Annals of Noninvasive Electrocardiology

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Intra- and Inter-Rater Reliability of Linear and Nonlinear Measures of Short-Term Heart Rate Variability Following Combat-Related Traumatic Injury

Rabeea Maqsood[1,2] | Susie Schofield[2] | Alexander N. Bennett[2,3] | Ahmed Khattab[1] | Anthony M. J. Bull[4] | Nicola T. Fear[5] | Christopher J. Boos[1,6] | for the ADVANCE Study

[1]Department of Medical Sciences and Public Health, Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, UK | [2]Faculty of Medicine, National Heart and Lung Institute, Imperial College London, London, UK | [3]Academic Department of Military Rehabilitation, Defence Medical Rehabilitation Centre, Loughborough, UK | [4]Centre for Injury Studies, Department of Bioengineering, Imperial College London, London, UK | [5]Academic Department of Military Mental Health and King's Centre for Military Health Research, King's College London, London, UK | [6]Department of Cardiology, University Hospitals Dorset, Poole Hospital, Poole, UK

Correspondence: Rabeea Maqsood (rmaqsood@bournemouth.ac.uk)

Received: 13 December 2023 | **Revised:** 4 May 2024 | **Accepted:** 2 August 2024

Funding: This study is a part of RM's PhD studentship—jointly funded by Bournemouth University and the ADVANCE study, UK. The ADVANCE study is funded through the ADVANCE Charity. Key contributors to the charity are the Headley Court Charity (principal funder), HM Treasury (LIBOR Grant), Help for Heroes, Nuffield Trust for the Forces of the Crown, Forces in Mind Trust, National Lottery Community Fund, Blesma—The Limbless Veterans, the UK Ministry of Defense, and the Office for Veterans' Affairs (OVA).

Keywords: combat injury | HRV | military | reliability | RMSSD

ABSTRACT

Background: Heart rate variability (HRV) is a marker of autonomic function. However, the reliability of short-term HRV measurement in individuals with combat-related traumatic injury (CRTI) remains undetermined.

Methods: An intra- and inter-rater reliability study was conducted using a subsample ($n=35$) of British servicemen with CRTI enrolled in the ongoing ADVANCE study. A five-minute epoch of single-lead electrocardiogram data collected during spontaneous breathing was used to measure HRV. HRV analyses were independently performed by two examiners using Kubios. Intraclass correlation coefficient (ICC), standard error of measurement (SEM), minimum detectable change (MDC), and coefficient of variance were calculated for linear [root mean square of successive difference (RMSSD), standard deviation of NN interval, low-frequency, high-frequency, total power] and nonlinear (SD1-2, acceleration and deceleration capacities, sample entropy) measures. Bland–Altman %plots were used to assess bias in intra- and inter-rater HRV data.

Results: The mean age of participants was 39.3 ± 6.3 years. An excellent ICC score of 0.9998 (95% CI 0.9997, 0.9999) was observed for intra-rater analyses of RMSSD, and similar excellent ICC scores were seen for all other HRV measures. The inter-rater reliability analyses produced an excellent ICC score (range 0.97–1.00). Comparatively, frequency-domain measures produced higher MDC% and SEM% scores than time-domain and nonlinear measures in both inter- and intra-rater analyses. The Bland–Altman plots revealed relatively higher bias for frequency-domain and nonlinear measures than time-domain measures.

Conclusion: ECG-related short-term HRV measures were reliable in injured servicemen under spontaneous breathing. However, the reliability appeared better with the time-domain measure than frequency-domain and nonlinear measures in this sample.

---

https://doi.org/10.1111/anec.70008

1 | Introduction

The variation between consecutive R–R intervals on the QRS complex represents heart rate variability (HRV) and has been conventionally captured using continuous electrocardiograms (ECG) (Malik 1996). HRV data can be recorded over differing time periods—which vary from ultra-short term (<5 min), short-term (5–10 min), long-term (up to 24 h) (Shaffer and Ginsberg 2017), and very long-term (≥24 h). Most of the existing HRV literature comprises short-term HRV as its recording is less time-consuming than its long-term counterpart (Shaffer and Ginsberg 2017). There is early evidence to suggest that combat-related traumatic injury (CRTI) may be associated with lower parasympathetic tone (Maqsood, Schofield, et al. 2023) but its relationship to affect physical and mental post-trauma rehabilitation remains uncertain. Accurate and reliable HRV measurement is fundamental to a better understanding of these relationships.

Results have suggested that the reliability of short-term HRV measures appears to be lower in clinical populations compared with healthy adults (Sandercock, Bromley, and Brodie 2005). Following this, the reliability of short-term HRV measurement has been widely investigated across a broad spectrum of clinical populations including individuals with diabetes (Bassi et al. 2018), post-COVID (Almeida et al. 2022), chronic back pain (Penha et al. 2023), and chronic obstructive pulmonary disease (Santos-de-Araújo et al. 2023). However, the reliability of short-term HRV measurement using the gold standard of continuous ECG acquisition in individuals with CRTI remains unaddressed. Wounded combat veterans and personnel are a unique population whose HRV profile warrants exploration as this group has been reported to have higher cardiovascular risk following combat injury (Boos et al. 2019). Consequently, there is a need to better understand the impact of CRTI on their HRV as this may affect their recovery and rehabilitation. The majority of the previous research relating to traumatic injury and HRV has been focused on exploring the effects of either acute trauma (King et al. 2009; Norris et al. 2005) or specific traumatic injuries such as traumatic brain injury (Pinto et al. 2024). Thus, there is a dearth of research into HRV and combat injury, especially no study reporting the reliability of short-term HRV measures in wounded combat veterans and personnel.

Previously, we have explored the reliability of the root mean square of successive difference (RMSSD), a time-domain measure of HRV over an ultra-short-term recording (up to 16 s) in individuals with CRTI (Maqsood, Khattab, et al. 2023). The reliability of the frequency-domain and nonlinear measures in ultra-short-term HRV analyses (10 s) is debatable. As frequency domain measures require longer recordings (>10 s) for reliable measurement (Kim, Seok, and Shin 2021), the reliability of these measures could not be determined in our previous work on CRTI and HRV due to ultra-short HRV data (Maqsood, Khattab, et al. 2023). Hence, there is a need for a contemporary study to establish the consistency and reliability of short-term HRV measurements up to quality standards in this target population.

This study aims to examine the intra- and inter-rater reliability of short-term HRV measures obtained from the conventional ECG source in individuals with CRTI. It was hypothesized that frequency-domain measures of short-term HRV would show lower reliability than time-domain and nonlinear measures.

2 | Methodology

2.1 | Study Design

This study was an intra- and inter-rater reliability study of 35 adults (Operation HERRICK, Afghanistan 2003-2014) enrolled in the ongoing *A*rme*D* Ser*V*ices Tr*A*uma and Rehabilitatio*N* Out*C*om*E* (ADVANCE) prospective cohort study and who were assessed at their first follow-up visit ($n = 1053$) (Bennett et al. 2020). The data for the present study were collected between 2019 and 2021. The present study was conducted following the guidelines for reporting reliability and agreement studies (GRRAS) (Kottner et al. 2011). A summary of the study design is presented in Figure 1.

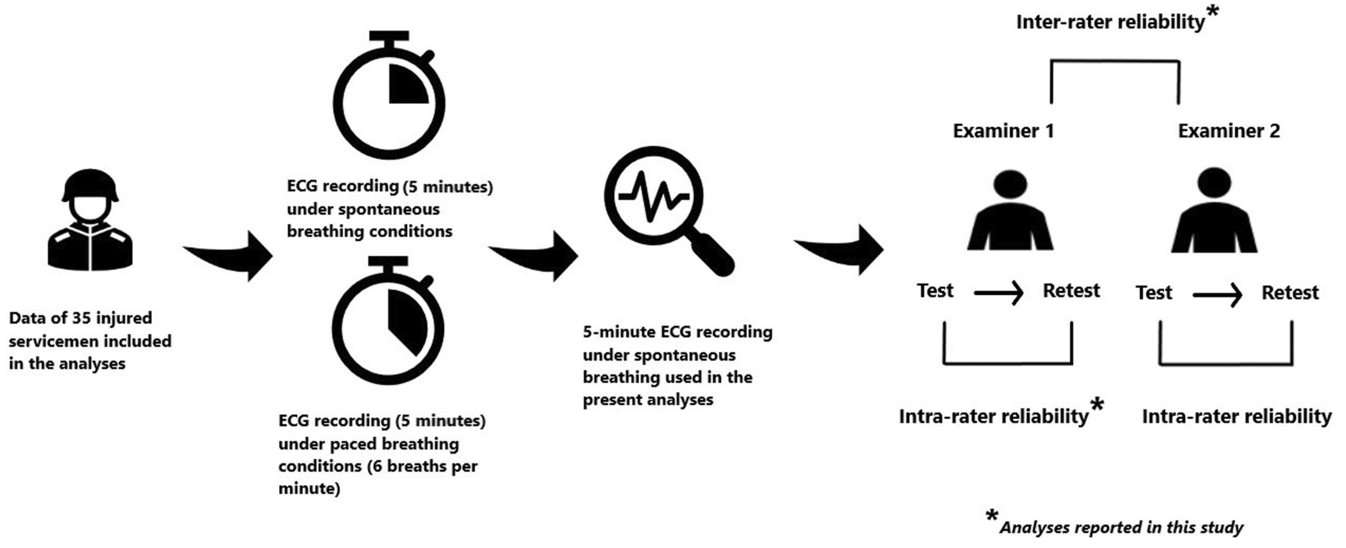

FIGURE 1 | Summary of study design and reliability protocol.

2.2 | Participants

All included participants had sustained a serious physical CRTI while on deployment. The time elapsed since injury at the first follow-up visit was approximately 11 years (Maqsood, Schofield, et al. 2023). These injured servicemen required aeromedical evacuation followed by medical treatment and rehabilitation at a UK hospital. None of the participants had a history of cardiovascular, renal, or liver disease prior to recruitment in the ADVANCE study. The ADVANCE study has full ethics approval from the UK Ministry of Defence Research Ethics Committee (protocol no: 357/PPE/12) and the full protocol can be found elsewhere (Bennett et al. 2020).

2.3 | Sample Size

Based on previous studies with clinical populations (Almeida et al. 2022; Penha et al. 2023), the sample size was calculated using an online calculator (Arifin 2023). The following criteria were applied for the calculation: minimum acceptable intraclass correlation coefficient (ICC) value of 0.40, expected ICC of 0.75, alpha error set to 5%, and power at 80% as used previously (Santos-de-Araújo et al. 2023); however, the sample loss was set to zero as not applicable in the present study. A minimum sample size of 28 participants was estimated. Based on this, a total of 35 participants were randomly selected using the “sample” command out of the first 248 participants with CRTI included in the first follow-up of the ADVANCE study at the time of the present analysis.

2.4 | Data Collection

HRV data were acquired using a single-lead ECG device, Bittium Faros™ 180 (Mega Motion Faros 180 recorder: Mega Electronics Ltd., Pioneerinkatu, Finland). All data were collected in the supine position on a hospital bed in a temperature-controlled and quiet room during the daytime. Participants were encouraged to relax and refrain from talking, moving, and sleeping during the recording. HRV data were collected over a continuous 5-minute recording period during which the participants were encouraged to breathe normally. The 5 min of the ECG recording were manually selected by each examiner at the end of an approximately 10-min window of spontaneous breathing during continuous ECG monitoring.

2.5 | HRV Analysis Protocol

Both examiners (R. M. and C. J. B.) had previous experience in HRV analysis and individually analyzed the signals using the Kubios HRV Premium Software version (3.5) (Biosignal Analysis and Medical Imaging Group, Department of Physics, University of Kuopio, Kuopio, Finland). The noise level was set to medium with automatic correction, which was further supplemented by visual inspection of the entire 5-min recording period. Ectopic beats were manually excluded by the examiners. Both examiners were blinded to each other's timestamps for spontaneous and paced breathing protocols. There was a wash-out period of 1 week between the analyses by both examiners to minimize bias.

Linear and nonlinear measures of HRV were measured according to the Task Force guidelines (Malik 1996). The RMSSD and the standard deviation of NN intervals (SDNN) were included as the time-domain measures of HRV. The absolute power of the low-frequency (LF) band (0.04–0.15 Hz) and high-frequency (HF) band (0.15–0.4 Hz) were included from the frequency domain along with the total power (TP) using the fast Fourier transform (FFT) method. Nonlinear measures included: in the Poincare plot, the standard deviation perpendicular to the line of identity (SD1), in the Poincare plot, the standard deviation along the line of identity (SD2), acceleration capacity (AC), deceleration capacity (DC), and sample entropy (SampEn).

2.6 | Statistical Analyses

All data were reported as mean ± standard deviation or number and percentages. For intra- and inter-rater reliability analyses, ICC, standard error of measurement (SEM), minimum detectable change (MDC), and coefficient of variance (CoV) were calculated as described previously (Almeida et al. 2022).

The strength of the ICC was graded according to the following recognized criteria—ICC scores of <0.5 (poor), 0.5–0.75 (moderate), 0.75–0.9 (good), and >0.9 (excellent) were interpreted accordingly (Koo and Li 2016). Following the guidelines of Koo and Li (2016), a two-way mixed effects model with absolute agreement was applied for the calculation of ICC in intra-rater reliability whereas a two-way random effects model with absolute agreement was followed for the inter-rater reliability ICC analysis.

The MDC was included as a measure of the smallest amount of change that is attributed to the method being tested and not to the measurement error (Haghayegh et al. 2020). We calculated MDC using the formula: $1.96 \times \text{SEM} \times \sqrt{2}$ (Weir 2005). MDC% was calculated as MDC ×100/average of two means. The SEM was calculated as a measure of random variation in the measurement score when no real change has taken place, using the formula: standard deviation of means $\times \sqrt{1 - \text{ICC}}$ (Almeida et al. 2022; Weir 2005). SEM% was calculated using the formula: SEM ×100/average of two means. CoV was calculated as standard deviation/mean ×100 and reported as a percentage (Almeida et al. 2022).

Bland–Altman analyses (Bland and Altman 2010) were performed to assess the average bias between the HRV scores calculated by the two independent examiners (inter-rater agreement) and repeated testing by the same examiner (intra-rater agreement). Most of the HRV measures indicated heteroscedasticity on the Bland–Altman plot with absolute values. Considering the increasing standard deviation with concentration, the percent plot with the Bland–Altman analysis was used using the absolute values as recommended (Dewitte et al. 2002). Average difference (mean bias %) and 95% confidence interval (CI) limits of agreement (LoA) were reported (Chatfield et al. 2023). All statistical tests were performed in Stata (V 17.0; StataCorp LLC).

3 | Results

In total, 35 participants were included in this study with no data loss. The mean age of participants at the time of their first follow-up visit was 39.3 ± 6.3 years. The majority of participants were White (88.5%). Amputation was reported in 34% of participants. Further anthropometric characteristics of the participants can be seen in Table 1 and the HRV scores can be found in Table 2.

3.1 | Intra-Rater Reliability (RM vs. RM)

This was graded excellent with an ICC score of 1.00 for all HRV measures. For time-domain measures, the SEM% ranged from 0.17% to 1.21% whereas for frequency-domain measures, the range was 2.40% to 11.34%. The nonlinear measures of HRV showed lower SEM% ranging from −0.81% to 1.80%. The MDC% ranged from 0.48% to 3.36% for time-domain measures while 6.66% to 31.42% for frequency-domain measures. A comparative lower range of MDC% score was observed for nonlinear measures of HRV from −2.24% to 4.99%. The lowest CoV values were observed for time-domain measures (0.10% to 0.77%) followed by nonlinear measures (−0.74% to 0.99%), and frequency-domain measures (1.22% to 2.03%) (Table 3).

TABLE 1 | Demographic and anthropometric characteristics of the sample.

	Mean ± SD or number (%)
Number	35
Age at first follow-up assessment, years	39.34 ± 6.38
Rank (NS-SEC)	
Junior	25 (71.43)
Senior	6 (17.14)
Commissioned officers	4 (11.43)
NISS	13 (9, 22)
Amputation	
Yes	12 (34.29)
No	23 (65.71)
Ethnicity	
White	31 (88.57)
Others	4 (11.42)
Smoker	
Non-smoker	15 (42.86)
Ex/light smoker	20 (57.14)
Height, cm	178.22 ± 8.79
Abdominal circumference, cm	101.05 ± 12.35
Hip circumference, cm	104.17 ± 10.02

Note: Data are shown as mean and standard deviation or number and percentages or median and inter-quartile range.
Abbreviations: cm; centimeters, NISS; New Injury Severity Score; NS-SEC, National Statistics Socio-Economic Classification—rank at sampling.

3.2 | Inter-Rater Reliability (RM vs. CJB)

Overall, the ICC ranged from 0.97 to 1.00 reflecting excellent agreement between the examiners. For time-domain measures, the SEM% scores ranged from 0.30% to 2.02% whereas MDC% ranged from 0.83% to 5.59%. For frequency-domain measures, SEM% scores ranged from 4.21% to 23.46% whereas MDC% ranged from 11.66% to 65.02%. Regarding nonlinear measures, SEM% ranged from –1.99% to 2.59% with an MDC% range of –5.53% to 7.19%. The CoV was found to be comparatively higher in the frequency-domain measures (range 3.22% to 5.56%) than in the time-domain (0.21% to 1.58%) and nonlinear measures (–1.65% to 1.83%) (Table 4).

Overall, the SEM% and MDC% scores were found to be higher for frequency-domain measures as compared to time-domain and nonlinear measures of HRV in both intra- and inter-rater analyses. This was further confirmed by the Bland–Altman analysis revealing relatively higher bias for frequency-domain and nonlinear measures as compared to time-domain measures of HRV in both intra- and inter-rater analyses (Tables 5 and 6, Figures 2 and 3).

4 | Discussion

The objective of the present study was to explore the intra- and inter-rater reliability of short-term HRV recordings (5 min) performed at rest during spontaneous breathing in individuals who sustained CRTI. While all HRV measures produced excellent ICC scores in intra- and inter-rater reliability analyses, frequency-domain measures displayed lower reliability (indicated by higher SEM% and MDC% scores) as compared to time-domain and nonlinear measures of HRV.

While these findings are not directly comparable to the existing literature due to the heterogeneity in the population (individuals with CRTI vs. those with other clinical conditions), type of HRV recording device (ECG vs. HR monitors), recording period (2 min vs. 5 min), breathing protocol (spontaneous vs. paced); however, some consistent trends have emerged. Almeida et al. (2022) reported higher SEM, MDC, and CoV scores for frequency-domain measures in a population with COVID-19 in their inter/intra-rater reliability study. Putting the findings in context, in the intra-rater reliability analysis, RMSSD produced an SEM% score of 1.21% whereas its frequency-domain counterpart (HF power) doubled with an SEM of 2.40%—as an indication of random variation in scores without any physiological changes.

For both inter- and intra-rater reliability, SEM% scores were ≤5% which has been deemed to be “very good” (Ostelo et al. 2004) except for LF power in the inter-rater analysis (23.46%, considered “negative”) and intra-rater analysis (11.34%, considered “doubtful”). In both inter- and intra-rater reliability analyses, mean heart rate produced the lowest MDC% score whereas LF produced the highest MDC% score. The inferior reliability of the

TABLE 2 | HRV scores in individuals with CRTI in the supine position.

	Examiner 1 (RM)		Examiner 2 (CJB)	
HRV measure	**Test mean ± SD**	**Retest mean ± SD**	**Test mean ± SD**	**Retest mean ± SD**
Time domain measures				
Mean HR, BPM	61.56 ± 10.67	61.61 ± 10.65	61.60 ± 10.74	61.46 ± 10.82
RMSSD, ms	37.48 ± 32.19	37.41 ± 31.96	37.51 ± 32.31	37.35 ± 32.36
SDNN, ms	38.69 ± 23.78	38.56 ± 23.70	38.46 ± 23.12	37.90 ± 23.48
Frequency domain measures				
LF power, ms^2	1001.07 ± 1481.95	1039.82 ± 1696.13	928.92 ± 1085.20	870.23 ± 963.82
HF power, ms^2	853.58 ± 2058.97	855.53 ± 2050.53	866.96 ± 2119.42	862.34 ± 2135.83
Total power, ms^2	1950.95 ± 3563.66	1990.62 ± 3768.57	1894.97 ± 3207.63	1822.08 ± 3091.27
Nonlinear measures				
SD1, ms	26.55 ± 22.81	26.50 ± 22.65	26.57 ± 22.90	26.46 ± 22.94
SD2, ms	47.21 ± 26.05	46.99 ± 26.07	46.73 ± 24.74	45.95 ± 25.28
AC, ms	−34.40 ± 27.85	−34.37 ± 27.79	−34.59 ± 28.31	−34.47 ± 28.53
DC, ms	35.40 ± 32.15	35.07 ± 31.32	35.11 ± 31.68	34.95 ± 31.39
SampEn	1.62 ± 0.25	1.62 ± 0.24	1.62 ± 0.25	1.62 ± 0.25

Note: Data are shown as mean and standard deviation.
Abbreviations: AC, acceleration capacity; BPM, beats per minute; DC, deceleration capacity; HF, high frequency; HR, heart rate; LF, low frequency; ms, millisecond; ms^2, millisecond square, RMSSD, root mean square of successive differences; SampEn, sample entropy; SD1, standard deviation of short-term RR intervals; SD2, standard deviation of long-term RR intervals; SDNN, standard deviation of NN intervals; TP, total power.

TABLE 3 | Intra-rater reliability of HRV measures in individuals with CRTI in the supine position.

HRV measure	**ICC* (95% CI)**	**SEM**	**SEM (%)**	**MDC**	**MDC (%)**	**CoV (%)**
Time domain measures						
Mean HR, BPM	0.9999 (0.9998–0.9999)	0.11	0.17	0.30	0.48	0.10 ± 0.16
RMSSD, ms	0.9998 (0. 9997–0. 9999)	0.45	1.21	1.26	3.36	0.59 ± 0.81
SDNN, ms	0.9997 (0.9995–0.9998)	0.41	1.06	1.14	2.95	0.77 ± 1.27
Frequency domain measures						
LF power, ms^2	0.9947 (0.9896–0.9973)	115.68	11.34	320.66	31.42	2.03 ± 3.26
HF power, ms^2	0.9999 (0.9999–0.9999)	20.55	2.40	56.95	6.66	1.22 ± 1.85
Total power, ms^2	0.9990 (0.9982–0.9995)	115.93	5.88	321.35	16.31	1.26 ± 1.89
Nonlinear measures						
SD1, ms	0.9998 (0.9997–0.9999)	0.32	1.21	0.89	3.36	0.59 ± 0.81
SD2, ms	0.9997 (0.9994–0.9998)	0.45	0.96	1.25	2.66	0.77 ± 1.49
AC, ms	0.9999 (0.9998–0.9999)	0.28	−0.81	0.77	−2.24	−0.74 ± 0.90
DC, ms	0.9996 (0.9992–0.9998)	0.63	1.80	1.76	4.99	0.99 ± 1.17
SampEn	0.9971 (0.9944, 0.9985)	0.01	0.82	0.04	2.29	0.61 ± 0.86

Note: Data are shown as mean and standard deviation.
*Two-way mixed effect model with absolute agreement, reporting average ICC.
Abbreviations: AC, acceleration capacity; BPM, beats per minute; CoV, coefficient of variation; DC, deceleration capacity; HF, high frequency; HR, heart rate; HRV, heart rate variability; ICC; intraclass correlation coefficient; LF, low frequency; MDC, minimum detectable change; ms, millisecond; ms^2, millisecond square, RMSSD, root mean square of successive differences; SampEn, sample entropy; SD1, standard deviation of short term RR intervals; SD2, standard deviation of long term RR intervals; SDNN, standard deviation of NN intervals; SEM, standard error of measurement; TP, total power.

TABLE 4 | Inter-rater reliability of HRV measures in individuals with CRTI in the supine position.

HRV measure	ICC*, 95% CI	SEM	SEM (%)	MDC	MDC (%)	CoV (%)
Time domain measures						
Mean HR, BPM	0.9997 (0.9995–0.9998)	0.19	0.30	0.51	0.83	0.21 ± 0.31
RMSSD, ms	0.9998 (0.9996–0.9999)	0.46	1.22	1.26	3.37	0.96 ± 1.08
SDNN, ms	0.9989 (0.9979–0.9994)	0.78	2.02	2.16	5.59	1.58 ± 2.27
Frequency domain measures						
LF power, ms^2	0.9689 (0.9387–0.9842)	226.36	23.46	627.44	65.02	5.56 ± 6.35
HF power, ms^2	0.9997 (0.9994–0.9998)	36.19	4.21	100.30	11.66	3.22 ± 3.60
Total power, ms^2	0.9966 (0.9933–0.9982)	148.77	5.40	412.36	14.96	3.66 ± 3.94
Nonlinear measures						
SD1, ms	0.9998 (0.9996–0.9999)	0.32	1.22	0.90	3.37	0.96 ± 1.08
SD2, ms	0.9977 (0.9956–0.9988)	1.22	2.59	3.38	7.19	1.83 ± 2.61
AC, ms	0.9994 (0.9990–0.9997)	0.69	−1.99	1.91	−5.53	−1.65 ± 1.33
DC, ms	0.9996 (0.9993–0.9998)	0.64	1.81	1.77	5.02	1.71 ± 1.81
SampEn	0.9846 (0.9697–0.9922)	0.03	1.92	0.09	5.32	1.62 ± 2.05

Note: Data are shown as mean and standard deviation.
Abbreviations: AC, acceleration capacity; BPM, beats per minute; CoV, coefficient of variation; DC, deceleration capacity; HF, high frequency; HRV, heart rate variability; ICC; intraclass correlation coefficient; LF, low frequency; MDC, minimum detectable change; ms, millisecond; ms^2, millisecond square; RMSSD, root mean square of successive differences; SampEn, sample entropy; SD1, standard deviation of short term RR intervals; SD2, standard deviation of long term RR intervals; SEM, standard error of measurement; SDNN, standard deviation of NN intervals; TP, total power.
*Two-way random effect model with absolute agreement, reporting average ICC.

TABLE 5 | Summary of the Bland–Altman analysis for the inter-rater reliability of HRV measures (RM and CJB).

HRV (RM and CJB)	Mean difference ± SD (%)	95% Limits of agreement (upper and lower limits)	95% CI (mean% difference)
HR, BPM	−0.04 ± 0.53	−1.10, 1.01	−0.23, 0.13
RMSSD, ms	−0.01 ± 2.07	−4.05, 4.07	−0.70, 0.72
SDNN, ms	0.35 ± 3.91	−7.32, 8.03	−0.98, 1.70
LF, ms^2	0.23 ± 12.01	−23.31, 23.78	−3.89, 4.36
HF, ms^2	−0.50± 6.86	−13.95, 12.94	−2.86, 1.85
TP, ms^2	−0.07 ± 7.66	−15.10, 14.94	−2.71, 2.55
SD1, ms	0.01 ± 2.07	−4.05, 4.07	−0.70, 0.72
SD2, ms	0.59 ± 4.50	−8.23, 9.41	−0.95, 2.13
SampEn	0.38 ± 3.69	−6.87, 7.63	−0.89,1.65
DC, ms	0.66 ± 3.48	−6.16, 7.49	−0.53, 1.85

Note: Data presented for absolute value and as mean bias ± SD, interpreted in percentage.
Abbreviations: BPM, beats per minute; CI, confidence interval; DC, Deceleration Capacity; HF, High Frequency; HR, Heart Rate; HRV, heart rate variability; LF, Low Frequency; LoA, limits of agreement; ms, millisecond; ms^2, millisecond square, RM & CJB (analysis of the same data by two examiners: RM and CJB); RMSSD, root mean square of successive differences; SD1, standard deviation of short-term RR intervals; SD2, standard deviation of long term RR intervals; SDNN, standard deviation of NN intervals; SampEn, sample entropy; TP, total power.

frequency-domain measure may be attributed to spontaneous breathing conditions during our data collection process as it has been suggested that the frequency-domain measures offer better reproducibility under paced breathing conditions (Gisselman, D'Amico, and Smoliga 2020).

Our findings also indicate that the HRV data analyzed by the same examiner (intra-rater reliability) may be more reliable than that analyzed by two examiners (inter-rater reliability) given the lower CoV values in intra-rater analyses (Tables 3 and 4). There is no consensus on an acceptable CoV (Sandercock, Bromley, and Brodie 2005); nonetheless, the agreement between the examiners was excellent given the ICC scores.

This is the first study to have explored the reproducibility and reliability of HRV measures from all domains in veterans and

TABLE 6 | Summary of the Bland–Altman analysis for the intra-rater reliability of HRV measures (RM1 and RM2).

HRV (RM1 and RM2)	Mean difference ± SD (%)	95% Limits of agreement (upper and lower limits)	95% CI (mean% difference)
HR, BPM	−0.09 ± 0.26	−0.61, 0.41	−0.18, −0.008
RMSSD, ms	0.24 ± 1.40	−2.51, 3.00	−0.23, 0.73
SDNN, ms	0.44 ± 2.07	−3.62, 4.50	−0.27, 1.15
LF, ms^2	−0.92 ± 5.37	−11.46, 9.60	−2.77, 0.91
HF, ms^2	−0.39 ± 3.13	−6.53, 5.75	−1.46, 0.68
TP, ms^2	−0.62 ± 3.17	−6.84, 5.58	−1.71, 0.45
SD1, ms	0.24 ± 1.40	−2.51, 3.00	−0.23, 0.73
SD2, ms	0.60± 2.30	−3.90, 5.12	−0.18, 1.39
SampEn	0.11 ± 1.50	−2.84, 3.06	−0.40, 0.62
DC, ms	0.56 ± 2.10	−3.57, 4.69	−0.16, 1.28

Note: Data presented for absolute value and as mean bias ± SD, interpreted in percentage.
Abbreviations: BPM, beats per minute; CI, confidence interval; DC, Deceleration Capacity; HF, high frequency; HR, heart rate; HRV, heart rate variability; LF, Low Frequency; LoA, limits of agreement; ms, millisecond; ms^2, millisecond square; RM1 and RM2 (test–retest data analyzed by the same examiner, RM); RMSSD, root mean square of successive differences; SD1, standard deviation of short-term RR intervals; SD2, standard deviation of long term RR intervals; SampEn, sample entropy; SDNN, standard deviation of NN intervals; TP, total power.

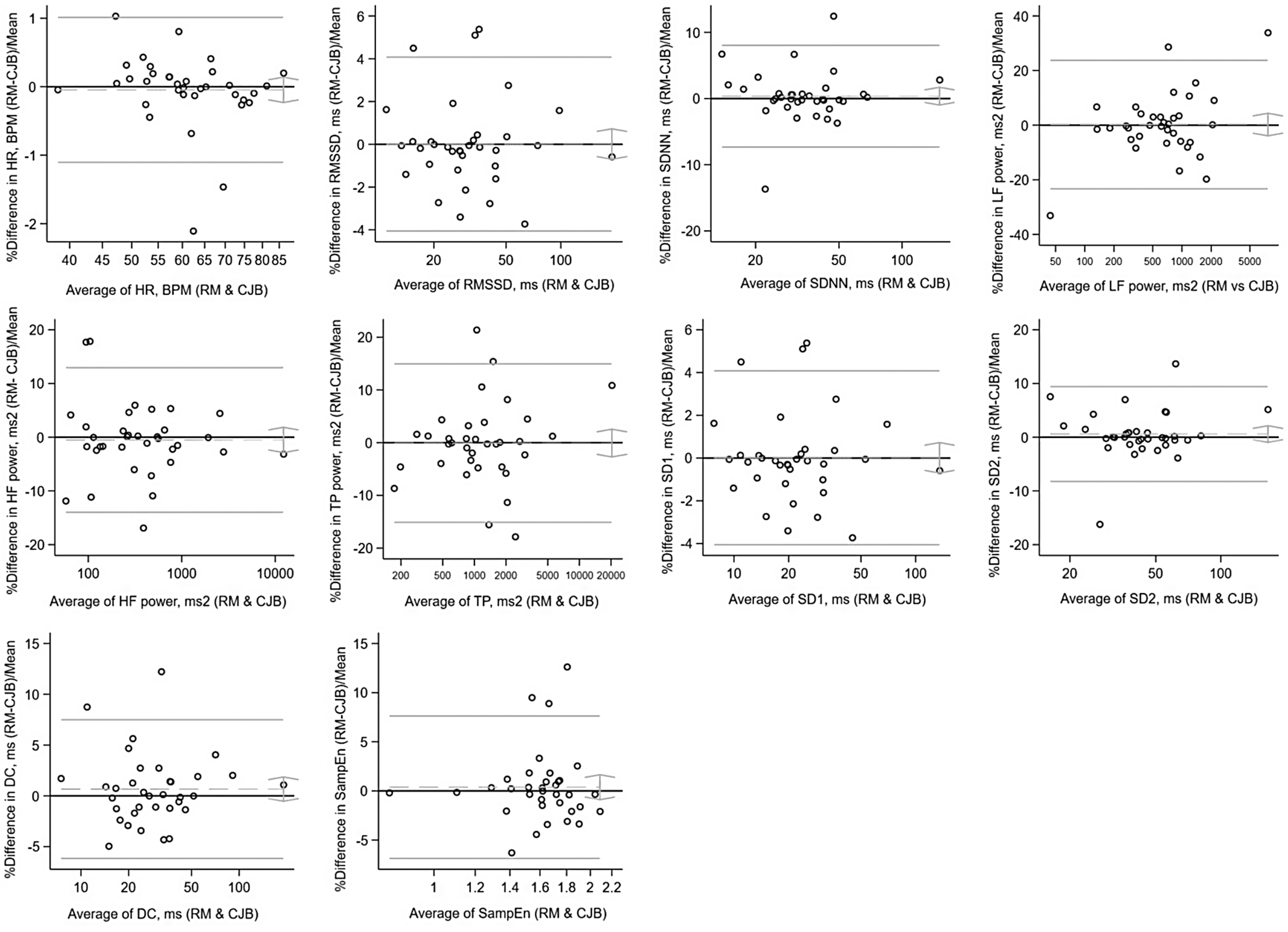

FIGURE 2 | The Bland–Altman (percent) plots for inter-rater reliability analyses: Absolute values have been used in the Bland-Altman plots. The *x*-axis represents the mean of the HRV index from examiner 1 and examiner 2 (RM+CJB/2), and the *y*-axis represents the percentage of the difference in HRV index between examiner 1 and examiner 2 (100* RM-CJB)/ mean. Grey dotted lines denote mean bias (%), and grey solid lines are 95% confidence intervals of bias (lower and upper limits of agreements). AC, acceleration capacity; BPM, beats per minute; DC, deceleration capacity; HF, high frequency; HR, heart rate; ms, millisecond; ms^2, millisecond square; LF, low frequency; RMSSD, root mean square of successive differences; Sample, sample entropy; SD1, standard deviation of short-term RR intervals; SD2, standard deviation of long-term RR intervals; SDNN, standard deviation of NN intervals; TP, total power.

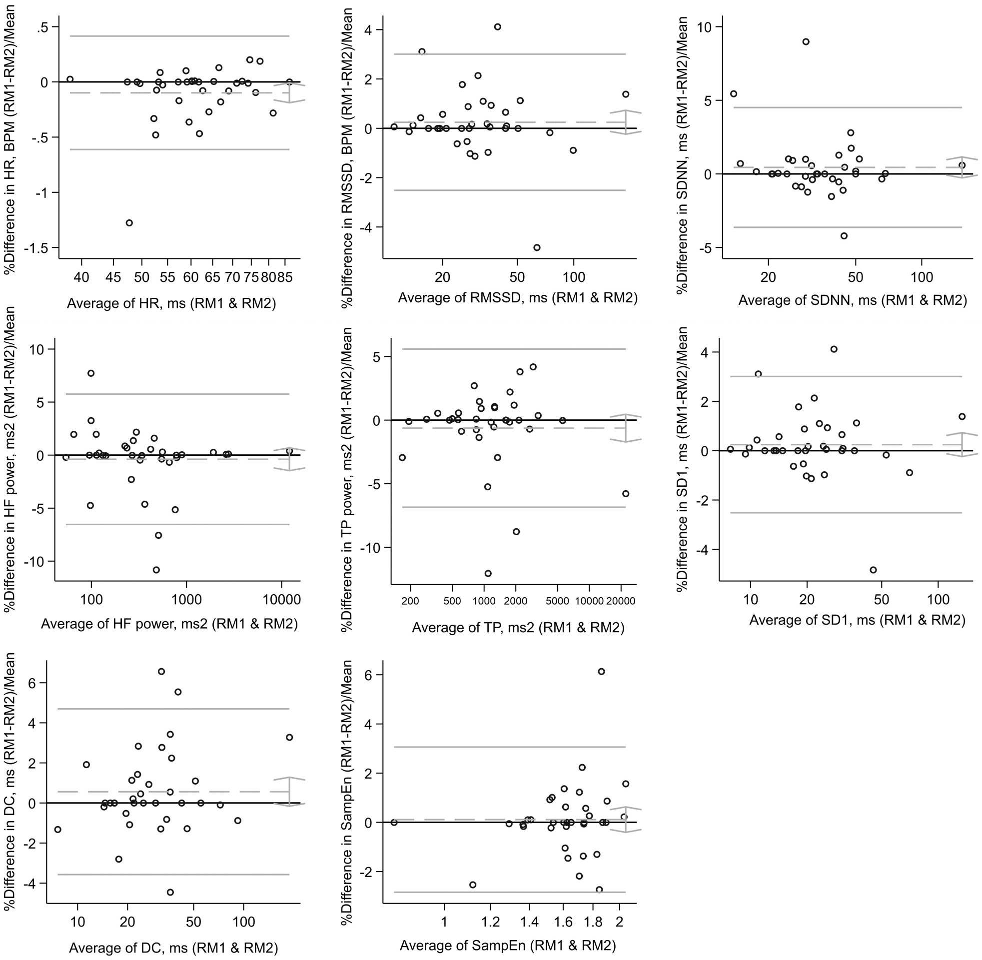

FIGURE 3 | The Bland–Altman (percent) plots for intra-rater reliability analyses: Absolute values have been used in the Bland-Altman plots. The *x*-axis represents the mean of the HRV index from test-rest data from the single examiner (RM1+RM2/2), and the *y*-axis represents the percentage of the difference in HRV index between test-rest data from the single examiner (100* RM1-RM2)/ mean. Grey dotted lines denote mean bias (%), and grey solid lines are 95% confidence intervals of bias (lower and upper limits of agreements). AC, acceleration capacity; BPM, beats per minute; DC, deceleration capacity; HF, high frequency; HR, heart rate; ms, millisecond; ms^2, millisecond square; LF, low frequency; RMSSD, root mean square of successive differences; Sample, sample entropy; SD1, standard deviation of short-term RR intervals; SD2, standard deviation of long-term RR intervals; SDNN, standard deviation of NN intervals; TP, total power.

personnel who sustained CRTI. Both examiners followed the same HRV analysis protocol with consensus on the data analysis filters and performed the "retest" analyses of HRV with a "wash-out" period of 7 days after the first "test" to minimize bias; however, this period did not seem to significantly affect the results and is in line with the literature (Sandercock, Bromley, and Brodie 2005). Our sample size was also based on the previous studies with clinical populations instead of healthy subjects. This is crucial because using the reliability coefficients reported for healthy subjects may lead to underestimated sample size, affecting the power of the study (Sandercock, Bromley, and Brodie 2005). Furthermore, the difference in the transition time between the two breathing protocols (spontaneous and paced) was insignificant between the two examiners and did not affect the HRV score (Spearman's rho = −0.06, $p = 0.72$). Lastly, the blinded inter-rater HRV analysis protocol is another key strength of this study.

Our findings should be interpreted under the context of a few limitations. Our sample consisted of predominantly White male participants with CRTI, limiting the generalizability of the findings for women, other ethnicities, and healthy populations. Despite slightly skewed data, we reported means and standard deviation to be consistent with the calculation of ICC, MDC, and SEM. This may have introduced bias. The reliability was reported for short-term data (5 min) which may not be applicable for ultra-short and long-term HRV data. The influence of breathing on the reliability of linear and nonlinear measures of HRV was not analyzed in the present study as it was beyond the scope of this study. Moreover, although "low" beat correction has been recommended for optimal HRV analysis (Gisselman, D'Amico, and Smoliga 2020), this study employed automatic correction, further supplemented by manual correction upon visual inspection of signals whenever needed. This was decided on the overall optimal quality of ECG recordings. Lastly, the Bland–Altman percent plot for AC could not be included given AC's negative values.

Our findings offer implications for future research and practice. HRV is a well-recognized noninvasive measure of autonomic function and lower HRV is strongly associated with worse cardiovascular outcomes (Fang, Wu, and Tsai 2020). Consequently, a greater understanding of the reliability of HRV measurement following traumatic injury and CRTI is crucial for the accurate interpretation of HRV data. Our study is unique in providing exploratory evidence on the reliability of both linear and nonlinear measures of HRV in injured combat veterans and personnel. It is anticipated that this may help other clinical and HRV researchers decide on the selection of HRV variables most suited for their research depending on the question and population of interest. Following this, we also invite other researchers to expand on our research and reproduce the reported reliability in other underrepresented "non-healthy" and vulnerable populations. Within the scope of military research, this study is timely as HRV evaluation is gaining traction in post-trauma rehabilitation studies given its feasibility and objectivity in measuring autonomic balance. While time-domain measures may offer more reliability than other measures, we suggest that the decision to report either time-domain, frequency-domain, or nonlinear measures should be guided by the research question and the breathing protocol used during the data collection.

5 | Conclusion

In individuals with CRTI, short-term HRV measures offer acceptable intra- and inter-rater reliability. Frequency-domain measures were indicated to have comparatively lower reliability than time-domain and nonlinear measures of HRV.

Author Contributions

R.M., A.K., C.J.B., A.N.B. contributed to the study conception and design. Material preparation was done by R.M. Heart Rate Variability analysis was performed by R.M. and C.J.B. Statistical analyses were performed by R.M. and supervised by S.S. The first draft of the manuscript was written by R.M. and revised by C.J.B. A.K. and C.J.B. overall supervised the study. A.N.B., S.S., N.T.F., A.M.J.B., and A.K. critically reviewed the manuscript and suggested improvements. All co-authors approved the final version of the manuscript.

Acknowledgments

We wish to thank all the research staff at Stanford Hall who helped with the ADVANCE study, including Emma Coady, Grace Blissitt, Melanie Chesnokov, Daniel Dyball, Sarah Evans, Guy Fraser, Nicola Goodman, Alison Hever, Meliha Kaya-Barge, Jocelyn Keshet-Price, Eleanor Miller, Steven Parkes, Bharti Patel, Samantha Paul, David Pernet, Vlad Pop, Helen Prentice, Ursula Pucilowska, Stefan Sprinckmoller, Jodie Stevenson, Lalji Varsani, Anna Verey, Molly Waldron, Owen Walker, Farheen Dairkee, Tasarla White, Seamus Wilson, and Severija Juškaitė.

Ethics Statement

This study had full ethical approval from the UK Ministry of Defense Research Ethics Committee (protocol no. 357/PPE/12). All participants in this study undertook full informed written consent. This study was conducted in compliance with the Declaration of Helsinki (1964).

Consent

The authors have nothing to report.

Conflicts of Interest

A. N. B. is a serving member of the Royal Air Force. N. T. F. is a trustee of Help for Heroes and is part funded by a grant from the UK Ministry of Defense. The views expressed are those of the authors and not necessarily those of the UK Ministry of Defense.

Data Availability Statement

Only the authorized authors (R. M., S. S., C. J. B., A. N. B.) had access to the data of this study. Given the sensitive nature of the participants, data have not been made widely available. Requests for data will be considered on a case-by-case basis and subject to the UK Ministry of Defense clearance. More information can be found at: https://www.advancestudydmrc.org.uk/.

References

Almeida, L. V. D., A. D. Santos-de-Araújo, R. C. Cutrim, et al. 2022. "Intra-and Interrater Reliability of Short-Term Measurement of Heart Rate Variability on Rest in Individuals Post-COVID-19." *International Journal of Environmental Research and Public Health* 19, no. 20: 13587.

Arifin, W. N. 2023. Sample size calculator. Retrieved from: https://wnarifin.github.io/ssc/ssicc.html.

Bassi, D., A. D. Santos-de-Araújo, P. F. Camargo, et al. 2018. "Inter and Intra-Rater Reliability of Short-Term Measurement of Heart Rate Variability on Rest in Diabetic Type 2 Patients." *Journal of Medical Systems* 42: 1–7.

Bennett, A. N., D. M. Dyball, C. J. Boos, et al. 2020. "Study Protocol for a Prospective, Longitudinal Cohort Study Investigating the Medical and Psychosocial Outcomes of UK Combat Casualties From the Afghanistan war: The Advance Study." *BMJ Open* 10, no. 10: e037850.

Bland, J. M., and D. G. Altman. 2010. "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement." *International Journal of Nursing Studies* 47, no. 8: 931–936.

Boos, C. J., N. De Villiers, D. Dyball, A. McConnell, and A. N. Bennett. 2019. "The Relationship Between Military Combat and Cardiovascular Risk: A Systematic Review and Meta-Analysis." *International Journal of Vascular Medicine* 2019: 1–14.

Chatfield, M. D., T. J. Cole, H. C. de Vet, L. Marquart-Wilson, and D. M. Farewell. 2023. "Blandaltman: A Command to Create Variants of Bland–Altman Plots." *Stata Journal* 23, no. 3: 851–874.

Dewitte, K., C. Fierens, D. Stockl, and L. M. Thienpont. 2002. "Application of the Bland–Altman Plot for Interpretation of Method-Comparison Studies: A Critical Investigation of its Practice." *Clinical Chemistry* 48, no. 5: 799–801.

Fang, S. C., Y. L. Wu, and P. S. Tsai. 2020. "Heart Rate Variability and Risk of All-Cause Death and Cardiovascular Events in Patients With Cardiovascular Disease: A Meta-Analysis of Cohort Studies." *Biological Research for Nursing* 22, no. 1: 45–56.

Gisselman, A. S., M. D'Amico, and J. M. Smoliga. 2020. "Optimizing Intersession Reliability of Heart Rate Variability—The Effects of Artifact Correction and Breathing Type." *Journal of Strength & Conditioning Research* 34, no. 11: 3199–3207.

Haghayegh, S., H. A. Kang, S. Khoshnevis, M. H. Smolensky, and K. R. Diller. 2020. "A Comprehensive Guideline for Bland–Altman and Intra Class Correlation Calculations to Properly Compare Two Methods of Measurement and Interpret Findings." *Physiological Measurement* 41, no. 5: 055012.

Kim, J. W., H. S. Seok, and H. Shin. 2021. "Is Ultra-Short-Term Heart Rate Variability Valid in Non-Static Conditions?" *Frontiers in Physiology* 12: 596060.

King, D. R., M. P. Ogilvie, B. M. Pereira, et al. 2009. "Heart Rate Variability as a Triage Tool in Patients With Trauma During Prehospital Helicopter Transport." *Journal of Trauma and Acute Care Surgery* 67, no. 3: 436–440.

Koo, T. K., and M. Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15, no. 2: 155–163.

Kottner, J., L. Audigé, S. Brorson, et al. 2011. "Guidelines for Reporting Reliability and Agreement Studies (GRRAS) Were Proposed." *International Journal of Nursing Studies* 48, no. 6: 661–671.

Malik, M. 1996. "Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use: Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology." *Annals of Noninvasive Electrocardiology* 1, no. 2: 151–181.

Maqsood, R., A. Khattab, A. N. Bennett, and C. J. Boos. 2023. "Reliability of Carotid-Femoral Arterial Waveforms for the Derivation of Ultra-Short Term Heart Rate Variability in Injured British Servicemen: An Inter-Rater Reliability Study." *PLoS One* 18, no. 9: e0290618.

Maqsood, R., S. Schofield, A. N. Bennett, et al. 2023. "Relationship Between Combat-Related Traumatic Injury and Ultrashort Term Heart Rate Variability in a UK Military Cohort: Findings From the ADVANCE Study." *BMJ Military Health*: e002316.

Norris, P. R., J. A. Morris Jr., A. Ozdas, E. L. Grogan, and A. E. Williams. 2005. "Heart Rate Variability Predicts Trauma Patient Outcome as Early as 12 h: Implications for Military and Civilian Triage." *Journal of Surgical Research* 129, no. 1: 122–128.

Ostelo, R. W., H. C. De Vet, D. L. Knol, and P. A. Van Den Brandt. 2004. "24-Item Roland-Morris Disability Questionnaire was Preferred out of Six Functional Status Questionnaires for Post-Lumbar Disc Surgery." *Journal of Clinical Epidemiology* 57, no. 3: 268–276.

Penha, L. M. B., A. Pontes-Silva, A. D. Santos-de-Araújo, et al. 2023. "Reliability of the Heart Rate Variability Registered Through Polar Cardio Frequency Meter in Individuals With Chronic Low Back Pain." *Journal of Chiropractic Medicine* 22: 180–188.

Pinto, S. M., B. Wright, S. Annaswamy, et al. 2024. "Heart Rate Variability (HRV) After Traumatic Brain Injury (TBI): A Scoping Review." *Brain Injury* 38: 1–22.

Sandercock, G. R., P. D. Bromley, and D. A. Brodie. 2005. "The Reliability of Short-Term Measurements of Heart Rate Variability." *International Journal of Cardiology* 103, no. 3: 238–247.

Santos-de-Araújo, A. D., D. Bassi-Dibai, P. F. Camargo, et al. 2023. "Inter- and Intrarater Reliability of Short-Term Measurement of Heart Rate Variability on Rest in Chronic Obstructive Pulmonary Disease (COPD)." *Heart & Lung* 62: 64–71.

Shaffer, F., and J. P. Ginsberg. 2017. "An Overview of Heart Rate Variability Metrics and Norms." *Frontiers in Public Health* 5: 258.

Weir, J. P. 2005. "Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM." *Journal of Strength & Conditioning Research* 19, no. 1: 231–240.