Advances in Applied Physiology
2022; 7(2): 26-33
http://www.sciencepublishinggroup.com/j/aap
doi: 10.11648/j.aap.20220702.12
ISSN: 2471-9692 (Print); ISSN: 2471-9714 (Online)

Heart Rate Variability: A Longitudinal Comparison of Commercial Devices for Individual and Group Stress-Response

Kaela Hierholzer[1, *], Robert Briggs[1, 2], Michael Tolston[3], Nicholas Mackowski[1], Jason Eckerle[1], Maegan O'Connor[1], Kristyn Barrett[1], Roger Smith[1, 2], Adam Strang[1]

[1]Strong Lab, Air Force Research Laboratory, Wright Patterson Air Force Base, Fairborn, USA

[2]Department of the Air Force, Washington D.C., USA

[3]711 Human Performance Wing, Fairborn, USA

Email address:

kaela.hierholzer.ctr@us.af.mil (Kaela Hierholzer), robert.briggs.8@us.af.mil (Robert Briggs), Michael.tolston.1@us.af.mil (Michael Tolston), nick@nickmackowski.com (Nicholas Mackowski), maegan.oconnor.ctr@us.af.mil (Maegan O'Connor), Kristyn.barrett.ctr@us.af.mil (Kristyn Barrett), roger.smith.35@us.af.mil (Roger Smith), Jason.eckerle.1@us.af.mil (Jason Eckerle), adam.strang.1@us.af.mil (Adam Strang)

[*]Corresponding author

To cite this article:

Kaela Hierholzer, Robert Briggs, Michael Tolston, Nicholas Mackowski, Jason Eckerle, Maegan O'Connor, Kristyn Barrett, Roger Smith, Adam Strang. Heart Rate Variability: A Longitudinal Comparison of Commercial Devices for Individual and Group Stress-Response. *Advances in Applied Physiology*. Vol. 7, No. 2, 2022, pp. 26-33. doi: 10.11648/j.aap.20220702.12

Received: August 29, 2022; **Accepted**: September 17, 2022; **Published**: October 11, 2022

Abstract: The collection of heart rate variability (HRV) for health and performance observations have become prominent. However, each wearable device has proprietary algorithms that govern methods and timing of HRV capture and subsequent analysis. The purpose of this study was to evaluate HRV metrics taken from three, commonly used commercial wearables, and identify reliability and relationships to one another over time. *Methods:* Twenty-five subjects (18 males; 7 females) with ages ranging from 23 to 41 years (32.70 ± 4.65 years) were included in this study. These subjects were participants in a 12-week exercise intervention study. Each subject was equipped with a Whoop Strap (v2.0), the Garmin Fenix 5 Smartwatch and chest strap, and the Omegawave chest strap and sensor. *Statistical Analysis:* Between and within-subject correlations were calculated as well as average correlations, descriptive and inferential statistics, and the resultant z-score, which was transformed back into a correlation. Intraclass correlation coefficients (ICC) were calculated. Finally, linear mixed models were used to evaluate trends in HRV. *Results:* Within-subject correlations (0.24 ± 0.27) were lower than between-subjects correlations (0.54 ± 0.43), t (35) = -4.02, $p < 0.001$. Garmin HRV Stress, Whoop RMSSD, Omegawave SDNN, and Omegawave RMSSD yielded an ICC between 0.65 and 0.75. Garmin All-day stress, Garmin prior all-day stress, and Omegawave LF/HF yielded an ICC of 0.30 and 0.37. To test the effects of day of the week on HRV, we fitted linear mixed models to HRV metrics from three of the identified communities related to ICC: Omegawave RMSSD (moderate to high ICC), Omegawave LF/HF (low to moderate ICC), and Whoop recovery score (very low ICC). There was a main effect of gender on Omegawave RMSSD ($p = 0.020$) and a negative effect of day of the week ($p = 0.030$). Day of the week was the only significant predictor of Whoop recovery score ($p < 0.001$). *Conclusion:* The correlations of HRV values remain more consistent when assessed at similar times of the day, rather than being device dependent. Regardless of which wearable device is considered, HRV measures should be collected at a specific time each day for the best reliability. When creating an individualized or group exercise program, the human performance specialist should be aware that fatigue may become increasingly evident during the course of each week (e.g. individuals demonstrably fatigued by Friday may exhibit physiological indicators of relative recovery by Monday).

Keywords: Heart Rate Variability, Wearables, RMSSD, SDNN

1. Introduction

Heart rate variability (HRV) is defined as the beat-to-beat variation in heart rate (HR) [1]. Research findings indicate that HRV can be used for physiological stress and recovery monitoring [2, 3]. Resting or post-exercise HRV can be used for the identification of various training adaptations, which can also be observed by monitoring autonomic nervous system (ANS) activity [3]. For instance, increases in fitness and exercise performance are thought to be associated with increases in vagal-related indices of HRV, while negative adaptations to training are generally associated with overtraining and subsequent reductions in vagal-related indices of HRV [4]. ANS cardiac dysregulation underlies the manifestation and perpetuation of symptoms of poor health. HRV commonly predicts morbidities from common mental (e.g., stress, depression, anxiety) and physical disorders (e.g., inflammation, chronic pain, insomnia, fatigue, etc) [5]. The evaluation of HRV allows coaches and practitioners to closely monitor and regulate the balance of training and recovery in an ongoing effort to improve fitness and performance [6] while maintaining general wellness.

The spotlight on HRV monitoring in recent years has resulted in increased demand for non-invasive and efficient means to assess this physiological response. The commercial industry has responded with a range of fitness wearables capable of capturing and recording HRV variables. [7]. However, most of these wearables use “black-box” and proprietary HRV-based metrics, limiting data transparency (i.e., it is unclear if and how traditional HRV variables are combined or transformed to create proprietary metrics). Therefore, there is a need for more evidence supporting the reliability/validity of any of the many popular commercial devices, since such data is lacking. Furthermore, there are likely individual differences in HRV metrics that need to be considered when evaluating them for reliability and validity. Thus, it is prudent to examine HRV baseline values and adaptations in an individual, while comparing across a cohort of individuals to better capture relationships between HRV metrics and their associations with health and performance.

Common & Derivative HRV Metrics

Multiple complex variables can be evaluated when assessing HRV. Time between normal heartbeats—the time interval between R peaks of the QRS complex [8]—is known as NN. NN intervals derive from the “normal” R-R interval. In other words, these intervals represent normal cardiac timing and are intended to be free from artifact. In addition, the standard deviation of NN intervals (SDNN) are measured in milliseconds and represent the variability of the IBI segment. [8]. Both sympathetic and parasympathetic nervous systems play a role in SDNN, a metric that correlates closely with low frequency (LF) power. LF power is hypothesized to be indicative of an index of vagal-cardiac nerve traffic [8]. SDNN is the gold standard for HRV metric cardiac risk stratification when measured over a 24-hour period [8]. While SDNN is highly correlated with LF power; the root mean square of successive RR interval differences (RMSSD) reflects the beat-to-beat variance (in milliseconds) over ~5 minutes and is highly correlated with high frequency (HF) power [8]. RMSSD is thought to be more heavily influenced by the parasympathetic nervous system (i.e. vagal traffic) (PNS) as compared to SDNN [8].

Low frequency power is typically recorded over a minimum 2-minute period [9] and reflects baroreceptor activity during resting conditions [10]. LF power is derived from the low-frequency band (0.04-0.15 Hz) and is governed primarily by the sympathetic nervous system (SNS) [8]. High frequency (HF) power is derived from the high-frequency band (0.15-0.4 Hz) and is primarily regulated by the PNS [8]. The LF to HF ratio (LF/HF) is the ratio between low and high frequency power on the HRV spectrum and is an estimate of the activity ratio of the SNS and PNS under controlled conditions [8]. Though somewhat controversial, the variance associated with measurement conditions such as duration of data sample collection, SNS/PNS interactions, and SNS/PNS activation under various test conditions; LF/HF Ratio is believed to represent the sympatho-vagal balance with the SNS impacting the LF power and PNS influencing HF power [11, 9, 8]. In this model, low LF/HF ratio reflects parasympathetic dominance, while a high LF/HF ratio is indicative of sympathetic dominance—as seen in fight or flight behaviors or PNS withdrawal [8].

In addition to the variables listed above, there are specific collected measures that are unique to commercial wearables. The Garmin Fenix 5 watch (Garmin Ltd, Olathe, KS) captures a large amount of information, including location from the global positioning system (GPS), movement from an accelerometer, and heart rate (labeled as pulse rate). The Garmin watch is intended to provide the user with information to help guide their daily training and provides an array of variables and relevant features to do so, including recovery time, training load, HRV stress, all-day stress, and many more [12]. The Garmin Fenix 5 is a watch-based wrist-worn wearable that was accompanied with a chest strap to collect biostatistical information across a 24-hour period, using electrodes and infrared LEDs to record the most variables of the three products compared in this study [13]. The variables from Garmin highlighted in this paper include HRV stress and all-day stress. All-day stress is a proprietary algorithm that incorporates an individual’s sleep, daily stress, and physical stress [14]. The wrist-worn Whoop Strap tracks an individual’s sleep, recovery, and strain based on variables such as: resting heart rate (RHR), respiratory rate, HRV, and sleep (Whoop Inc, Boston, MA). The proprietary variable from Whoop used in this study is recovery score. Whoop claims to be the gold standard in sleep tracking for wearable devices, as it captures individual biostatistics throughout the day, and measures sleep, recovery, and strain overnight [15]. The algorithm for recovery score utilizes HRV and is suggested to help an individual make appropriate training decisions based on their recovery score. The third device used in this study was the Omegawave (Omegawave Ltd, Espoo, Finland), which utilizes a chest strap and electrodes and claims to assess both the brain and heart over a 4-minute timespan while the individual is supine and at rest [16]. The Omegawave system analyzes the biological data from the central nervous system (CNS) and HRV to provide a "readiness status" that can help the individual determine how to optimally train for that day [16]. Omegawave includes three omega sensors and three ECG sensors. The omega sensors were designed to measure DC potential—a physiological response related to ultraslow (<0.5 Hz) brain wave activity — while the ECG sensors measure the cardiac and metabolic systems. DC-potential reportedly captures the resting cortical activity of the brain and helps detect CNS fatigue. Like the technologies used in other wearable devices, Omegawave takes the data collected and calculates proprietary variables, such as aerobic index, fatigue, and readiness. In contrast to the Garmin and the Whoop systems, Omegawave is collected once daily, over a pre-exercise 4-minute duration [16].

For this study, the subjects wore the Whoop Strap (Whoop, Boston, MA), the Garmin Fenix 5 Smartwatch and accompanying chest strap (Garmin, Olathe, KS), and the Omegawave chest strap and sensor (Omegawave, Espoo, Finland) daily over the course of a 12-week exercise training regimen. The purpose of the current study was to investigate HRV metrics via three separate devices at varying times and their subsequent relationship – or lack thereof – by analyzing the inter-reliability, as well as evaluate HRV over time. Our primary hypotheses are as follows: 1) We expect to find the highest correlations between readings from the same device to be the Omegawave metrics taken pre-workout; 2) Similarly, we anticipate the highest correlations from different devices to be between Omegawave and Garmin HRV stress taken pre-workout since these variables were captured at similar times each day (immediately preceding the workout); 3) Furthermore, we hypothesize the Garmin pre-workout HRV and the all-day stress variable to yield a moderate correlation between readings of the same device, taken at various time points throughout the day/night; 4) Lastly, taking readings at different times is expected to yield a lower correlation between HRV metrics across the three wearable devices. This fourth expectation is due to the data being collected at varying times throughout the day/night, and the differences in data collected from each device, since each device relies on its own proprietary algorithms.

As previously mentioned, the various devices that collect and analyze HRV, HR, and the derivatives each use different calculations in their respective algorithms. Due to the "black-box" nature of each wearable's propriety algorithms and the lack of current literature in this area, research is necessary to compare these devices, and their reliability in predicting readiness and identifying fatigue. Such information can arm the coach/practitioner to refine training parameters based on enhanced understanding of collected HRV metrics.

2. Methods

2.1. Subjects

Subjects included a convenience sample of healthy, active-duty military adults, who were recruited for participation in STRONG lab exercise sessions. The STRONG lab is an exercise laboratory in the Air Force Research Laboratories (Wright-Patterson Air Force Base, Ohio) designed to optimize and research human performance. Inclusion criteria for this study required subjects to be active-duty service members between the ages of 18 and 45. Subjects who were unable/unwilling to commit to participating in this study for 14 consecutive weeks (12 training weeks and 1 week each for baseline and post-training testing) were excluded from the study. In addition, those who were currently on a medical or pregnancy profile, currently breastfeeding, taking prescribed blood pressure medication, or undergoing hormone therapy were also excluded. Also, candidates who were unwilling to discontinue herbal dietary supplements, performance supplements, or other substances which contain ingredients that could affect cardiovascular response with exercise were excluded. Finally, those with a history of abdominal hernia surgery, and those suffering from a musculoskeletal injury, or cardiovascular/respiratory disease were excluded. Twenty-five subjects (18 males) with ages ranging from 23 to 41 years (32.62 ± 4.56 years) finished the study with complete data across all devices. To have successfully completed the study, each subject was required to complete at least 80% of the prescribed exercise sessions.

Table 1. *Demographics.*

Gender	18 males; 7 females
Age (years)	32.62 ± 4.56
Height (in)	68.59 ± 2.89
Weight (lb)	180.96 ± 31.89
Body Fat (%)	29.16 ± 7.45
Resting HR (bpm)	70.20 ± 15.57
VO_{2max} (mL/kg/min)	47.68 ± 11.46

2.2. Design

The experimental design on this study was an observational research study using repeated measures with the following factors: gender, training week, and day of week.

2.3. Methodology

The subjects participated in a 12-week exercise training program, with exercises scheduled five days a week (excluding holidays). The five-day training program was comprised of 45–60-minute training sessions with 3 days of strength training and two days of cardiovascular training. Due to incomplete data, cardiovascular training days were excluded from the analysis. The circuit included various strengthening exercises with minimal rest period to simultaneously target the subject's cardiovascular fitness/stamina and strength. Each strength training day was designed to target all major muscle groups with core work performed on cardiovascular training days. The goal of the training program was to improve general strength and fitness. HRV data were collected using a Whoop Strap (Whoop, Boston, MA), the Garmin Fenix 5 Smartwatch and strap (Garmin, Olathe, KS), and the Omegawave chest strap and sensor (Omegawave, Espoo, Finland). During the study, the subjects were instructed to wear the Fenix 5 (chest strap was worn during exercise only) and the Whoop all day for the entirety of the study. The Omegawave was only worn just prior to exercise sessions to collect biometrics and then was removed (in accordance with manufacturer recommendations).

2.4. Statistical Analysis

The Omegawave analysis consisted of industry standard metrics which analyze NN intervals. These metrics include: SDNN, RMSSD, and the LF/HF ratio. The metrics obtained daily from the Whoop Strap were "Recovery Score" and RMSSD. Both Whoop metrics were captured during sleep the night before each exercise session. From the Fenix 5 we obtained the "HRV Stress" metric just before each exercise session and "All-Day Stress" metric, which was derived from 24-hour wear/use of the device. Data from the following variables were log-normal and log-transformed prior to analysis: Whoop RMSSD, Omegawave SDNN, Omegawave RMSSD, and Omegawave LF/HF.

Due to the absence of RPE reporting during cardiovascular training days these days were excluded from analysis. The first three weeks of HRV data were excluded from analysis since some individuals did not begin consistently wearing their device until week two and prior results indicated that the correlational structure among some survey responses did not stabilize until this time [17]. Between-subjects analyses were calculated based on aggregated data and within-subject analysis of variations using the repeated measures correlation function available in R [18]. Statistical significance was set to p-value ≤ 0.05. To calculate average correlations, the Fisher z-transformation was performed, descriptive and inferential statistics were calculated, and the resultant z-score was transformed back into a correlation. Intra-class correlation coefficients (ICC) were calculated to estimate the variance in each variable attributable to subject-specific variation. Finally, linear mixed models were used to evaluate trends in HRV. All statistical procedures were conducted using R, version 3.6.2 (R Core Team, Vienna, Austria), except mean and standard deviations which were calculated in Excel (Microsoft Inc., Redmond, WA, USA).

3. Results

Within-subject correlations (0.24 ± 0.27) were generally lower than between-subjects correlations (0.54 ± 0.43), t (35) = -4.02, $p < 0.001$. In other words, covariations of HRV metrics around participants' mean values were less strongly correlated than overall average HRV values from participants around grand means. There was a moderate positive relationship among between- and within-subjects correlations (r = 0.65). For between-subjects correlations, significant relationships were found in 18 out of 28 possible pairs of HRV variables. To identify patterns of connectivity, community detection was conducted on the threshold between-subjects correlation matrix using the Louvain method [19]. Three communities were identified. The first community consisted of Garmin HRV Stress, Whoop RMSSD, Omegawave SDNN, and Omegawave RMSSD. While the ICC for these variables were not entered into the clustering algorithm, they do appear to be related to the magnitude. The first communities' ICC was between 0.65 and 0.75 (see Table 2 and Figure 1 left panel). The second community consisted of Garmin all day stress, Garmin prior all-day stress, and Omegawave LF/HF. The ICC for these metrics ranged from 0.30 to 0.37 (see Table 2 and Figure 1 left panel). The final community was Whoop Recovery Score, which did not correlate significantly with any of the other variables and yielded a very low ICC of 0.07 (see Table 2 and Figure 1 left panel). The lines in the left panel of Figure 1 indicate positive (green) or negative (red) correlations and the thickness is proportional to the magnitude of correlation, while the middle panel displays the proportion of variance that is accounted for in each variable by subject specific variability, and the right panel displays the within subjects correlation around their own means. These values indicate that clusters in the between-subjects HRV metrics were partitioned in accordance with the amount of subject-specific variability, with three classes (moderate to high, low to moderate, and very low ICC values).

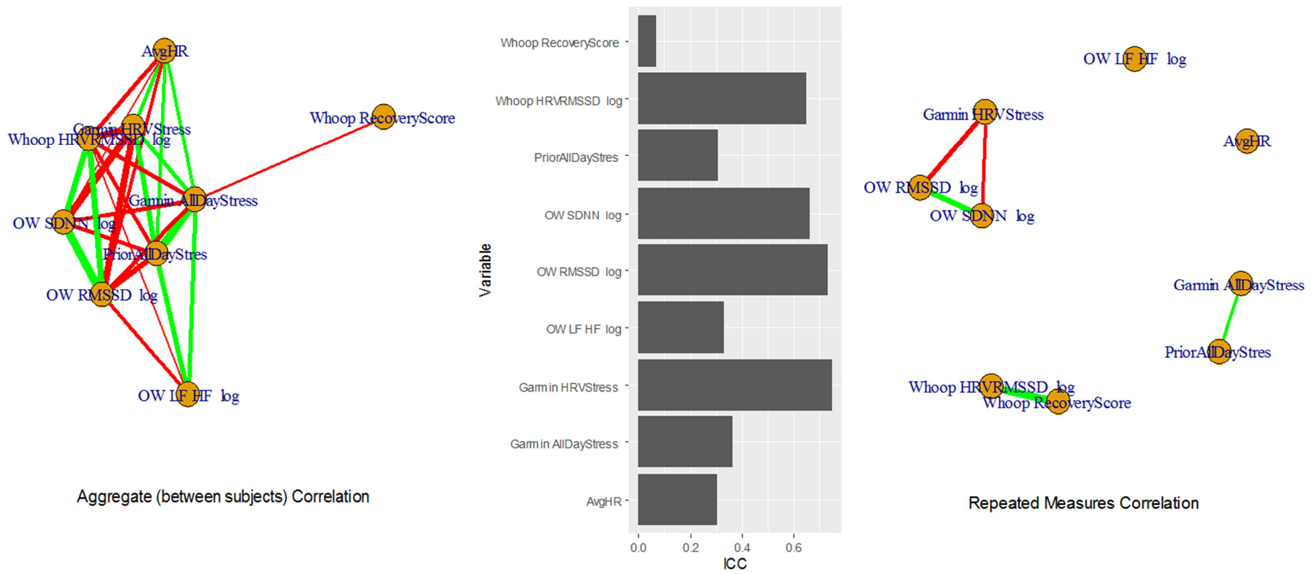

Figure 1. *Between Subject Correlations (left), ICC proportions of Variance (center), and Within Subject Correlations (right).*

Table 2. *HRV Metrics in the Between Subjects Correlation Matrix.*

Variable	ICC
Garmin HRV Stress, Whoop RMSSD, Omegawave SDNN, Omegawave RMSSD	0.65-0.75
Garmin All-Day Stress, Garmin Prior All-Day Stress, Omegawave LF/HF	0.30-0.37
Whoop Recovery Score	0.07

*Note: The clustering structure appears to be related to the magnitude of the ICC for the variables, but ICC was not entered into the clustering algorithm.

For repeated-measures correlations, significant relationships were found between Omegawave SDNN, Omegawave RMSSD, and Garmin HRV Stress. Whoop recovery score and Whoop RMSSD were also significantly related.

To test the effects of day of the week on HRV, we fit linear mixed models to HRV metrics from variables from each of the identified communities: Omegawave RMSSD (moderate to high ICC), Omegawave LF/HF (low to moderate ICC), and Whoop recovery score (very low ICC). The factors and covariates included in models predicting were week (to capture longer trends in HRV over the course of the study), week^2 (to capture quadratic trends), day of the week (to capture weekly cycles), gender, age, and all possible interactions of these variables. These variables were also included as random effects. The best-fitting, parsimonious model was selected using likelihood ratio tests of nested models.

There was a main effect of gender on Omegawave RMSSD. Males, on average, demonstrated a higher RMSSD, (p = 0.016) (see Table 3). There was a negative effect of day of the week: RMSSD consistently started out high on Mondays and ended lower on Fridays, (p = 0.032) (see Table 3). This trend can be seen in Figure 2, below.

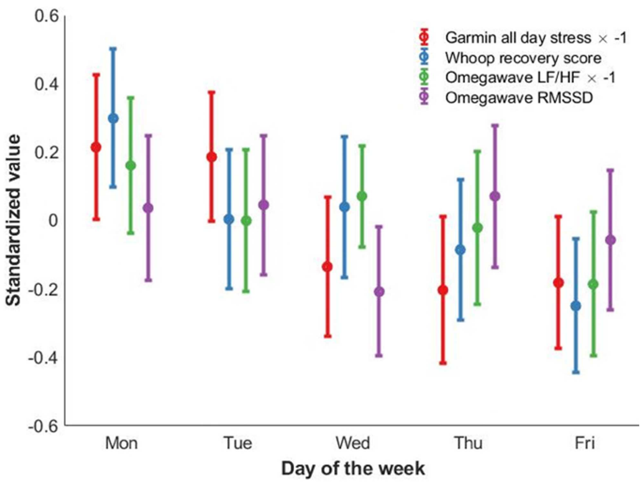

Figure 2. *HRV Metric Comparison 1-Week Trend.*

There was also a significant quadratic trend over the course of the training protocol; RMSSD decreased slightly during the first half of the study but increased slightly during the second half (p = 0.001) (see Table 3). For Omegawave LF/HF, there was a significant effect of day of week. This measure consistently started lower on Mondays and ended higher on Fridays (p = 0.002; see Table 4). Table 4 also displays there was a significant week by day of week interaction (p = 0.001), indicating that the weekly recovery trend increased over time for Omegawave LF/HF. Day of the week was the only significant predictor of Whoop recovery score, which started high Mondays and ended lower on Fridays (p < 0.001) (see Table 5).

Table 3. *Omegawave RMSSD.*

Predictors	Estimates	CI	p
(Intercept)	3.3	2.94 – 3.65	<0.001
Training Week	0.39	-0.23 – 1.01	0.222
Training Week^2	1.01	0.39 – 1.63	0.001
Day of Week	-0.02	-0.04 – -0.00	0.032
GenderMale [1]	0.52	0.10 – 0.94	0.016
Random Effects			
σ^2	0.1		
τ_{00} Subject	0.22		
ICC	0.7		
N Subject	25		
Observations	468		
Marginal R^2 / Conditional R^2	0.160 / 0.747		

Table 4. *Omegawave LF/HF.*

Predictors	Estimates	CI	p
(Intercept)	0.3	0.05 – 0.54	0.017
Day of Week	0.07	0.03 – 0.11	0.002
Training Week	-0.04	-0.09 – 0.01	0.113
Day of Week × Training Week	0.02	0.01 – 0.04	0.01
Random Effects			
σ^2	0.48		
τ_{00} Subject	0.31		
τ_{11} Subject. Training Week	0		
ρ_{01} Subject	0.48		
ICC	0.4		
N Subject	25		
Observations	468		
Marginal R^2 / Conditional R^2	0.021 / 0.417		

Table 5. *Whoop Recovery Week.*

Predictors	Estimates	CI	p
(Intercept)	60.07	56.25 – 63.88	<0.001
Day of Week	-2.46	-3.78 – -1.14	<0.001
Random Effects			
σ^2	439.36		
τ_{00} Subject	27.37		
ICC	0.06		
N Subject	25		
Observations	468		
Marginal R^2 / Conditional R^2	0.026 / 0.083		

4. Discussion

The primary aim of this study was to determine the inter-reliability of HRV metrics captured during various time-points from Omegawave, Whoop, and Garmin. The results show that reliability of Omegawave RMSSD was moderately acceptable, while reliability of Whoop Recovery Score was low due to a low ICC and a lack of correlation with other variables between subjects. We believe this low ICC from Whoop indicates that its value has been normalized around individual means. Concerns regarding proprietary metrics include lacking information pertinent to how specific metrics were calculated, and lacking means to regenerate or evaluate such calculations for accuracy due to uncertainty, thus leading to speculation regarding score derivation. This uncertainty can make it difficult to ascertain the meaning of one individual's response relative to other individuals.

This study identified that overall correlations from various wearable devices around a grand mean (0.54 ± 0.43) were larger than the correlations around an individual's mean assessed at varying timepoints (0.24 ± 0.27). This supports our expectation and prior evidence that timing is important in terms of HRV collection, as substantial individual variation is evident and typical throughout the day. Thus, where possible, we recommend that the coach/practitioner be consistent in timing of capturing such measures, as this will likely increase confidence in the value of these metrics. In contrast, capturing these metrics at differing times throughout the day should lower the user's confidence in the reliability of the metric for guiding exercise programs or determining readiness of an individual. Moderate correlation noted among various devices, indicates that despite propriety algorithms making it difficult to determine precisely what variables go into each calculation, there are relationships existing among the device, thus increasing face validity of the metrics. With only moderate correlations, caution should be implemented when attempting to interpret these findings to guide training. The coach/practitioner should be cognizant of the fact that this study demonstrated that devices worn over a 24-hour period appeared to be significantly more beneficial for tracking trends in HRV and associated metrics vs those that are only worn for a few minutes a day just prior to exercise.

In our prior work, we analyzed the relationship between HRV metrics from these devices and found that Omega Wave LF/HF was the best overall predictor for physical performance on a given day and that Garmin All Day stress was better than chance [23]. This current study indicates that measures from the Whoop (recovery score and RMSSD) were not significantly related to physical performance. In other words, this wearable did not independently identify an individual's capacity to perform on any given day. Therefore, we suggest that the practitioner should take caution when using this wearable to make decisions about exercise. This device should not be considered as a reliable, independent predictor to govern exercise when used in isolation; and instead, should be considered in context when combined with other wearables' data or additional information from the individual.

Findings in the current study also indicate that Omega Wave LF/HF is the most sensitive variable to training effects and weekly stress patterns, in that it showed improvements in HRV over the course of the entire training program, changes within week values, and an interaction in which the change in weekly recovery increased over time. However, these broader conclusions about the sensitivity of these metrics are limited by the fact that these findings come from the same data sample. Further investigation with these devices in a larger sample appears warranted to further refine our findings.

Our results indicate that both gender and day of the week impacted Omegawave RMSSD. While all subjects started out with higher HRV values on Monday and decreased over the week, men typically demonstrated higher HRV values than women (Figure 3). This finding was expected, as previous research reported similar findings, suggesting that male HRV values are commonly higher than female [20] when comparing individuals of similar fitness level. In separate studies conducted by Unetani et al. (1998), and Saleem, et al. (2012) results indicated that females consistently exhibited lower HRV values compared to males [21].

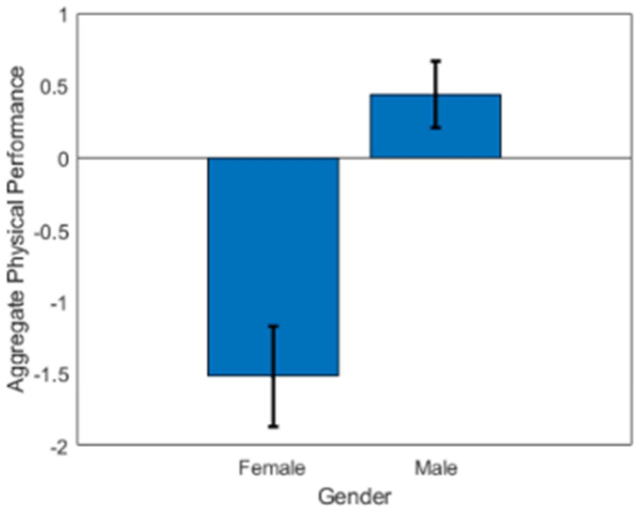

Figure 3. *Aggregate physical performance from projections through performance variables as a function of gender.*

While not novel to this study, we understand that training is one of the primary factors that impacts HRV; specifically, volume, intensity, new stimuli, and work-to-rest ratio [22]. The authors of this report suggest that proper sleep, hydration, and a training plan that allows for adequate recovery are required to progressively and consistently increase HRV [22]. In recent years, the military has more frequently integrated holistic approaches to recognize and consider a combination of factors such as sleep, hydration, nutrition, and recovery as being critical components of multimodal training strategies. Newly implemented initiatives such as holistic health and fitness (H_2F), Preservation of the Force and Family (POTFF), and integrated operational support teams (iOST) consisting of multi-disciplinary teams embedded within specific operational teams to offer a multi-faceted approach to optimal health and wellness. We believe that such approaches are likely to benefit the warfighter while reducing the paradoxical unintended side effect of injury which might occur during training efforts intended to improve fitness. Based on HRV trends identified and outlined in this study, we suggest that such efforts are beneficial and should continue in order to optimize training and performance while mitigating inherent risks associated with training and high-effort performance efforts to improve fitness.

One of the most interesting findings in this cohort was that our subjects showed evidence of fatigue during the course of the week. This trend was captured by multiple devices and metrics included in this study. In addition to a decrease in HRV, there was also a decrease in Whoop Recovery Score, and Garmin All Day Stress score throughout the week, with scores rebounding to begin the subsequent week. This finding is somewhat intuitive, since many individuals subjectively report fatigue throughout the course of the week, as Friday approaches, with renewed vigor and energy at the onset of a subsequent week, but we do think that this finding captured from multiple devices is important. Since exercise specialists typically monitor and regulate training and recovery balance in an effort to optimize fitness and performance, the impact of "Friday fatigue" along with apparent recovery by Monday of each week is critical and should be considered when designing exercise parameters for this population. Fatigue was apparent, due to the nature of a five-day, Monday through Friday progressive strength training program among active-duty military participants who were concurrently required to perform the various responsibilities and tasks of their operational requirements over the course of the week. This finding is somewhat expected, but with multiple devices capturing this trend, we believe this finding warrants further investigation and likely extends beyond a military population. Further investigation can help determine optimal dosage of exercise to yield intended benefits while minimizing excessive fatigue with its inherent injury risk and potential impact on performance and health in multiple domains.

A persistent trend identified by the Garmin All Day Stress metric was apparent in this military cohort throughout the course of training (Figure 2). A steady decline emerged throughout the course of the week, but then on Friday it began to rebound slightly. To properly interpret this finding, it is important to note that our population faces many different stressors throughout the week aside from training. Therefore, the decrease throughout the week may be influenced by mental fatigue, physical fatigue, work related stress, or personal reasons. In turn, the increase observed going into the weekend could be due to multiple factors, including an enhanced mental state or elevated mood as service members approached the end of the work week or perceived an upcoming break from training intensity. This study has found that practitioners can detect macro-level (group) trends in HRV that correspond with the work week. While this may not be beneficial for detecting changes in any one individual on any given day for purposes of intervention, it can be useful for macro-level training/planning. We suggest that the utilization of daily wellness surveys in addition to a wearable device could potentially help explain factors related to a subject's fatigue (individually and group-trends), and recovery over time, and warrants further investigation.

There were some limitations identified in this study. One limitation was the small sample size and high dropout rate in the sample. This emphasizes the challenges of recruiting participants in military populations for long-term training studies and should be considered in future studies. In addition, we required subjects to complete at least 80% of intended workouts in order to be retained for analysis. Compliance at this level is relatively rare for exercise studies of this duration, particularly among military personnel (citations). The data that we collected ascertains that this subgroup of individuals was highly compliant with device wear included in this study, as well as exercise attendance, thus increasing our confidence in the results. Further studies with larger sample size, while maintaining high exercise compliance are critical to further understand the relationships identified in this study.

Our analyses showed that Omegawave RMSSD yielded acceptable reliability compared to Whoop and Garmin. Whoop Recovery Score had low reliability represented by a low ICC for Day of Week Whoop Recovery Score and lacked correlation with additional variables between subjects. Our analyses also showed a trend in Garmin All Day Stress, decreasing as the week progresses, then an up-trend at the end of the week.

5. Conclusion

In the studied sample there appears to be significant variability between three separate devices evaluated in terms of accurate prediction of readiness and fatigue. Algorithmic, proprietary data, unspecified for the consumer, further complicates the variability of these devices. Coaches and practitioners should consider this when relying on a single device for determining parameters for upcoming workouts. Further research might help determine which combination of wearables and/or which combination of subjective information reported from the individual should be combined with wearable device output in order to optimize a training regimen for the individual. Regardless of device used, a trend of increasing fatigue over the course of the week emerged. This information should be considered when designing an exercise program as being cognizant of this information is likely to help mitigate injury risk, avoid chronic burnout, and optimize performance. In conclusion, the Omegawave provides adequate reliability for HRV metrics and performance testing. In order to better track overall readiness for our active-duty military personnel, it is suggested that a daily wellness survey be used in addition to a HRV device due to the outside stressors our military population faces, such as physical fatigue, personal stressors, mental fatigue, and work-related stressors. The added information would likely help determine the reason for the various trends and relationships identified in this study.

6. Recommendations for Future Work

It is recommended that more research be performed on this topic. Specifically, further investigation into the effects of exercise dosage, timing, and fatigue should be explored due to the significant impact on training and performance. There is a necessity to conduct this research with a larger cohort and updated versions of these (and perhaps other) wearable devices to increase confidence in the results obtained in this study. Finally, a self-reported "daily wellness" survey paired with the collected HRV metrics might provide additional insight into the potential impact of factors linked to daily activities and general wellness trends on HRV readings.

References

[1] Billman, G. E. (2011). Heart rate variability – a historical perspective. Clinical and Translational Physiology, 2 (96), 1-13.

[2] Barrero, A., Schnell, F., Carrault, G., Kervio, G., Matelot, D., Carre, F., Lahaye, S. L. (2019). Daily fatigue-recovery balance monitoring with heart rate variability in well-trained female cyclists on the Tour de France circuit. PLOS ONE, 14 (3).

[3] Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. Sports Medicine, 43, 773-781.

[4] Task Force of The European Society of Cardiology. (1996). Heart rate variability. European Heart Journal, 17, 354-381.

[5] Jung, W., Jang, K., & Lee, S. (2019). Heart and brain interaction of psychiatric illness: A review focused on heart rate variability, cognitive function, and quantitative electroencephalography. Clinical Psychopharmacology and Neuroscience, 17 (4), 459-474.

[6] Dong, J. (2016). The role of heart rate variability in sports physiology (review). Experimental and Therapeutic Medicine, 11 (5), 1531-1536.

[7] Schuurmans, A. T., de Loof, P., Nijhof, K. S., Rosada, C., Sholte, R. H., Popma, A., & Otten, R. (2020). Validity of the Empatica E4 wristband to measure heart rate variability (HRV) parameters: A comparison to electrocardiography (ECG). Journal of Medical Systems, 44 (190), 1-11.

[8] Shaffer, F. & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. Frontiers in Public Health, 5 (258).

[9] Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. Frontiers in Psychology, 5, 1040.

[10] McCraty, R. & Shaffer, F. (2015). Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. Global Advances in Health and Medicine, 4 (1), 46-61.

[11] Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. Frontiers in Physiology, 4, 26.

[12] Firstbeat Technologies Ltd. (2014). Stress and recovery analysis method based on 24-hour heart rate variability. White Paper by Firstbeat Technologies Ltd., 1-13.

[13] Garmin Ltd. (2021). Fenix® 5 Specs. Garmin. https://buy.garmin.com/en-US/US/p/552982#specs.

[14] Firstbeat Technologies Ltd. (2019). A sleep analysis method based on heart rate variability. White Paper by Firstbeat Technologies Ltd., 1-7.

[15] Whoop, Inc. (2021). Experience Whoop. WHOOP. https://www.whoop.com/experience/.

[16] Omegawave, Inc. (2021). Omegawave. Omegawave. https://www.omegawave.com.

[17] Tolston, M. T., O'Connor, M., Mackowski, N., & Strang, A. J. (2019). Evaluating the structural and temporal aspects of responses to a short daily wellness questionnaire. Poster presented at the 2019 INFORMS Annual Meeting. Seattle, WA.

[18] Bakdash, J. Z. & Marusich, L. R. (2107). Repeated measures correlation. Frontiers in Psychology, 8, 456.

[19] Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 10, P10008.

[20] Unetami, K. Singer, D. H., McCraty, R. & Atkinson, M. (1998). Twenty-four hour time domain heart rate variability and heart rate: Relations to age and gender over nine decades. Journals of American College of Cardiology, 31 (3), 593-601.

[21] Saleem, S., Hussain, M. M., Majeed, S. M., & Khan, M. A. (2012). Gender differences of heart rate variability in healthy volunteers. Journal of the Pakistan Medical Association, 62 (5), 422.

[22] Van Deusen, M. (2019, September 30). Everything You Need to Know About Heart Rate Variability (HRV). WHOOP. https://www.whoop.com/thelocker/heart-rate-variability-hrv/.

[23] Hierholzer, K., Briggs, R., Tolston, M., Mackowski, N., O'Connor, M., Barrett, K., Strang, A. (2022). Heart Rate Variability Metrics from Commercial Devices Predicts Strength and Cardiovascular Performance in a Military Cohort. American Journal of Sports Science, 10 (1), 14.