Validation of Wearable Derived Heart Rate Variability and Oxygen Saturation from the Garmin's Health Snapshot

Kieran Williams[1], Alexandra Jamieson[1], Nishi Chaturvedi[1], Alun Hughes[1], Michele Orini[1]

[1] MRC Unit for Lifelong Health and Ageing, Institute of Cardiovascular Science, University College London, London, United Kingdom

Abstract

The new Garmin 'Health Snapshot' feature claims to measure resting heart rate (RHR), heart rate variability (HRV) and oxygen saturation (SpO_2) over a two-minute period, providing a "glimpse of overall cardiovascular status". This is the first study to investigate the accuracy of the feature in healthy adults (n=27, 63% male, mean ± SD age = 21.9 ± 6.7). Slower respiratory rates are known to increase HRV, therefore two respiratory rates (normal and controlled) were incorporated within the protocol. Reference measures for RHR and HRV metrics (RMSSD and SDNN) were derived from an electrocardiogram (ECG), whereas reference SpO_2 was determined using a Pulse Oximeter. Health Snapshot accuracy was quantified using Pearson's/Spearman's (cc_p/cc_s) correlation coefficients, Bland-Altman plots and mean absolute error (MAE). Health Snapshot estimations of RHR produced almost perfect correlation (0.99), MAE < 2% and narrow limits of agreement. Under normal breathing, both HRV metric estimations produced good correlation (cc_p>0.82). SpO_2 estimation was relatively poor with ~16.7% of Garmin estimations < 95%, despite all references ≥ 98%. HRV metric estimations were less accurate during controlled breathing, because wearable-derived HRV was slightly underestimated for lager HRV values.

1. Introduction

Consumer-grade wearables such as smartwatches have become increasingly popular in recent years, with the number of global users predicted to reach over 229.5 million by 2027. Cardiovascular disease (CVD) remains the leading cause of morbidity and mortality worldwide; however, wearables may be a novel solution to tackle this [1]. The main limitation associated with the clinical use of smartwatches is the uncertainty regarding the accuracy of their measurements. Cardiovascular (CV) healthcare is very likely to benefit from the validation of consumer-grade wearables, given many devices measure CV-system related metrics such as resting heart rate (RHR), heart rate variability (HRV) and peripheral oxygen saturation (SpO_2) [2]. Most wearable devices achieve this using photoplethysmography (PPG), an optical signal that detects small variations in blood volume caused by each heartbeat using the intensity of light reflected from the skin capillaries, which is proportional to the blood volume and its light absorption. Nevertheless, PPG is susceptible to various sources of error including motion artifacts, darker skin tone and obesity [3], [4].

Garmin 'Health Snapshot' is a new feature available on a variety of Garmin devices which claims to measure RHR, HRV and SpO_2 in a two-minute period. Users are instructed to "sit comfortably and hold still" for the duration of measurements, presumably to provide optimal measurement conditions [4].

RHR is a valuable metric clinically. There are strong associations between RHR and CVD morbidity and mortality; higher RHRs increase the relative risk of myocardial infarction (MI), stroke and heart failure, as well as both abnormally low and high RHRs associated with atrial fibrillation (AF) [5], [6]. Heart rate variability (HRV) is also important, functioning as a marker of the autonomic nervous system (ANS) [7]. HRV measures the variation in the interval between successive heartbeats, specifically consecutive R-R intervals on an electrocardiogram (ECG) [7]. The parameters of HRV used in our study included the "standard deviation of all normal R-R intervals" (SDNN) and "root mean square of successive differences between adjacent R-R intervals" (RMSSD). RMSSD represents short-term vagal-changes in HRV, reflecting "beat-to-beat" variations in HR however, SDNN is a longer-term metric influenced by both (sympathetic and parasympathetic) branches of the ANS [7]. It is well-established that low HRV is associated with an increased risk of CV events [6], [8]. Finally, arterial oxygen saturation (SaO_2) can be estimated non-invasively using "dual-wavelength PPG" to produce SpO_2 values. The global Covid-19 pandemic appears to have prompted the incorporation of SpO_2 measurements into many smartwatches, however the accuracy of many is unknown with very few devices appearing to have published validation studies [9].

2. Methods

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.237

2.1. Study design and protocol

Twenty-seven healthy adults between the ages of 18 and 54 were recruited to participate in our observational study (17 (63%) male, mean ± SD age = 21.9 ± 6.7 years). Age, sex, height, and weight were self-reported by study participants. The Fitzpatrick scale (FPS) classification was applied to participants in order to quantify skin tone [10] (Table 1).

Table 1. Baseline characteristics. Data reported as median [interquartile range] and frequency (%).

Characteristics	Results
Sex (male)	17 (63%)
Age (years)	20 [20-21]
Height (cm)	177 [172-182]
Weight (kg)	75.0 [62.0-78.8]
BMI (kg/m^2)	23.1 [21.0-25.5]
Fitzpatrick Scale	4 [2-6]

All measurements were conducted in a quiet room to provide a relaxed environment, optimal for recording quality. The study consisted of two measurement periods. The first period had participants breathe normally, while the second period required participants to control their breathing (breaths every five seconds). The slower, controlled breathing was used to investigate the effect of increased respiratory sinus arrhythmia (RSA) on Health Snapshot accuracy. Health Snapshots, ECG recordings and pulse oximetry took place simultaneously within the study. Health Snapshots were recorded via a Garmin Venu 2S GPS smartwatch (Garmin Ltd, KS, USA), worn by subjects on their left wrist. ECG recordings were conducted using a 3-lead 180° eMotion Faros ECG device (Mega Electronics, Finland). Finally pulse oximetry readings were obtained using an industry standard pulse oximeter (Omron, UK) placed on the right index finger of subjects.

SpO_2 readings were recorded from the pulse oximeter at the start of measurement periods and the ECG recording was started twenty seconds before the Health Snapshot to allow the ECG trace to stabilise. Health Snapshot data was saved and automatically synced to an iPhone 13 Pro (Apple, Cupertino CA, USA) via the Garmin Connect app. The procedure was replicated for measurements under controlled breathing conditions however, participants were verbally prompted to take breaths every five seconds.

2.2. HR and HRV analysis

Raw ECG traces were exported from the Faros ECG device to a MacBook Pro, 13 inch, 2020 (Apple, Cupertino CA, USA) as European Data Format (.EDF) files. Raw ECG traces were processed using a dedicated Graphical User Interface (EPMApp) in MATLAB R2022b (MathWorks Inc., Massachusetts, USA) and custom written MATLAB scripts. ECG R-wave peaks were labelled and visually inspected, allowing manual correction of any wrongly identified complex. Average HR was expressed in beats per minute. Indices of HRV were expressed in milliseconds. These were the standard deviation of N-N intervals (SDNN) and the root mean square of successive differences (RMSSD).

2.3. Statistical analysis

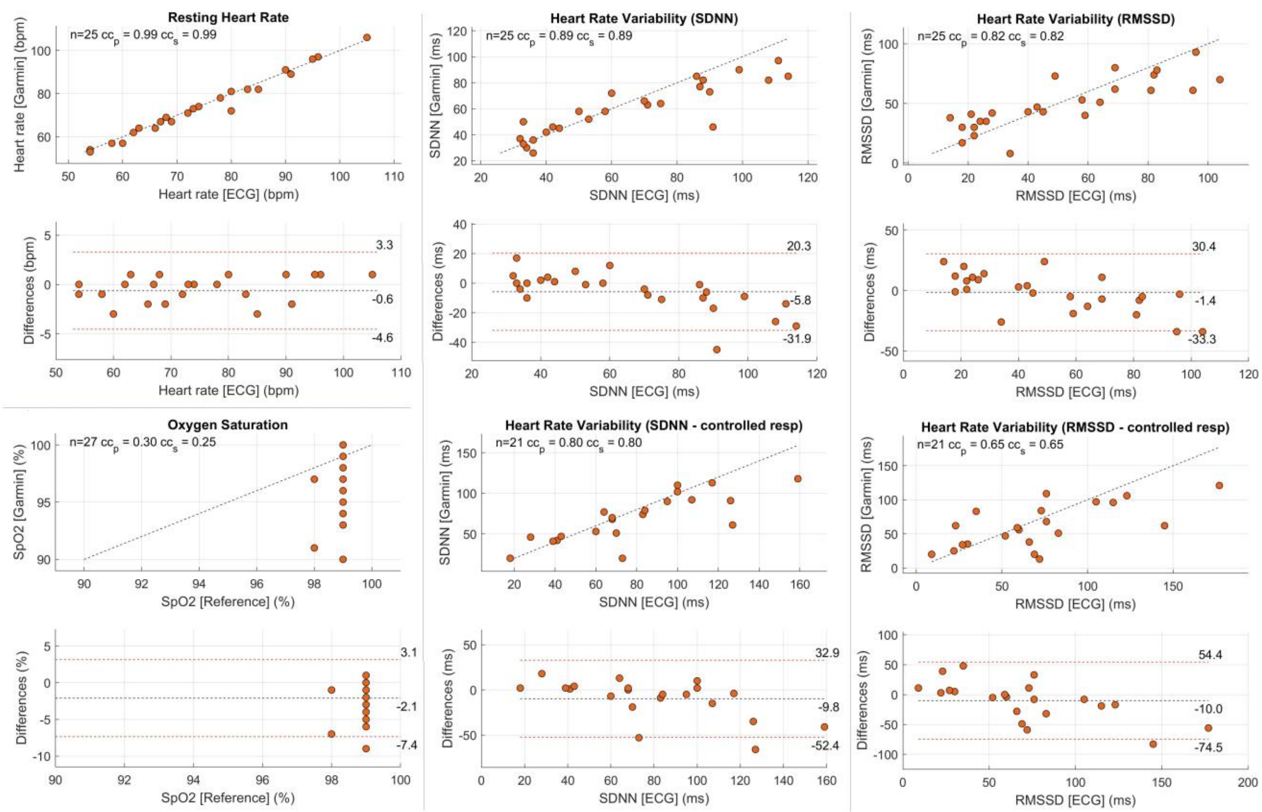
Baseline characteristics of participants were presented as median and interquartile range (IQR) for continuous variables, and frequency (%) for categorical data. Health Snapshot metrics were compared with references using Pearson's (cc_p) and Spearman's (cc_s) correlation coefficients, Bland-Altman plots to graphical represent agreement (mean bias and limits of agreement), mean absolute error (MAE) and mean absolute percentage error (MAPE). Correlation coefficients were interpreted according to recommendations used in *Theurl et al., 2023* [11]. Coefficients of <0.50, 0.50-0.75, 0.76-0.90 and >0.90 were interpreted as poor, moderate, good and excellent respectively. SDNN and RMSSD were analysed separately and according to breathing conditions when measured.

In sub-group analysis, we assess whether inaccuracies, measured as the absolute difference between estimates and reference values, differed by gender or skin tone using the Wilcoxon ranksum test. Skin tone was assessed using the Fitzpatrick Scale (FPS), which ranges from Type I ("pale white skin that always burns and never tans") to Type VI ("black skin that never burns, but tans"). For the purposes of our study, data was grouped into FPS ≤ 3 and FPS > 3. All p values for the above were measured using the Wilcoxon Rank Sum test. P values < 0.05 were considered statistically significant. All statistical analysis was completed using MATLAB.

3. Results

Reference and wearable derived RHR and HRV measures are reported in Table 2, while differences between estimated and reference measured are reported in Table 3. MAPE was below 2% for RHR, while it ranges between 14% and 44% for HRV metrics. Correlation and Bland-Altman plot (Figure 1) show a very strong correlation and narrow limits of agreement (LoA) for RHR, and a strong correlation (between 0.82 and 0.89) between Garmin's and reference HRV measurements in normal breathing. The correlation was slightly reduced and the LoA in Bland-Alman plots slightly increased for controlled breathing, mainly due to underestimation at larger HRV values (Figure 1). Correlation for SpO2 was poor, with 16.7% of

Figure 1: Correlation plots and Bland-Altman plots showing agreement between reference and Garmin's Health Snapshot measures. Diagonal lines represent the unity line; Black and red horizontal lines represent the bias and limits of agreement, respectively.



wearable's SpO2 measures <95% despite all reference values were >98%.

Table 2. Median (interquartile range) for RHR, RMSSD and SDNN measured by the Health Snapshot and derived from ECG signals under normal breathing conditions.

Parameter	ECG (Reference)	Garmin (Health Snapshot)
RHR [bpm]	72.0 (62.2-82.2)	71.0 (62.5-81.8)
RMSSD [ms]	49.0 (24.5-69.8)	47.0 (35.8-68.2)
SDNN [ms]	60.0 (40.8-87.8)	58.0 (45.2-80.8)

Table 3. Difference between estimated and reference measures.

Parameter	N	MAE ± SD	MAPE ± SD
RHR [bpm]	27	1.1 ± 1.4	1.6 ±1.8
RMSSD-nb [ms]	27	12.4 ± 9.6	33.0 ± 37.4
RMSD-cb [ms]	21	25.0 ± 23.1	44.3 ± 47.2
SDNN-nb [ms]	27	9.4 ± 10.1	13.9 ± 13.1
SDNN-cb [ms]	21	14.9 ± 18.5	18.3 ± 20.7
SpO_2 [%]	27	2.5 ± 2.3	2.5 ± 2.4

cb: Controlled breathing; nb: Normal breathing.

4. Discussion

This study is the first to validate the Garmin 'Health Snapshot' feature. We investigated the accuracy of Garmin estimations for RHR, HRV (RMSSD and SDNN) as well as SpO_2 against clinical references.

Wearable-derived RHR produced almost perfect correlation and excellent agreement with the ECG reference (Figure 1). This is consistent with findings in other smartwatch models [11]–[13]. Clinically, individuals at increased risk of MI, Stroke and heart failure may be able to be identified within the community, simply due to remote smartwatch measures. Further studies may assess whether higher RHR values measured by the Health Snapshot are also predictive for the development of CVD. Compared with RHR, wearables generally estimate HRV parameters with lower levels of correlation and agreement [11]–[13]. This observation was also true for our study however under normal breathing conditions there was good correlation and agreement for RMSSD and SDNN estimations overall (Figure 1). The results of our study showed superior levels of agreement compared to most devices validated previously [11]–[13]. Exceptions included the Garmin Vivoactive 4 (not compatible with Health Snapshot) used by *Theurl et al.* and WHOOP3.0 validated by *Miller et al.* however, in both cases, none of the values used by researchers were directly calculated or displayed by the devices [11]–[13].

Our study also appears to be the first to investigate the effects of RSA on smartwatch estimations. Remarkably, Health Snapshot accuracy decreased when breathing was controlled at slower respiratory rates (Figure 1). SDNN estimations appeared to be affected to a lesser extent, with reduction in accuracy disproportionately affecting RMSSD. This may be due to the short-term nature of

RMSSD, thus is relies on the accurate measurement of each beat and is more sensitive to errors in PPG signal peak detection [7], [11]. Notably, reductions in SDNN agreement appeared to be driven by outliers at higher HRVs, therefore estimations by the Health Snapshot may still prove useful for identifying lower HRVs, associated with the development of CVD later in life and poor CV outcomes in ongoing CVD [8].

The lowest levels of accuracy were seen in SpO_2 estimation compared with the Pulse Oximeter with poor correlation and agreement (Figure 1). 16.7% of SpO_2 estimations were <95% (associated with increased all-cause mortality), despite all participants being young and healthy with reference readings $\geq$ 98%. Despite various smartwatches incorporating SpO_2 measurements into their devices following Covid-19, the accuracy of most remains unknown due to limited validation studies [9]. Our study was limited by an entirely healthy study population that resulted in a limited range in baseline SpO_2 references. Sub-group analysis found no differences in estimation accuracy by Health Snapshot due to gender or skin tone for all metrics in the study.

4. Conclusion

Health Snapshot' can be considered a leading feature amongst smartwatches displaying HR and HRV directly, but SpO_2 measurement requires improvement. Despite reduced accuracy compared to RHR, agreement for SDNN was particularly encouraging at lower HRVs, clinically relevant when determining CVD risk. SpO_2 estimation produced poorer correlation and agreement with references, however results were likely impacted by an entirely healthy study population and small sample size.

References

[1] M. Amini, F. Zayeri, and M. Salehi, "Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017," *BMC Public Health*, vol. 21, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/S12889-021-10429-0/TABLES/3.

[2] K. Bayoumy *et al.*, "Smart wearable devices in cardiovascular care: where we are and how to move forward," *Nat. Rev. Cardiol.*, vol. 18, no. 8, pp. 581–599, Aug. 2021, doi: 10.1038/s41569-021-00522-7.

[3] J. Fine *et al.*, "Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring," *Biosens. 2021, Vol. 11, Page 126*, vol. 11, no. 4, p. 126, Apr. 2021, doi: 10.3390/BIOS11040126.

[4] M. Orini, G. Guvensen, A. Jamieson, N. Chaturvedi, and A. D. Hughes, "Movement , sweating , and contact pressure as sources of heart rate inaccuracy in wearable devices," *2022 Comput. Cardiol.*, vol. 498, pp. 1–4, 2022, doi: 10.22489/CinC.2022.232.

[5] J. Mensah-Kane *et al.*, "No clinically relevant effect of heart rate increase and heart rate recovery during exercise on cardiovascular disease: A mendelian randomization analysis," *Front. Genet.*, vol. 12, p. 59, Feb. 2021, doi: 10.3389/FGENE.2021.569323/BIBTEX.

[6] M. Orini *et al.*, "Long-term association of ultra-short heart rate variability with cardiovascular events," *Scientific Reports.*, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-45988-2.

[7] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Front. Public Heal.*, vol. 5, no. September, pp. 1–17, 2017, doi: 10.3389/fpubh.2017.00258.

[8] S. Hillebrand *et al.*, "Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: Meta-analysis and dose-response meta-regression," *Europace*, vol. 15, no. 5, pp. 742–749, May 2013, doi: 10.1093/europace/eus341.

[9] Z. Zhang and R. Khatami, "Can we trust the oxygen saturation measured by consumer smartwatches?," *Lancet Respir. Med.*, vol. 10, no. 5, pp. e47–e48, May 2022, doi: 10.1016/S2213-2600(22)00103-5.

[10] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Arch. Dermatol.*, vol. 124, no. 6, pp. 869–871, Jun. 1988, doi: 10.1001/archderm.124.6.869.

[11] F. Theurl *et al.*, "Smartwatch-derived heart rate variability: a head-to-head comparison with the gold standard in cardiovascular disease," *Eur. Hear. J. - Digit. Heal.*, vol. 4, no. 3, pp. 155–164, Jun. 2023, doi: 10.1093/ehjdh/ztad022.

[12] F. Sarhaddi *et al.*, "A comprehensive accuracy assessment of Samsung smartwatch heart rate and heart rate variability," *PLoS One*, vol. 17, no. 12 December, p. e0268361, Dec. 2022, doi: 10.1371/journal.pone.0268361.

[13] D. J. Miller, C. Sargent, and G. D. Roach, "A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults," *Sensors*, vol. 22, no. 16, p. 6317, Aug. 2022, doi: 10.3390/s22166317.

Address for correspondence:
Name: Kieran Williams
Address: University College London, Gower Street, London,
E-mail address: kieran.williams.20@ucl.ac.uk