various wrist-wearable activity trackers. Of the 65 studies examined, however, only nine dealt with the evaluation of heart rate measurement accuracy. Among the 15 devices from seven different manufacturers examined, the performance of the Apple Watch stands out as the best device with a Mean Absolute Percentage Error (MAPE) between one and seven percent.

Among the nine studies are those by Dooley et al. (2017) and Gillinov et al. (2017). The latter investigated the measurement quality of the heart rates of an Apple Watch, Garmin Forerunner 225 and Fitbit Blaze in their study. In addition to a standard electrocardiogram, they also used a Polar chest strap. Depending on the type of exercise (treadmill, stationary bicycle or elliptical trainer), the accuracy varied. The Apple Watch performed best, followed by the Garmin and the Fitbit watch, with all watches performing best on the treadmill and worst on the elliptical trainer.

Dooley et al. (2017) similarly examined the heart rate measurement accuracy of an Apple, Fitbit, and Garmin smartwatch among 62 participants (58% female) from different ethnic populations. The test persons' heart rates were evaluated during a 10-minute seated period, 4 minutes of treadmill exercise at light, moderate and high intensity level, and another 10-minute seated recovery period against a Polar T31 chest strap. Again, the Apple Watch performed best with a MAPE between 1.14% and 6.70% followed by the Fitbit with 2.38% and 16.99% and the Garmin with 7.87% and 24.38% depending on the intensity level. Furthermore, the Apple Watch tended to underestimate the heart rate in general while the Garmin Watch showed a higher heart rate in particular. Several other studies also found significant differences in the measurement accuracy of the respective smartwatches at different intensity levels (Hajj-Boutros et al., 2023; Kim et al., 2022; Martín-Escudero et al., 2023; Nissen et al., 2022).

However, all of the mentioned studies have been performed in a controlled environment (e.g. treadmill, stationary bicycle or elliptical trainer). Nevertheless, a study by Cosoli et al. (2022) investigates the measurement accuracy of wearable devices of swimming athletes. In their study, they evaluated the effect of movement artifacts and water on the heart rate measurement accuracy. They examined two smartwatches (Polar Vantage V2 and Garmin Venu Sq) against a Polar H10 chest strap during dry conditions on a treadmill and in the water while swimming. Their results show that arm movements during swimming, as well as the water itself, have a significant impact on the accuracy of the smartwatches. In fact, the average deviation per measurement point was -18 bpm for the Polar watch and -57 bpm for the Garmin watch. However, the influence of everyday disturbance factors on the measurement accuracy of wearable devices remains unclear so far.

3. Approach

3.1. Participants

Ten (four female, six male) healthy individuals consented to participate in this study (see Table 1).

All test persons are non-smokers and physically fit. In addition, none of the participants are known to be taking any medications regarding cardiovascular diseases. Furthermore, all persons are right-handed, so wearing the device on the left wrist is considered as non-dominant wrist and while the right wrist is considered as dominant wrist.

Table 1. Participants Characteristics.

Characteristic	$\mu \pm \sigma$	Value Range
Age (Years)	32.80 ± 15.03	[21; 58]
Height (cm)	170.80 ± 8.13	[160; 190]
Weight (kg)	68.20 ± 15.73	[48; 103]

3.2. Procedure

This study aims to assess the accuracy of heart rate measurements obtained from three distinct smartwatches. Each participant underwent a total of six 20-minute recordings. Three recordings were conducted indoor on a stationary bike, while the other three took place outdoor on a conventional bike. To ensure comparability, given the slight variations in the duration of participants' rides, we standardized the data by extracting a precise 20-minute segment from each participant's recording. The end time of such a segment is always 60 seconds before the recording was stopped on one of the devices. Accordingly, the start time is determined to be 20 minutes before the defined ending of the respective segment. Additionally, this method effectively eliminates transitional periods, specifically those occurring right after the devices were attached to the participant, and just before the devices were removed from them. In this way, we ensured that 100% of the examined time window was recorded during the actual ride.

During one of the recordings, the participants wore two smartwatches (one on the left wrist and one on the right wrist) in addition to the Polar H10 chest strap, serving as a reference device. The smartwatches were worn alternately, ensuring that exactly two indoor as well as two outdoor recording sessions were available for each smartwatch from each participant. For each of these two recordings, the respective smartwatch was worn once on the dominant wrist and once on the non-dominant wrist. In this way, we could analyze the effect of the hand dominance on the performance of the smartwatch by comparing the accuracy and efficiency of the data collected from both wrists. Prior to a recording, the age, gender, height and weight of the participants was specified in the settings of the watches, as well as the wrist on which the watch was worn during the recording. To establish a realistic scenario, participants were given no specific instructions regarding the direction or speed of their outdoor cycling sessions.

In our study, while we did evaluate wearable devices that track heart rates, our research design and protocol ensured that there was no personal or identifiable information collected from the participants. The devices were used solely to capture anonymous heart rate data. Furthermore, participants were informed about the purpose of the study, and their involvement was purely voluntary. No interventions or treatments were applied, making the risk to participants negligible. We designed our study according to the WMA Declaration of Helsinki (World Medical Association, 2022).

3.3. Devices

We choose an Apple Watch Ultra, a Garmin Enduro 2 and a Polar Pacer Pro as our testing devices. All three watches are from different manufacturers in the wearable devices market segment and represent the latest generation of commercially available smartwatches. While the Apple Watch Ultra and the Garmin Enduro 2 are high-priced watches, the Polar Pacer Pro is in the lower to mid-price segment. In addition, Apple is considered the market leader with a share of around 30% of global smartwatch shipments, while Garmin is listed as the fifth largest smartwatch manufacturer and Polar is relegated to the back of the market (Counterpoint Research, 2023). To the best of the authors' knowledge, no other study has examined any of these three devices at the time of conducting this study.

As a reference device we use the Polar H10 chest strap. It is considered and used as a benchmark in many studies (Cosoli et al., 2022; Hajj-Boutros et al., 2023; Montalvo et al., 2022). For all participants, we attached the chest belt according to the manufacturer's recommendations and slightly moistened the electrodes. All recordings of the corresponding smartwatches have been evaluated against the recording of the chest strap, where the heart rate recorded by the Polar H10 was treated as benchmark.

Reasons for this include the precise heart rate measurement accuracy of the chest strap, which has been demonstrated in several studies: Gilgen-Ammann et al. (2019) investigate in their study the RR interval signal quality of an electrocardiogram Holter monitor and a Polar H10 chest strap at rest and during activities. They observed that the signal quality of the Holter monitor decreased significantly with increasing activity level while that of the chest strap remained constantly

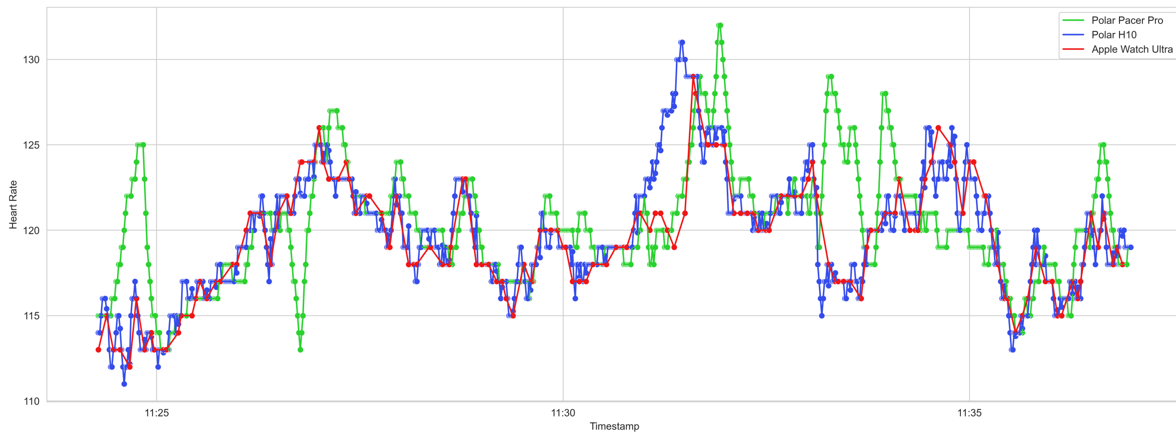


Figure 1. Excerpt of a sample indoor recording (Apple Watch Ultra left, Polar Pacer Pro right).

very high, which is why they recommend the use of such a chest strap as a reference device for future research. In their research, Montalvo et al. (2022) came to similar conclusions and also advocated the use of the Polar H10 chest strap if no electrocardiogram is available. Furthermore, Delgado-Gonzalo et al. (2015) provide evidence that the measurement accuracy of the chest strap is also robust enough for outdoor applications.

4. Results

4.1. Data Preparation

To evaluate the accuracy of each device's heart rate measurement, the agreement of each smartwatch with the Polar H10 chest strap was measured using several metrics. Since the Apple Watch is the only device that does not record the heart rate every second, we adjusted the benchmark of the chest strap to the Apple Watch for the evaluation. In doing so, we determined the average heart rate of the chest strap between the previous measurement of the Apple Watch and the current measurement for each measured value of the Apple Watch: If there is an Apple Watch measurement at time t, then the chest strap measurement against which we evaluate the Apple Watch is

$$X_t = \frac{1}{s}\sum_{i=0}^{s-1} X_{t-i},$$

where s is the number of seconds that have passed since the last Apple Watch measurement and X_{t-i} is the Polar H10 chest strap measurement i seconds ago. On average, the interval between two measurements is five seconds.

4.2. Evaluation

Since we compare the measurements of the other two smartwatches on a second-by-second basis with the chest strap, about five times more data points are evaluated for the Garmin and Polar watch than for the Apple Watch. Adjusting the two watches to the same level as the Apple Watch did not show a significant difference. An excerpt of a sample recording is depicted in Figure 1.

To determine the level of agreement between a watch and the Polar H10 chest strap, we calculated Lin's Concordance Correlation Coefficient (CCC) and performed a Bland and Altman (1986) analysis. Furthermore, we assess the agreement by the CCC according to McBride (2005) in four categories: Almost perfect (CCC $>$ 0.99), substantial (0.95 to 0.99), moderate (0.90 to 0.95) and poor (CCC $<$ 0.90).

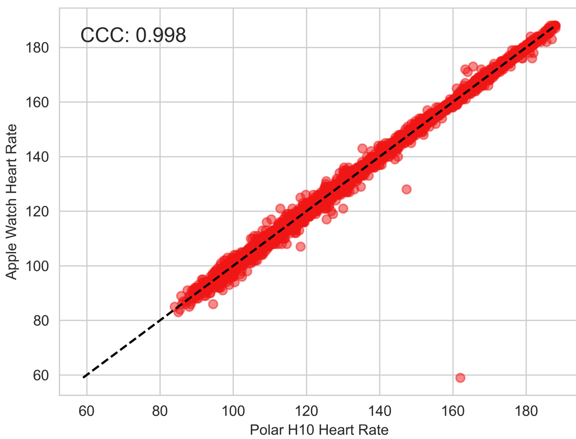
Overall, the Apple Watch Ultra has the highest agreement with the reference device. Both indoors and outdoors, it achieves an almost perfect correlation regardless of the side of the wrist on which it is worn (see Table 2).

The performance of the Garmin watch is similar to that of the the Apple Watch, but slightly inferior. The correlation with the chest strap is also almost perfect. Nevertheless, the dispersion is noticeably higher than that of the Apple Watch (see Figure 3 (a) and (b)). For example, the average difference of the Apple Watch is 0.01 and that of the Garmin Watch is -0.45. The standard deviation upwards and downwards is also significantly higher than that of the Apple Watch.

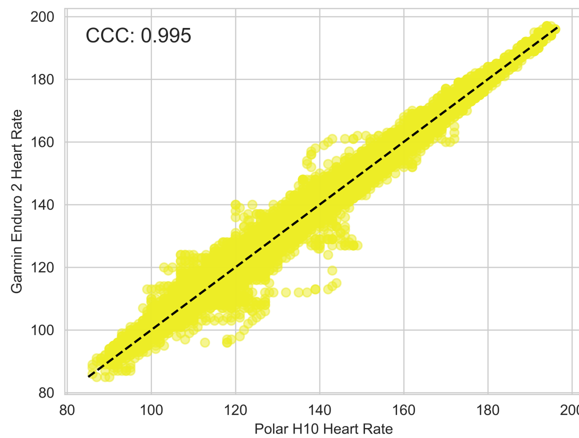
The Polar Pacer Pro performs significantly worse than the other two devices. Following the definition of McBride (2005) the performance can be classified as ”substantial”. Especially during the tests while the

Table 2. CCC between the three smartwatches and the Polar H10 chest strap in dependence of the different scenarios. Cells with a green background highlight high values, while yellow to red color schemes indicate a lower level of agreement.

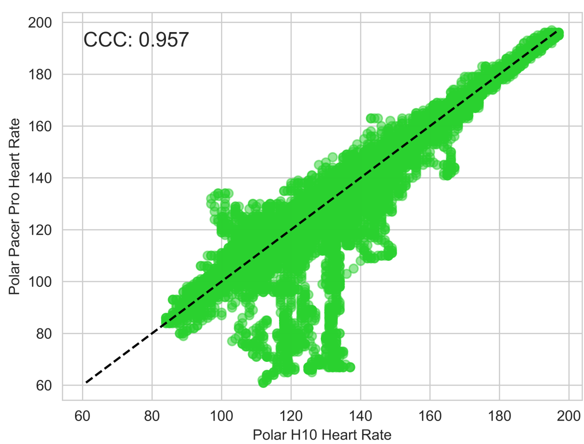
Device	Indoor		Outdoor		Overall
	Left	Right	Left	Right	
Apple Watch Ultra	0.999	0.994	0.998	0.999	0.998
Garmin Enduro 2	0.997	0.996	0.993	0.992	0.995
Polar Pacer Pro	0.990	0.863	0.989	0.968	0.957



(a) Apple Watch Ultra

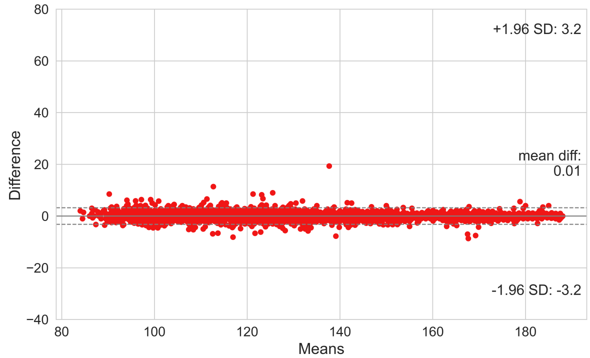


(b) Garmin Enduro 2

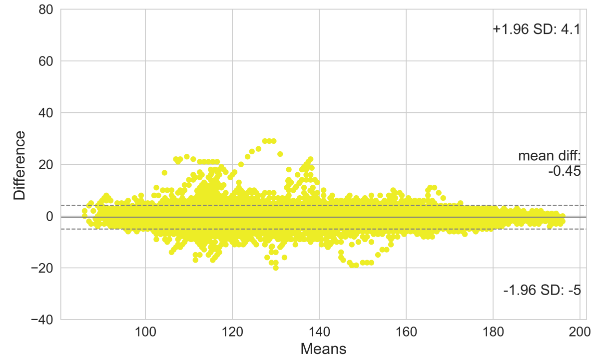


(c) Polar Pacer Pro

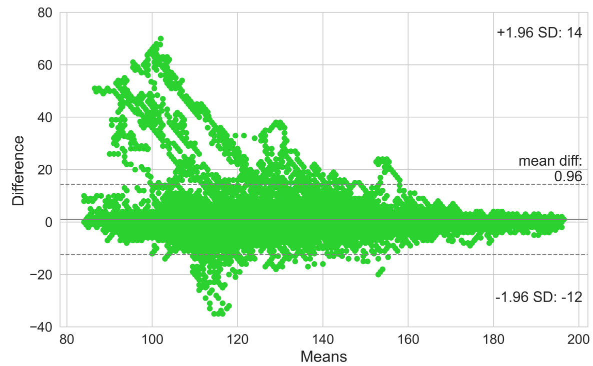
Figure 2. CCC overall agreement of the different smartwatches to the Polar H10.



(a) Apple Watch Ultra



(b) Garmin Enduro 2



(c) Polar Pacer Pro

Figure 3. Overall Bland-Altman analysis. Solid line represents the mean heart rate difference while dashed lines indicate the 95% confidence interval.

participants were sitting on a stationary bike, there are large differences in the measurements with regard to wearing the watch on the dominant or non-dominant wrist across all persons, such that the measurement accuracy on the dominant wrist can be described as poor. This effect can also be observed in a weakened form in the outdoor environment. Although the measurement accuracy of the Polar watch can generally be classified as ”substantial”, it still has an average deviation of one beat per minute per measurement and also a significantly increased variance. Especially in the range of low heart rates between 90 and 120 beats per minute, deviations of up to 70 beats occurred. Moreover, it was observed that strong deviations occurred particularly frequently during the first minute after the start of recording.

Except for the outdoor measurements of the Apple Watch, we find that the watches tend to perform worse (although usually only insignificantly) on the dominant wrist than on the non-dominant arm. This might be due to increased motion artifacts in the dominant arm (Fine et al., 2021), since the test persons were not prohibited from taking their hands off the handlebar during their ride, for example to take a drink or wipe their sweat with a towel. Since these movements occur more easily and more frequently on the stationary bike than outdoor in road traffic, this also explains why this effect is observed more strongly with the Polar Pacer Pro indoor than outdoor. Furthermore, other disturbing factors, such as bumps in the asphalt experienced when cycling outdoors, do not seem to have a significant effect on the watches’ measurement accuracy.

5. Discussion and Conclusion

The purpose of this study was to assess the measurement accuracy of heart rate recordings from different wearable devices in both controlled and uncontrolled environments, with a specific focus on external factors encountered in everyday life situations. By conducting a comparative study with ten healthy individuals and evaluating heart rate recordings during indoor and outdoor cycling, we aimed to bridge the gap between laboratory research and real-life application of these devices.

Our findings provide valuable insights into the measurement accuracy of these devices under real-world conditions. The Apple Watch Ultra and Garmin Enduro 2 demonstrated high correlation during regular cycling activities, both indoors and outdoors. These two smartwatches exhibited an almost perfect agreement with the reference chest strap, indicating their reliability in capturing accurate heart rate measurements. The Polar Pacer Pro also performed reasonably well, although it showed slightly lower agreement compared to the Apple Watch Ultra and Garmin Enduro 2. These results are consistent with previous studies that have highlighted the accuracy of Apple Watches in controlled settings.

While the Apple Watch Ultra and Garmin Enduro 2 boast commendable accuracy, their higher cost presents a significant barrier to their widespread adoption in clinical settings. Given that cost-effectiveness is often a priority in medical environments, medical staff may hesitate to use these devices for routine heart rate monitoring, especially when budget constraints are present. The Polar Pacer Pro, despite its marginally lower performance, might be perceived as a more viable option due to its presumably more affordable price point. Clinicians and decision-makers should weigh the trade-offs between the accuracy of heart rate measurements and financial feasibility when selecting wearable devices for patient care.

Importantly, our study contributes to the current literature by considering real-world variability and complexities that could affect the measurement accuracy of wearable devices. By incorporating outdoor activities with potential motion artifacts, such as bumps in the road or hand signals during cycling, we provide a more comprehensive evaluation of these devices. Our results suggest that unobserved factors that may occur during outdoor cycling activities and affect the accuracy of the devices by interfering with the optimal position of the device on the skin did not significantly affect the accuracy of the Apple Watch Ultra and Garmin Enduro 2. This finding suggests that these devices can be reliable options for individuals engaging in outdoor sports or activities with similar motion patterns.

Our results highlight the importance of conducting studies in real-life scenarios to obtain a comprehensive understanding of the performance of wearable devices. While controlled experiments provide valuable insights into device accuracy, they do not capture the variability and complexities of everyday life situations. Our study addresses this gap by evaluating devices in both controlled and uncontrolled environments, providing a more generalizable assessment of their reliability.

Our study is not free of limitations: Primarily, the utilization of a notably small sample size may lack the necessary statistical power to identify significant findings. Furthermore, although our study centered on heart rate recordings during cycling, it does not encompass a variety of everyday situations. As we did not explicitly measured external factors, we cannot ensure that our data encompass the full spectrum of variations that might affect heart rate measurement accuracy. For example, since we have recorded our data on warm summer days, it only includes a narrow temperature range. This limitation means that our findings may not reflect all possible everyday situations and should be taken with caution. Future research would benefit from examining additional activities and explicitly measuring a variety of external factors such as motion artifacts, ambient temperature, hydration levels, individual fitness levels, and even emotional states and analyzing their correlation with the accuracy of heart rate measurement from such wearable devices. This would contribute to provide a more holistic picture of the impact of external factors on the heart rate measurement accuracy of wearable devices.

In conclusion, this study highlights the importance of evaluating wearable devices in real-life situations to assess their measurement accuracy and reliability. Our findings show that both the Apple Watch Ultra and the Garmin Enduro 2 exhibited near-perfect performance during cycling activities. It is worth noting that such activities might plausibly introduce motion artifacts, even though they were not directly measured in this research. These devices, therefore, present a viable option for individuals aiming for accurate heart rate monitoring during outdoor sporting engagements. Furthermore, researchers might consider these devices as potential platforms for the development of state-of-the-art innovations in health applications, such as early detection of epileptic seizures or panic attacks. As wearable technology continues to advance, further research and development efforts should focus on addressing the limitations identified in this study to enhance the overall performance and usability of these devices in various real-life contexts.

References

Ates, H. C., Yetisen, A. K., Güder, F., & Dincer, C. (2021). Wearable devices for the detection of COVID-19. *Nature Electronics*, *4*(1), 13–14. https://doi.org/10.1038/s41928-020-00533-1

Bland, J. M., & Altman, D. (1986). Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Cosoli, G., Antognoli, L., Veroli, V., & Scalise, L. (2022). Accuracy and Precision of Wearable Devices for Real-Time Monitoring of Swimming Athletes. *Sensors*, *22*(13), 4726. https://doi.org/10.3390/s22134726

Counterpoint Research. (2023). Global Smartwatch Shipments Grow 12% YoY in 2022; Price Polarization Seen in Demand. Retrieved June 5, 2023, from https://www.counterpointresearch.com/global-smartwatch-shipments-grow-yoy-2022/

Davis-Martin, R. E., Alessi, S. M., & Boudreaux, E. D. (2021). Alcohol Use Disorder in the Age of Technology: A Review of Wearable Biosensors in Alcohol Use Disorder Treatment. *Frontiers in Psychiatry*, *12*, 642813. https://doi.org/10.3389/fpsyt.2021.642813

Delgado-Gonzalo, R., Parak, J., Tarniceriu, A., Renevey, P., Bertschi, M., & Korhonen, I. (2015). Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 430–433. https://doi.org/10.1109/EMBC.2015.7318391

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR mHealth and uHealth*, *5*(3), e34. https://doi.org/10.2196/mhealth.7043

Dunn, J., Runge, R., & Snyder, M. (2018). Wearables and the medical revolution. *Personalized Medicine*, *15*(5), 429–448. https://doi.org/10.2217/pme-2018-0044

Fine, J., Branan, K. L., Rodriguez, A. J., Boonya-ananta, T., Ajmal, Ramella-Roman, J. C., McShane, M. J., & Coté, G. L. (2021). Sources of Inaccuracy in Photoplethysmography for Continuous Cardiovascular Monitoring. *Biosensors*, *11*(4), 126. https://doi.org/10.3390/bios11040126

Germini, F., Noronha, N., Debono, V. B., Philip, B. A., Pete, D., Navarro, T., Keepanasseril, A., Parpia, S., Wit, K. d., & Iorio, A. (2022). Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature. *Journal of Medical Internet Research*, *24*(1), e30791. https://doi.org/10.2196/30791

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *European Journal of Applied Physiology*, *119*(7), 1525–1532. https://doi.org/10.1007/s00421-019-04142-5

Gillinov, A. M., Etiwy, M., Gillinov, S., Wang, R., Blackburn, G., Phelan, D., Houghtaling, P., Javadikasgari, H., & Desai, M. Y. (2017). Variable Accuracy of Commercially Available Wearable Heart Rate Monitors. *Journal of the American College of Cardiology*, *69*(11, Supplement), 336. https://doi.org/10.1016/S0735-1097(17)33725-7

Hajj-Boutros, G., Landry-Duval, M.-A., Comtois, A. S., Gouspillou, G., & Karelis, A. D. (2023). 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. *European Journal of Sport Science*, *23*(2), 165–177. https://doi.org/10.1080/17461391.2021.2023656

Hall, C. S., Fottrell, E., Wilkinson, S., & Byass, P. (2014). Assessing the impact of mHealth interventions in low- and middle-income countries – what has been shown to work? *Global Health Action*, *7*(1), 25606. https://doi.org/10.3402/gha.v7.25606

Henriksen, A., Mikalsen, M. H., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., & Grimsgaard, S. (2018). Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *Journal of Medical Internet Research*, *20*(3), e9157. https://doi.org/10.2196/jmir.9157

Khan, Y., Ostfeld, A. E., Lochner, C. M., Pierre, A., & Arias, A. C. (2016). Monitoring of Vital Signs with Flexible and Wearable Medical Devices. *Advanced Materials*, *28*(22), 4373–4395. https://doi.org/10.1002/adma.201504366

Kim, C., Kim, S. H., & Suh, M. R. (2022). Accuracy and Validity of Commercial Smart Bands for Heart Rate Measurements During Cardiopulmonary Exercise Test. *Annals of Rehabilitation Medicine*, *46*(4), 209–218. https://doi.org/10.5535/arm.22050

Martín-Escudero, P., Cabanas, A. M., Dotor-Castilla, M. L., Galindo-Canales, M., Miguel-Tobal, F., Fernández-Pérez, C., Fuentes-Ferrer, M., & Giannetti, R. (2023). Are Activity Wrist-Worn Devices Accurate for Determining Heart Rate during Intense Exercise? *Bioengineering*, *10*(2), 254. https : / / doi . org / 10 . 3390 / bioengineering10020254

McBride, G. B. (2005). A proposal for strength-of-agreement criteria for lin's concordance correlation coefficient. *NIWA client report: HAM2005-062*, *45*, 307–310.

Mishra, T., Wang, M., Metwally, A. A., Bogu, G. K., Brooks, A. W., Bahmani, A., Alavi, A., Celli, A., Higgs, E., Dagan-Rosenfeld, O., Fay, B., Kirkpatrick, S., Kellogg, R., Gibson, M., Wang, T., Hunting, E. M., Mamic, P., Ganz, A. B., Rolnik, B., … Snyder, M. P. (2020). Pre-symptomatic detection of COVID-19 from smartwatch data. *Nature Biomedical Engineering*, *4*(12), 1208–1220. https://doi.org/10.1038/s41551-020-00640-6

Montalvo, S., Martinez, A., Arias, S., Lozano, A., Gonzalez, M. P., Dietze-Hermosa, M. S., Boyea, B. L., & Dorgo, S. (2022). Commercial Smart Watches and Heart Rate Monitors: A Concurrent Validity Analysis. *The Journal of Strength & Conditioning Research*. https://doi.org/10.1519/JSC.0000000000004482

Nissen, M., Slim, S., Jäger, K., Flaucher, M., Huebner, H., Danzberger, N., Fasching, P. A., Beckmann, M. W., Gradl, S., & Eskofier, B. M. (2022). Heart Rate Measurement Accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: Device Evaluation Study. *JMIR Formative Research*, *6*(3), e33635. https://doi.org/10.2196/33635

Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., Balasubramanian, V., Russo, A. M., Rajmane, A., Cheung, L., Hung, G., Lee, J., Kowey, P., Talati, N., Nag, D., Gummidipundi, S. E., Beatty, A., Hills, M. T., Desai, S., … Turakhia, M. P. (2019). Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. *New England Journal of Medicine*, *381*(20), 1909–1917. https://doi.org/10.1056/NEJMoa1901183

Research and Markets. (2023). Connected Medical Device Market - Growth, Trends, COVID-19 Impact, and Forecasts (2023-2028). Retrieved June 1, 2023, from https : / / www . researchandmarkets . com / reports / 4622734 / connected-medical-device-market-growth

Weenk, M., Goor, H. v., Frietman, B., Engelen, L. J., Laarhoven, C. J. v., Smit, J., Bredie, S. J., & Belt, T. H. v. d. (2017). Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study. *JMIR mHealth and uHealth*, *5*(7), e7208. https://doi.org/10.2196/mhealth.7208

World Medical Association. (2022). WMA Declaration of Helsinki – Ethical Principles For Medical Research Involving Human Subjects. Retrieved August 29, 2023, from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/