ESC European Society of Cardiology

European Heart Journal - Digital Health (2023) **4**, 155–164
https://doi.org/10.1093/ehjdh/ztad022

ORIGINAL ARTICLE

Smartwatch-derived heart rate variability: a head-to-head comparison with the gold standard in cardiovascular disease

Fabian Theurl [1], Michael Schreinlechner [1], Nikolay Sappler [1], Michael Toifl[1], Theresa Dolejsi [1], Florian Hofer[1], Celine Massmann[1], Christian Steinbring[1], Silvia Komarek[2,3], Kurt Mölgg[2,3], Benjamin Dejakum[2,3], Christian Böhme[2], Rudolf Kirchmair[1], Sebastian Reinstadler [1], and Axel Bauer [1,*]

[1]Department of Internal Medicine III—Cardiology and Angiology, Medical University of Innsbruck, Anichstr. 35, Innsbruck 6020, Austria; [2]Department of Neurology, Medical University of Innsbruck, Anichstr. 35, Innsbruck 6020, Austria; and [3]Research Centre on Vascular Ageing and Stroke (VASCage), Anichstr. 5a, Innsbruck 6020, Austria

Received 22 December 2022; revised 14 March 2023; online publish-ahead-of-print 23 March 2023

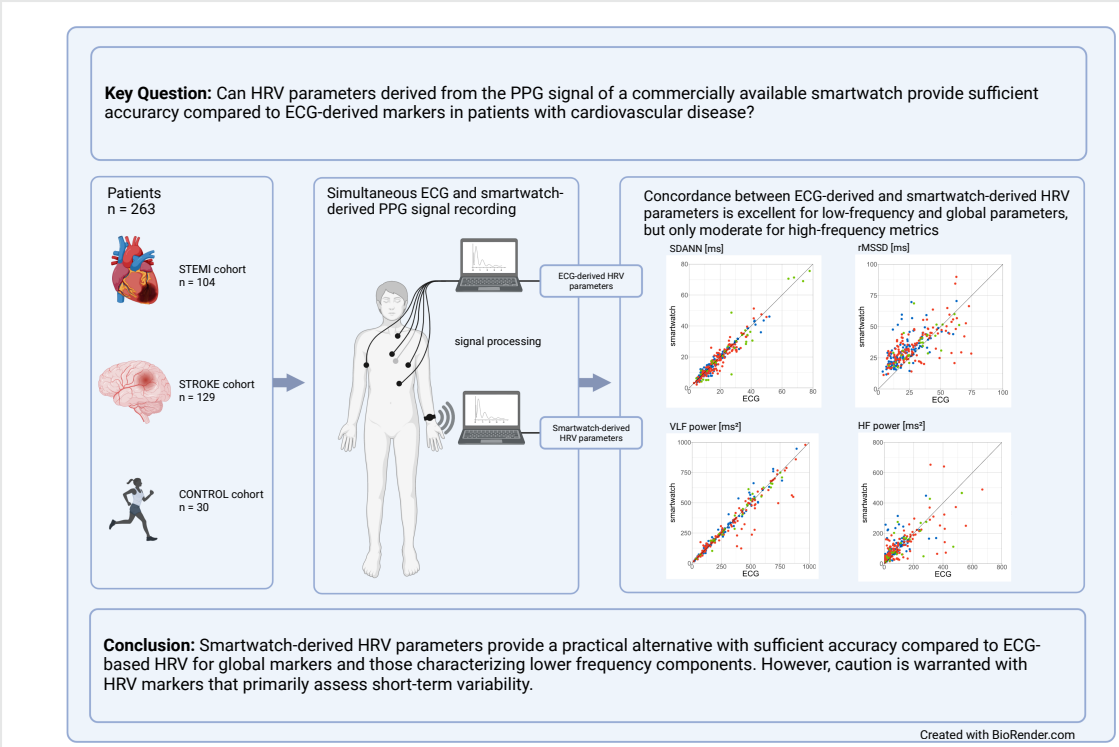
Aims We aimed to investigate the concordance between heart rate variability (HRV) derived from the photoplethysmographic (PPG) signal of a commercially available smartwatch compared with the gold-standard high-resolution electrocardiogram (ECG)-derived HRV in patients with cardiovascular disease.

Methods and results We prospectively enrolled 104 survivors of acute ST-elevation myocardial infarction, 129 patients after an ischaemic stroke, and 30 controls. All subjects underwent simultaneous recording of a smartwatch (Garmin vivoactive 4; Garmin Ltd, Olathe, KS, USA)-derived PPG signal and a high-resolution (1000 Hz) ECG for 30 min under standardized conditions. HRV measures in time and frequency domain, non-linear measures, as well as deceleration capacity (DC) were calculated according to previously published technologies from both signals. Lin's concordance correlation coefficient (ρ_c) between smartwatch-derived and ECG-based HRV markers was used as a measure of diagnostic accuracy. A very high concordance within the whole study cohort was observed for the mean heart rate ($\rho_c = 0.9998$), standard deviation of the averages of normal-to-normal (NN) intervals in all 5min segments (SDANN; $\rho_c = 0.9617$), and very low frequency power (VLF power; $\rho_c = 0.9613$). In contrast, detrended fluctuation analysis (DF-α1; $\rho_c = 0.5919$) and the square mean root of the sum of squares of adjacent NN-interval differences (rMSSD; $\rho_c = 0.6617$) showed only moderate concordance.

Conclusion Smartwatch-derived HRV provides a practical alternative with excellent accuracy compared with ECG-based HRV for global markers and those characterizing lower frequency components. However, caution is warranted with HRV markers that predominantly assess short-term variability.

* Corresponding author. Tel: +43 512 504 25621, Fax: +43 512 504-25622, Email: axel.bauer@i-med.ac.at

Graphical Abstract



Keywords Heart rate variability • Deceleration capacity • Smartwatch • Wearables

Introduction

Heart rate variability (HRV) reflects the variation of heartbeat intervals and is an established non-invasive tool to assess the functional status of the cardiac autonomic nervous system.[1] Dysfunction of the autonomic nervous system, often characterized by sympathetic overactivity and vagal withdrawal, plays a key role in the pathogenesis of various cardio- and cerebrovascular diseases and is therefore associated with an impaired prognosis.[2,3] Accordingly, HRV measures have been proposed to predict risk in patients with cardiovascular diseases, including myocardial infarction, chronic heart failure, diabetes mellitus, ischaemic stroke, and others.[4–7]

Starting from the traditional standard HRV measures in the time and frequency domain, a variety of non-standard measures[8] have been developed, including, among others, markers of short-term complexity,[9] entropy,[10] and Poincare' plot analysis,[11] deceleration capacity (DC),[12] and heart rate turbulence.[13] Although some of these measures have shown promising results as risk predictors in large clinical studies, HRV has received little adoption in clinical practice so far.

Even a 5 min electrocardiogram (ECG) recording allows HRV analysis.[1] For the purpose of risk prediction, however, HRV analysis is usually performed on longer ECG recordings, especially if HRV parameters are assessed that capture the lower-frequency components of HRV. This may limit the applicability of HRV in daily life, especially when HRV is intended to be used as a continuous monitoring tool in an outpatient setting. With certain limitations, HRV can also be calculated from photoplethysmographic (PPG) pulse-wave recordings.[14] The increasing penetration of wearables with integrated PPG sensors in the population could therefore make PPG-based HRV analysis using commercially available smart devices a promising future approach. Two small studies in healthy populations demonstrated the general feasibility of smartwatch-derived HRV markers.[15,16] However, the accuracy of PPG-based HRV derived from a smartwatch in larger populations suffering from cardio- or cerebrovascular diseases who might benefit from continuous risk monitoring is unknown.

Therefore, the aim of this work was to determine the accuracy of standard and non-standard HRV markers calculated from the PPG recordings of a commercially available smartwatch compared with ECG-derived HRV metrics in a relevant number of patients with cardio- and cerebrovascular diseases.

Methods

Study design and participants

We conducted a prospective, observational, single-centre study to evaluate the concordance of standard and non-standard HRV markers measured by a commercially available smartwatch compared with a standard, high-resolution ECG recording.

The study included three different cohorts presenting to our cardiologic outpatient clinic between June 2021 and July 2022: patients after ST-elevation myocardial infarction (STEMI cohort), patients after ischaemic stroke (STROKE cohort), and subjects without known structural

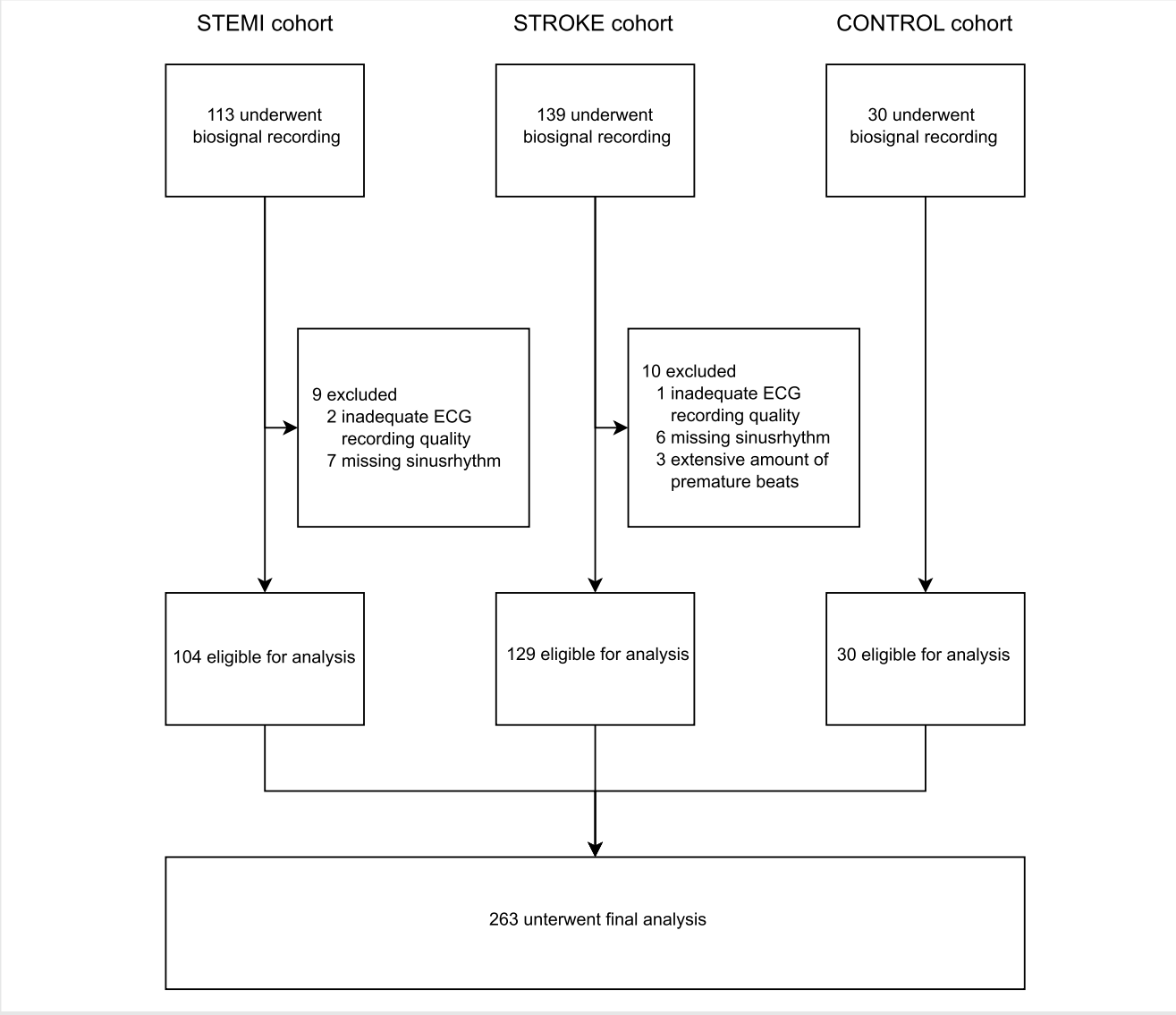


Figure 1 Consort diagram.

cardiovascular disease (CONTROL cohort). The study was approved by the local ethics committee in Innsbruck, Austria. All participants provided written informed consent.

Inclusion criteria for the STEMI cohort were patients presenting with their first STEMI treated with primary percutaneous coronary intervention within 24 h of symptom onset. Inclusion criteria for the STROKE cohort were patients suffering from a computed tomography– or magnetic resonance imaging–confirmed acute ischaemic stroke with symptom onset within 30 days prior hospital admission. The CONTROL cohort included subjects without known history of manifest structural cardiovascular disease. General inclusion criteria were age ≥18 years and presence of sinus rhythm. Subjects with recordings with frequent premature ventricular beats (>10% of total beats) or low quality were excluded (*Figure 1*).

Recordings and signal processing

All subjects underwent simultaneous ECG and PPG recording over 30 min. Recordings were made under standardized conditions in supine position and with spontaneous breathing. External stimuli (noise, etc.) were reduced to a minimum and patients were instructed to relax.

ECGs were recorded in Frank lead configuration (orthogonal, XYZ) using the Octal Bio Amp (ADInstruments, Dunedin, New Zealand) connected to the PowerLab 16/35 (ADInstruments, Dunedin, New Zealand) with a sampling rate of 1 kHz. The detection of QRS complexes was done using a Pan Tompkins–based algorithm[17] to obtain successive beat-to-beat intervals for each recording. All beat annotations were checked by an experienced technician and corrected, if necessary. In particular, ectopic beats were eradicated according to current measurement standards.[1] Exclusion of noisy segments was not necessary, as detection of QRS complexes was possible in all segments. ECG signal processing was done using a customized software (SMARTlab 1.5) developed with MATLAB (The MathWorks Inc., Natick, MA, USA). PPG signals were recorded using the PPG sensor of a Garmin vivoactive 4® smartwatch (Garmin Ltd, Olathe, KS, USA) worn by the subjects on the right arm. Proper fit of the smartwatches was checked by an experienced study nurse. Beat-to-beat intervals were calculated automatically using the built-in PPG signal processing software[18] without manual correction. Data were transferred to the signal processing unit by a customized Bluetooth interface.

Heart rate variability assessment

For all analyses, an adaptive threshold filter[19] was applied to the ECG- and PPG-based series of heartbeat intervals.

Table 1 Baseline characteristics

	Total ($n = 263$)	STEMI cohort ($n = 104$)	STROKE cohort ($n = 129$)	CONTROL cohort ($n = 30$)
Age (years)	61 (52–71)	59 (52–64)	67 (57–75)	44 (26–58)
Sex				
Male	192 (73%)	86 (83%)	90 (70%)	16 (53%)
Female	71 (27%)	18 (17%)	39 (30%)	14 (47%)
BMI (kg/m^2)	25.6 (23.4–28.2)	26.2 (24.1–28.7)	25.6 (23.7–28.1)	22.9 (21.1–25.0)
History of stroke	130 (49%)	1 (1%)	129 (100%)	0 (0%)
Ischaemic heart disease	130 (49%)	104 (100%)	26 (20%)	0 (0%)
History of myocardial infarction	118 (45%)	104 (100%)	14 (5%)	0 (0%)
Hypertension	153 (58%)	57 (55%)	91 (71%)	5 (17%)
Diabetes mellitus	40 (15%)	14 (13%)	26 (20%)	0 (0%)
Dyslipidaemia	195 (74%)	93 (89%)	95 (74%)	7 (23%)
Chronic kidney disease	16 (6%)	5 (5%)	11 (9%)	0 (0%)
COPD	10 (4%)	4 (4%)	6 (5%)	0 (0%)
PAD	6 (2%)	2 (2%)	4 (3%)	0 (0%)
LVEF (%)	56 (49–60)	50 (43–56)	59 (56–62)	—
NT-proBNP (ng/L)	678 (255–1460)	982 (542–1942)	157 (76–315)	—

Continuous data as median (IQR); categorical data as proportions (%).
BMI, body mass index; COPD, chronic obstructive pulmonary disease; LVEF, left ventricular ejection fraction; PAD, peripheral artery disease.

Standard HRV measures in time and frequency domain were calculated according to the Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology.[1] Time domain measures included mean heart rate, the standard deviation of normal-to-normal (NN) intervals (SDNN), the standard deviation of the averages of NN-intervals in all 5 min segments (SDANN), HRV triangular index (HRVi), and the square mean root of the sum of squares of adjacent NN interval differences (rMSSD). Frequency-domain measures included the powers in the very low frequency (VLF; 0.0033–0.04 Hz), low frequency (LF; 0.04–0.15 Hz), and high frequency (HF; 0.15–0.4 Hz) ranges. In addition, the HRV power contents in arbitrary overlapping segments of 0.01 Hz width were calculated from 0.01 to 0.4 Hz to assess association between targeted frequency and respective concordance.

Non-linear HRV metrics included detrended fluctuation analysis (DF-α1 and DF-α2) and Poincaré plots with calculated longitudinal and transversal standard deviation (SD1 and SD2).[9,11]

DC was calculated according to previously published technologies using phase-rectified signal averaging (PRSA).[12] In principle, heartbeat segments around instances of heart rate decelerations are rectified and averaged to obtain the so-called PRSA signal. The central part of the PRSA signal is quantified by Haar wavelet analysis. The PRSA algorithm allows several adjustments, such as varying the number of consecutive beat-to-beat intervals used to define heart rate decelerations. In the present study, we used two previously validated configurations: $t = 1$ (DC_{t1})[12] and $t = 4$ (DC_{t4}).[20]

The calculation of all markers was done in CRAN R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria). Rodriguez-Linares *et al.*'s package RHRV[21] was used to calculate standard HRV, DF-α1 and DF-α2, as well as SD1 and SD2. For the calculation of DC, a customized script developed by the authors was used.

Statistical analyses

Continuous data are presented as median (interquartile range) and compared using the Wilcoxon–Mann–Whitney rank sum test or as mean (standard deviation) and compared using *t*-test, as appropriate. Categorical data are summarized as numbers (frequencies; %) and compared using the χ^2 test. Concordance between PPG- and ECG-derived metrics was assessed using Lin's concordance correlation coefficient (ρ_c), intraclass correlation coefficient (ICC; two-way mixed effects model), mean absolute error (MAE), mean absolute percentage error (MAPE), and Bland–Altman plots with calculated bias and limits of agreement (LoA). The ICC was interpreted according to previous recommendations, with values <0.5, 0.5–0.75, 0.76–0.9, and >0.9 being indicative of poor, moderate, good, and excellent reliability, respectively.[22] In addition, HRV parameters were categorized into preserved and depressed HRV. The 25th percentile of each ECG-based HRV metric was used as a cut-off value. Cohen's kappa values were calculated to assess agreement, with kappa values of <0.2, 0.2–0.4, 0.41–0.6, 0.61–0.8, and >0.81, indicating slight, fair, moderate, substantial, and almost perfect agreement.[23] The correlation between powers in spectral frequencies and Lin's concordance correlation coefficient was assessed using Spearman's rank correlation coefficient (*r*). For all analyses, a two-sided *P*-value <0.05 was considered statistically significant. Statistical analyses and plots were created using CRAN R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Of the 282 subjects undergoing simultaneous ECG and PPG recordings, 263 subjects met inclusion criteria and were enrolled (*Figure 1*). Of these, 104 had STEMI, 129 had stroke, and 30 belonged to the CONTROL cohort. Of the overall cohort, median age was 61 (IQR 52–71) years and 71 (27%) were females. The median recording length was 32.1 (IQR 32.1–32.5) min. In the STEMI and STROKE cohorts, recordings were performed 3 (IQR 3–4) days and 5 (IQR 4–8) days after the index events, respectively. *Table 1* shows the characteristics of the study cohorts. Descriptive statistics of ECG- and smartwatch-derived HRV metrics are depicted in *Table 2*. HRV measures were significantly lower in the STEMI and STROKE cohorts compared with the CONTROL cohort ($P < 0.05$ for all, except for DF-α2). Poincare plots for each of the three cohorts are displayed in Supplementary material online, *Figure S1*.

Concordances between ECG- and PPG-derived HRV metrics are shown in *Figure 2* and Supplementary material online, *Table S1*. Highest concordances were observed for mean heart rate ($\rho_c = 0.9998$), SDANN ($\rho_c = 0.9617$), VLF power ($\rho_c = 0.9613$), and SD2 ($\rho_c = 0.9523$). In contrast, SD1 ($\rho_c = 0.6617$), rMSSD ($\rho_c = 0.6617$), and DF-α1 ($\rho_c = 0.5919$) showed the lowest concordances. For HF

Table 2 Descriptive statistics

	All		STEMI cohort		STROKE cohort		CONTROL cohort	
	Smartwatch	ECG	Smartwatch	ECG	Smartwatch	ECG	Smartwatch	ECG
HR (bpm)	64.9 (58.8–72.4)	64.8 (58.7–72.3)	64.7 (59.5–73.5)	64.7 (59.5–73.3)	64.8 (57.8–71.5)	64.8 (57.7–71.4)	67.2 (60.6–73.3)	67.1 (60.6–73.0)
DC_{t1} (ms)	6.3 (4.5–8.3)	5.9 (3.7–8.8)	5.8 (4.1–7.5)	5.6 (3.4–7.4)	6.4 (4.5–8.1)	5.8 (3.5–8.4)	8.6 (7.3–11.5)	9.6 (7.9–12.6)
DC_{t4} (ms)	7.0 (5.6–8.7)	6.6 (4.9–8.4)	6.6 (5.2–8.2)	6.2 (4.5–8.0)	7.0 (5.6–8.7)	6.5 (4.8–8.3)	8.3 (7.1–9.1)	7.8 (6.9–9.1)
SDNN (ms)	42.6 (30.9–52.0)	40.3 (27.9–51.4)	39.8 (29.0–47.1)	34.8 (26.9–45.4)	42.9 (34.3–53.0)	42.6 (32.9–54.1)	54.7 (39.3–75.0)	55.8 (36.0–81.8)
SDANN (ms)	16.6 (10.8–25.0)	17.0 (11.1–25.6)	13.7 (9.5–22.0)	14.2 (10.0–22.2)	17.9 (11.3–24.6)	18.3 (12.9–25.7)	24.7 (11.1–34.2)	26.1 (12.9–36.8)
HRVi	10.6 (7.9–13.1)	10.7 (7.6–13.8)	9.9 (7.5–12.4)	9.4 (7.3–12.5)	10.6 (8.0–13.1)	10.9 (8.1–13.7)	13.7 (10.0–16.6)	14.4 (10.9–18.7)
rMSSD (ms)	30.3 (24.1–41.7)	23.5 (15.1–35.6)	29.2 (22.0–35.8)	20.1 (13.1–28.4)	30.4 (24.6–43.3)	24.1 (16.2–39.8)	37.2 (27.0–49.7)	34.7 (20.6–46.5)
VLF power (ms)	256.5 (139.5–491.0)	261.6 (127.7–492.0)	221.9 (126.6–388.5)	225.0 (117.5–379.8)	260.1 (139.7–498.0)	270.5 (124.4–518.3)	446.7 (238.3–603.0)	450.8 (237.9–604.8)
LF power (ms)	147.4 (72.4–277.0)	110.1 (54.4–253.5)	112.3 (62.4–224.5)	84.2 (34.0–164.6)	158.9 (76.3–274.8)	129.8 (59.9–287.5)	333.9 (117.5–460.0)	379.6 (111.0–601.3)
HF power (ms)	77.6 (44.1–153.3)	59.5 (25.6–133.4)	66.1 (38.2–124.2)	40.9 (18.9–84.2)	84.5 (45.2–151.5)	66.5 (28.1–159.3)	162.7 (64.7–275.2)	140.4 (67.1–304.4)
SD1 (ms)	21.4 (17.0–29.5)	16.6 (10.7–25.2)	20.6 (15.6–25.3)	14.2 (9.3–20.1)	21.5 (17.4–30.6)	17.1 (11.5–28.1)	26.3 (19.1–35.2)	24.5 (14.6–32.9)
SD2 (ms)	55.5 (40.0–68.7)	53.7 (38.0–69.1)	50.8 (37.5–62.2)	47.2 (35.4–60.8)	55.6 (43.1–69.3)	55.9 (42.5–71.1)	70.5 (51.1–98.5)	75.7 (49.1–110.2)
DF-α1 (ms)	1.0 (0.9–1.1)	1.1 (0.9–1.3)	1.0 (0.9–1.1)	1.1 (1.0–1.3)	1.0 (0.9–1.1)	1.1 (0.9–1.3)	1.0 (0.8–1.0)	1.0 (0.9–1.1)
DF-α2 (ms)	0.9 (0.8–1.0)	0.9 (0.8–1.0)	0.9 (0.8–1.0)	0.9 (0.8–1.0)	0.9 (0.8–1.0)	0.9 (0.8–1.0)	0.9 (0.8–1.0)	0.9 (0.8–1.0)

DC, deceleration capacity; DF-α, detrend fluctuation analysis; ECG, electrocardiogram; HF, high frequency; HR, mean heart rate; HRVi, heart rate variability triangular index; LF, low frequency; rMSSD, square mean root of the sum of squares of adjacent normal-to-normal interval differences; SD1, transversal standard deviation in Poincare plot; SD2, longitudinal standard deviation in Poincare plot; SDNN, standard deviation of normal-to-normal intervals; SDANN, standard deviation of the averages of normal-to-normal intervals in all 5 min segments; VLF, very low frequency.

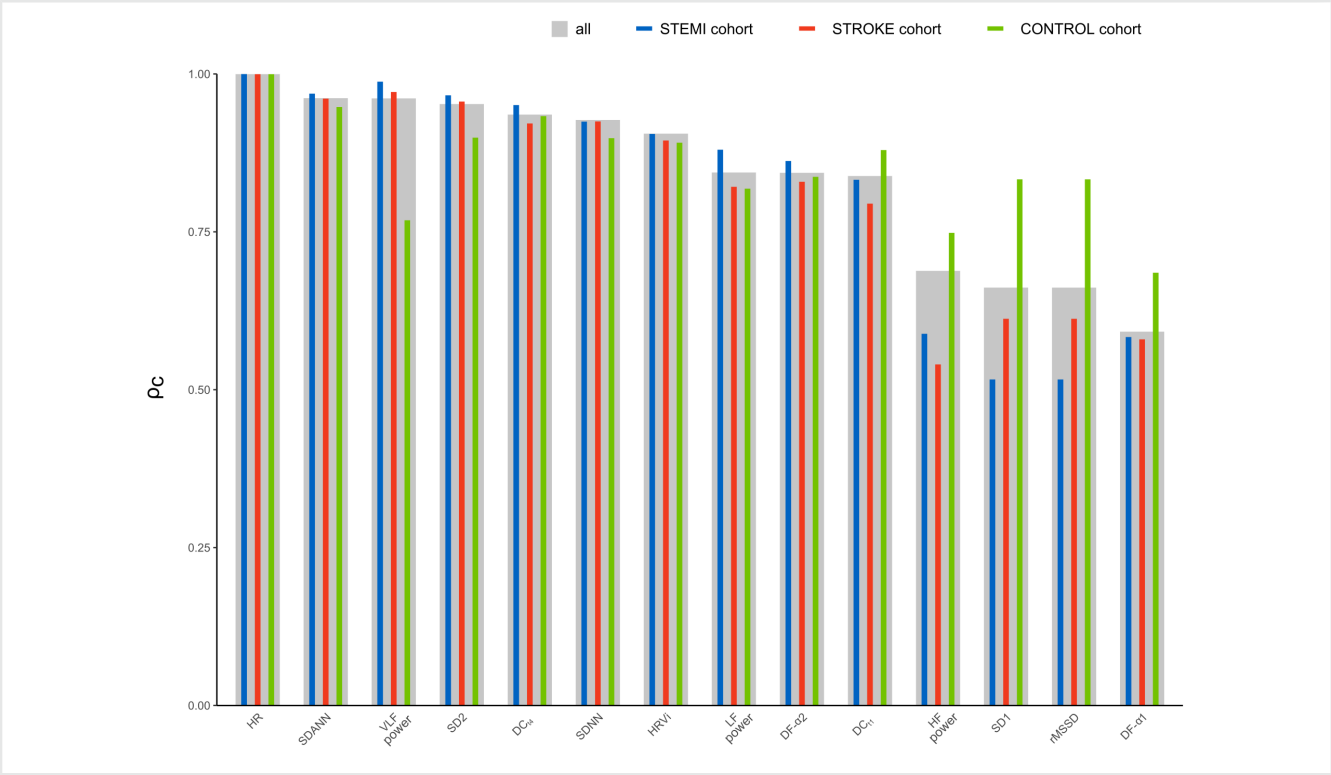


Figure 2 Agreement between ECG-derived and smartwatch-derived HRV metrics.

power, SD1, rMSSD, and DF-α1 concordances were lower in STEMI and STROKE cohorts compared with controls. For all other markers, concordances were similar throughout the three cohorts. Intraclass correlation coefficients ranged from 0.6318 for DF-α1 to 0.9998 for mean heart rate (see Supplementary material online, *Table S1*). Agreement was found to be excellent for 7 (50%), good for 3 (21%), moderate for 4 (29%), and poor for none of the markers. *Figure 3* displays Bland–Altman plots and Supplementary material online, *Figure S2* correlation plots for all parameters. No systematic deviation could be observed with biases close to zero for all analyses. Cohen's kappa values for distinction between preserved and depressed HRV are listed in Supplementary material online, *Table S2*. Almost perfect agreement was found for VLF and SD2, while DC_{t4}, SDNN, SDANN, HRVi, and DF-α2 exhibited substantial agreement. On the contrary, only slight agreement was found for rMSSD and SD1.

Additionally, concordance between ECG- and PPG-derived power content in arbitrary overlapping frequency bands were calculated, to evaluate a possible relationship between HRV spectral frequency and respective concordance. As shown in *Figure 4*, an inverse relationship between HRV spectral frequency and concordance between ECG- and PPG-derived HRV metrics was observed ($r = -0.94$, $P < 0.0001$). This trend was apparent in all three cohorts.

Discussion

The results of our study show that HRV measures obtained from the PPG signals of a commercially available smartwatch can have high agreement with those derived from high-resolution ECGs in patients with cardio- and cerebrovascular diseases as well as controls. However, this applies only to measures that either reflect global variability or quantify variability in the low frequency ranges. Lower concordance, especially in disease, is shown by HRV markers that quantify short-term variability in the higher frequency range.

Two previous studies compared smartwatch-derived HRV markers with ECG-derived markers. Hernando *et al.*[15] investigated the validity of the Apple Watch to estimate several time- and frequency-domain HRV measures in 20 healthy subjects. Reported concordance correlation coefficients on a beat-to-beat level were as high as 0.989, indicating excellent agreement. In contrast, Miller *et al.*[16] reported conflicting results when investigating the validity of HRV assessment by 6 different wearable devices, including 3 smartwatches, in 53 healthy subjects. Intraclass correlation coefficient in one of the 3 smartwatches studied (Garmin Forerunner®) was as low as 0.41 and 0.24 for estimating mean heart rate and rMSSD, respectively, indicating poor agreement. On the contrary, ICC for the 2 other smartwatches investigated were as high as 0.96 and 0.67 for the Apple Watch® and 0.93 and 0.65 for the Polar Vantage®. The results of our study (ICC > 0.99 and 0.69 for heart rate and rMSSD, respectively) were similar to those reported in the study by Miller for Apple Watch® and Polar Vantage®. Our study, however, differs from the aforementioned studies not only by a much larger sample size but also by several important other aspects. Patients in our study were significantly older (61 (IQR 52–71) years vs. 25.4 ± 5.9 years) and suffered from cardio- and cerebrovascular diseases. Since HRV may be impaired in cardiovascular disease, it is important to investigate the concordance between two HRV measurement modalities in these patients. And indeed, our results suggest that the agreement of smartwatch- and ECG-based HRV reflecting short-term variability is lower in patients after infarction or stroke compared with control subjects. Finally, HRV metrics in previous studies were limited to traditional HRV metrics, while our study evaluated a broader spectrum of HRV measures, including markers of short-term complexity, Poncairé plot analysis, and DC.

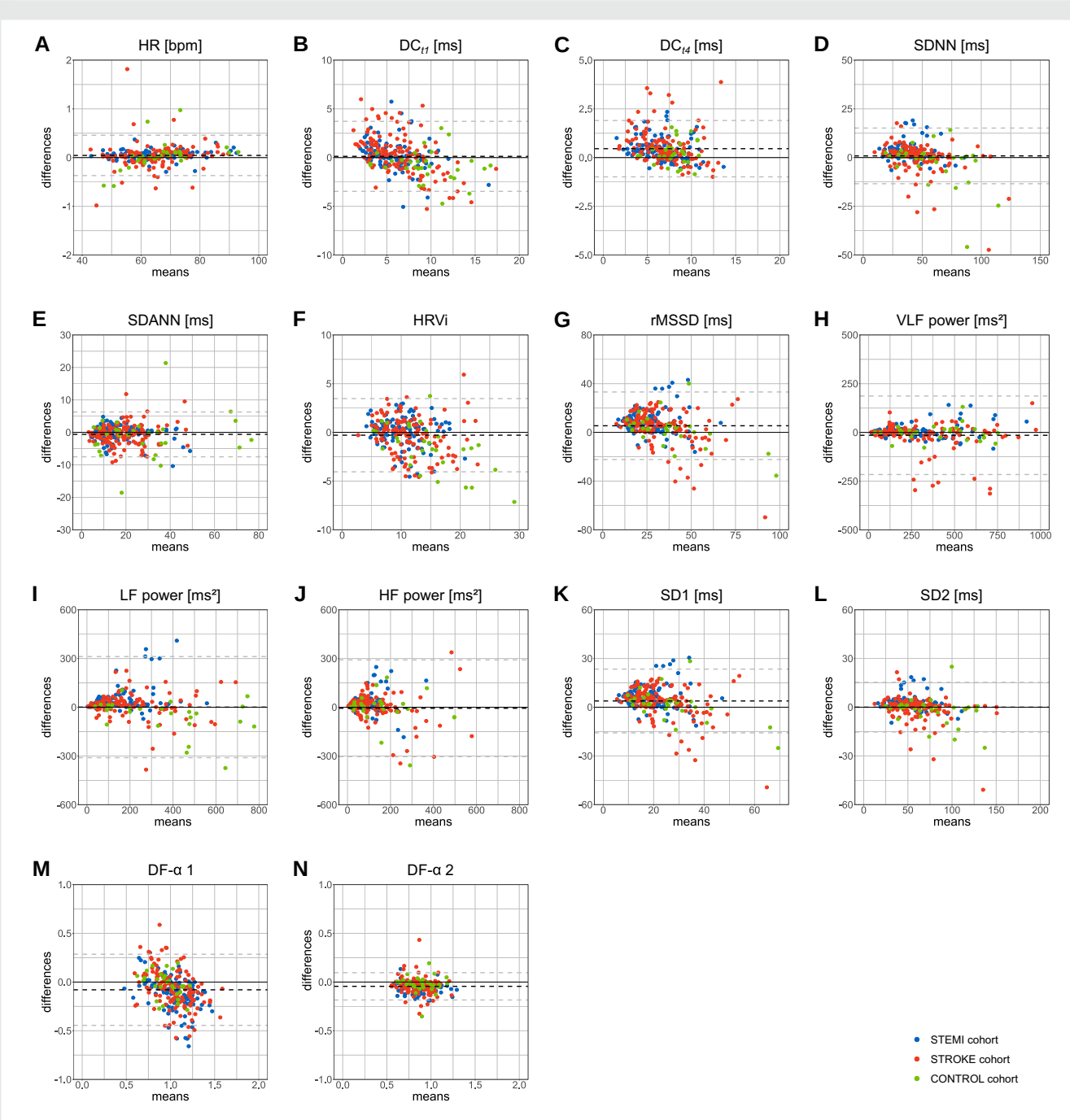


Figure 3 Bland–Altman plots with calculated bias (dashed black line) and limits of agreement (dashed grey lines) for mean heart rate (*A*), DC$_{t1}$ (*B*), DC$_{t4}$ (*C*), SDNN (*D*), SDANN (*E*), HRVi (*F*), rMSSD (*G*), VLF power (*H*), LF power (*I*), HF power (*J*), SD1 (*K*), SD2 (*L*), DF-α1 (*M*), and DF-α2 (*N*).

Basically, inaccuracies between smartwatch- and ECG-derived HRV markers can arise for several reasons. First, ECG and PPG measure different physiological signals with electrical activity and pulse waves, respectively. Based on the influence of various factors such as arterial stiffness and blood pressure on the pulse transit from the heart to the measurement site, practical consequences of HRV analysis were investigated in healthy volunteers already 30 years ago.[24] Thereby physiological discordances in pulse-wave-derived HRV analysis were described, affecting primarily the HF spectra. The same phenomena can now be observed three decades later in the results of our study in the new setting of wearable smart devices. Second, inaccuracies could be due to inaccurate peak detection. QRS morphology is almost identical within an ECG recording, and due to the R-peak's steep slope, its detection is accurate within a few milliseconds. On the contrary, peak detection within the PPG signal is limited by the variability and shallower slope of the signal compared to the ECG. Due to methodological reasons, beat annotations in the ECG signal could be visually inspected and corrected manually, while peak positions determined by the built-in algorithm of the smartwatch could not be revised. Third, some of the differences may be

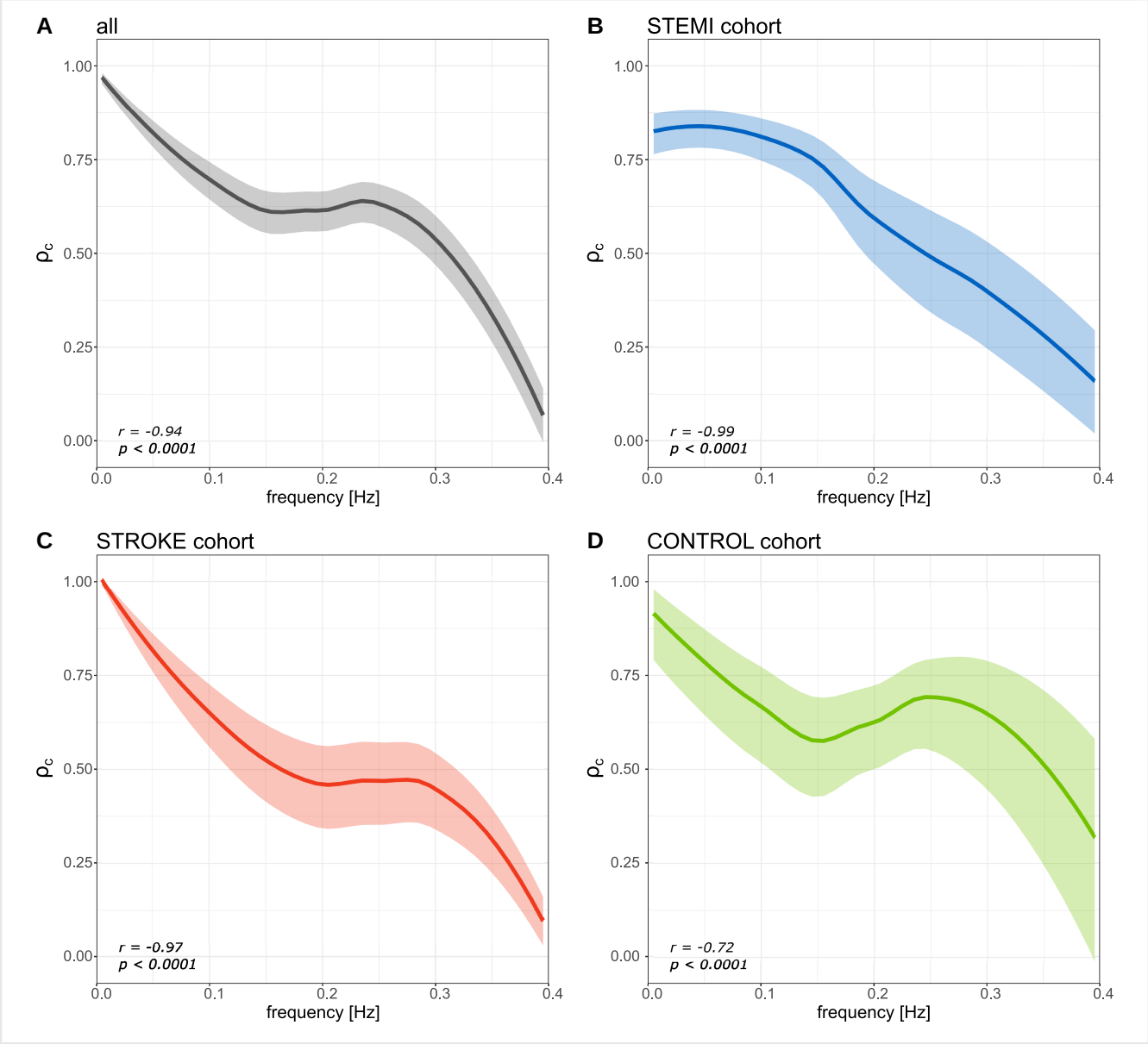


Figure 4 Concordance as a function of target frequency in HRV power spectra: concordance between ECG-derived and smartwatch-derived power content in arbitrary overlapping segments of 0.01 Hz width for all participants (*A*) as well as STEMI cohort (*B*), STROKE cohort (*C*), and CONTROL cohort (*D*) alone. *r*, Spearman's rank correlation coefficient; ρ_c, Lin's concordance correlation coefficient.

due to inaccuracies of the smartwatch's PPG sensor itself, caused by contact errors, lower sampling frequency, and others. Although these sources of error could affect all evaluated HRV markers, their influence on the respective concordance varies substantially. Markers reflecting short-term variability in the higher frequency ranges depend on accurate measurement of each single beat and are therefore more susceptible to single missing or inaccurate values. In contrast, markers quantifying variability in lower frequency ranges are more robust to these influences, resulting in generally excellent agreement. Therefore, these markers seem more promising for future implementation in smartwatch-based HRV technologies.

The implementation of HRV into commercially available wearables might have important clinical implications as it may extend the use of HRV to the outpatient setting and, in perspective, enable the vision of continuous risk assessment. This could benefit patients with chronic diseases and dynamic disease progression such as heart failure, myocardial infarction, or ischaemic heart disease, where early intervention in the subclinical stage is crucial. However, the results of our study show that the selection of the right parameter is crucial for this purpose. According to our results, clinical implementation of smartwatch-derived HRV should focus on low-frequency and global variability parameters, while caution should be exercised with high-frequency parameters. However, for the goal of risk stratification in cardiovascular patients, this does not imply a significant limitation. Numerous clinical studies could demonstrate high prognostic values for markers SDNN,[25] SDANN,[5] VLF,[25] LF,[25] DF-α2,[26] and DC,[12] for which good agreements were found in our study. However, future studies need to show which of these markers—either alone or in

combination—are best suited for smart-device-based continuous risk assessment.

The limitations of our study should be recognized. First, this study does not reflect a real-life scenario, since recordings were performed under standardized conditions in supine position at rest and the proper fit of the smartwatches was checked by an experienced study nurse. Thus, motion artefacts, contact errors, and external stimuli were reduced to a minimum and algorithms used in this study cannot be unmodified applied to data from real-life settings. Second, recordings were limited to 30 min. The predictive value of most HRV markers is better validated for longer recordings. Third, patients with conditions other than sinus rhythm and frequent premature complexes were excluded, which may limit the clinical use of HRV. For exploratory purpose concordance analysis without exclusion of recordings with multiple ectopic beats is depicted in Supplementary material online, Table S3. Fourth, the study examined PPG signals originating from only one smartwatch manufacturer. Although PPG signals from devices of different manufacturers are likely comparable, confirmation is needed on this.

Conclusions

In summary, our study provides sufficient evidence that calculation of HRV markers from a smartwatch's PPG signal is feasible in resting conditions. The concordance between smartwatch-derived and high-resolution ECG-derived HRV is excellent for low-frequency parameters but only moderate for high-frequency metrics, especially in patients with known cardio- or cerebrovascular diseases. Future studies are needed to validate these results in real-life settings. However, only prospective clinical studies will be able to prove the added value of a smartwatch-based HRV assessment.

Lead author biography

Fabian Theurl is currently working as a PhD student and first-year resident under the supervision of Prof. Axel Bauer at the Department of Cardiology, Medical University of Innsbruck, Austria. The research group focuses on risk stratification via non-invasive biosignal analysis with special focus on ECG-based markers of cardiac autonomic function/dysfunction and their transferability to smart wearable devices.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Funding

This study was supported by the excellence initiative VASCage (Research Centre on Vascular Ageing and Stroke, project number 868624) of the Austrian Research Promotion Agency FFG (COMET programme, Competence Centers for Excellent Technologies) funded by the Federal Ministry for Climate Protection, Environment, Energy, Transport, Innovation and Technology (Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie); the Federal Ministry for Labour and Economy; and the federal states Tyrol (via Standortagentur), Salzburg, and Vienna (via Vienna Business Agency). The trial was further supported by grants from the 'Austrian Society of Cardiology (Österreichische Kardiologische Gesellschaft)', 'Tiroler Wissenschaftsförderung', and 'Austrian Science Fund' [FWF grant KLI 772].

Conflict of interest: None declared.

Data availability

All data is included in the manuscript file.

References

1. Task Force of the ESC. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Circulation* 1996;**93**:1043–1065.
2. Corr P, Yamada K, Witkowski F, Fozzard H, Haber E, Jennings R, *et al. Mechanisms Controlling Cardiac Autonomic Function and Their Relation to Arrhythmogenesis*. New York: Raven Press; 1986.
3. Lown B, Verrier RL. Neural activity and ventricular fibrillation. *N Engl J Med* 1976;**294**: 1165–1170.
4. Kleiger RE, Miller JP, Bigger JT, Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol* 1987;**59**: 256–262.
5. Adamson PB, Smith AL, Abraham WT, Kleckner KJ, Stadler RW, Shih A, *et al.* Continuous autonomic assessment in patients with symptomatic heart failure: prognostic value of heart rate variability measured by an implanted cardiac resynchronization device. *Circulation* 2004;**110**:2389–2394.
6. Gerritsen J, Dekker JM, TenVoorde BJ, Kostense PJ, Heine RJ, Bouter LM, *et al.* Impaired autonomic function is associated with increased mortality, especially in subjects with diabetes, hypertension, or a history of cardiovascular disease: the Hoorn study. *Diabetes Care* 2001;**24**:1793–1798.
7. Yperzeele L, van Hooff RJ, Nagels G, De Smedt A, De Keyser J, Brouns R, *et al.* Heart rate variability and baroreceptor sensitivity in acute stroke: a systematic review. *Int J Stroke* 2015;**10**:796–800.
8. Sassi R, Cerutti S, Lombardi F, Malik M, Huikuri HV, Peng CK, *et al.* Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. *Europace* 2015;**17**:1341–1353.
9. Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 1995;**5**: 82–87.
10. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 2000;**278**:H2039–H2049.
11. Huikuri HV, Seppanen T, Koistinen MJ, Airaksinen J, Ikaheimo MJ, Castellanos A, *et al.* Abnormalities in beat-to-beat dynamics of heart rate before the spontaneous onset of life-threatening ventricular tachyarrhythmias in patients with prior myocardial infarction. *Circulation* 1996;**93**:1836–1844.
12. Bauer A, Kantelhardt J, Barthel P, Schneider R, Mäkikallio T, Ulm K, *et al.* Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. *Lancet* 2006;**367**:1674–1681.
13. Schmidt G, Malik M, Barthel P, Schneider R, Ulm K, Rolnitzky L, *et al.* Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction. *Lancet* 1999;**353**:1390–1396.
14. Vescio B, Salsone M, Gambardella A, Quattrone A. Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short- and long-term recordings. *Sensors (Basel)* 2018;**18**:844.
15. Hernando D, Roca S, Sancho J, Alesanco A, Bailon R. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Sensors (Basel)* 2018;**18**:2619.
16. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. *Sensors (Basel)* 2022;**22**: 6317.
17. Pan J, Tompkins WJ. A real-time QRS detection algorithm. *IEEE Trans Biomed Eng* 1985; **32**:230–236.
18. MacDonald PR, Kulach CJ. Heart rate monitor with time varing linear filtering. US Patent. Garmin Switzerland GmbH; 2016.
19. Vila J, Palacios F, Presedo J, Fernandez-Delgado M, Felix P, Barro S. Time-frequency analysis of heart-rate variability. *IEEE Eng Med Biol Mag* 1997;**16**:119–126.
20. Eick C, Rizas KD, Meyer-Zurn CS, Groga-Bada P, Hamm W, Krethet F, *et al.* Autonomic nervous system activity as risk predictor in the medical emergency department: a prospective cohort study. *Crit Care Med* 2015;**43**:1079–1086.
21. Rodriguez-Linares L, Vila X, Lado M, Mendez A, Otero A, Garcia C. RHRV: heart rate variability analysis of ECG data. R package version 4.2.6. 2020. https://CRAN.R-project.org/package=RHRV.

22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;**15**:155–163.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–174.
24. Janssen MJA, Swenne CA, de Bie J, ter Heide H, van Bemmel JH, Rompelman O. Methods in heart rate variability analysis: may the ventricular or the pulse rhythm be used as a substitute for the atrial rhythm? *High Blood Press* 1994;**3**:23–29.
25. Mäkikallio TH, Barthel P, Schneider R, Bauer A, Tapanainen JM, Tulppo MP, *et al.* Prediction of sudden cardiac death after acute myocardial infarction: role of Holter monitoring in the modern treatment era. *Eur Heart J* 2005;**26**: 762–769.
26. Tapanainen JM, Thomsen PE, Køber L, Torp-Pedersen C, Mäkikallio TH, Still AM, *et al.* Fractal analysis of heart rate variability and mortality after an acute myocardial infarction. *Am J Cardiol* 2002;**90**:347–352.