EXPLORING THE UTILITY OF UNOBTRUSIVE METHODS OF ACQUIRING HEART RATE VARIABILITY AT REST AND FOLLOWING EXERCISE

by

WARD C. DOBBS

MICHAEL R. ESCO, COMMITTEE CHAIR
MICHAEL V. FEDEWA, COMMITTEE CO-CHAIR
HAYLEY V. MACDONALD
JONATHAN E. WINGO
RANDALL E. SCHUMACKER
PHILIP A. GABLE

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Kinesiology
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2018

ABSTRACT

Use heart rate variability (HRV) for monitoring cardio-autonomic perturbations in response to physical stimuli has increased in popularity and demand. However, the utility of measures of HRV to reflect acute changes in muscle recovery from exercise is controversial. As criterion short-term HRV recordings are performed with a gold standard electrocardiogram (ECG), less obtrusive methods have been developed for improved practicality of HRV measures. Three studies were performed to investigate the accuracy of less obtrusive methods for acquiring HRV and the utility of HRV for tracking changes in muscular performance recovery. The first study involved a systematic review and meta-analysis on the accuracy of portable devices for acquiring HRV. Twenty-three studies yielded 301 effects and revealed that HRV measures acquired from portable devices differed from those obtained from ECG (*ES*=0.23, 95% *CI*: 0.05, 0.42), although this effect was small and highly heterogeneous (I^2=78.6%, 95% *CI:* 76.2%, 80.7%). Moderator analysis revealed that HRV metric (p<0.001), position (p=0.033), and biological sex (β=0.45, 95% *CI:* 0.30, 0.61; p<0.001), but not portable device, modulated the degree of absolute error. Within metric, absolute error was significantly higher when expressed as SDNN (*ES*=0.44) compared to any other metric but was no longer significantly different after a sensitivity analysis removed outliers. Likewise, the error associated with the tilt/recovery position was significantly higher than any other position and remained significantly different without outliers in the model. In the second study, the time course in recovery between criterion short-term HRV measures and acute muscular performance 72 hours following an exhaustive bout of resistance training was investigated. All HRV metrics had a significant interaction with

muscular performance (*performance*) over time ($p < .01$) indicating change scores in *performance* and HRV following the physiological stressor were not parallel and did not track. Mean change scores in all HRV metrics significantly differed from *performance* across time ($p < .05$), except the standard deviation of all normal-to-normal R-R intervals (SDNN), low frequency power (LF), and the standard deviation of long-term HRV from the Poincaré plot (SD2) at the 0.5-hr mark, and high frequency power (HF) at the 24-hr time point. Furthermore, repeated measures correlation analysis indicated a lack of intra-individual association between the change in *performance* and HRV over time (all < .45). In the third study, the agreement between ultra-short and criterion short-term HRV measures surrounding a bout of exhaustive resistance training was investigated. Results displayed the highest levels of agreement from the log-transformed (ln) root mean square of successive R-R differences (lnRMSSD) [LOA = -0.91–0.69, *ICC* = .91, $p = .082$, *ES* = 0.15] and the standard deviation of the points through the width of the plot (lnSD1) [LOA = -0.90 – 0.72, *ICC* = .91, $p = .156$, *ES* = 0.13] compared to all other metrics.

DEDICATION

Dedicated to my parents and siblings who have provided me with never ending support throughout my life and the work ethic to complete this dissertation. Thank you for all you have done, I cannot express how thankful I am to have all of you in my life.

LIST OF ABBREVIATIONS AND SYMBOLS

1RM	one-repetition maximum
ANS	autonomic nervous system
BL	baseline
BMI	body mass index
CE	constant error
Cl	confidence interval
cm	centimeter
CMJ	counter movement jump
ECG	electrocardiogram
ES	effect size
FFT	fast fourier transform
HF	high frequency power in normalized units
HR	heart rate
hr	hour
HRV	heart rate variability
ICC	intraclass correlation coefficient
kg	kilogram
LF	low frequency power
LF:HF	ratio of low frequency power to high frequency power
ln	natural log transformation

LOA	limits of agreement
M	mean
min	minute
NN	normal to normal
nu	normalized units
PNS	parasympathetic nervous system
PPG	photoplethysmography
r	Pearson product moment correlation
RMSSD	root mean square of successive R-R differences
s	second
SD	standard deviation
SD1	standard deviation of short-term HRV from the Poincaré plot
SD2	standard deviation of long-term HRV from the Poincaré plot
SDNN	standard deviation of all n intervals
SE	standard error
SMD	standardized mean difference
SNS	sympathetic nervous system
$STARD_{HRV}$	standard for reporting diagnostic accuracy studies guidelines for HRV research
TP	total power

ACKNOWLEDGMENTS

There are several people who assisted in the completion of this dissertation, there contributions were truly invaluable. First I would like to thank my committee. My chair, Dr. Michael Esco and my co-chair, Dr. Michael Fedewa, your support and mentorship throughout this process was tremendous. You both provided me with invaluable guidance toward pursuing my doctoral degree. To Dr. Jonathan Wingo, Dr. Hayley MacDonald, Dr. Randall Schumacker and Dr.Philip Gable, thank you for all your support throughout my doctoral experience. It has been my pleasure to work with you and I could not have asked for a better committee.

Second, I would like to acknowledge the participants in this study. It takes a special individual to endure the physical demands of the back squat protocol utilized within this dissertation. Third, I would like to thank the research team I have had the privilege to work with as I am truly thankful for their assistance. In particular, I would like to recognize Dano Tolusso and Clifton Holmes. These two individuals were incredibly reliable and provided tremendous insight and support throughout the data collection process. Without them, I would not have been able to complete this dissertation.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

As the autonomic nervous system (ANS) responds to threats to homeostasis, physical and mental, measures of cardio-autonomic perturbations through heart rate variability (HRV) have become a valuable tool for researchers and practitioners alike. HRV refers to the variability in consecutive heart beats and is considered a non-invasive marker of cardiovascular autonomic control [1] which is a balance between the parasympathetic nervous system (PNS) and sympathetic nervous system (SNS). Therefore, the evaluation of these branches has become a useful indicator of health, and disease [2, 3]. To this end, HRV has emerged as an objective physiological marker for monitoring athletic performance and responses to exercise training [4-6].

Measurements of HRV are traditionally performed in short-term 10-min recordings (5-min recording following a 5-min stabilization period) with the gold standard of an electrocardiogram (ECG) [1]. However, with the growing interest in HRV monitoring, less obtrusive methods, such as, ultra-short recordings (1-min recording following a 1-min stabilization period) have been derived [7, 8]. Ultra-short recordings have shown acceptable agreement against traditional short-term ECG recordings previously recommended by the HRV Task Force [1]. This shortened method of HRV data collection is beneficial to the practitioner as it allows for increased practicality [8].

Conventionally, HRV obtained from heart rate data from an ECG requires specialized software for calculation which can be expensive and require a certain level of technical

knowledge for interpretation [1]. Thus, HRV measurements have primarily been limited to laboratory or clinical settings. However, as wearable technology continues to increase in popularity [9, 10] advances in technology have provided many commercially available systems which include HRV as a feature. From as early as 2003, the ability of these various portable devices to extract HRV data has been validated against the gold standard of an ECG [11-13]. However, the error in HRV measures associated with the various portable devices as a whole has yet to be quantified.

Recently, there has been a strong interest for the utilization of HRV monitoring in the team setting in the attempt to prevent non-functional overreaching or worse, overtraining [14]. Typically, longitudinal HRV monitoring has been performed in aerobic demanding events (e.g., soccer) [15, 16]. Therefore, less is understood about the utility of HRV monitoring for acute responses to resistance exercise. Currently, only one study has investigated the ability of HRV to monitor acute muscular performance changes [17]. The respective authors determined that the frequency domain HRV metric high frequency power (HF) mirrored muscular performance over a 72-hr period following a high intensity bout of resistance training. However, only frequency derived metrics were utilized within the analysis. As such, the utility of time domain HRV metrics, such as the root mean square of successive R-R intervals (RMSSD), which are simpler to derive and readily available on most portable device applications, have yet to be evaluated.

While less obtrusive measures of HRV have shown promise as a means for acquiring non-invasive measures of cardio-autonomic perturbations, their utility and accuracy is still not fully understood. Therefore, the purposes of this dissertation were to: quantify the accuracy of HRV measures extrapolated from portable devices, determine if HRV can be used as a non-invasive objective marker of muscular performance recovery, and determine whether ultra-short

recordings provide accurate HRV measures obtained in response to a bout of strenuous resistance training. The specific aims were as follows:

Study 1: Perform a systematic review and meta-analysis of HRV literature to examine the agreement between HRV measures recorded using a portable HRV device and an ECG, and to identify study-level moderators that can explain a meaningful proportion of the observed variance between HRV measurement techniques.

Study 2: To determine which metric of HRV from frequency-domain, time-domain, and Poincaré plotting, best reflects recovery of muscular performance up to 72-hr following a strenuous bout of resistance exercise in strength-trained participants.

Study 3: To determine the agreement between ultra-short and traditional short-term measures of HRV on linear and non-linear metrics at rest and 72-hrs following a strenuous bout of resistance exercise.

CHAPTER 2

THE ACCURACY OF ACQUIRING HEART RATE VARIABILITY FROM PORTABLE DEVICES: A SYSTEMATIC REVIEW AND META-ANALYSIS

ABSTRACT

Advancements in wearable technology have provided practitioners and researchers the ability to conveniently measure various health and/or fitness indices. Specifically, portable devices have been devised for convenient recordings of heart rate variability (HRV). Yet, their accuracies remain questionable. Therefore the objective was to quantify the accuracy of portable compared to electrocardiography (ECG) for measuring a multitude of HRV metrics and to identify potential moderators of this effect. This meta-analysis was conducted in accordance with the PRISMA (*P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses) statement. Articles published before July 29, 2017 were located *via* four electronic databases using a combination of the terms related to *HRV* and *validity*. Separate effect sizes (*ES*), defined as the absolute standardized difference between the HRV value recorded using the portable device compared to ECG, were generated for each HRV metric (10 metrics analyzed in total). A multivariate multi-level model, incorporating random-effects assumptions, was utilized to quantify the mean *ES* and 95% confidence interval (*CI*) and explore potential moderators. Twenty-three studies yielded 301 effects and revealed that HRV measures acquired from portable devices differed from those obtained from ECG (*ES*=0.23, 95% *CI*: 0.05, 0.42), although this effect was small and highly heterogeneous (I^2=78.6%, 95% *CI:* 76.2%, 80.7%).

Moderator analysis revealed that HRV metric (p<0.001), position (p=0.033), and biological sex (β=0.45, 95% *CI:* 0.30, 0.61; p<0.001), but not portable device, modulated the degree of absolute error. Within metric, absolute error was significantly higher when expressed as SDNN (*ES*=0.44) compared to any other metric but was no longer significantly different after a sensitivity analysis removed outliers. Likewise, the error associated with the tilt/recovery position was significantly higher than any other position and remained significantly different without outliers in the model. Our results suggest that HRV measures acquired using portable devices demonstrate a small amount of absolute error when compared to ECG. However, this small error is acceptable when considering the improved practicality and compliance of HRV measures acquired through portable devices in the field setting. Practitioners and researchers should consider the cost-benefit along with the simplicity of the measure when attempting to increase compliance in acquiring HRV measures.

INTRODUCTION

Advances in technology have produced an array of portable devices that can measure health and fitness outcomes under ambulatory conditions. Indeed, the American College of Sports Medicine found that "wearable technology" was the top fitness trend in 2017 [1] and is expected to remain among the top trends for 2018 [2]. Despite the popularity of "wearable technology", few portable devices are capable of accurately measuring physiologic responses to training and/or tracking athletic performance. Particularly needed are portable devices that can precisely gauge recovery status and readiness to perform.

Heart rate variability (HRV) has emerged as an objective physiological marker for monitoring athletic performance and responses to exercise training [3-5]. HRV refers to the variability between successive heart beats and is considered a non-invasive marker of

cardiovascular autonomic control [6]. Traditionally, HRV is determined by obtaining heart rate data from an electrocardiogram (ECG) and then using specialized software for calculation [6]. These processes are relatively expensive and require a certain degree of technical knowledge for interpretation. Thus, the measurement of HRV has largely been limited to laboratory or clinical settings. However, due to advances in technology and increasing interest in portable monitoring devices, many commercially available systems now routinely include HRV as a feature and have been validated as early as 2003 [7].

Several field investigations have examined the utility of HRV measures obtained using portable devices as a tool to monitor training responses and athletic performance [3, 8-10]. Researchers found that HRV was sensitive to changes in training load and able to predict performance outcomes among athletes participating in competitive soccer [8, 9, 11], futsal [12], football [13] and cross-country [14]. However, because heart rate (HR) is regulated by several different control mechanisms, one of which is central command, a number of physiologic factors (e.g., pulmonary ventilation, circulation, and endocrine regulation) and external factors (e.g., sleep quality, nutrition, psychological stressors and exercise) can also influence HRV [15]. Thus, the day-to-day changes in HRV has become a marker of homeostatic perturbation during athletic training [12].

Although the validity of many portable HRV devices has been established [16], these studies vary widely in terms of experimental design and methods used to assess HRV. For instance, HRV has been assessed at rest [17-19], during exercise [20-22] and in differing body positions (i.e., seated [23-25], supine [17, 18, 26-28], standing [29, 30] and tilted [18, 31]). Some portable systems measured HRV under ultra-shortened conditions of approximately one minute [26, 27], which is substantially less than the short-term recordings (5 minutes) found in

traditional recommendations [6]. Furthermore, the validity of these devices depends on the metric used to quantify HRV (see Table 2.1 for common HRV metrics). For example, time domain measures acquired through portable devices have been found to possess good agreement with ECG measures (i.e., root mean square of successive R-R differences (RMSSD) [17, 30]), while some metrics attained through spectral analysis have shown poor agreement with ECG (i.e., low frequency (LF) [32, 33]). These factors are important to consider when measuring HRV, independent of device. Furthermore, because of the substantial variations among studies (e.g., position, recording time, metric, etc.), it is unclear whether these factors also influence the validity of portable HRV devices.

The overall objective of this paper is to provide an overview of the HRV literature and a quantitative interpretation of HRV measures obtained using portable devices, which in turn, will inform practitioners and coaches alike about the most convenient yet accurate method for monitoring athletes outside the laboratory. As such, the purposes of this systematic review and meta-analysis of HRV literature was to examine the agreement between HRV measures recorded using a portable HRV device and an ECG, and to identify study-level moderators that can explain a meaningful proportion of the observed variance between HRV measurement techniques.

METHODS

Search strategy

This systematic review and meta-analysis was conducted in accordance with the PRISMA (*P*referred *R*eporting *I*tems for *S*ystematic *R*eviews and *M*eta-*A*nalyses) Statement [34]. Potentially qualifying reports published before July 29, 2017 were identified with a Boolean search strategy using combinations of the terms: *heart rate variability*, *HRV, valid,* and *validity*.

Four electronic databases, Physical Education Index, PubMed, Scopus, and SPORTDiscus were searched from their inception. Duplicate records were removed and all original records were reviewed against inclusion/exclusion criteria. Manual searches of the reference lists from included articles were reviewed for additional publications not discovered during the electronic database search.

Study Selection

Included articles were limited to publications that (1) peer-reviewed publications, (2) were available in the English language, (3) compared HRV obtained in a clinical or laboratory setting using an ECG to those obtained using a portable device, and (4) reported HRV outcomes for both the portable device and ECG assessment method. Potential records were excluded if they: (1) were non-peer reviewed, (2) provided a review, meta-analysis, position statement, or proposed study design, or (3) did not provide adequate information from which an effect size (*ES*) could be calculated.

Methodological Quality Assessment

The methodological quality of the included studies was assessed using a modified version of the Standard for Reporting Diagnostic Accuracy Studies (STARD) Guidelines [35, 36] which adopted recommendations provided by Quintana et al. [37] and Laborde et al. [38] on HRV research. The modified STARD for HRV research ($STARD_{HRV}$) included 25 items equating to a total of 25 possible points (see Table 2.2 and 2.3 for $STARD_{HRV}$ details). All 23 studies were initially reviewed and coded by the one author (WCD); ≈25% of the final sample were selected at random and reviewed by a second author (CJH) to assure consistency in the coding process.

Discrepancies were addressed by the two authors finalizing the individual quality indices of the $STARD_{HRV}$. The overall methodological quality, which was gauged as the percentage of items satisfied on the $STARD_{HRV}$ assessment tool out of the total possible points (25-points), and individual study quality items were examined in moderator analysis to further investigate sources of potential heterogeneity among the effect estimates.

Study Outcomes and Effect Size Calculation

The absolute value of the standardized mean difference (SMD) effect size (*ES*) was used to quantity the accuracy of portable devices compared to electrocardiography (ECG) for measuring a multitude of HRV metrics. *ESs* were defined as the mean (M) difference in HRV values between the portable and ECG devices divided by the standard deviation (*SD*) of the ECG measure, adjusted for small sample bias [39, 40]. The square root of the squared SMD *ES* was computed to display the absolute difference or (error) between the portable device and ECG measures of HRV [41]. This computation was chosen to enhance the identification of poor agreement which can be masked when averaging effects which overestimate HRV values with effects that underestimate HRV values [42]. The *ES* was interpreted as small (0.2), medium (0.5), and large (0.8) [39] in terms of the magnitude of error between the portable device and ECG. Thus, an *ES* value closer to 0 suggests greater agreement between the portable device and ECG measures of HRV. For studies that failed to report mean HRV values for portable device or ECG, or did not provide a measure of variability for mean values, the corresponding author was contacted and missing data were requested [20, 21, 43, 44]. We were unable to obtain missing or unreported data for two studies [21, 44], and thus, they were excluded from our final sample. Data were independently extracted by the authors (WCD & MVF) and a two-way (effects ×

raters) intraclass correlation coefficient for agreement was calculated to examine inter-rater reliability for calculated effects. Discrepancies were addressed by a third investigator (MRE) and resolved, increasing the intraclass correlation to 100%, prior to aggregating effects.

To represent the degree of agreement between HRV measurement methods, a multivariate/multi-level random-effects model with reduced maximum likelihood estimation was used to generate an overall mean absolute *ES* and 95% confidence interval (*CI*), across all HRV metrics. A multi-level model was used according to standard procedures to adjust for between-study variance and the correlation between effects nested within studies [45, 46].

Inconsistencies in *ES* estimates were quantified using the I^2 statistic and its 95% *CIs* and interpreted as the level of heterogeneity among studies [47, 48]. I^2 values were categorized as low, moderate, and high levels of heterogeneity based on tentative thresholds corresponding to values of 25%, 50%, or 75%, respectively.

Moderator Analyses

In the presence of moderate to high levels of heterogeneity, we evaluated the influence of several *a priori* study-level moderators (e.g., HRV metric, portable device, body position, recording time) on the degree of agreement between HRV measurement methods. In an attempt to identify individual sources of potential heterogeneity, each moderator was evaluated using a multi-level moderator model that accounted for the non-independent effects from multiple outcomes provided by each study. Furthermore, significant study-level moderators were combined in a single multiple-moderator model to identify relationships associated with potential heterogeneity.

Assessment of Potential Biases

Potential bias was assessed through visual inspection of funnel plots for asymmetries in the *ES* distribution to identify potential outliers and Egger's test [49]. Outliers were defined as any effect that fell outside of the upper bound of the 95% *CI*s. We also evaluated the potential for publication bias using the fail-safe N+ test [50]. Furthermore, a sensitivity analysis was completed to determine the effect of potential outliers on our results.

Statistical Computing

Study summary statistics and *ESs* were calculated using Microsoft Excel (Microsoft Office Professional Plus 2016). Analyses and plot construction were performed in R version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria) using the metafor package functions: *rma.mv, fsn, funnel,* and *forest.rma* [51, 52]. Aggregate-level data characteristics were presented as mean (*M*) ± *SD* or *M* (95% *CIs*) unless otherwise noted. Statistical significance was determined *a priori* at α=.05.

RESULTS

After removing duplicates, our search yielded 5,233 original records for review. Of these, 22 studies met our inclusionary criteria [7, 17-20, 22-33, 43, 53-56]. One additional study published after the original search [57] was identified through the manual search and included in the current analysis, yielding 23 total studies. A flowchart of study selection is provided in Figure 2.1.

On average, studies included 28 ± 27 participants (range: 7 to 137), compared a total of 20 devices to ECG using HRV recordings of 5.9 ± 4.6 minutes (range: 1 to 20). Across all 23

studies, aggregate-level data from participants aged 29.7 ± 11.6 who were normal weight (BMI: 23.6 ± 1.5) consisting of 18.3 % women were included in our meta-analysis (a detailed description of the study, sample and HRV recording details are summarized for each of the 23 studies in Table 2.4).

Methodological Study Quality: STARD$_{HRV}$

Methodological quality utilizing the STARD$_{HRV}$ for the 23 respective studies ranged from 62% to 92% (79±8%). Three studies scored above 90%, eight scored between 80-89%, nine scored between 70-79%, and three studies scored below 70% according to the STARD$_{HRV}$. Studies were most likely to satisfy the following items: explanation of how comparison calculations were performed (Q14, 96%), provided *M* ± *SD* along with another measure of estimate of precision (Q21, 96%), fully description of study protocol (Q24, 96%), identifying as a validation study (Q1, 100%), systematically organization of the abstract (Q2, 100%), provided background information and intended use (Q3, 100%), performed a with-in subject design (100%), described how each HRV metric was calculated along with the software used (Q18, 100%), and provided implications for practice (Q23, 100%).

Contrarily, studies were least likely to acknowledge how sample size was derived (Q6, 0%), provide an explanation of how missing data was handled along with percentage missing (Q15, 50%), identifying study objectives along with hypotheses (Q4, 57%), denote whether breathing rate was controlled (Q13, 57%), and identifying sources of funding (Q25, 57%). An item-by-item summary of the methodological study quality utilizing the STARD$_{HRV}$ for the respective 23 studies can be viewed in Table 2.5.

Overall Agreement between HRV Measures Obtained using Portable Devices vs. ECG

The cumulative results from 301 effects collected from 23 studies (10.8 ± 6.2 effects per study) published between 2003 and 2018 indicated a small *ES*, representing absolute error, in HRV measures acquired from portable devices compared to ECG (*ES*=0.23, 95%*CI*: 0.05, 0.42); this mean effect was highly heterogeneous (I^2=78.6%, 95%*CI:* 76.2%, 80.7%) with sampling error accounting for 21.4% of the observed variance. Variability amongst effects are presented in ascending order in the forest plots in Figures 2.2.1, 2.2.2, and 2.2.3.

Moderator Analysis

Pre-specified *a priori* moderator variables were analyzed separately to determine their individual influence on the degree of absolute error between HRV measurement methods. Subgroup comparison analyses revealed that HRV metric (p<0.001) and body position during HRV recording (p=0.033), but not portable device (p=0.468) modulated the degree of absolute error (see Table 2.6 for subgroup comparisons). A significant degree of error was observed between measurement methods when HRV was expressed as SDNN, RMSSD, and HF. However, further analyses within metric revealed that the magnitude of absolute error between portable devices and ECG was significantly greater SDNN measures compared to the degree of error associated with all other HRV metrics except SD1 and SD2. In addition, RMSSD, LF, and HF measures displayed greater absolute error compared to LF:HF ratio; no other relationships between metrics significantly differed (Table 3). Studies that recorded HRV in the supine, seated, and tilt/recovery position were independently associated with significant levels of absolute error. Within positon analyses revealed that the magnitude of absolute error between

portable devices and ECG was greater for recordings performed in tilt/recovery than supine, sitting, active (all $p<0.05$) and standing ($p=0.053$) positions.

Sample characteristics for the most part did not significantly modulate the degree of error between HRV measurement methods with the exception of biological sex, gauged as the percentage of women present in each study sample and evaluated as a continuous variable. Multi-level meta-regression analysis revealed that the degree of absolute error between HRV measurement methods was larger among studies involving a greater number of women participants (higher percentage of total sample) (β=0.45, 95% *CI* 0.30, 0.61, R^2 = 5.9%, $p<0.001$). Lastly, methodological study quality, gauged as a summary score of items that were fully satisfied (25 possible points), did not a significantly moderate the magnitude of absolute error between HRV measurement methods.

Multiple Moderator Model

When significant moderators were combined into a single model (i.e., metric within position within device, with the inclusion of percentage of women in study sample) and evaluated simultaneously, heterogeneity among the *ESs* was reduced but a moderate level remained (I^2=56.6%, 95%*CI:* 50.6%, 61.8%). The model provided 12 relationships with significant associations with the absolute error, including biological sex (percentage of woman in study sample). Aside from biological sex, the 11 other significant relationships were associated with a combination of three devices (i.e., Polar RS800CX, iPhone4S, Motorola Droid), two positions (supine and tilt/recovery), and several linear and frequency metrics (SDNN, RMSSD, HF, and LF) which are displayed in Table 2.7.

Assessment of Biases

Publication bias was acquired using a funnel plot, shown in Figure 2.3, and Egger's test [49]. The funnel plot was created by plotting the treatment effects against standard error [58], as recommended when using the a standardized mean difference *ES* [58-60]. Through visual inspection of the funnel plot, 26 of the 301 effects (8.6%) were recognized as potential outliers, falling outside the 95% *CI*. All potential outliers were represented in four of the 23 studies [18, 19, 31, 55]; five of the 20 notable portable devices, which include the iPhone 4s (n=2) and Motorola Droid (n=2) [18], Polar S810 (n=4) [55], Polar RS800 (n=6) [19], and the Polar RS800CX (n=12) [31]; two positions, supine (n=15) and tilt/recovery (n=11); 12 in the frequency domain and 14 in the time domain. Visual determinations were confirmed Egger's test suggested the presence of publication bias (p<0.001). A fail-safe N^+ represents the minimal number of additional null effects from multiple studies of average sample size needed to reach a similar null conclusion [50, 61]. Results of the fail-safe N^+ indicated 3,051 null effects would be required to overturn this significant finding.

Furthermore, a sensitivity analysis removing all 26 potential outliers yielded a similar significant result (*ES*=0.10, 95% *CI* 0.05 to 0.15, p<0.001). However, this effect was significantly smaller compared to the original overall *ES*, and eliminated the observed heterogeneity (I^2=0%), as well as publication bias, as observed in the reduced funnel plot in Figure 2.4. Furthermore, subgroup comparisons with outliers removed (n=275) can be viewed in Table 2.8 and the multiple moderator model with outliers removed (n=275) can be viewed in Table 2.9.

DISCUSSION

The purpose of this systematic review and meta-analysis was to quantify the absolute error associated with HRV measures acquired by portable devices compared to the gold standard of an ECG. Error was quantified as the absolute standardized difference, a non-directional approach to estimating error associated with HRV metrics. The cumulative results from 23 publications indicated a significant yet, small amount of absolute error in HRV measures acquired from portable devices compared to ECG. There was a significant amount of heterogeneity in the absolute error amongst HRV metrics, position and percent female associated with the aggregate sample.

The percentage of females per sample was found to be significantly associated with the absolute error of HRV measures acquired from portable devices. In essence, there was a rise in error with an increase in female participants within a given sample. This finding supports that of Wallen et al. [19], who found a higher error rate in females compared to males when acquiring HRV measures through the Polar RS800 compared to an ECG. However, of the fourteen studies which included female participants [7, 19, 25-27, 30-33, 43, 54-57], Wallen et al. [19] was the only study which presented findings relative to biological sex. Therefore, associations (e.g., metric, position, recording time) attributing greater error within the moderator cannot be elaborated upon from the aggregate data. Conversely, of the 26 identified outliers, 21 of the effects included female participants (42% n=4, 55% n=12, 100% n=5). When the multiple moderator model was performed after removal of outliers (k=257), biological sex was no longer significantly associated with absolute error.

In regard to position, error did not significantly differ between the supine, seated, and standing positions or when participants were active (cycling). In contrary, the tilt/recovery

position represented a period following an orthostatic challenge and showed significant increase in error compared to all other positions (*ES* increased by 0.32 - 0.40), and remained significantly higher after the removal of outliers (see Table 2.8). Of the 20 effects which were represented an HRV measurement in the tilt/recovery setting, 10 were derived from a pulse wave variability through photoplethysmography (PPG) from the camera of cell phones [18]. A review by Schäfer and Vagedes [62] reported the utilization of PPG provides good agreement at rest, but dissipated with activity due to increased artifact. This increase in artifact or "noise" has also been communicated as a potential limitation to HRV recordings during exercise [63]. This may have accounted for some of the error associated tilt/recovery position. Thus, practitioners should recognize the importance of utilizing a stabilized position during resting conditions when acquiring HRV from PPG. Furthermore, regardless of the non-significant difference in error noted above, it has been identified that HRV measures acquired from different positions (i.e., supine and standing) are independent and not interchangeable [64]. Therefore, practitioners should also choose a consistent position during HRV recordings to ensure appropriate longitudinal monitoring.

Heterogeneity with HRV metric displayed SDNN, a marker to overall variability, as the metric associated with the greatest amount of error (medium effect) compared to all other metrics which could be classified as small or below. Contrarily, this relationship was no longer significant after the removal of the 26 outliers (see Table 2.8). However, as metrics which primarily reflect the parasympathetic nervous system activity (i.e., RMSSD, HF, and SD1) are typically of interest to researchers and practitioners whom utilize HRV measurements for monitoring performance, it is important to note they did not significantly differ. In practicality, this warrants favor for the linear metric RMSSD and the non-linear metric SD1 as they have

been shown to be less susceptible to breathing rate [65]. Thus, allowing practitioners to acquire a measure of parasympathetic activity in a less controlled environment. In addition, compared to spectral analysis, these metrics can also be calculated in excel [66], further improving the practicality of portable device as a tool for monitoring/tracking performance through perturbations in HRV without the need for an ECG in a laboratory or sophisticated software.

When a larger model was derived in an attempt to identify sources of error associated with specific devices, three devices were identified as significant when including possible relationships between metric and position. However, Egger's test denoted a significant amount of publication bias suggesting the extreme effects may have been derived from study level differences. This was confirmed when the outliers were removed, as Egger's test became non-significant and heterogeneity became non-significant. Nevertheless, the absolute error remained significant but was reduced to trivial with a tight range. Thus, differences in the overall mean effect of the absolute error associated with HRV metrics and position acquired by portable devices may have been driven by study level differences. For example, the Polar RS800CX was utilized in two studies [31, 32] yet the absolute error associated with the same device drastically differed between the two studies. Vasconcellos et al. [32] reported four effects in the supine position which ranged from 0.25 to 0.36 while Montaño et al. [31] reported 15 effects which ranged from 0.21 to 104.13, 12 of which were outliers. Both studies utilized Kubios software in their respective analyses of linear and frequency derived metrics in the supine position (Montaño et al. [31] also included a tilt positon and recovery from an orthostatic challenge), yet the difference was notable through our analysis in the quantification of absolute error.

It is pertinent to note that the quantification of absolute error which was performed in this analysis could be considered a conservative method. However, it was the intent to determine the

degree of error regardless of direction in order to derive our findings which is unique in its application and a key strength of the current analysis. In regard to previous example, Montaño et al. [31] concluded the Polar RS800CX provided a valid assessment of HRV indices as the Pearson's correlations between the ECG and the Polar device exceeded 0.75. However, when assessing this relationship through the calculation of absolute error, the magnitude of the difference between the two devices do not appear to agree (*ES* ranged from 0.21 – 104.13). Contrarily, it is not the intent to disregard the findings of Montaño et al. [31], as one could argue the importance of a repeatable measure even if the absolute values differ by a certain degree on average as a correction factor may be made, but strong evidence for repeatability alone does not infer validity [42, 67]. However, the quantification of the absolute error associated HRV metrics derived from portable devices compared to ECG from 23 studies suggests a small - trivial absolute error may be present on average. In the context of monitoring autonomic perturbations in response to athletic performance and training regiments, this small – trivial error may be acceptable by practitioners for use in the field setting.

Furthermore, as this analysis synthesized the amount of absolute error associated with HRV metrics acquired by portable devices, it is imperative to recognize the non-significant relationships. For example, 20 different devices including a combination of traditional heart rate monitors, PPG, and one derived from a non-commercialized shirt, were found to not significantly differ in the amount of absolute error. Of these devices, 11.3% (k=34) of the effects were derived from the utilization of applications available on cell phones. These applications utilize the ultra-short method (one-minute stabilization followed by one minute recording) for acquiring HRV which has been previously validated [12, 14, 27, 68], and further recognized within this analysis as recording time was found to be non-significant. This denotes an increase of accessibility for

practitioners and researchers to perform objective athletic monitoring of cardiac-autonomic modulations observed through HRV on a large team scale in the field setting [69]. For example, portable devices have been utilized to monitor HRV indices in elite rowers [70], futsal players [71], collegiate soccer players [8, 9, 11], American football players [13], swimmers [72], and to guide training regiments [73]. Furthermore, the small-trivial amount of absolute error associated with the portable devices included in this analysis suggest practitioners and researchers can continue to utilized the benefit of accessibility, along with acceptable accuracy, provided by portable devices outside the confines of an ECG typically utilized inside a laboratory.

In note to the practitioner, the cost-benefit for utilizing portable devices for acquiring HRV measures takes president. It has been demonstrated that meaningful interpretations of longitudinal HRV data is improved with an increase in the signal-to-noise ratio [66]. This can be accomplished through weekly averages of consistent day-to-day recordings which have been shown to be superior to isolated measures when identifying meaningful changes in performance [3, 74, 75]. Thus, it is the goal of the practitioner to find a means of acquiring HRV which promote the greatest compliance by athletes to ensure measures on a day-to-day basis. Given the amount of error did not significantly differ amongst devices, recording time, or metrics indicative of parasympathetic activity, one cannot disregard the potential benefit of improved compliance through the utilization of applications available on cellular phones. These applications do not require athletes to wear any additional equipment (i.e., chest strap), provide measures in ultra-short time segments, and may be cost-effective for monitoring on a team basis compared to other portable team-based systems. Additionally, these applications may provide a psychometric questionnaire which can also be completed by athletes on a daily basis and assist practitioners in monitoring the overall profile of the athlete [8, 66, 76].

Limitations

All systematic reviews and meta-analyses have methodological limitations, and the current study is no different. Despite searching four electronic databases, only 17 of the original 22 publications (77.3%) identified for inclusion in the current analysis could be identified using a traditional keyword search, yielding a search sensitivity less than the desired sensitivity outlined in previous research [77-79]. By manually searching the reference lists of relevant review articles and publications identified through the electronic search, the authors were able to identify six additional publications that were included in the final analysis highlighting the importance of including a manual reference search in a systematic review in order to completely exhaust the published literature [80]. Furthermore, only published data extracted from peer-reviewed journals was included in the current study. Excluding data presented in conference proceedings and non-peer-reviewed publications may have influenced the overall mean ES and estimates of heterogeneity. However, the likelihood that including information collected from these sources significantly changed the outcome of the current analysis is small. These results do not relate to the specifics of error corrections technology within the respective portable devices. Therefore, the suggestion of acceptable absolute error is applicable for the use of athletic monitoring and cannot be generalized into clinical populations which may require accurate ectopic beat identification.

Future Research

With the continuous advancements in technology there will inevitably be more devices in need of validation in the future (e.g., Oura ring). Therefore the need for validation of portable devices will be ongoing. Also, as present in this analysis, study level differences exist during the

validation of devices. Thus, there is a need to validate instruments in various environments and conditions regardless if previous research has been performed in an attempt to validate a particular device. As previously mentioned, biological sex was a significant moderator in the associated error within this analysis. However, due to the pooled results in the majority of the studies, we were unable to confidently identify attributes which may minimize error specific to biological sex. Therefore, there is a need for research to determine if there are characteristics (e.g., metric, position) which may be biological sex dependent in regards to the accuracy of an HRV measure.

CONCLUSIONS

There have been several systematic reviews regarding the accuracy of HRV derived from portable devices in comparison to the "gold standard" of an ECG [16, 81]. To our knowledge, this is the first quantitative analysis specifically addressing the error associated with HRV metrics the various portable devices. The conservative approach to quantifying the absolute error objectively across all 23 studies yielded a small but significant mean effect which was not dependent on which portable device was utilized. Error did not differ between supine, seated, and standing positions or when active. However, a non-resting tilt/recovery position which induces an orthostatic challenge may provide a significantly higher amount of error and should therefore not be utilized. Likewise, metrics RMSSD, HF, SD1 which represent parasympathetic activity did not significantly differ. Due to the lower influence of breathing rate, and non-requirement of high priced software, RMSSD and SD1 may be preferable. Furthermore, the mean effect was additionally reduced when removing the bias of a small cluster of effects but remained significant with a close 95% *CI*. Thus, these findings indicate that portable devices for

acquiring HRV provide accurate measures when compared to lab-based ECG. Additionally, practitioners and researchers should feel confident in the current validated portable technology and should consider the cost-benefit of acquiring HRV measures through validated cell phone applications. The acceptable accuracy, along with the simplicity of the measure (one-minute recording through PPG), may improve compliance by the athlete on the day-to-day basis. Allowing the practitioner to derive weekly averages in the attempt to and appropriately identify/interpret meaningful changes in HRV during longitudinal monitoring of the athlete.

REFERENCES

1. Thompson WR. Worldwide survey of fitness trends for 2017. ACSM's Health & Fitness Journal. 2016;20(6):8-17.

2. Thompson WR. Worldwide survey of fitness trends for 2018: the CREP edition. ACSM's Health & Fitness Journal. 2017;21(6):10-9.

3. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. Sports Med. 2013;43(9):773-81.

4. Sandercock GRH, Brodie DA. The use of heart rate variability measures to assess autonomic control during exercise. Scand J Med Sci Spor. 2006;16(5):302-13.

5. Silva VP, Oliveira NA, Silveira H, Mello RGT, Deslandes AC. Heart rate variability indexes as a marker of chronic adaptation in athletes: a systematic review. Annals of Noninvasive Electrocardiology (Online). 2015;20(2):108-18.

6. Malik M, Camm AJ, Bigger Jr JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17(3):354-81.

7. Radespiel-Tröger M, Rauh R, Mahlke C, Gottschalk T, Muck-Weymann M. Agreement of two different methods for measurement of heart rate variability. Clin Auton Res. 2003;13(2):99-102.

8. Flatt AA, Esco MR, Nakamura FY, Plews DJ. Interpreting daily heart rate variability changes in collegiate female soccer players. J Sports Med Phys Fitness. 2017;57(6):907-15.

9. Flatt AA, Esco MR. Evaluating individual training adaptation with smartphone-derived heart rate variability in a collegiate female soccer team. J Strength Cond Res. 2016;30(2):378-85.

10. Senthinathan A, Mainwaring LM, Hutchison M. Heart rate variability of athletes across concussion recovery milestones: a preliminary study. Clin J Sport Med. 2017;27(3):288-95.

11. Esco MR, Flatt AA, Nakamura FY. Initial weekly HRV response is related to the prospective change in VO2max in female soccer players. Int J Sports Med. 2016;37(6):436-41.

12. Nakamura FY, Pereira LA, Esco MR, Flatt AA, Moraes JE, Abad CCC, et al. Intraday and interday reliability of ultra-short-term heart rate variability in rugby union players. J Strength Cond Res. 2017;31(2):548-51.

13. Flatt AA, Esco MR, Allen JR, Robinson JB, Earley RL, Fedewa MV, et al. Heart rate variability and training load among national collegiate athletic association division 1 college football players throughout spring camp. J Strength Cond Res. 2017.

14. Flatt AA, Esco MR. Heart rate variability stabilization in athletes: towards more convenient data acquisition. Clin Physiol Funct Imaging. 2016;36(5):331-6.

15. Berntson GG, Bigger JT, Jr., Eckberg DL, Grossman P, Kaufmann PG, Malik M, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology. 1997;34(6):623-48.

16. Board L, Ispoglou T, Ingle L. Validity of telemetric-derived measures of heart rate variability: a systematic Review. J Exerc Physiol. 2016;19(6):64-84.

17. Barbosa MP, da Silva NT, de Azevedo FM, Pastre CM, Vanderlei LC. Comparison of Polar(R) RS800G3 heart rate monitor with Polar(R) S810i and electrocardiogram to obtain the series of RR intervals and analysis of heart rate variability at rest. Clin Physiol Funct Imaging. 2016;36(2):112-7.

18. Bolkhovsky JB, Scully CG, Chon KH, editors. Statistical analysis of heart rate and heart rate variability monitoring through the use of smart phone cameras. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE; 2012: IEEE.

19. Wallen MB, Hasson D, Theorell T, Canlon B, Osika W. Possibilities and limitations of the Polar RS800 in measuring heart rate variability at rest. Eur J Appl Physiol. 2012;112(3):1153-65.

20. Hernando D, Garatachea N, Almeida R, Casajús JA, Bailón R. Validation of heart rate monitor Polar RS800 for heart rate variability analysis during exercise. J Strength Cond Res. 2016.

21. Kingsley M, Lewis MJ, Marson RE. Comparison of Polar 810 s and an ambulatory ECG System for RR interval measures during progressive exercise. Int J Sports Med. 2005(26):39-44.

22. Romagnoli M, Alis R, Guillen J, Basterra J, Villacastin J, Guillen S. A novel device based on smart textile to control heart's activity during exercise. Autstrals Phys Eng S. 2014;37(2):377-84.

23. Weippert M, Kumar M, Kreuzfeld S, Arndt D, Rieger A, Stoll R. Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. Eur J Appl Physiol. 2010;109(4):779-86.

24. Chellakumar PJ, Brumfield A, Kunderu K, Schopper AW. Heart rate variability: comparison among devices with different temporal resolutions. Physiol Meas. 2005;26(6):979-86.

25. Russoniello CV, Zhirnov YN, Pougatchev VI, Gribkov EN. Heart rate variability and biological age: implications for health and gaming. Cyberpsychol Behav Soc Netw. 2013;16(4):302-8.

26. Esco MR, Flatt AA, Nakamura FY. Agreement between a smartphone pulse sensor application and electrocardiography for determining lnRMSSD. J Strength Cond Res. 2017;31(2):380-5.

27. Flatt AA, Esco MR. Validity of the ithlete smart phone application for determining ultra-short-term heart rate variability. J Hum Kinet. 2013;39:85-92.

28. Vanderlei LCM, Silva RA, Pastre CM, Azevedo FMD, Godoy MF. Comparison of the Polar S810i monitor and the ECG for the analysis of heart rate variability in the time and frequency domains. Braz J Med Biol Res. 2008;41(10):854-9.

29. Gamelin FX, Berthoin S, Bosquet L. Validity of the Polar S810 heart rate monitor to measure R-R intervals at rest. Med Sci Sports Exerc. 2006;38(5):887-93.

30. Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol. 2016;116(3):563-71.

31. Montaño A, Brown F, Credeur DP, Williams MA, Stoner L. Telemetry-derived heart rate variability responses to a physical stressor. Clin Physiol Funct Imaging. 2017;37(4):421-7.

32. Vasconcellos FV, Seabra A, Cunha FA, Montenegro RA, Bouskela E, Farinatti P. Heart rate variability assessment with fingertip photoplethysmography and Polar RS800cx as compared with electrocardiography in obese adolescents. Blood Press Monit. 2015;20(6):351-60.

33. Nunan D, Jakovljevic DG, Donovan G, Hodges LD, Sandercock GR, Brodie DA. Levels of agreement for RR intervals and short-term heart rate variability obtained from the Polar S810 and an alternative system. Eur J Appl Physiol. 2008;103(5):529-37.

34. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:332-6.

35. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ open. 2016;6(11):e012799.

36. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. standards for reporting of diagnostic accuracy. Clin Chem. 2003;49(1):1-6.

37. Quintana DS, Alvares GA, Heathers JA. Guidelines for reporting articles on psychiatry and heart rate variability (GRAPH): recommendations to advance research communication. Transl Psychiatry. 2016;6:e803.

38. Laborde S, Mosley E, Thayer JF. Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. Front Psychol. 2017;8:213.

39. Cohen J. A power primer. Psychol Bull. 1992;112(1):155-9.

40. Hedges L, Olkin I. Statistical methods for meta-analysis. 6 ed. San Diego, CA: Academic Press; 1985. 79-201 p.

41. Crossingham IR, Nethercott DR, Columb MO. Comparing cardiac output monitors and defining agreement: a systematic review and meta-analysis. J Intensive Care Soc. 2016;17(4):302-13.

42. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. PLoS One. 2012;7(5):e37908.

43. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart rate variability recording with smart phone photoplethysmographic, Polar H7 chest strap and electrocardiogram methods. Int J Sports Physiol Perform. 2017:1-17.

44. Boos CJ, Bakker-Dyos J, Watchorn J, Woods DR, O'Hara JP, Macconnachie L, et al. A comparison of two methods of heart rate variability assessment at high altitude. Clin Physiol Funct Imaging. 2016.

45. Hox J. Multilevel analysis: techniques and applications. 2 ed. Marcoulides G, editor. New York, NY: Taylor & Francis; 2010. 205-32 p.

46. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J Educ Behav Stat. 1998;23(4):323-55.

47. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.

48. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ: British Medical Journal. 2003;327(7414):557.

49. Egger M, Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.

50. Rosenberg M. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. Evolution. 2005;59(2):464-8.

51. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1-48.

52. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.

53. Gamelin FX, Baquet G, Berthoin S, Bosquet L. Validity of the Polar S810 to measure R-R intervals in children. Int J Sports Med. 2008;29(2):134-8.

54. Heathers JAJ. Smartphone-enabled pulse rate variability: An alternative methodology for the collection of heart rate variability in psychophysiological research. Int J Psychophysiol. 2013;89:297-304.

55. Nunan D, Donovan G, Jakovljevic DG, Hodges LD, Sandercock GR, Brodie DA. Validity and reliability of short-term heart-rate variability from the Polar S810. Med Sci Sports Exerc. 2009;41(1):243-50.

56. Porto LGG, Junqueira J. Comparison of time-domain short-term heart interval variability analysis using a wrist-worn heart rate monitor and the conventional electrocardiogram. Pacing and Clinical Electrophysiology. 2009;32(1):43-51.

57. Guzik P, Piekos C, Pierog O, Fenech N, Krauze T, Piskorski J, et al. Classic electrocardiogram-based and mobile technology derived approaches to heart rate variability are not equivalent. Int J Cardiol. 2018;258:154-6.

58. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343(d4002).

59. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J R Stat Soc A Sta. 1988;151(3):419-63.

60. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol. 2001;54(10):1046-55.

61. Rosenthal R. Meta-analytic procedures for social research. 2 ed: SAGE Publications; 1991.

62. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability?: a review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol. 2013;166(1):15-29.

63. Buchheit M, Mendez-Villanueva A. Improbable effect of carbohydrate diet on cardiac autonomic modulation during exercise. Eur J Appl Physiol. 2010;109(3):571-4.

64. Schmitt L, Regnard J, Millet GP. Monitoring fatigue status with HRV measures in elite athletes: an avenue beyond RMSSD? Front Physiol. 2015;6:343.

65. Penttila J, Helminen A, Jartti T, Kuusela T, Huikuri HV, Tulppo MP, et al. Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns. Clin Physiol. 2001;21(3):365-76.

66. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol. 2014;5:73.

67. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231-40.

68. Esco MR, Flatt AA. Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. J Sports Sci Med. 2014;13(3):535-41.

69. Altini M, Amft O. HRV4Training: large-scale longitudinal training load analysis in unconstrained free-living settings using a smartphone application. Conf Proc IEEE Eng Med Biol Soc. 2016;2016:2610-3.

70. Plews DJ, Laursen PB, Buchheit M. Day-to-day heart-rate variability recordings in world-champion rowers: appreciating unique athlete characteristics. Int J Sports Physiol Perform. 2017;12(5):697-703.

71. Nakamura FY, Flatt AA, Pereira LA, Ramirez-Campillo R, Loturco I, Esco MR. Ultra-short-term heart rate variability is sensitive to training effects in team sports players. J Sports Sci Med. 2015;14(3):602-5.

72. Clemente-Suarez VJ, Arroyo-Toledo JJ. The use of autonomic modulation device to control training performance after high-intensity interval training program. J Med Syst. 2018;42(3):47.

73. Nuuttila OP, Nikander A, Polomoshnov D, Laukkanen JA, Hakkinen K. Effects of HRV-guided vs. predetermined block training on performance, HRV and serum hormones. Int J Sports Med. 2017;38(12):909-20.

74. Plews DJ, Laursen PB, Kilding AE, Buchheit M. Heart rate variability in elite triathletes, is variation in variability the key to effective training? a case comparison. Eur J Appl Physiol. 2012;112(11):3729-41.

75. Plews DJ, Laursen PB, Kilding AE, Buchheit M. Evaluating training adaptation with heart-rate measures: a methodological comparison. Int J Sports Physiol Perform. 2013;8(6):688-91.

76. Flatt AA, Esco MR, Nakamura FY. Individual heart rate variability responses to preseason training in high level female soccer players. J Strength Cond Res. 2017;31(2):531-8.

77. Lemeshow AR, Blum RE, Berlin JA, Stoto MA, Colditz GA. Searching one or two databases was insufficient for meta-analysis of observational studies. J Clin Epidemiol. 2005;58(9):867-73.

78. Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol. 2008;61(4):357-64.

79. Papaioannou D, Sutton A, Carroll C, Booth A, Wong R. Literature searching for social science systematic reviews: consideration of a range of search techniques. Health Info Libr J. 2010;27(2):114-22.

80. Linder SK, Kamath GR, Pratt GF, Saraykar SS, Volk RJ. Citation searches are more sensitive than keyword searches to identify studies using specific measurement instruments. J Clin Epidemiol. 2015;68(4):412-7.

81. Sammito S, Böckelmann I. Options and limitations of heart rate measurement and analysis of heart rate variability by mobile devices: a systematic review. Herzschrittmachertherapie & Elektrophysiologie. 2016;27(1):38-45.

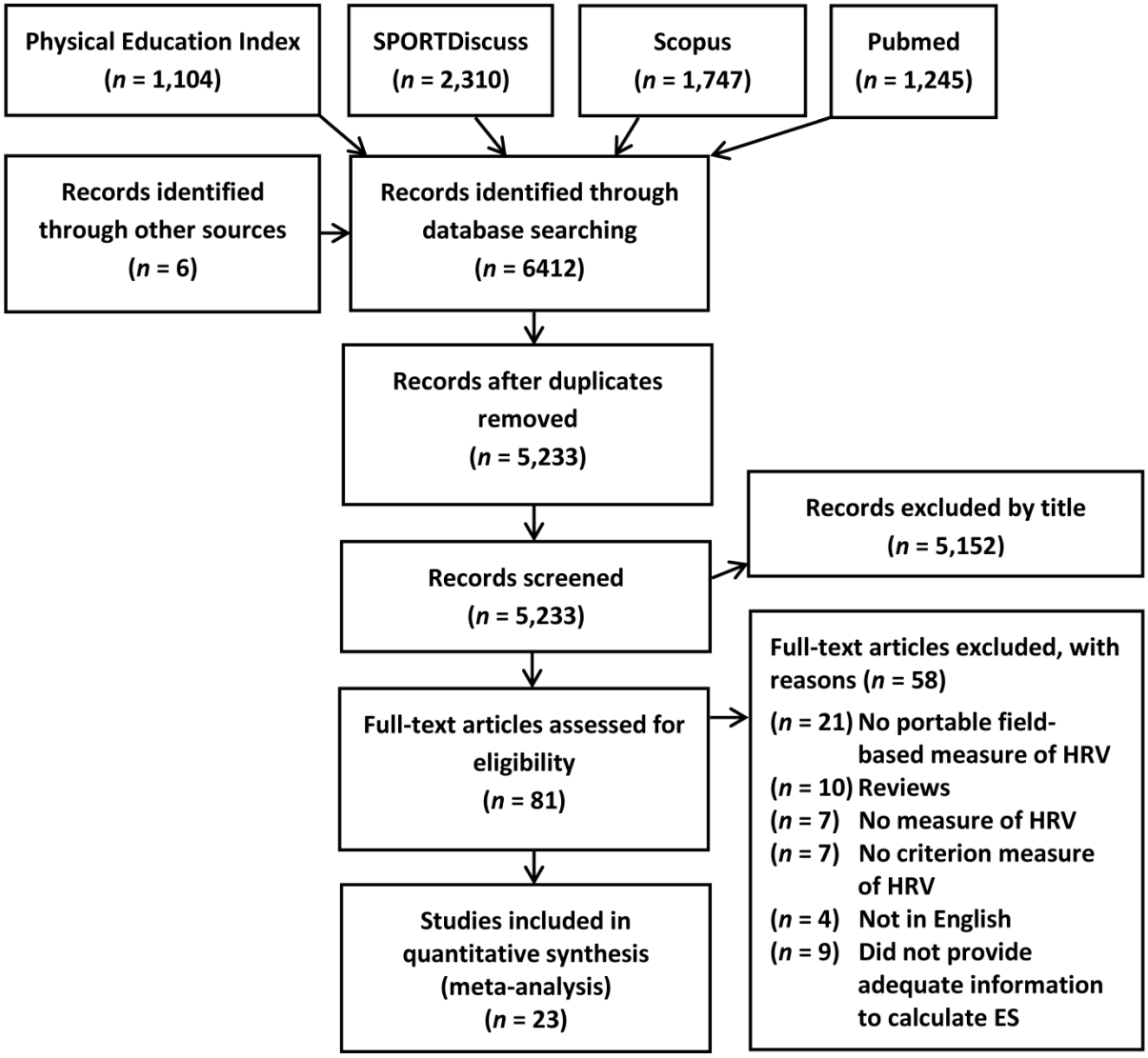

Figure 2.1. Flowchart of study selection. *ES*, effect size; *n*, number of studies.

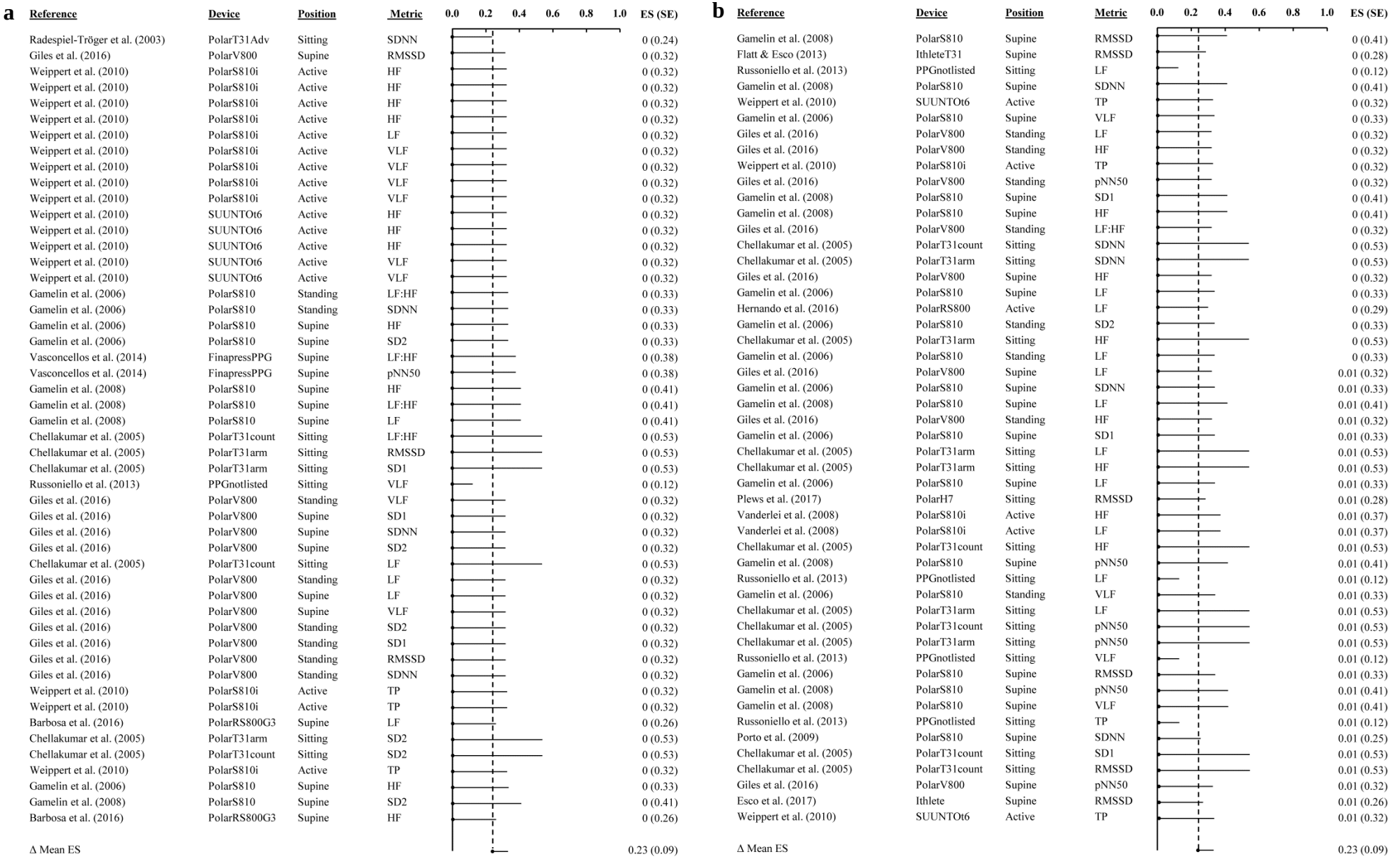

Figure 2.2.1 Forest plot of the absolute standardized effect size weighted by the inverse variance, displayed in ascending order. (a) Effects 1-50; (b) Effects 51-100. *ES*, effect size; HF, high frequency; LF, low frequency; LF:HF, LF to HF ratio; PPG, photoplethysmography; pNN50, percentage value of consecutive RR intervals that differ by more than 50 milliseconds; RMSSD, square root of the mean squared differences between normal adjacent RR intervals; SDNN, standard deviation of the mean of all normal RR intervals; *SE*, standard error; TP, total power; VLF, very low frequency.

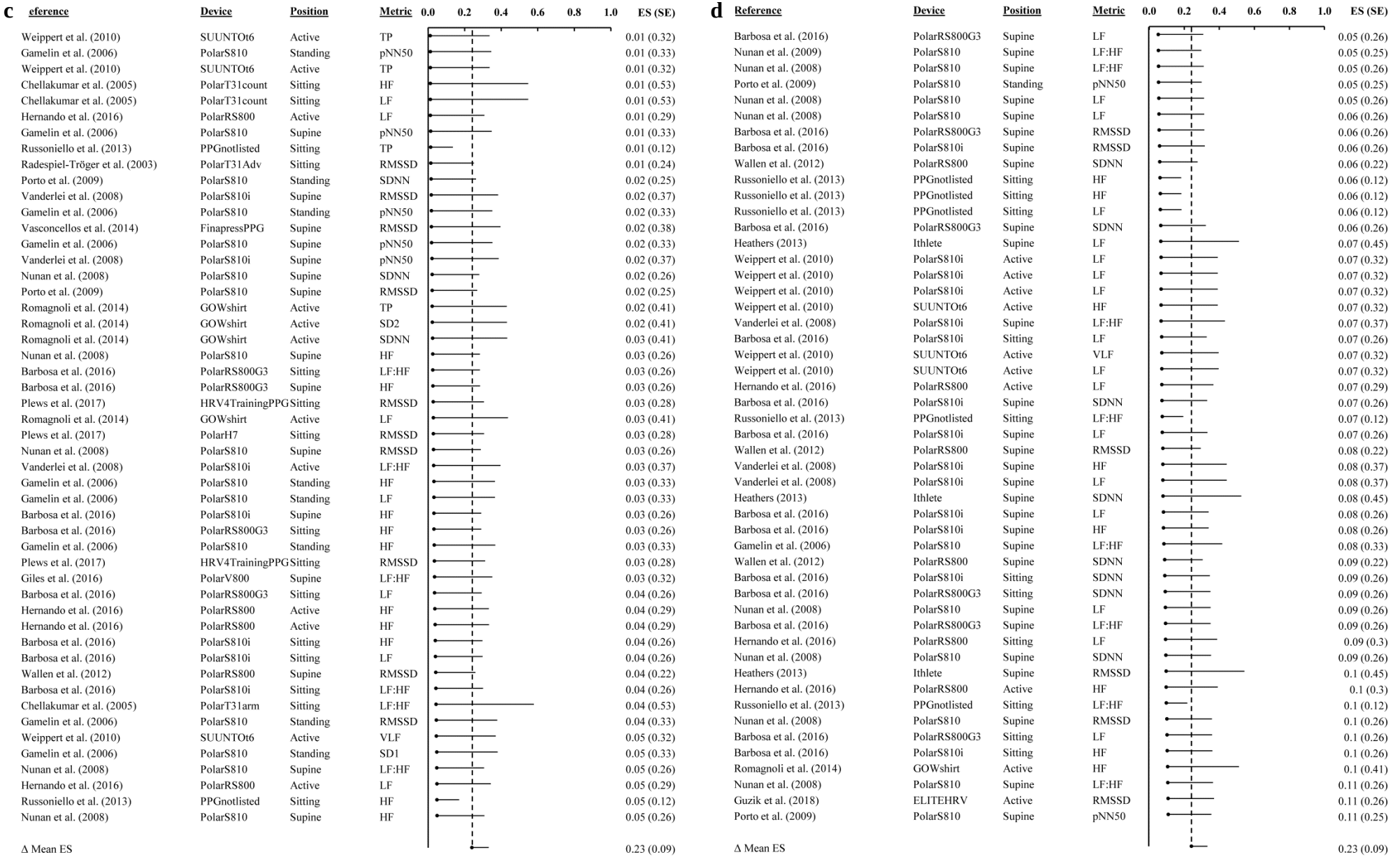

Figure 2.2.2 Forest plot of the absolute standardized effect size weighted by the inverse variance, displayed in ascending order. (c) Effects 101-150; (d) Effects 151-200. *ES*, effect size; HF, high frequency; LF, low frequency; LF:HF, LF to HF ratio; PPG, photoplethysmography; pNN50, percentage value of consecutive RR intervals that differ by more than 50 milliseconds; RMSSD, square root of the mean squared differences between normal adjacent RR intervals; SDNN, standard deviation of the mean of all normal RR intervals; *SE*, standard error; TP, total power; VLF, very low frequency.

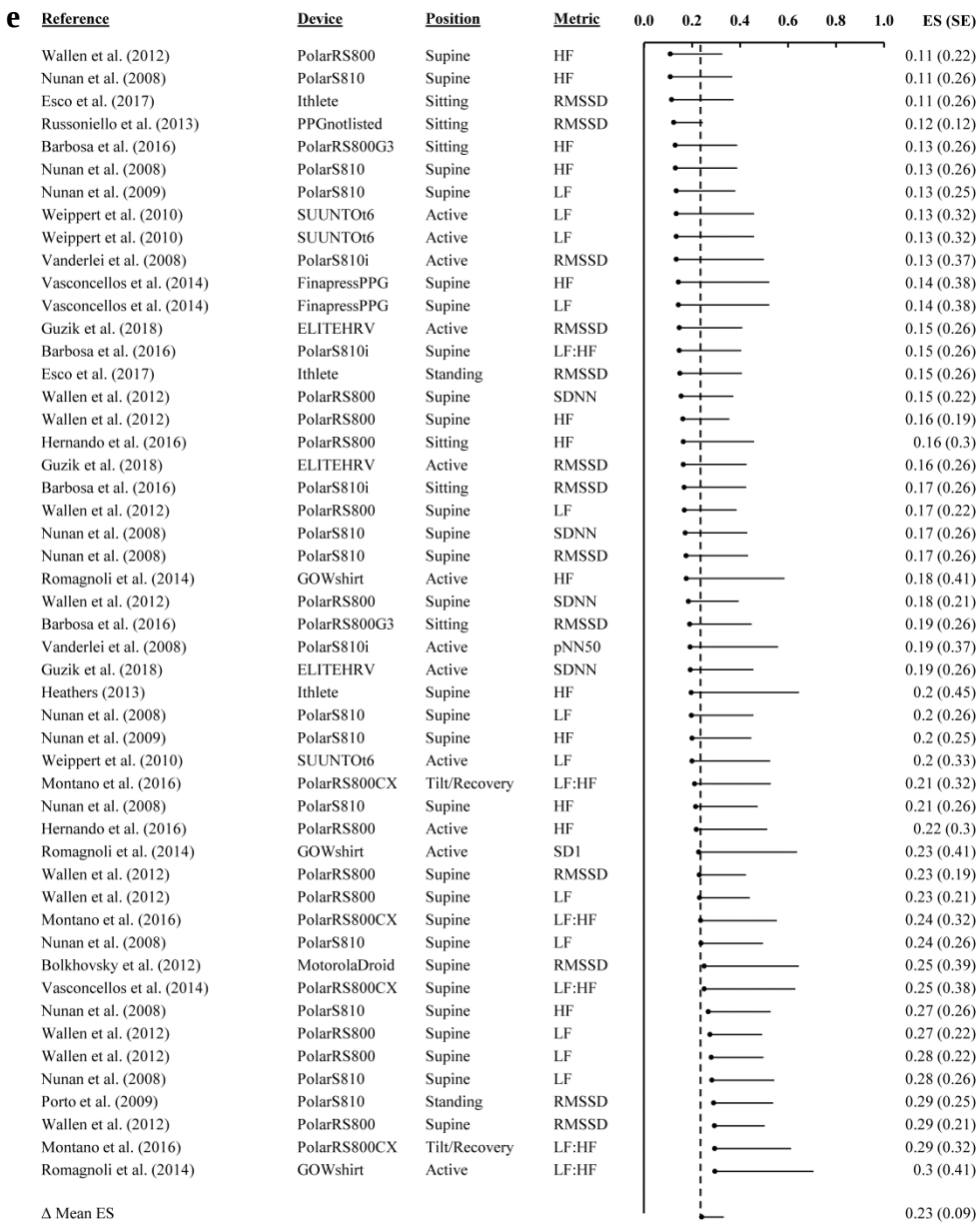

Figure 2.2.3 Forest plot of the absolute standardized effect size weighted by the inverse variance, displayed in ascending order. (e) Effects 201-250, (f) Effects 251-301 [note the change in scale in]. †, significant effect; *ES*, effect size; HF, high frequency; LF, low frequency; LF:HF, LF to HF ratio; PPG, photoplethysmography; pNN50, percentage value of consecutive RR intervals that differ by more than 50 milliseconds; RMSSD, square root of the mean squared differences between normal adjacent RR intervals; SDNN, standard deviation of the mean of all normal RR intervals; *SE*, standard error; TP, total power; VLF, very low frequency.

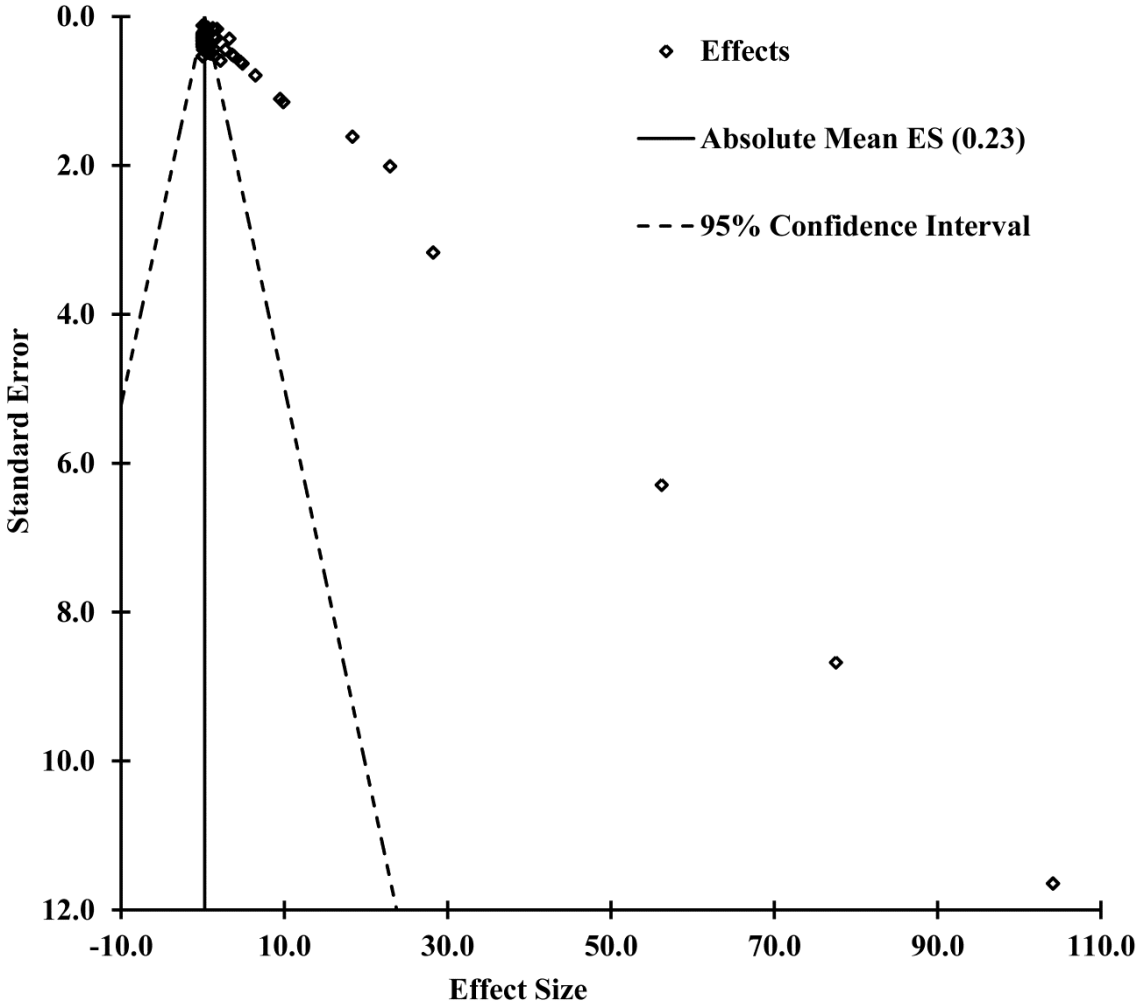

Figure 2.3. Funnel plot of absolute standardized effect size vs. standard error from the full model (*n*=301).

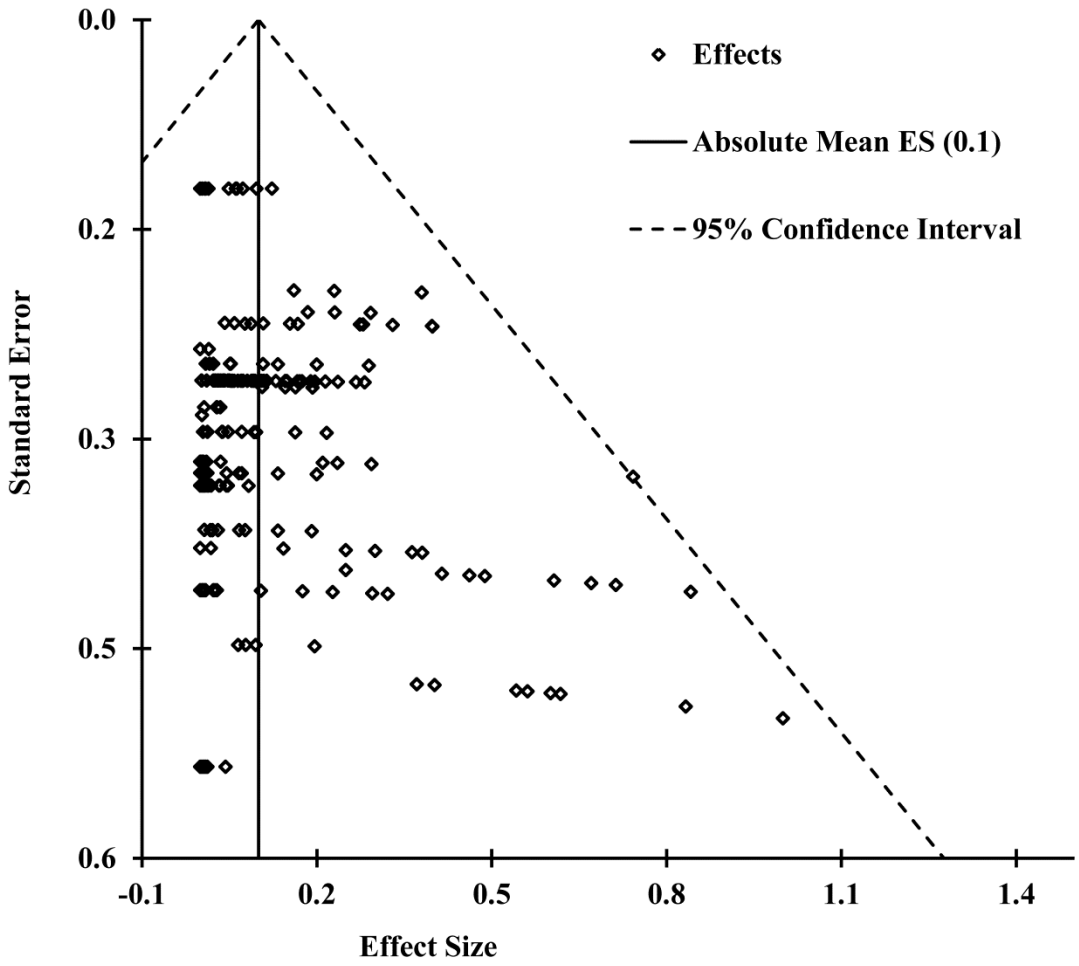

Figure 2.4. Funnel plot of absolute standardized effect size vs. standard error with outliers removed (*n*=275).

Table 2.1. Common heart rate variability metrics included in this meta-analysis and the corresponding reflection of the autonomic nervous system.

Acquisition	Metric	Description	Autonomic Reflection
Time-domain	*SDNN*	Standard deviation of all normal-normal (R-R) intervals	PNS and SNS activity
	pNN50	Percent of normal R-R intervals more than 50ms	PNS activity
	RMSSD	Root mean square of successive differences	PNS activity
Frequency-domain	TP	Total Power (<0.4 Hz)	Overall autonomic variability
	VLF	Very low frequency (<0.04 Hz)	*Thermoregulatory cycles
	LF	Low frequency (0.05 – 0.15 Hz)	*Mix of PNS and SNS activity
	HF	High frequency (0.15 – 0.4 Hz)	PNS activity
	LF:HF	Ratio of low frequency to high frequency	*SNS-to-PNS balance
Non-linear	SD1	Standard deviation of the width of the Poincaré plot	Short term PNS and SNS activity
	SD2	Standard deviation of the length of the Poincaré plot	Long term PNS and SNS activity

Abbreviations: PNS, parasympathetic nervous system; SNS, sympathetic nervous system; *, more research is needed on autonomic reflection

Table 2.2. Standard for Reporting Diagnostic Accuracy Studies Guidelines for Heart Rate Variability Research **(**$STARD_{HRV}$) Methodology Study Quality Assessment Tool for Primary-Level Evidence.

Section & Topic	No	Item	Point Value	Reported on page #
TITLE OR ABSTRACT				
	1	Identification as a study of validation.	/1	
ABSTRACT				
	2	Structured summary of study objective, design, methods, results, and conclusions.	/1	
INTRODUCTION				
	3	Scientific and practical background, including the intended use of the index device/software	/1	
	4	Study objectives and hypotheses described.	/1	
METHODS				
Study design	**5**	Study uses within-subject design.	/1	
	6	Intended sample size and how it was determined (e.g. G*Power 3)	/1	
Participants	**7**	Eligibility criteria including specific restrictions (medical use, gender, age, activity level or BMI).	/1	
Pre-Testing	**8**	Pre-testing guidelines reported (e.g., limitations to caffeine, alcohol, physical activity etc.).	/1	
Test methods	**9**	Setup of reference standard and index device described in sufficient detail to allow replication (e.g. hardware/software such as brand, electrode configuration, etc.).	/1	
	10	Description of environmental conditions (e.g. temperature, humidity, lights on or off, time of day) and posture.	/1	
	11	A stabilization period prior to sampling was described.	/1	
Interbeat interval	**12**	The raw sampling rate and length of collection are described.	/1	
	13	Acknowledgment of breathing (e.g. controlled or not controlled).	/1	
Analysis	**14**	Description of how estimates or comparison measures were calculated (e.g. *ES*, LOA, Pearson's r or *ICC*).	/1	
	15	Reasons for missing data, along with percentage missing (e.g., equipment, persistent ectopy) and how it was handled.	/1	
	16	Interbeat artifact identification method (e.g. algorithm, manual inspection).	/1	
	17	Artifact cleaning methods and percentage of beats corrected.	/1	
	18	Description of metrics used and software/script for HRV calculation (log transformation etc.).	/1	
	19	Specification of frequency bands used and how they were calculated (e.g. Fast Fourier Transform or Autoregressive modelling).	/1	
RESULTS				
Participants	**20**	Baseline demographics of participants.	/1	
Test results	**21**	Mean ± *SD* along with at least one estimates of precision (e.g. LOA, Pearson's r or ICC).	/1	
DISCUSSION				
	22	Study limitations, (e.g., sources of potential bias, confounding variables, statistical uncertainty, and generalisability).	/1	
	23	Implications for practice, including the intended use.	/1	
OTHER INFORMATION				
	24	Where the full study protocol can be accessed if not fully described.	/1	
	25	Sources of funding and other support; role of funders.	/1	
			/25	

Abbreviations: BMI, body mass index; *ES*, effect size; *ICC*, intra-class correlation; LOA, limits of agreement; *SD*, standard deviation

Table 2.3. Item Description of the Standard for Reporting Diagnostic Accuracy Studies Guidelines for Heart Rate Variability Research **(**$STARD_{HRV}$) Methodology Study Quality Assessment Tool for Primary-Level Evidence.

Item	Explanation of Quality Item and Scoring Details	Source
1	The research study should identify as a validation study of a novel device as referenced to "gold standard" ECG. *Worth "1" point.*	**Modified, STARD 2015 item 1**
2	The abstract should be formatted to include the study objectives, design, methods, results, and conclusions in systematic order. *All must be present to receive a score of "1".*	**Modified, STARD 2015 item 2**
3	*The background along with the intended use should both be described in order to receive a score of "1" point.*	**Modified, STARD 2015 item 3**
4	The study objectives and hypotheses should be stated. *Worth "0.5"point each, 1 point total.*	**STARD 2015 item 4**
5	The Study must be within-subject design. *Worth "1" point*.	**Modified, Laborde et al. 2017**
6	An explanation of how the intended sample size was determined should be provided (e.g. G*Power 3). *Worth* "1" point.	**Modified, STARD 2015 item 18**
7	Acknowledge of participant eligibility criteria or any special considerations must be provided. *Worth "1" point*.	**Modified, STARD 2015 item 6**
8	Pre-testing guidelines should be provided in a specific manner to allow for study replication. *Worth "1" point*.	**New Addition by Authors**
9	The setup of the reference standard and index device should be described in sufficient detail to allow replication. *Detailed description of both is required to receive a score of "1". Partially or only ONE device being fully described results in a score of "0.5".*	**Modified, STARD 2015 item 10a-10b, and GRAPH item 2a**
10	The time of day and environmental conditions during sampling should be described. *Worth "0.5" each, 1 points total.*	**Modified, GRAPH item 2b**
11	A stabilization period (e.g. 5min) prior to sampling HRV should be addressed. Worth "1" point.	**New Addition by Authors**
12	The raw sampling rate and length of collection should be described. *Worth "0.5" point each, 1 point total.*	**Modified, GRAPH item 2b**
13	Breathing rate should be acknowledged if it was controlled or not (e.g. breathing 12 beats/min or freely breathing). *Worth "1" point.*	**Modified, Laborde et al. 2017**
14	A description or reference of how comparison calculations were performed should be provided. *Worth "1" point.*	**New Addition by Authors**
15	An explanation of missing data, along with the % missing and how it was handled should be provided. *A total of "1" point. If partial information is provided, give a score of "0.5".*	**Modified, STARD 2015 item 17**
16	A description should be provided on how the interbeat artifacts were identified. *Worth "1" point.*	**Modified, GRAPH item 3b**
17	The artifact cleaning method and the percentage of beats corrected should be fully described. *Worth "0.5" points each, 1 point total.*	**Modified, GRAPH item 3d**

Item	Explanation of Quality Item and Scoring Details	Source
18	A description of each metric used and how they were calculated (e.g. log transformed), along with the software used should be provided. *Each worth "0.5"point, 1point total.*	**Modified, GRAPH item 4a**
19	An explanation of which frequency bands were measured and how they were analysed should be provided (e.g. Fast Fourier Transform or Autoregressive modelling). *Worth "1" point.*	**Modified, GRAPH item 4b and Laborde et al. 2017**
20	The baseline demographics of participants (e.g. age, sex, BMI, height, weight) should be provided. *Worth "1" point.*	**Modified, STARD 2015 item 20**
21	The mean and standard deviation for each device and metric, along with one estimates of precision (e.g. LOA, Pearson's r or ICC) should be reported. *Worth "1" point. Provide partial points "0.5" if only one was provided.*	**New Addition by Authors**
22	Study limitations. *Worth "1" point.*	**(Modified, STARD 2015 item 26**
23	Implications for practice. *Worth "1" point.*	**Modified, STARD 2015 item 27**
24	Access to full study protocol. *Worth "1" point.*	**STARD 2015 item 29**
25	Sources of funding. *Worth "1" point.*	**STARD 2015 item 30**

* The source of the $STARD_{HRV}$ quality item is provided, and if applicable, describes whether the item was used in its original form, modified, or newly added by authors.
STARD, Standard for Reporting Diagnostic Accuracy Studies Guidelines
GRAPH, Guidelines for Reporting Articles on Psychiatry and Heart rate variability
Abbreviations: BMI, body mass index; *ICC*, intra-class correlation; LOA, limits of agreement

Table 2.4. Description of selected studies (*k*=23) examining the validity of measures of heart rate variability from portable heart rate monitors compared to a criterion electrocardiogram.

Author (year)	Sample Characteristics					HRV Protocol			Analysis of HRV			Results of comparison to ECG
	N (% female)	Mean age (male/ female yrs)	Mean BMI (male/ female kg/m^2)	Popu-lation	PA level	Device(s) compared to ECG	Position during measure	Length of recording	Time-domain	Freq-domain	Non-linear	
Bolkhovsky (2012)	9 & 13 (0.0)	NR	NR	NR	NR	iPhone 4s, Motarola Droid	Supine, tilt	2 min, 5 min	RMSSD, SDNN	LF, HF, LF:HF	None	Strong correlation b/w both smart phone devices and ECG, but significant differences in were present
Chellakumar (2005)	7 (0.0)	23.5	23.8†	NR	NR	Polar chestbelt, SenseWare PRO armband	Seated	5 min	Mean RR, SDNN, mean HR, RMSSD, pNN50	LF, HF	SD1, SD2, ApEn	Strong correlation (r > 0.99) b/w methods
Barbosa (2016)	30 (0.0)	20.3	24.16	NR	NR	Polar S800G3, Polar S810i	Supine, seated	30 min supine, 30 min seated	RMSSD, SDNN	LF, HF	None	No significant differences b/w methods
Esco (2017)	30 (50.0)	24.8/21.3	27.1†	Athletic	NR	ithlete infra-red pulse finger sensor	Supine, seated, standing	55 sec supine, 55 sec seated, 55 sec standing	lnRMSSD	None	None	Statistically significant difference b/w methods for sitting and supine, but very trivial effect size for all conditions
Flatt (2013)	25 (32.0)	26.7/25.4	27.7/23.0	NR	NR	ithlete smart phone app	Supine	55 sec	lnRMSSD	None	None	No significant differences b/w methods
Gamelin (2006)	18 (0.0)	27.1	25.5	Healthy	Active	Polar S810	Supine, standing	10 min, 7 min	RMSSD, SDNN, pNN50,	LF, HF, VLF	SD1, SD2	Valid measure for HRV analysis in the supine position, but caution should be taken RMSSD and SD1 in the standing position
Gamelin (2008)	12 (0.0)	9.6	17.4†	Healthy	NR	Polar S810	Supine	10 min	RMSSD, SDNN, pNN50	LF, HF, VLF	SD1, SD2	Strong correlation b/w methods
Giles (2016)	20 (15.0)	28.7	24.5†	NR	NR	Polar V800	Supine, seated	10 min, 7 min	SDNN, RMSSD, pNN50	LF, HF, VLF	SD1, SD2, SaEn	Polar V800 is a valid tool for detecting RR intervals at rest

Guzik (2018)	29 (48.3)	24.7	NR	Healthy	NR	ELITEHRV	Supine, following mental stress, standing	5 min	RMSSD, SDNN	None	None	HRV values from the ECG and mobile device were significantly different and could not be accounted for by the setting or biological sex
Heathers (2013)	10 (40.0)	25.5	NR	NR	NR	ithlete infra-red pulse finger sensor	Seated	5 min	Mean RR, SDNN, RMSSD	LF, HF	None	Smartphone pulse rate variability is an accurately reflects changes in autonomic output
Hernando (2016)	23 (0.0)	34.8	23.5†	Healthy	Active	Polar RS800	Seated, cycle ergometer (0-40, 40-60, 60-80, 80-100%)	5 min, individual variability by stage	None	LF, HF	None	Good agreement throughout the exercising protocol with increased differences at shorter RR intervals. HF power showed lower agreement at intensities above 60% $\dot{V}O_2$
Montaño (2017)	20 (55.0)	26.0	24.7	Healthy	NR	Polar RS800CX	Supine, tilt, recovery	5min	Mean RR, RMSSD, SDNN	LF, HF, LF:HF	None	Excellent agreement between devices at all positions and displayed by a strong correlation, RMSSD and HF acquired form telemetry can be used as a valid and reliable measure of HRV
Nunan (2008)	33 (42.4)	34/48 (median)	24.9†	NR	Active	Polar S810	Supine	5min	RR count, mean RR, lnSDNN, lnRMSSD	LF, HF	None	Polar S810 provides a valid measure of HRV
Nunan (2009)	33 (42.4)	34/48 (median)	24.9†	NR	Active	Polar S810	Supine	5 min	Mean RR, lnSDNN, lnRMSSD	LF, HF	None	Polar S810 can be used to record and interpolate RR intervals
Plews (2017)	26 (26.9)	31	23.7	Recreational to elite athletes	Active to highly active	HRV4TTraining, Polar H7 chest strap	Seated	5 min	RMSSD	None	None	Both devices have acceptable agreement with ECG when collecting HRV
Porto (2009)	33 (54.5)	26.1	23.3	Clinical	Active	Polar S810	Supine, standing	5 min supine, 5 min standing	RR count, Mean RR, SDNN, pNN50, RMSSD	None	None	Polar S810 is highly agreeable with ECG for RR interval analysis

Radespiel-Tröger (2003)	36 (38.9)	27.4	NR	NR	NR	Polar T31 Advantage	Seated	3 min	RMSSD, SDNN	LF, HF, VLF, TP	None	Time domain measures of HRV showed good agreement while all measures from the portable device were highly correlated with ECG measures
Romagnoli (2014)	12 (0.0)	60.8	21.6	Clinical	Active	GOW small textile shirt	Cycle ergometer exercise	3 min segments	RMSSD, SDNN,	LF, HF, LF:HF, TP	SD1, SD2	GOW system can be utilized to monitor HR during exercise but not as a clinical tool for HRV
Russoniello (2013)	137 (50.3)	31.0	NR	NR	NR	USB Pulse Wave Sensors	Seated	5 min	RMSSD	LF, HF, LF:HF, TP, VLF	None	PPG accuracy is comparable to ECG measures
Vanderlei (2008)	15 (0.0)	20.9	24.3	NR	NR	Polar S810i	Supine, seated on exercise ergometer	20 min supine, 20 min exercise on erg	Heart beats, pNN50, RMSSD	LF, HF	None	Polar S810i is a reliable tool for calculating HRV in the time and frequency domains
Vasconcellos (2015)	14 (21.4)	15	25.2/26.4	Obese insulin resistant	6 months of activity prior	Polar RS800CX, PPG Finapress	Seated	5 min	RMSSD, pNN50	LF, HF, LF:HF	None	HRV from Polar RS800CX and PPG provide accurate measures in comparison to ECG
Wallen (2012)	341 (59.2)	52.0/53.0	NR	NR	NR	Polar RS800-PPT 5	Supine	5 min	SDNN, RMSSD	HF, LF	None	Polar RS800 did not correct errors, and did not produce valid results for women, specifically women over age 60. May not be appropriate for a clinical setting.
Weippert (2010)	19 (0.0)	24 (median)	22.7†	Healthy	Active	Polar 810i, Suunto t6	Supine, seated, walking, upper limb isometric	3 min each	None	LF, HF, TP, VLF	None	The three devices can be used interchangeable for RR interpolation, but the LoA between HRV frequency parameters were unacceptable

†, calculated from data provided; %, percentage; yrs, years; ApEn, approximate entropy; b/w, between; BMI, body mass index; kg/m, kilograms/meters; ECG, electrocardiogram; HF, high frequency; HR, heart rate; HRV, heart rate variability; LF, low frequency; lnRMSSD, log transformed RMSSD; LoA, limits of agreement; *N*, study population size; NR, not reported; PA, physical activity; pNN50, percentage value of consecutive RR intervals that differ by more than 50 milliseconds; RR, RR interval; RMSSD, square root of the mean squared differences between normal adjacent RR intervals; SaEn, sample entropy; SD1, dispersion of points perpendicular to the line of identity; SD2, dispersion of points along the line of identity; SDNN, standard deviation of the mean of all normal RR intervals; TP, total power; VLF, very low frequency.

Table 2.5. Item-by-item summary of methodological study quality for the included studies (*k*=23) using a version of the Standard for Reporting Diagnostic Accuracy Studies modified for the use of heart rate variability ($STARD_{HRV}$).

Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	**Rating**
Barbosa et al. (2016)	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	**88%**
Bolkhovsky et al. (2012)	1	1	1	½	1	0	0	0	1	½	0	1	0	1	½	½	½	1	1	0	1	1	1	1	0	**62%**
Chellakumar et al. (2005)	1	1	1	½	1	0	0	1	1	½	1	1	0	1	0	1	0	1	1	1	1	1	1	1	0	**72%**
Esco et al. (2017)	1	1	1	½	1	0	1	1	1	0	1	0	1	1	0	0	0	1	n/a	1	1	0	1	1	1	**73%**
Flatt & Esco (2013)	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	n/a	0	1	1	1	1	0	**88%**
Gamelin et al. (2006)	1	1	1	½	1	0	1	1	½	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	**80%**
Gamelin et al. (2008)	1	1	1	½	1	0	1	1	½	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	**84%**
Gileset et al. (2016)	1	1	1	½	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**90%**
Heathers et al. (2013)	1	1	1	½	1	0	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	**78%**
Hernando et al. (2016)	1	1	1	½	1	0	1	1	½	½	1	1	0	1	0	1	0	1	1	1	½	1	1	1	1	**76%**
Montaño et al. (2017)	1	1	1	½	1	0	1	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	**78%**
Nunan et al. (2008)	1	1	1	½	1	0	1	1	1	½	0	1	0	1	1	1	½	1	1	1	1	0	1	1	0	**74%**
Nunan et al. (2009)	1	1	1	½	1	0	1	1	1	½	1	½	0	1	1	1	½	1	1	1	1	0	1	1	1	**82%**
Plews et al. (2017)	1	1	1	½	1	0	0	0	1	0	1	½	1	1	½	1	½	1	n/a	1	½	1	1	1	1	**73%**
Porto et al. (2009)	1	1	1	½	1	0	1	1	1	1	1	1	1	1	1	1	1	1	n/a	1	1	1	1	1	0	**90%**
Radespiel-Tröger et al. (2003)	1	1	1	½	1	0	0	0	1	½	0	½	1	1	0	1	½	1	1	1	1	0	1	1	0	**64%**
Romagnoli et al. (2014)	1	1	1	½	1	0	1	1	1	1	n/a	1	n/a	1	½	1	½	1	1	1	1	1	1	1	1	**89%**
Russoniello et al. (2013)	1	1	1	1	1	0	1	1	1	½	0	1	1	0	0	1	0	1	1	1	1	1	1	1	1	**78%**
Vanderleiet al. (2008)	1	1	1	½	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	**86%**
Vasconcellos et al. (2015)	1	1	1	½	1	0	1	1	1	1	1	1	1	1	0	½	½	1	1	1	1	1	1	1	1	**86%**
Wallen et al. (2012)	1	1	1	½	1	0	1	1	1	½	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**92%**
Weippert et al. (2010)	1	1	1	½	1	0	0	0	1	½	1	1	n/a	1	0	1	1	1	1	1	1	0	1	1	1	**75%**
Guzik et al. (2018)	1	1	1	½	1	0	0	0	½	0	1	1	0	1	0	1	1	1	n/a	1	1	1	1	0	1	**67%**

n/a, not applicable due to the study design.

Table 2.6. Individual moderator analyses for categorical variables of interest (*n*=301 total effects).

Moderator	*n*	%	*ES*	*SE*	*p*-value
Position*					
Supine	140	46.5	0.24	0.09	.009
Tilt/Recovery	20	6.6	0.56	0.17	.001**
Sitting	56	18.6	0.21	0.11	.043
Standing	27	9.0	0.22	0.12	.060
Active	58	19.3	0.16	0.14	.254
Metric*					
SDNN	34	11.3	0.44	0.11	<.001**
RMSSD	49	16.3	0.26	0.10	.012
pNN50	16	5.3	0.22	0.13	.097
LF	65	21.6	0.19	0.10	.064
HF	66	21.9	0.24	0.10	.021
LF:HF	29	9.6	0.05	0.11	.656
TP	11	3.7	0.14	0.12	.250
VLF	15	5.0	0.15	0.12	.209
SD1	8	2.7	0.21	0.17	.214
SD2	8	2.7	0.18	0.17	.286
Device					
EliteHRV	4	1.3	0.16	0.45	.727
Finapress PPG cuff	5	1.7	0.81	0.36	.026
GOW Shirt	9	3.0	0.14	0.45	.756
HRV4Training PPG	2	0.7	0.03	0.47	.947
iPhone4S	10	3.3	0.82	0.46	.074
Ithlete PPG	7	2.3	0.1	0.33	.761
Ithlete Chest Strap	1	0.3	<0.01	0.52	.995
Motorola Droid	10	3.3	0.75	0.45	.094
Polar H7	2	0.7	0.02	0.47	.969
Polar RS800	34	11.3	0.29	0.31	.352
Polar RS800CX	20	6.6	1.25	0.32	<.001
Polar RS800G3	14	4.7	0.04	0.27	.875
Polar S810	70	23.3	0.2	0.2	.316
Polar S810i	40	13.3	0.06	0.25	.827
Polar T31 Advantage	2	0.7	<0.01	0.46	.987
Polar T31 Armband	10	3.3	<0.01	0.46	.985
Polar T31 Counter	10	3.3	<0.01	0.46	.988
Polar V800	22	7.3	0.04	0.46	.935
SUUNTOt6	16	5.3	0.09	0.27	.743
PPG Not Identified	13	4.3	0.05	0.43	.918

*, significant omnibus test; **, significantly different from all other measure within the variable; %, proportion of effects accounted for; *ES*, estimated absolute standardized mean difference effect size; *n*, number of effects; PPG, photoplethysmography; *SE*, standard error.

Table 2.7. Significant moderating relationships within the multiple moderator model (*n*=301 total effects).

Position	**Device**	**Metric**	***ES***	***SE***	**95% *CI***	***p*-value**
Supine	Polar RS800CX	HF	1.29	0.56	0.20 - 2.38	.02
Supine	Polar RS800CX	LF	1.91	0.53	0.87 - 2.95	< .001
Supine	Polar RS800CX	RMSSD	1.826	0.50	0.85 – 2.8	< .001
Supine	Polar RS800CX	SDNN	2.68	0.67	1.37 – 3.99	< .001
Supine	iPhone4s	SDNN	1.50	0.74	0.04 – 2.96	.006
Tilt/Recovery	Polar RS800CX	HF	65.9	5.55	55.02 - 76.78	< .001
Tilt/Recovery	Polar RS800CX	LF	10.87	1.15	8.61 - 13.13	< .001
Tilt/Recovery	Polar RS800CX	RMSSD	3.31	0.59	2.16 - 4.47	< .001
Tilt/Recovery	Polar RS800CX	SDNN	2.31	0.54	1.25 - 3.37	< .001
Tilt/Recovery	iPhone4s	SDNN	2.18	0.79	0.64 - 3.73	.006
Tilt/Recovery	Motorola Droid	SDNN	1.41	0.68	0.08 - 2.74	.038

CI, confidence interval; *ES*, estimated absolute standardized mean difference effect size; HF, high frequency power; LF, low frequency power; RMSSD, square root of the mean squared differences between normal adjacent RR intervals; SDNN, standard deviation of the mean of all normal RR intervals; *SE*, standard error.

Table 2.8. Individual moderator analyses for categorical variables of interest (*n*=275 total effects).

Moderator	*n*		*ES*	*SE*	*p*-value
Position*					
Supine	125	41.5	0.13	0.03	<.001
Tilt/Recovery	9	3.0	0.51	0.13	<.001**
Sitting	56	18.6	0.05	0.03	.079
Standing	27	9.0	0.04	0.06	.531
Active	58	19.3	0.06	0.04	.144
Metric					
SDNN	25	8.3	0.07	0.06	.232
RMSSD	44	14.6	0.13	0.04	.004
pNN50	16	5.3	0.07	0.09	.462
LF	60	19.9	0.11	0.04	.005
HF	60	19.9	0.11	0.04	.004
LF:HF	28	9.3	0.12	0.05	.029
TP	11	3.7	0.04	0.07	.583
VLF	15	5.0	0.04	0.07	.582
SD1	8	2.7	0.05	0.14	.693
SD2	8	2.7	0.02	0.14	.861
Device*					
EliteHRV	4	1.3	0.15	0.13	.247
Finapress PPG cuff	5	1.7	0.06	0.17	.720
GOW Shirt	9	3.0	0.14	0.14	.321
HRV4Training PPG	2	0.7	0.03	0.20	.873
iPhone4S	8	2.7	0.61	0.17	<.001
Ithlete PPG	7	2.3	0.10	0.12	.436
Ithlete Chest Strap	1	0.3	<0.01	0.28	.991
Motorola Droid	8	2.7	0.55	0.14	<.001
Polar H7	2	0.7	0.09	0.20	.927
Polar RS800	28	9.3	0.19	0.04	<.001
Polar RS800CX	8	2.7	0.29	0.13	.018
Polar RS800G3	14	4.7	0.06	0.07	.353
Polar S810	66	21.9	0.07	0.04	.042
Polar S810i	40	13.3	0.05	0.05	.269
Polar T31 Advantage	2	0.7	<0.01	0.17	.965
Polar T31 Armband	10	3.3	<0.01	0.17	.960
Polar T31 Counter	10	3.3	<0.01	0.17	.968
Polar V800	22	7.3	0.04	0.07	.598
SUUNTOt6	16	5.3	0.05	0.03	.189
PPG Not Identified	13	4.3	0.05	0.08	.559

*, significant omnibus test; **, significantly different from all other measure within the variable; %, proportion of effects accounted for; *ES*, estimated absolute standardized mean difference effect size; *n*, number of effects; PPG, photoplethysmogrophy; *SE*, standard error.

Table 2.9. Significant moderating relationships within the multiple moderator model after removal of outliers (*n*=275 total effects).

Position	**Device**	**Metric**	***ES***	***SE***	**95% *CI***	***p*-value**
Supine	iPhone4s	RMSSD	1.00	0.50	0.02 – 1.98	.046
Supine	Motorola Droid	LF	0.84	0.41	0.04 – 1.64	.04
Supine	Polar RS800	LF	0.24	0.10	0.04 – 0.44	.017

CI, confidence interval; *ES*, estimated absolute standardized mean difference effect size; HF, high frequency power; LF, low frequency power; RMSSD, square root of the mean squared differences between normal adjacent RR intervals; SDNN, standard deviation of the mean of all normal RR intervals; *SE*, standard error.

CHAPTER 3

TIME COURSE IN RECOVERY BETWEEN HEART RATE VARIABILITY AND MUSCULAR PERFORMANCE FOLLOWING STRENUOUS RESISTANCE EXERCISE

ABSTRACT

The aim of this study was to examine the time course in recovery between multiple HRV metrics and muscular performance over a 72-hour period following an exhaustive bout of resistance training. Eight resistance trained males completed 5 laboratory visits within a 7-day period. The first visit involved short-term HRV recordings followed by a familiarization of a battery of non-fatiguing performance measures (vertical jump, back squat bar velocity, isokinetic leg extension and flexion, and isometric mid-thigh pull), and a one-repetition max test of the back squat. Participants returned to the laboratory 48-hr later for baseline measures, immediately followed by an exhaustive back squat protocol (8-sets of 10-repetitions with 2-min rest). The HRV and muscular performance measures were replicated at 0.5, 24, 48, and 72-hours post-exhaustive squat protocol. Recovery scores in HRV and muscular performance tests were calculated as change scores (baseline - trial / baseline). The battery of muscular performance measures were averaged for each trial and classified as the "*performance*" variable. A multivariate profile analysis between recovery scores for each HRV metric and *performance* was computed which included a test of parallelism and multivariate analysis of variance (MANOVA). Furthermore, a repeated measures correlation was performed on HRV and *performance* over time to determine the intra-individual association. All HRV metrics had a significant interaction with *performance* over time ($p < .01$) indicating change scores in

performance and HRV following the physiological stressor were not parallel. MANOVA results showed that mean change scores in all HRV metrics significantly differed from *performance* ($p < .05$) across time, except the standard division of all normal-to-normal R-R intervals (SDNN), low frequency power (LF), and the standard deviation of long-term HRV from the Poincaré plot (SD2) at the 0.5-hr mark, and high frequency power (HF) at the 24-hr time point. Repeated measures correlation analysis indicated a lack of intra-individual association between the change in *performance* and HRV over time (all < .45). Recovery in HRV measures following an exhaustive bout of resistance training was not associated with muscular performance recovery.
Key Words: Heart Rate Variability, Muscular Performance Recovery, Resistance Training, Autonomic Nervous System, Multivariate Analysis

INTRODUCTION

The ability to monitor the fatigue-recovery relationship is an invaluable aspect of performance when monitoring individuals in clinical and competitive settings. The continuous mental and physical stressors of training constantly tax human homeostasis. This ongoing conflict is constantly regulated through the autonomic nervous system (ANS). Overall ANS activity is a balance between the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS); therefore, the evaluation of these branches has become a useful indicator of health and disease [1, 2]. From the late 1940's, heart rate variability (HRV) has emerged as a common method of ANS evaluation [3], assessed by measuring the variations in instantaneous heart rate (HR) and normal – normal (NN) or (RR) intervals across QRS complexes [4]. Measurement of HRV is performed by analyzing HR data gathered using electrocardiography

(ECG) or commercially available mobile devices and proprietary software. Over the past 2 decades, HRV interpretation has been utilized to evaluate readiness to perform [5, 6].

HRV values can be calculated through linear or non-linear analyses. Linear analyses include the use of time domain metrics which are perhaps the simplest to quantify and have been utilized for reliable athletic monitoring in the detection of change in aerobic capacity or non-functional overreaching and overtraining [7-10]. Likewise, spectral analyses of respiratory frequencies have been utilized for monitoring HRV recovery post exercise [11-18]. Non-linear HRV analysis via Poincaré plotting (acquired by plotting each RR interval against the previous RR intervals) has also provided valuable insight into autonomic modulation resulting from physical stressors and training [19-21].

The relationship between cardio-autonomic perturbations following acute bouts of resistance training has also been investigated [13, 22, 23]. For example, Heffernan et al. [13] determined that cardiac parasympathetic activity is reduced to a greater degree 30-min following an acute resistance training compared to acute endurance training. Similarly, Teixeira et al. [24] also found decreased high frequency power (HF) and increased low frequency power (expressed as normalized units [LFnu]) 20-min following a bout of acute resistance training, suggesting a blunting of parasympathetic reactivation and increased sympathetic activation following resistance training. However, more research is needed to identify the relationship between the time course in internal (HRV) and external (muscular performance) recovery following a physical strain in order to further the understanding of the physiology of recovery.

To our knowledge, only one study has assessed the use of HRV as a marker of performance recovery over a 72-hr period [18]. In the study, the relationship between spectral analysis of HRV and weightlifting performance in trained weightlifters at four time points (3, 24,

48, and 72 hrs) following a high intensity resistance training stimulus was evaluated. Spectral analysis of HRV reflected performance recovery, as measured by one repetition maximum (1RM) values at the 72-hr time point. As such, the authors provided evidence that HRV may be utilized by practitioners as a useful indicator of recovery of resistance training performance. However, as there are multiple HRV metrics which are reflective of the PNS, SNS, or a combination, the debate as to which metric provides the greatest relationship with acute performance recovery has yet to be resolved.

With regards to utility, the use of spectral analysis requires an understanding of sophisticated algorithms which may not be readily accessible to practitioners in the field. Time-domain measures such as the root mean square of successive RR interval differences (RMSSD) can be calculated with an Excel spreadsheet and have shown promise for longitudinal analysis of fatigue and recovery in athletes, which can be utilized by practitioners [25]. Likewise, the utilization of non-fatiguing objective performance measures (e.g. vertical jump) serve as practical methods of monitoring external recovery in athletes [26]. Therefore, the purpose of this study was to determine which metric of HRV from frequency-domain, time-domain, and Poincaré plotting, best reflected recovery of muscular performance up to 72-hrs following a strenuous bout of resistance exercise in strength-trained participants. As parasympathetic indices of HRV have been utilized for monitoring readiness to perform and throughout longitudinal monitoring of athletes, it was hypothesized that HRV metrics which primarily reflect PNS activity (RMSSD, HF, and SD1) would yield the strongest correlation with changes in muscular performance.

MATERIALS AND METHODS

Experimental Design

The experimental protocol consisted of a single-group repeated measures design, in which each participant served as their own control (see Figure 3.1 for a visual depiction of the repeated measures design). Baseline measures of HRV and muscular performance were acquired over a 48-hr period, prior to an exhaustive bout of resistance training. HRV and muscular performance were assessed at four time points (0.5, 24, 48, and 72 hours) over a 72-hr period following the exhaustive exercise bout. The resistance training bout (8 sets, 10 repetitions, 70% 1RM) consisted of a modified back squat protocol previously described by Bartolomei et al. [27]. All measurements were performed between 5:00 and 10:00 am as current recommendations suggest HRV measures should be acquired early in the morning, as close to awakening as possible [28].

Subjects

Nine healthy males volunteered to participate in the study. One participant was removed from the collection as he was unable to complete the exercise protocol. Therefore, 8 men [mean (*SD*) age = 23.3 (3.9) y, mass = 90.4 (10.9) kg; height = 180.8 (8.3) cm, body mass index = 27.7 (3.3) kg/m^2, percent body fat estimated from skinfold = 13.0 (4.3) %, 1RM = 158.6 (34.2) kg, relative strength = 1.7 (0.2) kg/body weight] who were non-smokers; free from cardiovascular, pulmonary, metabolic disease and had been involved in a moderate-to-vigorous training program for over 6-years participated in the study. A power analysis revealed that a sample of 7 participants would be sufficient to detect a small effect (0.3) for HRV outcome variables when compared across time, assuming power = 0.80 and alpha = 0.05 [29]. All participants signed an

informed consent approved by the Institutional Review Board of The University of Alabama prior to participating.

Procedures

Participants were instructed to refrain from ingesting alcohol, stimulants (e.g. caffeine), and non-prescription drugs and to avoid strenuous activity 12-hr prior to the initial visit and throughout the data collection process. Participants were also instructed to attend each laboratory session in a fasted state. A 24-hr history questionnaire was completed prior to every trial to ensure adherence throughout the study. Experimental trials were collected at the same time of each day (± 0.25 hours).

Heart Rate Variability

On the first day (familiarization day), subjects arrived to the laboratory upon wakening and were instructed to complete paperwork (i.e., informed consent, 24-hr history questionnaire, medical history questionnaire). Next, participants were fitted with a modified lead-II ECG configuration utilizing three Ag/AgCl surface electrodes as previously described [30]. The right mid-clavicular notch served as the negative electrode, the fifth intercostal on the left mid-axillary line represented the positive electrode, and the left anterior-superior iliac crest served as the ground electrode. Each site was cleaned with an alcohol swab prior to electrode placement. Electrodes were connected to a data-collection device (Biopac MP100 Goletta, CA) which was interfaced with a laptop computer for data acquisition. Participants were then instructed to sit with their legs at a 90° angle and with their hands resting on their thighs in an enclosed dim lit controlled environmental chamber (24°C, 20% relative humidity) to reduce noise and external

influence [31]. A 5-min acclimation period (sitting still) preceded a 5-min HRV recording (10-mins total) to ensure stability of the measure [4]. Software (Acknowledge v 3.9, BIOPAC, Goletta, CA, USA) was used to collect ECG signals at a sampling rate of 2000 Hz to warrant appropriate sampling of the fiducial point of each QRS complex. No attempt was made to control for breathing rate as it has been shown to potentially bias the interpretation [31].

On the familiarization day only, anthropometric and body composition data were assessed following the HRV measurements. Participant height was measured to the nearest 0.1cm using stadiometer (SECA 67310, SECA©, Chino, CA), body weight was measured to the nearest 0.1kg using a digital scale (Tanita BWB-800, Tanita©, Arlington Heights, IL) , and 3 site skinfold thicknesses (chest, abdomen, and thigh sites) were attained by a trained research assistant [32]. Following anthropometric and body composition measures, participants performed a standardized warm-up previously described by Bartolomei et al. [27] prior to executing the physical performance measures. After completion of the warm-up, 3-min rest was allocated preceding the non-fatiguing muscular performance protocol.

Non-Fatiguing Performance Indicators

The non-fatiguing protocol began with a counter-movement vertical jump (CMJ), which was assessed by participants jumping on a portable force plate (Kistler 9286BA 10kN, Switzerland). Participants performed three jumps with 30-s rest between each jump and were instructed to jump as high as possible while maintaining their hands on their hips. Jump height (cm) was calculated by software (Measurement, Analysis & Reporting Software, Kistler Group, Switzerland; MARS) with the highest value of the three jumps recorded.

Following the CMJ measurements, participants performed three back squat repetitions at 70% 1RM (estimated during the familiarization trial) with 2-min rest between each set. First, a validated linear position transducer (GymAware PowerTool, GymAware, Kinematic Performance Technology, Canberra, Australia) was attached to the unloaded barbell [33]. Participants were then instructed to perform a full squat with the unloaded barbell. The top position (i.e., full extension of the hips and knees) and the full squat position (i.e., flexing of the knee and hip with femur parallel to the ground) were recorded on the GymAware software [34]. Once the corresponding position was recorded, the bar was loaded with 70% of the participant's 1RM and the 3 trial repetitions commenced. Squat depth was assessed by a certified strength and conditioning specialist between all repetitions to ensure reproducibility throughout the investigation. The highest peak bar velocity of the 3 repetitions was recorded [34, 35].

After the bar velocity assessment, participants were seated on an isokinetic dynamometer (Humac Norm Computer Sports Medicine Inc., Stoughton, MA). Three maximal isokinetic leg extension measurements were acquired on the participant's dominant leg at an angular velocity of 180°/s of concentric contraction [27]. A 30-s rest was allotted between each attempt and the highest peak torque (N·m) of the three contractions was recorded.

Lastly, participants were situated over force plates inside a power rack where they performed 2 isometric mid-thigh pulls lasting 5-s each, with 3-mins rest between each attempt. If the 2 attempts differed by 250 N or greater, a 3rd attempt was performed [36]. Participants were instructed to assume the position of the second pull of a power clean with their hands fixated onto the bar with straps. Hand width, knee angle, and hip angle were measured for reproducibility at every trial. Knee and hip angle were approximately 125° and 145° as this position has been shown to produce the greatest amount of force [37]. The highest value of peak

force (N) over the 5-s pull, and peak rate of force development ($N \cdot s^{-1}$) through a 250-ms window, were acquired by a customized program (LabVIEW, National Instruments, Version 17.0) [27, 38].

After completing the non-fatiguing muscular performance measures, the participant's 1RM back squat was established. The 1RM values were determined through an incremental increase of intensity within several trials, allowing 3 to 5 minutes rest between each attempt as described by Hoffman [39]. Task progression involved three warm up sets, 5-10 repetitions at ~50% and three repetitions at ~85% followed by one repetition at 90-95% 1RM. The remaining attempts were allocated toward attaining a 1RM value.

Forty-eight hours following the familiarization trial, participants returned to the lab and repeated all measures previously described with the exception of anthropometric and 1RM back squat assessments. The non-fatiguing muscular performance values from this trial were utilized as baseline (BL) values. The average HRV measures between the familiarization and pre-fatiguing trial were computed to represent the BL values for HRV as the values between the two trials did not significantly differ for any metric ($p>.05$). Immediately following the pre-fatiguing trial, participants performed the exhaustive back squat protocol as previously described [27]. In the event that all 10 repetitions could not be accomplished within a set, a 30-s recovery was allocated before the remaining repetitions were completed. If 3 failed repetitions occurred within 1 set, the 2-min inter-set rest period commenced and the load was reduce by 5% of the participant's 1RM for the remainder of the squat protocol. If residual repetitions were present after the completion of set 8, the participant was allotted a 30-s rest before completing the remaining repetitions. This ensured that the total volume (80 repetitions) was consistent across all participants.

Following the exhaustive protocol, participants rested passively for 30-min s before completing HRV and non-fatiguing muscular performance measures in the order previously specified. These values were utilized to represent the 0.5-hr trial. Participants returned to the laboratory on three additional occasions over a 72-hr period to perform HRV and non-fatiguing performance measures which represent the 24, 48, and 72-hr trials.

HRV Analysis/cleaning

After collection of HRV data, missing and ectopic beats were identified through visual inspection of the tachogram and replaced with interpolated values as previously recommended [4]. Once reduced, HRV data were analyzed using HRV software (Kubios v 3.0.2, Biosignal Analysis and Medical Imaging Group, Department of Physics, University of Kuopio, Kuopio, Finland) [40]. Time domain measures were quantified as the standard deviation of all NN intervals (SDNN) and RMSSD. A Fast Fourier Transform (FFT) algorithm was utilized to obtain power spectral analysis estimates of HRV (ms to hertz [Hz]) through detection of very low-frequency (VLF) < 0.04 Hz, low-frequency (LF) 0.04 – 0.15 Hz and HF 0.15 -0.4 Hz bands and total power (TP) reported as raw (ms^2) or normalized units (nu) [4]. These metrics were computed as:

$$\mathrm{LF(nu)} = \frac{\mathrm{LF}}{\mathrm{TP-VLF}} \times 100 \qquad \mathrm{HF(nu)} = \frac{\mathrm{HF}}{\mathrm{TP-VLF}} \times 100$$

Poincaré plots were acquired by plotting each RR interval against previous RR intervals providing a 2-dimensional oval appearance rotated at a 45° angle [19]. The standard deviation of the points through the width of the plot (SD1), represents short-term HRV, while the standard

deviation of the points along the length of the plot (SD2) represents long-term HRV [41]. Inherently, HRV measures may be non-parametric. Therefore, all HRV measures were log transformed (ln) for better interpretation [4].

Recovery Status

To account for different units of measure across the various assessments (HRV and muscular performance), a recovery score was calculated as a percentage of baseline to determine percentage of recovery, assuming baseline measures represented 100 percent recovered. A recovery score was determined for each measure at each time point using the following computation:

$$\text{Recovery} = \frac{\text{Trial}}{Baseline} \times 100$$

For the muscular performance assessment only, multiple measures were pooled at each time point to represent a single measure of "*performance*", which was then examined in relation to with HRV metrics over the course of the 72-hr recovery period.

Statistical Analysis

A multivariate profile analysis between recovery scores for each HRV metric and *performance* was computed. This included a repeated measures test of parallelism and a multivariate analysis of variance (MANOVA). A non-significant *F*-test for the group*time interaction between recovery score means would indicate the recovery scores were parallel across time [42]. The data were restructured to a person-period data set to perform the MANOVA and assess whether the mean change in the respective HRV metric and *performance*

implied differed at equivalent time points. Furthermore, a one-way analysis of variance (ANOVA) was performed on each HRV metric and *performance* to describe the change from baseline across time. The assumption of normality was assessed through visual inspection of Q-Q plots of standardized residuals. Mauchly's test was performed to assess the assumption of sphericity in the repeated measures analysis. When sphericity was not violated, the univariate within-subject results were interpreted to improve power due to the low number of observations compared to the amount of repeated measures [43]. For the MANOVA, Levene's test of equality of variances was assessed at each time point. Additionally, a repeated measures correlation was calculated to assess the common intra-individual association between the recovery score in each HRV metric and *performance* [44, 45]. Statistical analyses were performed using Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA, USA), SPSS Statistics 23 (SPSS Inc. Chicago, IL) and the *rmcorr* function in R [46]. All statistical tests utilized an alpha level of 0.05.

RESULTS

The repeated measures analysis indicated that sphericity was not violated ($p > .05$) for any of the variables tested. Therefore, univariate repeated measure results were used to assess the interaction between *performance* and HRV metric over time. Visual inspection of the Q-Q plots revealed acceptable normality for all HRV metrics and *performance* at each time. Levene's test of equality of variance was violated for *performance* and LFnu at the 24-hr ($p = .026$) and 48-hr time point ($p = .035$) but was not violated for any other comparison at any time point. Therefore, caution is warranted with the interpreted mean differences between LFnu and *performance* at the 24 and 48-hr time points.

All HRV metrics had a significant interaction with *performance* over time ($p < .01$) indicating change scores in *performance* and HRV following the physiological stressor were not parallel. The MANOVA analysis revealed there were significant mean differences between recovery scores in *performance* and HRV over time. *Performance* recovery scores were significantly lower 0.5, 24, and 48-hrs post stimulus but were not significantly different at the 72-hr time point. Recovery scores for all HRV metrics, except for LFnu, significantly differed from baseline at the 0.5-hr post stimulus but did not significantly differ from baseline at any other time point. The relationship between the *performance* and the HRV metrics are visually depicted in Figure 3.2.1 and Figure 3.2.2. Furthermore, repeated measures correlation analysis indicated a lack of intra-individual association between the change in *performance* and HRV over time (all < .45). An example of the lack of intra-individual association is depicted in Figure 3.3 and represents two participants' recovery scores for RMSSD and *performance*. Results from the test of parallelism and the repeated measure correlation for each HRV metric are displayed in Table 4.1.

DISCUSSION

This study evaluated the relationship between recovery scores in pooled muscular performance (CMJ, bar velocity, isokinetic leg extension/flexion, and isometric mid-thigh pull) and the natural log transformation of HRV metrics (RMSSD, SDNN, HF, HFnu, LF, LFnu, SD1, and SD2) over a 72-hr period following an exhaustive back squat protocol. HRV collection, cleaning, and analysis were performed in a well-controlled environment as previously recommended [4], and *performance* measures carefully monitored for reproducibility by trained personnel to ensure internal validity. The value of this investigation was to determine the

relationship between cardio-autonomic responses and muscular performance recovery following a bout of exhaustive resistance training over a 72-hr period. The cumulative results of eight participants revealed that none of the recovery scores from the HRV metrics followed a similar profile as *performance* recovery scores and there was weak to moderate intra-individual association. Therefore, the results of the current study indicate that the natural log of HRV metrics acquired from short-term recordings in the seated position lacked association with *performance* and therefore did not follow a similar time course of recovery. Instead, HRV measures returned to baseline within 24-hr while *performance* was significantly reduced up to 48-hr following the exhaustive bout of resistance training, suggesting objective indicators of internal strain (HRV) precede the recovery of external indicators (*performance)*.

In comparison, Chen et al. [18] who also investigated the time course in recovery between measure of muscular performance and HRV over a 72-hr period, observed a restoration of muscular performance which preceded HRV recovery. Muscular performance and parasympathetic modulation, as represented through HF HRV, were both significantly increased 72-hr following a 2-hr bout of high intensity resistance training. However, HF decreased significantly 24-hrs following the resistance training bout which was not reflected by changes in muscular performance. Muscular performance recovery actually superseded HF at hour 48 as 1RM back squat, dead lift, and front squat were significantly increased from baseline. The discrepancy in these findings may be representative of the stressor stimulus and external recovery markers performed. Chen et al. [18] induced a high intensity stressor after a 2-week detraining taper period to which the trained power lifters in the respective study may have been acclimated. This may have induced a super compensation effect as observed through their significant improvement in muscular performance which is advantageous and was congruently

distributed by HF power at the 72-hr time point, but not at any other time point. Likewise, muscular performance measures were acquired through assessments of 1RM values. The impact of the repeated 1RM bouts may have resulted in further suppression of parasympathetic activity due to internal strain. Thus, the performance measures may have influenced the time course of internal recovery.

A key strength to this investigation was the utilization of a high-volume stressor designed to induce fatigue, a protocol which has previously been shown to significantly delay muscular performance recovery compared to a high intensity intervention [27], along with non-fatiguing measures of muscular performance to limit the potential impact of ongoing internal and external strain. In response, only one participant had *performance* values which returned to or superseded baseline measures (~104%). Thus, our analysis revealed that fatigue was significantly induced and prolonged throughout the data collection period, although the reduction was no longer significant at the 72-hr period. The significant reduction in performance was a notable strong point of this study as we sought to explore the association in the time course of recovery between HRV and muscular performance as a strong association would be an invaluable asset to practitioners in their efforts to prevent adverse training effects (e.g., non-functional over reaching or training induced injury). However, the trend in *performance* was not consistently associated with any measure of cardio-autonomic activity over the 72-hr period. Instead, all measures of HRV, except for LFnu, significantly dropped 0.5-hr post stimulus, which is congruent with current literature [13, 18, 23, 24], but returned within baseline values throughout the remainder of the collection. Furthermore, the intra-individual relationship, as represented through the repeated measures correlation, assists in the elaboration of the relationship between the HRV metrics and *performance* across time. With the largest correlation of approximately .41 (roughly

16% of the variance in *performance* accounted for by HRV) we can determine that changes in *performance* were not appropriately accounted for by any measure of HRV throughout the 72-hr period.

Interestingly, measures of total variability (SDNN, LF, and SD2) did not significantly differ from *performance* at the 0.5-hr time point. This may have been attributed to the continued sympathetic dominance following the back squat protocol as SDNN, LF, and SD2 are believed to represent both parasympathetic and sympathetic activity [47]. This is supported by the slight increase in LFnu which some suggest is representative of sympathetic modulation (see Figure 3.2.2) [4]. Thus, the heightened sympathetic activity 0.5-hrs following the exhaustive back squat protocol may have resulted in the attenuated decrement in HRV indices which represent a combination of both ANS branches compared to those which are indicative of PNS activity. Therefore, the increased SNS activity provides a plausible explanation for the non-significant difference (similar trend) between these HRV metrics and the reduction in *performance* 0.5-hr post strenuous exercise.

The relationship between the time course of internal and external physiological recover measures has important implication for the practitioner. Although typically employed in aerobic training, the utilization of HRV monitoring for guiding periodized training models and assessing readiness to perform has increased in popularity. For example, Kiviniemi et al. [6] observed a significant increase in peak oxygen uptake ($\dot{V}O_{2peak}$) and maximal running velocity following a four week aerobic training protocol in a group of men who utilized daily HRV values to determine the intensity of the training load (i.e., no change or increase from a baseline 10-day average HRV permitted a high intensity training load). These improvements were not observed in the traditional training group which did not utilize HRV measures to determine training load

and followed a predetermined training regimen. Similarly, Nuuttila et al. [48] observed a significant increase and large effect size (*ES* = .88) in vertical jump performance following an eight week block training protocol when administering HRV guided training (HRV values composed of three day averages) compared to pre-determined block training. However, these examples used average HRV measures for interpreting responses to training and do not address whether acute changes in HRV are related to acute changes in muscular performance metrics following resistance exercise. Therefore, less is known about the relationship between HRV and muscular performance recovery following resistance exercise and whether there is a viable application for the practitioner.

To further the body of literature in this regard, our results demonstrated a weak correlation between the time course in recovery between *performance* and HRV measures. This supports previous literature which suggests the utility of short-term HRV measures for monitoring readiness to perform may not be optimal within an isolated setting and should be a compilation of average weekly measures [28, 49, 50]. This is further illustrated in Figure 3.3 which displays an extremely similar time course in internal recovery (as represented through RMSSD) between two participants. Yet, external *performance* recovery drastically differed as one participant did not deviate from baseline and the other continually decreased throughout the 72-hr period. Thus, if the isolated RMSSD measures were used as a gauge for readiness to perform in this example, participant A may have benefited from a progressive training load while participant B would likely be subject to non-functional overreaching or training induced injury if a progressive training stimulus was allocated. Thus, practitioners should compile longitudinal athletic profiles, composed of objective (internal and external) and subjective measures, in an attempt to appropriately address individual responses to training and readiness to perform.

This study is not without limitations. It was the intent to recruit resistance trained men (all with over 6-years of resistance training and 1RM back squat ≥ 1.5 times body weight) so these results cannot be generalized to other participants (e.g., youth, untrained individuals, female participants, or clinical populations). Furthermore, this analysis was performed on acute muscular performance changes. Therefore, these results do not reflect potential usefulness for monitoring cardio-autonomic perturbations through chronic resistance training and longitudinal monitoring of an athletic profile.

With current advancements in technology and the increased availability of portable devices, which acquire HRV measures predominately through ultra-short (approximately 1-min) recordings, future research should investigate the utility of ultra-short recordings for monitoring acute muscular performance recovery and readiness to perform following resistance training. These are the applications most accessible to practitioners in the field setting. Furthermore, it has been identified that HRV measures acquired in different postural profiles (e.g., supine and standing) are independent [51]. Therefore, further research is needed to determine if the relationship between the time course of HRV and muscular performance recovery differs between body positions (e.g., supine or standing).

Conclusion

This study investigated the association in the time course of recovery between multiple HRV metrics and acute muscular performance following an exhaustive bout of resistance training. The results of the profile analysis determined that none of the HRV metrics followed a similar trend with *performance* across the 72-hr period as all measures of HRV had a significant interaction with *performance* ($p < .05$) and HRV recovery preceded *performance* recovery.

Furthermore, the repeated measures correlation represented a weak to moderate intra-individual association between the HRV metrics and *performance*. The findings are different from previous literature which suggests that HF measures of HRV mirror acute changes in muscular performance over a 72-hr period [18]. However, the results support the current body of literature which have demonstrated a decrease in PNS activity ~30-min following resistance exercise [13, 24]. Our results suggest isolated measures may be insufficient when administering HRV as a gauge of readiness to perform following an exhaustive bout of resistance training as HRV and muscular performance did not follow the same time course of recovery of the 72-hr period.

REFERENCES

1. Prinsloo GE, Rauch HG, Derman WE. A brief review and clinical application of heart rate variability biofeedback in sports, exercise, and rehabilitation medicine. Phys Sportsmed. 2014;42(2):88-99.

2. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol. 2010;141(2):122-31.

3. Malmo RB, Shagass C. Heart rate variability during a stress situation in psychiatric patients with and without frontal lobe operation. Rev Can Biol. 1948;7(1):188.

4. Malik M, Camm AJ, Bigger Jr JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17(3):354-81.

5. Pichot V, Roche F, Gaspoz JM, Enjolras F, Antoniadis A, Minini P, et al. Relation between heart rate variability and training load in middle-distance runners. Med Sci Sports Exerc. 2000;32(10):1729-36.

6. Kiviniemi AM, Hautala AJ, Kinnunen H, Tulppo MP. Endurance training guided individually by daily heart rate variability measurements. Eur J Appl Physiol. 2007;101(6):743-51.

7. Flatt AA, Esco MR. Evaluating individual training adaptation with smartphone-derived heart rate variability in a collegiate female soccer team. J Strength Cond Res. 2016;30(2):378-85.

8. Kiss O, Sydo N, Vargha P, Vago H, Czimbalmos C, Edes E, et al. Detailed heart rate variability analysis in athletes. Clin Auton Res. 2016;26(4):245-52.

9. Nakamura FY, Flatt AA, Pereira LA, Ramirez-Campillo R, Loturco I, Esco MR. Ultra-short-term heart rate variability is sensitive to training effects in team sports players. J Sports Sci Med. 2015;14(3):602-5.

10. Plews DJ, Laursen PB, Buchheit M. Day-to-day heart-rate variability recordings in world-champion rowers: appreciating unique athlete characteristics. Int J Sports Physiol Perform. 2017;12(5):697-703.

11. Terziotti P, Schena F, Gulli G, Cevese A. Post-exercise recovery of autonomic cardiovascular control: a study by spectrum and cross-spectrum analysis in humans. Eur J Appl Physiol. 2001;84(3):187-94.

12. Javorka M, Zila I, Balharek T, Javorka K. Heart rate recovery after exercise: relations to heart rate variability and complexity. Braz J Med Biol Res. 2002;35(8):991-1000.

13. Heffernan KS, Kelly EE, Collier SR, Fernhall B. Cardiac autonomic modulation during recovery from acute endurance versus resistance exercise. European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. 2006;13(1):80-6.

14. Buchheit M, Laursen PB, Ahmaidi S. Parasympathetic reactivation after repeated sprint exercise. Am J Physiol Heart Circ Physiol. 2007;293(1):H133-41.

15. Kaikkonen P, Rusko H, Martinmaki K. Post-exercise heart rate variability of endurance athletes after different high-intensity exercise interventions. Scand J Med Sci Sports. 2008;18(4):511-9.

16. Cornelissen VA, Verheyden B, Aubert AE, Fagard RH. Effects of aerobic training intensity on resting, exercise and post-exercise blood pressure, heart rate and heart-rate variability. J Hum Hypertens. 2010;24(3):175-82.

17. Michael S, Graham KS, Oam GMD. Cardiac autonomic responses during exercise and post-exercise recovery using heart rate variability and systolic time intervals-a review. Front Physiol. 2017;8.

18. Chen JL, Yeh DP, Lee JP, Chen CY, Huang CY, Lee SD, et al. Parasympathetic nervous activity mirrors recovery status in weightlifting performance after training. J Strength Cond Res. 2011;25(6):1546-52.

19. Shi P, Hu S, Yu H. Recovery of heart rate variability after treadmill exercise analyzed by lagged Poincare plot and spectral characteristics. Med Biol Eng Comput. 2017.

20. Mourot L, Bouhaddi M, Perrey S, Cappelle S, Henriet MT, Wolf JP, et al. Decrease in heart rate variability with overtraining: assessment by the poincare plot analysis. Clin Physiol Funct Imaging. 2004;24(1):10-8.

21. Mourot L, Bouhaddi M, Perrey S, Rouillon JD, Regnard J. Quantitative poincare plot analysis of heart rate variability: effect of endurance training. Eur J Appl Physiol. 2004;91(1):79-87.

22. Gonzalez-Badillo JJ, Rodriguez-Rosell D, Sanchez-Medina L, Ribas J, Lopez-Lopez C, Mora-Custodio R, et al. Short-term recovery following resistance exercise leading or not to failure. Int J Sports Med. 2016;37(4):295-304.

23. Kingsley JD, Figueroa A. Acute and training effects of resistance exercise on heart rate variability. Clin Physiol Funct Imaging. 2016;36(3):179-87.

24. Teixeira L, Ritti-Dias RM, Tinucci T, Mion Junior D, Forjaz CL. Post-concurrent exercise hemodynamics and cardiac autonomic modulation. Eur J Appl Physiol. 2011;111(9):2069-78.

25. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. Sports Med. 2013;43(9):773-81.

26. Twist C, Highton J. Monitoring fatigue and recovery in rugby league players. Int J Sports Physiol Perform. 2013;8(5):467-74.

27. Bartolomei S, Church DD, Arroyo E, Iii JAG, Varanoske AN, Wang R, et al. Comparison of the recovery response from high-intensity and high-volume resistance exercise in trained men. Eur J Appl Physiol. 2017:1-12.

28. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol. 2014;5:73.

29. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39:175-91.

30. Esco MR, Flatt AA, Nakamura FY. Agreement between a smartphone pulse sensor application and electrocardiography for determining lnRMSSD. J Strength Cond Res. 2017;31(2):380-5.

31. Laborde S, Mosley E, Thayer JF. Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. Front Psychol. 2017;8:213.

32. Jackson AS, Pollock ML. Practical Assessment of Body Composition. Phys Sportsmed. 1985;13(5):76-90.

33. Hori N, Andrews W. Reliability of velocity, force and power obtained from the Gymaware optical encoder during countermovement jump with and without external loads. J Aust Strength Cond. 2009;17(1):12-7.

34. Banyard HG, Nosaka K, Sato K, Haff GG. Validity of Various Methods for Determining Velocity, Force and Power in the Back Squat. Int J Sports Physiol Perform. 2017:1-25.

35. Garnacho-Castano MV, Lopez-Lastra S, Mate-Munoz JL. Reliability and validity assessment of a linear position transducer. J Sports Sci Med. 2015;14(1):128-36.

36. Beckham G, Mizuguchi S, Carter C, Sato K, Ramsey M, Lamont H, et al. Relationships of isometric mid-thigh pull variables to weightlifting performance. J Sports Med Phys Fitness. 2013;53(5):573-81.

37. Beckham GK, Sato K, Mizuguchi S, Haff GG, Stone MH. Effect of body position on force production during the isometric mid-thigh pull. J Strength Cond Res. 2017.

38. Haff GG, Ruben RP, Lider J, Twine C, Cormie P. A comparison of methods for determining the rate of force development during isometric midthigh clean pulls. J Strength Cond Res. 2015;29(2):386-95.

39. Hoffman JR. Athlete testing and program evaluation. In: Hoffman JR, editor. NSCA's guide to program design. Champaign, IL: Human Kinetics; 2012. p. 23-49.

40. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-aho PO, Karjalainen PA. Kubios HRV - Heart rate variability analysis software. Comput Meth Prog Bio. 2014;113(1):210-20.

41. Kamen PW, Tonkin AM. Application of the poincare plot to heart rate variability: a new measure of functional status in heart failure. Aust N Z J Med. 1995;25(1):18-26.

42. Schumacker RE. Using R with multivariate statistics. California, USA: Sage Publications; 2015.

43. Pituch KA, Stevens JP. Applied multivariate statistics for the social sciences. Sixth, editor. New York, NY: Routledge; 2016.

44. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: part 1-correlation within subjects. BMJ. 1995;310(6977):446.

45. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol. 2017;8(456):1-13.

46. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.

47. Houle MS, Billman GE. Low-frequency component of the heart rate variability spectrum: a poor marker of sympathetic activity. Am J Physiol. 1999;276(1 Pt 2):H215-23.

48. Nuuttila OP, Nikander A, Polomoshnov D, Laukkanen JA, Hakkinen K. Effects of HRV-guided vs. predetermined block training on performance, HRV and serum hormones. Int J Sports Med. 2017;38(12):909-20.

49. Plews DJ, Laursen PB, Kilding AE, Buchheit M. Evaluating training adaptation with heart-rate measures: a methodological comparison. Int J Sports Physiol Perform. 2013;8(6):688-91.

50. Plews DJ, Laursen PB, Le Meur Y, Hausswirth C, Kilding AE, Buchheit M. Monitoring training with heart rate-variability: how much compliance is needed for valid assessment? Int J Sports Physiol. 2014;9(5):783-90.

51. Schmitt L, Regnard J, Millet GP. Monitoring fatigue status with HRV measures in elite athletes: an avenue beyond RMSSD? Front Physiol. 2015;6:343.

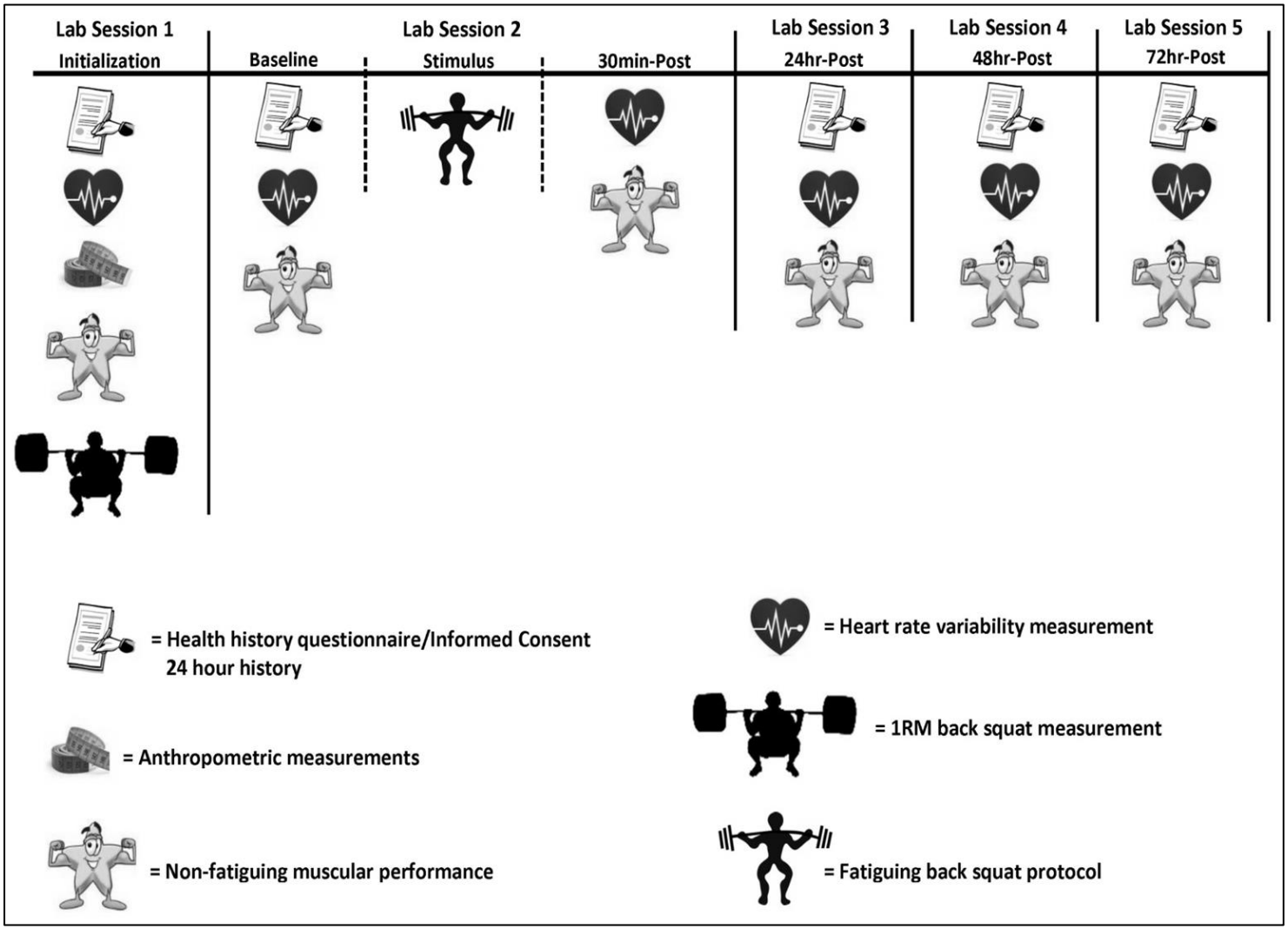

Figure 3.1. Repeated measures study design

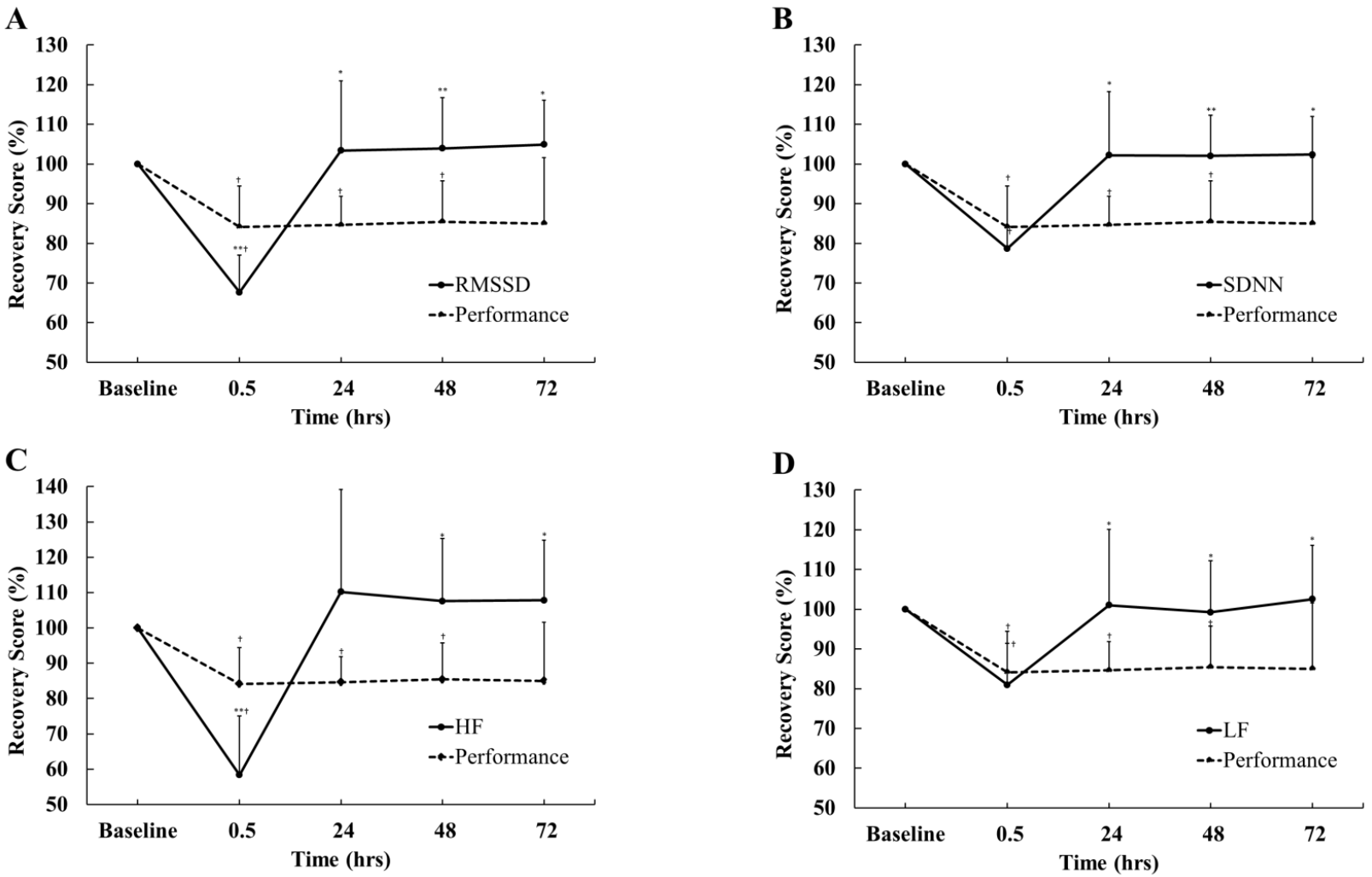

Figure 3.2.1 Recovery scores in *performance* and each log transformed HRV metric across the 72-hr time period. (A) root mean square of successive R-R differences; (B), standard division of all normal-to-normal R-R intervals; (C), High frequency power; (D), Low frequency power. Data is represented as mean (*SD*).

* Significantly different from *performance* at the .05 level

** Significantly different from *performance* at the .01 level

† Significantly different from baseline at the .05 level

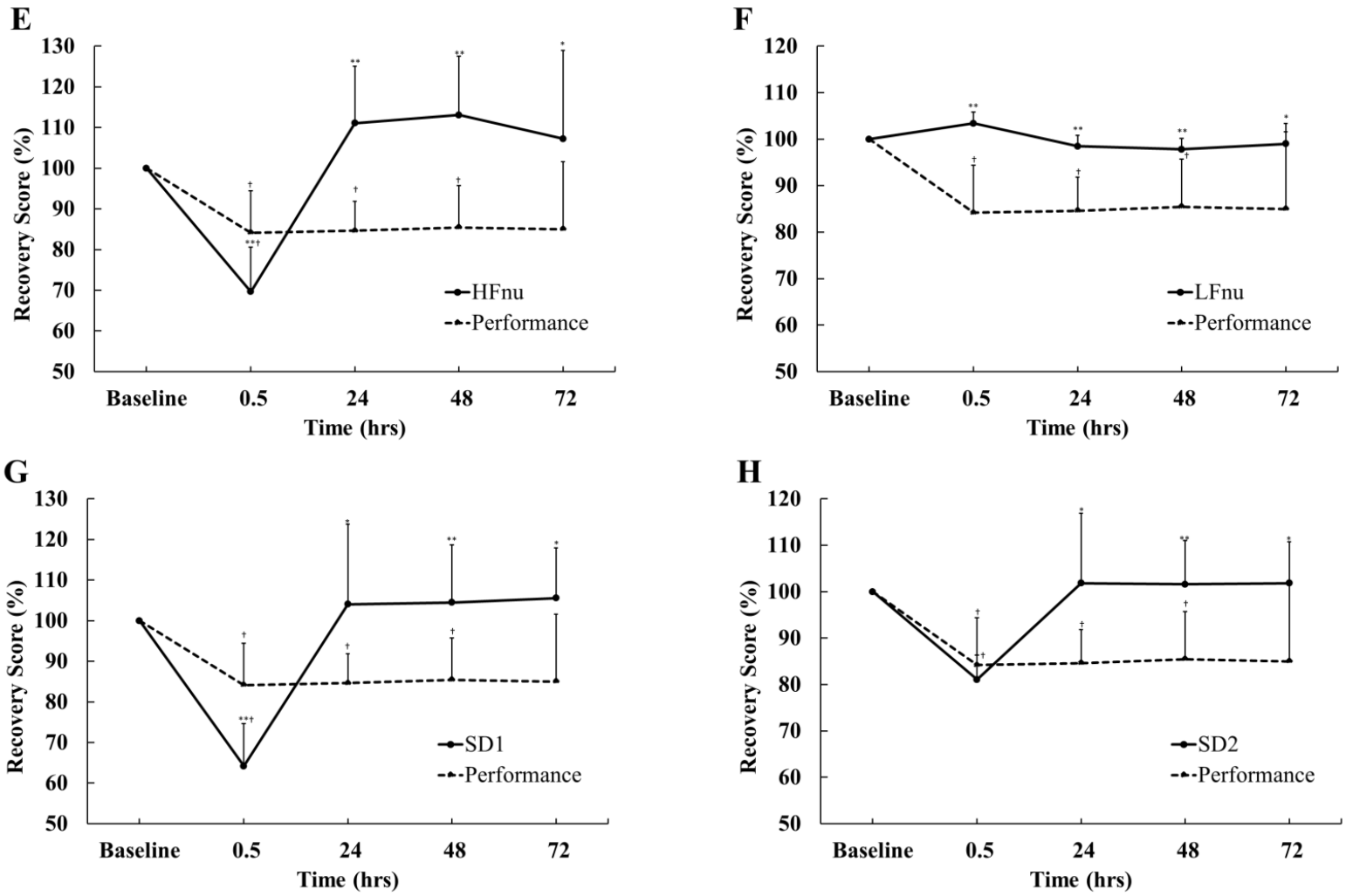

Figure 3.2.2. Recovery scores in *performance* and each log transformed HRV metric across the 72-hr time period. (E), High frequency power in normalized units; (F), Low frequency power in normalized units; (G), standard deviation of short-term HRV from the Poincaré plot; (H), standard deviation of long-term HRV from the Poincaré plot. Data is represented as mean (*SD*).
* Significantly different from *performance* at the .05 level
** Significantly different from *performance* at the .01 level
† Significantly different from baseline at the .05 level
* Significantly different from *performance* at the .05 level

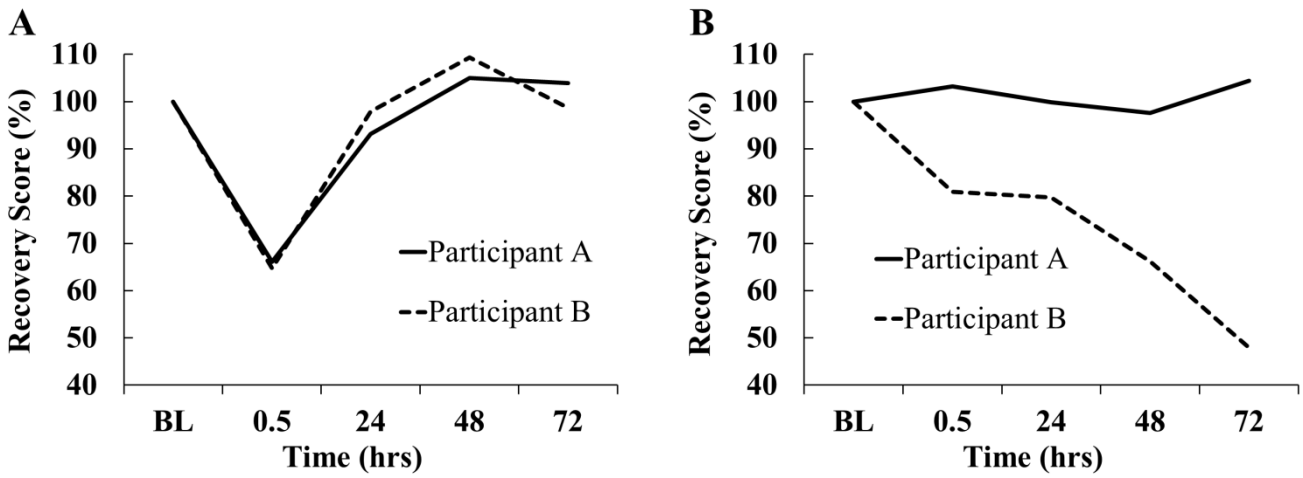

Figure 3.3. Example of the individual variability in recovery scores of the natural log of the root mean square of successive R-R differences (RMSSD) and *performance* across the 72-hr time period between two participants. (A), recovery score in RMSSD between the two individuals; (B), recovery score in *performance* between the two individuals.

Table 3.1. Results of the within subject interaction between the recovery scores of the log transformation of each HRV metric and *performance* over time along with the repeated measures correlation.

	Measure*Time	*r* (95% *CI*)
RMSSD	$F(4,28)=15.944$, $p<.001$	.319 (-.107, .646), $p=.120$
SDNN	$F(4,28)=7.805$, $p<.001$	.175 (-.255, .548), $p=.401$
HF	$F(4,28)=12.558$, $p<.001$	.206 (-.225, .569), $p=.323$
HFnu	$F(4,28)=12.844$, $p<.001$	.354 (-.168, .668), $p=.083$
LF	$F(4,28)=4.924$, $p=.004$	.047 (-.373, .450), $p=.825$
LFnu	$F(4,28)=5.679$, $p=.002$	-.406 (-.701, .008), $p=.044$
SD1	$F(4,28)=16.859$, $p<.001$	.317 (-.110, .645), $p=.123$
SD2	$F(4,28)=6.605$, $p=.001$	.133 (-.296, .517), $p=.527$

CI, confidence interval; HF, high frequency power; HFnu, high frequency power in normalized units; LF, low frequency power; LFnu, low frequency power in normalized units; Measure*Time, interaction between the specified HRV metric and *performance* over time; *p*, probability of two-tailed dependent t-test; *r*, repeated measures correlation; RMSSD, root mean square of successive R-R differences; SD1, standard deviation of short-term HRV from the Poincaré plot; SD2, standard deviation of long-term HRV from the Poincaré plot; SDNN, standard division of all normal-to-normal R-R intervals

CHAPTER 4

VALIDITY OF ULTRA-SHORT MEASURES OF HEART RATE VARIABILITY FOLLOWING RESISTANCE EXERCISE

ABSTRACT

The purpose of this study was to determine the level of agreement between the log transformed ultra-short and short-term recording of a multitude of heart rate variability (HRV) metrics acquired from a gold standard electrocardiogram (ECG) throughout six data collections surrounding an exhaustive resistance training protocol. Eight resistance trained males volunteered to participate in the study. Participants performed seated HRV measures 48-hr before and immediately prior to a strenuous resistance training protocol consisting of 8-sets of 10 back squats at an intensity of 70% 1RM. Collection of HRV measures was replicated on four additional occasions (0.5, 24, 48, and 72-hr post-exercise). Ultra-short HRV measures were acquired in a 1-min epoch following a 1-min stabilization period and short-term recordings were acquired in a 5-min epoch following a 5-min stabilization period. The level of agreement was quantified by two-tailed dependent t-test, two-way mixed intraclass correlation coefficient (*ICC*) for absolute agreement, Cohen's *d* effect size (*ES*), and the Bland Altman method for limits of agreement (LOA). The highest levels of agreement were displayed in the log-transformed (ln) root mean square of successive R-R differences (lnRMSSD) [LOA = -0.93– 0.72, *ICC* = .91, *ES* = -0.14, *p* = .082] and standard deviation of the points through the width of the plot (lnSD1) [LOA = -0.91 – 0.75, *ICC* = .91, *ES* = 0.13, *p* = .156] compared to all other metrics. Our results

support the use of ultra-short lnRMSSD and lnSD1 as they provide accurate measures compared to the criterion short-term recordings.

Key Words: Heart Rate Variability, Ultra-Short, Electrocardiogram

INTRODUCTION

The use of unobtrusive measures of heart rate variability (HRV) for monitoring autonomic modulation has grown in popularity and demand. HRV is the variation between consecutive heart beats and is considered a non-invasive marker of cardio-autonomic control [1]. Perturbations in cardio-autonomic activity are a reflection of ongoing disturbances to homeostasis which are regulated through the autonomic nervous system (ANS). Thus, monitoring of cardio-autonomic modulations, through HRV, has become a useful measure for evaluation of health and disease [2, 3]. By tradition, HRV is acquired through the collection of heart rate data from an electrocardiogram (ECG) over a 10-min epoch (short-term HRV), which is comprised of a 5-min stabilization period followed by a 5-min recording as recommended by the HRV Task Force [1].

Currently, there is a strong body of research supporting the use of more convenient method for acquiring HRV [4-13]. For example, ultra-short recordings (recordings less than one minute) have shown acceptable agreement against traditional 10-min ECG recordings [1]. This shortened method of data collection may be beneficial to practitioners as recordings can be acquired after a 1-min stabilization period allowing for increased practicality [11]. For example, ultra-short recordings of the log-transformed root mean square of successive R-R differences (lnRMSSD) by smartphone pulse finger sensor plethysmography have been validated against traditional ECG measures [6, 7, 10-12]. Similarly, researchers have found ultra-short recordings

of the standard deviation of all normal-to-normal (NN) intervals (SDNN) and standard deviation of the points through the width of the Poincaré plot (SD1), acquired through plotting each R-R interval against the previous R-R intervals, were a useful means of tracking autonomic modulation due to exercise [14-16]. However, the agreement of other HRV metrics, linear and non-linear, acquired through ultra-short analysis remains controversial. More so, the extent to which these metrics accurately reflect shifts in autonomic modulation over a 72-hr period post strenuous exercise has yet to be identified.

The utilization of accurate monitoring of autonomic modulation across a given time span is invaluable to practitioners in the fields of athletic monitoring and prescription. Consequently, there is a need to further validate the use of ultra-short recordings across the broad spectrum of HRV metrics at rest and following exhaustive physical exertion. Therefore, the purpose of this investigation was to determine the agreement between ultra-short and traditional short measures of HRV on linear and non-linear metrics over a 48-hr period prior to and 72-hr period post-strenuous exercise. Because of previous research indicating ultra-short term and non-linear methods of parasympathetic markers produce good agreement with short-term criterion [14, 17], it was hypothesized that the linear metric RMSSD and the non-linear metric SD1 acquired through ultra-short recordings would provide acceptable agreement with criterion short-term ECG measures.

METHODS

Experimental Design

A repeated measures design was utilized in which each participant served as his own control. HRV measures were assessed on each participant over a 72-hr period. Participants

performed HRV measures 48-hr before and immediately prior to a strenuous resistance training protocol consisting of 8-sets of 10 back squats at an intensity of 70% 1RM [18]. Collection of HRV measures was replicated on four additional occasions (0.5, 24, 48, and 72 hours) post resistance training protocol (six total collections). All measurements were performed between 5:00 and 10:00 am as current recommendations suggest HRV measures should be acquired early in the morning, as close to awakening as possible [19].

Subjects

Eight healthy men [mean (*SD*)] [age = 23.3 (3.9) y, mass = 90.4 (10.9) kg; height = 180.8 (8.3) cm, body mass index = 27.7 (3.3) kg/m^2, percent body fat estimated from skinfold = 13.0 (4.3) %] who were non-smokers; free from cardiovascular, pulmonary, and metabolic disease; and who had participated in a moderate-to-vigorous training program for over 6 years participated in the study. An *a priori* power analysis with G power (version 3.1) estimated that 34 HRV measures would provide adequate statistical power to detect a difference between traditional short and ultra-short recordings of HRV, based on a moderate effect size (0.5) assuming power = 0.80 and alpha = 0.05 [20]. Due to the repeated measures design of 6 data collections time points a sample size of 6 participants was sufficient. All participants signed an informed consent approved by the Institutional Review Board of The University of Alabama prior to participating.

Procedures

Participants were instructed to refrain from ingesting alcohol, caffeine, non-prescription drugs and strenuous activity 12-hrs prior to the initial visit and throughout the data collection

process. Participants were also instructed to attend each laboratory session in a fasted state. Adherence throughout the study was assessed by a 24-hr history questionnaire completed prior to every trial upon entrance into the laboratory. Experimental trials were collected at the same time each day (± 0.25 hours).

HRV Collection

On the first day (familiarization day), subjects arrived to the laboratory upon wakening and were instructed to complete paperwork (i.e., informed consent, 24-hr history questionnaire, medical history questionnaire). Following, participants were fitted with a modified lead-II ECG configuration using three Ag/AgCl surface electrodes [12]. Prior to electrode placement, each site was cleaned with an alcohol swab. The right mid-clavicular notch served as the negative electrode, the fifth intercostal on the left mid-axillary line represented the positive electrode, and the left anterior-superior iliac crest served as the ground electrode. Electrodes were connected to a data acquisition system (Biopac MP100 Goletta, CA) which was interfaced with a laptop computer for data acquisition. Participants were then instructed to sit in an enclosed dim lit controlled environmental chamber (24 °C, 20% relative humidity) with their legs in at a 90° angle and with their hands resting on their thighs for all HRV measures [21].

All HRV measures were collected over a 10-min period. Ultra-short recordings of HRV were acquired within the first 2-min of the ECG recording. The first minute served as stabilization period while minute two represented the ultra-short recording [11]. Each minute was marked on the data to reflect ultra-short recordings. For short HRV recordings, the first 5-min of the ECG recording served as an acclimation period and the last 5-min (5-10) served as the HRV collection as recommended by the Task Force [1]. Continuous sampling at a rate of 2000

Hz was utilized to ensure appropriate sampling of the fiducial point of each QRS complex throughout the 10-min period. Both epochs (ultra-short and short) were marked on the data software (Acknowledge v 3.9, BIOPAC, Goletta, CA, USA) and save for analysis at the completion of each trial. No attempt was made to control for breathing rate as it has been shown to potentially bias the interpretation [21].

Anthropometric and 1RM Assessment

On the familiarization day only, anthropometric and body composition data were assessed following the HRV measurements. Participant height was measured to the nearest 0.1cm using stadiometer (SECA 67310, SECA©, Chino, CA,) body weight was measured to the nearest 0.1kg using a digital scale (Tanita BWB-800, Tanita©, Arlington Heights, IL) , and 3 site skinfold thicknesses (chest, abdomen, and thigh sites) were attained with a caliper (Lange Skinfold Caliper, Beta Technology©, Santa Cruz, CA) by a trained research assistant [22]. After completing the anthropometric measures, the participants' 1RM back squat was established to acquire relative loads for the resistance training stimulus. The 1RM values were determined through an incremental increase of intensity within several trials, allowing 3 to 5 minutes rest between each attempt as described [23]. Task progression involved three warm up sets, 5-10 repetitions at ~50% and three repetitions at ~85% followed by one repetition at 90-95% 1RM. The remaining attempts were allocated toward attaining a 1RM value.

Resistance Training Protocol

Forty-eight hours following the familiarization trial, participants returned to the lab and repeated all measures previously described with the exception of anthropometrics and the 1RM

back squat (pre-fatiguing trial). Immediately following the pre-fatiguing trial, participants performed the fatiguing back squat protocol previously described. In the event that all 10 repetitions could not be accomplished within a set, a 30-s recovery was allocated before the remaining repetitions were completed. If three failed repetitions occurred within 1 set, the 2-min inter-set rest period commenced and the load was reduce by 5% of the participant's 1RM for the remainder of the squat protocol. If residual repetitions were present after the completion of set eight, the participant was allotted a 30-s rest before completing the remaining repetitions. This ensured that the total volume (80 repetitions) was consistent across all participants.

Following the fatiguing protocol, participants rested passively for 30-min before completing HRV measures as previously specified. These values were utilized to represent the 0.5-hr trial. Participants returned to the laboratory on three additional occasions over a 72-hr period to perform measures which represent the 24, 48, and 72-hr trials.

HRV Analysis/cleaning

Once HRV data collection was completed, missing and ectopic beats were identified through visual inspection of the tachogram and replaced with interpolated values as previously recommended [1]. Following artifact identification, all ECG recordings were transferred to software (Kubios, Biosignal Analysis and Medical Imaging Group, Department of Physics, University of Kuopio, Kuopio, Finland) for HRV analysis [24]. Time domain measures were quantified as the standard deviation of all NN intervals (SDNN); and root mean square of successive RR interval differences (RMSSD). A Fast Fourier Transform (FFT) algorithm was utilized to obtain power spectrum estimates of HRV (i.e., ms to hertz [Hz]). Spectral analysis was performed through detection of very low-frequency (VLF) < 0.04 Hz, low-frequency (LF)

0.04 – 0.15 Hz and high-frequency (HF) 0.15 -0.4 Hz bands and total power (TP) reported as raw (ms^2) or normalized units (nu) [1].

Poincaré plots were acquired by plotting each RR interval against previous RR intervals providing a two dimensional oval appearance rotated at a 45° angle [25]. SD1 represented short-term HRV and the standard deviation of the points along the length of the plot (SD2) represented long-term HRV [26]. Inherently, HRV measures may be non-parametric. Therefore, all HRV measures were log transformed (ln) for better interpretation of statistical analysis [1].

Statistical Analysis

The relationship between the traditional short and ultra-short HRV recordings was quantified by a two-tailed dependent t-test, and a two-way mixed intraclass correlation coefficient (*ICC*) for absolute agreement for all HRV metrics. A stringent cutoff of .90 was utilized for the *ICC* values to ensure acceptable agreement (approximately 81% variability accounted for) as previously utilized by Shaffer et al. [27]. Effect sizes (*ES*) were also calculated by Cohen's *d* statistic and Hopkin's scale of magnitude were utilized to interpret the *ES* as trivial (0-0.2), small (0.2-0.6), moderate (0.6-1.2), large (1.2-2.0), and very large (>2.0) [28]. Additionally, the Bland-Altman method was used to establish the limits of agreement (LOA) between the ultra-short and traditional short ECG recordings for each HRV metric [29]. The differences between the two methods (difference = ultra-short – traditional) were plotted against the average the two methods. The mean difference was calculated to identify the constant error (*CE*) and 95% confidence intervals (*CI*) of the difference were derived to represent the upper and lower LOA ($CE \pm 1.96*SD_{difference}$) [29, 30]. The quality of agreement was assessed by calculating the ratio of half the LOA divided by the mean of the paired measurements [31, 32].

The ratio was considered as good (<0.1), moderate (0.1-0.2), or insufficient (>0.2). Furthermore, the trend in error associated with ultra-short measures was established through assessing the Pearson product-moment correlation (*r*) between the observed error in each ultra-short recording against the mean value of each ultra-short and short-term recording. A Shapiro-Wilk test was performed to address the assumption of normality among all HRV metrics [33]. Statistical analyses were performed using a spreadsheet (Microsoft Excel 2010, Microsoft Corporation, Redmond, WA, USA) and software (SPSS version 23.0 Somers, NY, USA). A statistically significant difference was established with an alpha level of 0.05. Data are represented as mean (*SD*) unless otherwise noted.

RESULTS

All statistical data are represented in Table 4.1. The mean ultra-short and short HRV values significantly differed across lnLF, lnLFnu, and HFnu, but did not differ between parasympathetic indices lnRMSSD, lnHF, lnSD1 and estimates of total variability lnSDNN and lnSD2. Similarly, the *ES* of lnRMSSD, lnHF, and lnSD1 metrics were trivial, while all others were considered small. All *ICCs* between the ultra-short and short measures were significant ($p < 0.01$). However lnRMSSD, lnSD1, lnHF, and lnSDNN provided the strongest correlations, although lnRMSSD and lnSD1 were the only metrics to meet the .90 cutoff (*ICC* = .91). The weakest association amongst short and ultra-short log transformed values were displayed in lnLF and lnLFnu measures with an *ICC* equal to .61 and .52 respectively. Furthermore, the Bland-Altman analysis revealed consistent under estimation of HRV by ultra-short measures compared to short measures in all metrics except lnHFnu (Figures 4.1.1 and 4.1.2). The ratio of half the LOA and the paired means of the ultra-short and short measures demonstrated good agreement

for lnLFnu (0.09) and moderate agreement (0.11-0.20) for all other metrics. Furthermore, the trend in error was trivial in lnRMSSD, lnSDNN, lnHF, lnSD1, and lnSD2. However, a significant increase in error at lower mean values [(ultra-short + short)/2] was displayed in lnLF (r = .29, p = .049), lnHFnu (r = .31, p = .032) and lnLFnu (r = .73, p < .001). Visual depictions of the Bland-Altman analyses are displayed in Figure 4.1.1 and Figure 4.1.2. Normality assumptions were not violated within any of the log transformed metrics (p > .05).

DISCUSSION

This investigation assessed the level of agreement between the ultra-short and short term recording of lnRMSSD, lnSDNN, lnHF, lnHFnu, lnLF, lnLFnu, lnSD1 and lnSD2 acquired from a gold standard ECG method throughout six collections (two prior and four after an exhaustive resistance training protocol). When addressing these analyses as a whole, our original hypothesis was confirmed as parasympathetic indices lnRMSSD and lnSD1 provided tight limits of agreement, strong *ICCs,* trivial *ES* and non-significant mean difference when a physically exhaustive bout of resistance training is involved. However, lnHF, lnSDNN and lnSD2 are of note as they followed a similar relationship as lnRMSSD and lnSD1. The ratio within the Bland Altman analysis for lnSDNN and SD1were approximately equivalent to lnRMSSD and lnSD1, but lnHF was slightly larger. The *ICC* values, specifically for lnHF (.89), were also strong but not close to the cutoff of .90 suggesting a decrease in precision. However, when addressing the *ICCs* of lnSDNN, lnHFnu, and lnSD2, the associations between short and ultra-short recordings were >.75 which is a threshold that has been previously identified as acceptable agreement within HRV literature [34].

These findings support current literature which has found acceptable agreement between ultra-short and short recordings of RMSSD yet recommend longer recordings for frequency derived HRV metrics such as LF [1, 11, 17, 27, 35]. For example, Shaffer et al. [27] collected 5-min epochs from 38 college aged individuals and compared multiple time segments (up to 240 s) to the 5-min ECG recording utilizing the Pearson product-moment correlation with a cutoff of .90. The authors determined a 60-s epoch was sufficient to acquire accurate measures of RMSSD and SDNN as both displayed a strong association of .95. However, a 90-s epoch was required for LF, and 180-s were needed for LFnu, HF, and HFnu before a correlation of .90 or higher was achieved.

In contrast to the findings of Shaffer et al. [27], lnSDNN did not reach the cutoff of .90 for the *ICC* (.82) within our analysis. Thus, we cannot conclude acceptable agreement. Similar results were noted by Nussinovitch et al. [35] who found a very strong association between 1-min and 5-min ECG recordings in RMSSD (*ICC* = .96) but not SDNN (*ICC* = .86). Ultra-short measures of lnHF within our analysis showed a much higher association with 5-min recordings compared to Shaffer et al. [27] (r = .75). One explanation for this discrepancy is that our values were log transformed as measures of HRV tend to possess normality issues [1]. Thus, the increased accuracy of HF measures within our analysis may have been attributed to the reduction in bias due to the log transformation.

Fewer studies have evaluated the agreement between ultra-short and short recordings of non-linear indices SD1 and SD2. Gomes et al. [14] evaluated the relationship between ultra-short and short recordings of SD1 at rest with 35 healthy males and determined the values were significantly correlated (r = .78) and not significantly different (p = .20). Shaffer et al. [27] also examined SD1 and SD2 within their respective study and determined acceptable agreement

within 90-s epochs (SD1, r = .95; SD2, r = .97). However, a 60-s epoch for SD1 and SD2 was not provided within the authors' results. Our results elaborate on the current literature to further suggest ultra-short recordings of lnSD1 to be an acceptable surrogate for short-term recordings. As SD1 has been shown to be highly correlated with its counterpart RMSSD [36], it is not surprising that it also provided acceptable agreement. Similarly, SD2 has been shown to be highly correlated with LF, therefore the lower level of agreement may be attributed to the need for longer time segments than the 60-s epochs utilized in this study.

However, ultra-short recordings comprised of 60-s have increased in utility by the practitioner, particularly for extrapolations of lnRMSSD that may have been derived from a portable device. For example, Nakamura et al. [7] determined that ultra-short lnRMSSD recordings of 1-min preceded by a 1-min stabilization period, acquired from a Polar RS800cx, were able to detect changes due to training in 24 elite futsal players after 3 to 4-weeks of preseason training, to a similar degree as the criterion 5-min recording. Similarly, Flatt and Esco [9] revealed that 55-s recordings of lnRMSSD from the ithete application on an iPad2, via a Polar T-31 chest strap, provided tight LOA (-2.57 to 2.63) compared to 55-s from an ECG. This was a method which they later employed to monitor changes in resting cardiac autonomic indices through the first 3-weeks of a 5-week training camp in 12 female collegiate soccer players. The authors reported a significant correlation between the change in the coefficient of variation of lnRMSSD between week one and three, and the change in Yo-Yo Intermittent Recovery Test Level 1 performance between weeks one and five [10]. Thus, the increased practicality of ultra-short HRV recordings has provided useful applications for monitoring cardiac autonomic perturbations in the field setting. As our results suggest, ultra-short measures of lnRMSSD and lnSD1 have the strongest agreement with their respective criterion short-term values and provide

accurate measures of parasympathetic activity prior to and following an exhaustive bout of resistance training. These findings further promote the potential use of ultra-short recordings of these HRV metrics in a setting which may provide increased practicality to the practitioner.

This investigation is not without limitations. For example, only healthy resistance trained men were included in this analysis. Therefore, these results cannot be generalized into female, older adult, or clinical populations. Likewise, HRV measures were conducted in the seated position only and consequently do not represent plausible findings in supine or standing body positions. Furthermore, HRV collections were acquired through criterion ECG recordings in a well-controlled laboratory setting which do not represent ecological validity of HRV collections which occur in the field setting. However, our results support the utilization of ultra-short recordings which are intended to promote HRV monitoring in the field setting as shorter collection epochs may provide improved compliance by athletes. As the results of this study displayed acceptable agreement with parasympathetic indices across a period of several days which involved an exhaustive bout of resistance training, future research should investigate whether ultra-short HRV recordings can track changes in skeletal muscular performance.

CONCLUSION

In conclusion, ultra-short recordings of lnRMSSD and lnSD1 in the seated position provided the highest level of agreement to criterion short-term recordings compared to all other log transformed HRV metrics. Ultra-short extrapolations of lnHF power, lnSDNN, and lnSD2 also warrant note as they did not significantly differ from criterion measures and the respective *ICCs* and LOA (particularly lnSDNN and SD2) were at a level which some researchers may consider acceptable. Furthermore, our findings support previous literatures which suggest longer

recording periods are required for accurate depictions of HRV metrics which represent a combination parasympathetic and sympathetic influence (e.g., LF). When monitoring cardio-autonomic activity before and after a bout of exhaustive resistance training, accurate measures can be acquired from ultra-short HRV recordings of lnRMSSD and lnSD1.

PRACTICAL APPLICATIONS

HRV monitoring is increasing in popularity as a non-invasive tool for monitoring cardio autonomic responses to training and is best utilized when measures are acquired on a consistent basis. The high level of accuracy displayed from ultra-short recordings of RMSSD and the non-linear counterpart SD1 provide evidence to support their use as a more efficient means of collection compared to the criterion 5-min recordings as their measures were not affected by the resistance training session. Furthermore, the lnRMSSD is becoming one of the most utilized HRV metrics for applications in the field setting. Therefore, practitioners should exploit the opportunity to utilize such applications which have been validated against the criterion of an ECG.

REFERENCES

1. Malik M, Camm AJ, Bigger Jr JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17(3):354-81.

2. Prinsloo GE, Rauch HG, Derman WE. A brief review and clinical application of heart rate variability biofeedback in sports, exercise, and rehabilitation medicine. Phys Sportsmed. 2014;42(2):88-99.

3. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol. 2010;141(2):122-31.

4. Barbosa MP, da Silva NT, de Azevedo FM, Pastre CM, Vanderlei LC. Comparison of Polar(R) RS800G3 heart rate monitor with Polar(R) S810i and electrocardiogram to obtain the series of RR intervals and analysis of heart rate variability at rest. Clin Physiol Funct Imaging. 2016;36(2):112-7.

5. Bolkhovsky JB, Scully CG, Chon KH, editors. Statistical analysis of heart rate and heart rate variability monitoring through the use of smart phone cameras. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE; 2012: IEEE.

6. Nakamura FY, Pereira LA, Esco MR, Flatt AA, Moraes JE, Abad CCC, et al. Intraday and interday reliability of ultra-short-term heart rate variability in rugby union players. J Strength Cond Res. 2017;31(2):548-51.

7. Nakamura FY, Flatt AA, Pereira LA, Ramirez-Campillo R, Loturco I, Esco MR. Ultra-short-term heart rate variability is sensitive to training effects in team sports players. J Sports Sci Med. 2015;14(3):602-5.

8. Flatt AA, Esco MR, Nakamura FY, Plews DJ. Interpreting daily heart rate variability changes in collegiate female soccer players. J Sports Med Phys Fitness. 2017;57(6):907-15.

9. Flatt AA, Esco MR. Validity of the ithlete smart phone application for determining ultra-short-term heart rate variability. J Hum Kinet. 2013;39:85-92.

10. Flatt AA, Esco MR. Evaluating individual training adaptation with smartphone-derived heart rate variability in a collegiate female soccer team. J Strength Cond Res. 2016;30(2):378-85.

11. Flatt AA, Esco MR. Heart rate variability stabilization in athletes: towards more convenient data acquisition. Clin Physiol Funct Imaging. 2016;36(5):331-6.

12. Esco MR, Flatt AA, Nakamura FY. Agreement between a smartphone pulse sensor application and electrocardiography for determining lnRMSSD. J Strength Cond Res. 2017;31(2):380-5.

13. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart rate variability recording with smart phone photoplethysmographic, Polar H7 chest strap and electrocardiogram methods. Int J Sports Physiol Perform. 2017:1-17.

14. Gomes RL, Marques Vanderlei LC, Garner DM, Ramos Santana MD, de Abreu LC, Valenti VE. Poincare plot analysis of ultra-short-term heart rate variability during recovery from exercise in physically active men. J Sports Med Phys Fitness. 2017.

15. Aslani A, Aslani A, Kheirkhah J, Sobhani V. Cardio-pulmonary fitness test by ultra-short heart rate variability. J Cardiovasc Dis Res. 2011;2(4):233-6.

16. Chen X, Yao R, Yin G, Li J. Consecutive ultra-short-term heart rate variability to track dynamic changes in autonomic nervous system during and after exercise. Physiol Meas. 2017;38(7):1384-95.

17. Munoz ML, van Roon A, Riese H, Thio C, Oostenbroek E, Westrik I, et al. Validity of (ultra-)short recordings for heart rate variability measurements. PLoS One. 2015;10(9):e0138921.

18. Bartolomei S, Church DD, Arroyo E, Iii JAG, Varanoske AN, Wang R, et al. Comparison of the recovery response from high-intensity and high-volume resistance exercise in trained men. Eur J Appl Physiol. 2017:1-12.

19. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol. 2014;5:73.

20. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39:175-91.

21. Laborde S, Mosley E, Thayer JF. Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. Front Psychol. 2017;8:213.

22. Jackson AS, Pollock ML. Practical Assessment of Body Composition. Phys Sportsmed. 1985;13(5):76-90.

23. Hoffman JR. Athlete testing and program evaluation. In: Hoffman JR, editor. NSCA's guide to program design. Champaign, IL: Human Kinetics; 2012. p. 23-49.

24. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-aho PO, Karjalainen PA. Kubios HRV - Heart rate variability analysis software. Comput Meth Prog Bio. 2014;113(1):210-20.

25. Shi P, Hu S, Yu H. Recovery of heart rate variability after treadmill exercise analyzed by lagged Poincare plot and spectral characteristics. Med Biol Eng Comput. 2017.

26. Kamen PW, Tonkin AM. Application of the poincare plot to heart rate variability: a new measure of functional status in heart failure. Aust N Z J Med. 1995;25(1):18-26.

27. Shaffer F, Shearman S, Meehan ZM. The promise of ultra-short-term (UST) heart rate variability measurements. Biofeedback. 2016;44(4):229-33.

28. Hopkins WG. A scale of magnitude for effect statistics 2009 [Available from: http://www.sportsci.org/resource/stats/index.html.

29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

30. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. Journal of biopharmaceutical statistics. 2007;17(4):571-82.

31. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability?: a review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol. 2013;166(1):15-29.

32. Radespiel-Tröger M, Rauh R, Mahlke C, Gottschalk T, Muck-Weymann M. Agreement of two different methods for measurement of heart rate variability. Clin Auton Res. 2003;13(2):99-102.

33. Pituch KA, Stevens JP. Applied multivariate statistics for the social sciences. Sixth, editor. New York, NY: Routledge; 2016.

34. Montaño A, Brown F, Credeur DP, Williams MA, Stoner L. Telemetry-derived heart rate variability responses to a physical stressor. Clin Physiol Funct Imaging. 2017;37(4):421-7.

35. Nussinovitch U, Elishkevitz KP, Katz K, Nussinovitch M, Segev S, Volovitz B, et al. Reliability of ultra-whort ECG indices for heart rate variability. Ann Noninvasive Electrocardiol. 2011;16(2):117-22.

36. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health. 2017;5:258.

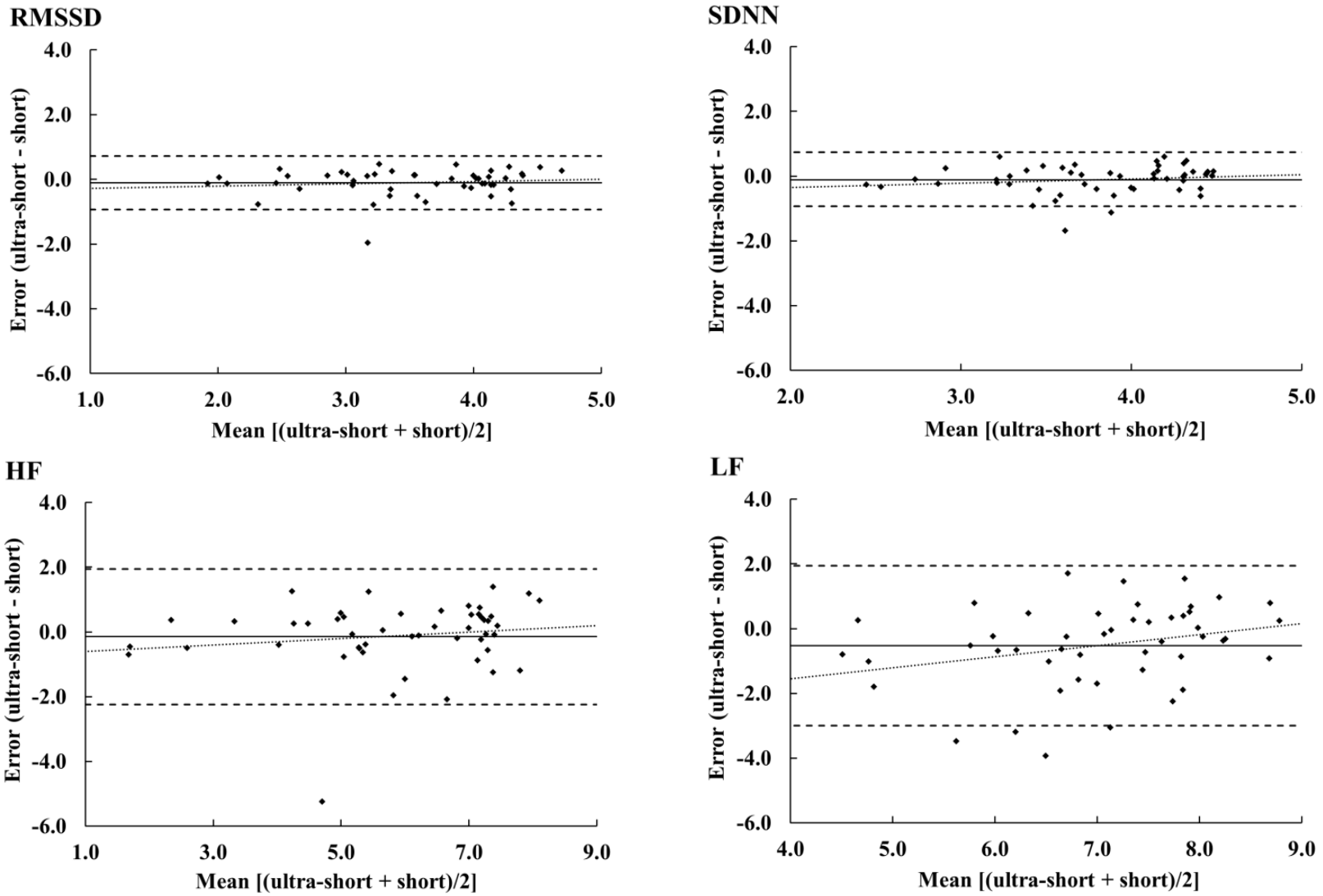

Figure 4.1.1. Bland Altman plots comparing the log transformation of the ultra-short and criterion short-term recordings from an electrocardiogram. The solid line represents the constant error, the outside dashed lines represent the 95% limits of agreement, and the dashed dotted regression line represents the trend in the error between the differences and the mean values. HF, high frequency power; LF, low frequency power RMSSD, root mean square of successive R-R differences; SDNN, standard deviation of all normal-to-normal R-R intervals.

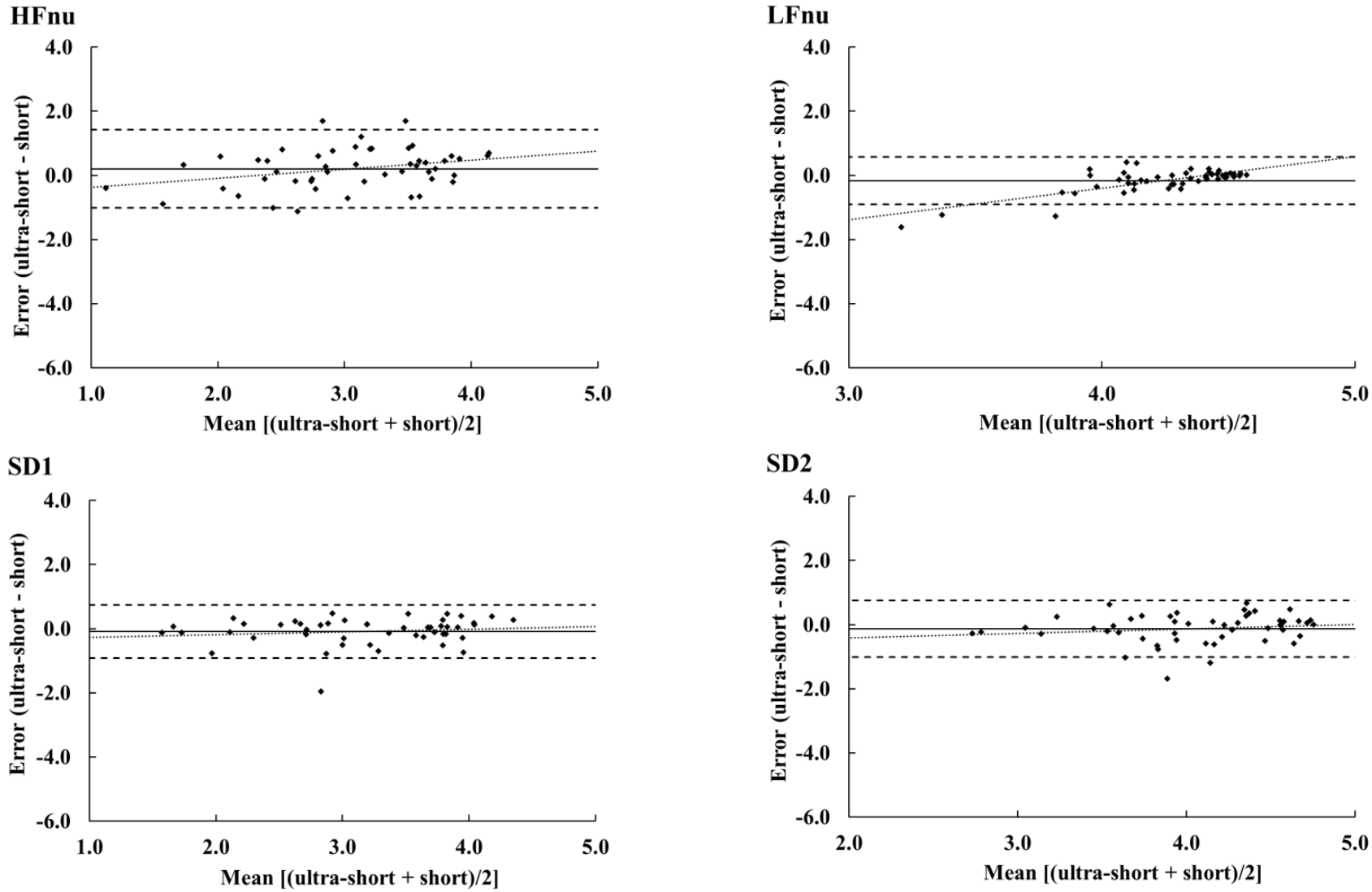

Figure 4.1.2. Bland Altman plots comparing the log transformation of the ultra-short and criterion short-term recordings from an electrocardiogram. The solid line represents the constant error, the outside dashed lines represent the 95% limits of agreement, and the dashed dotted regression line represents the trend in the error between the differences and the mean values. HFnu, high frequency power in normalized units; LFnu, low frequency power in normalized untis; SD1, standard deviation of short-term HRV from the Poincaré plot; SD2, standard deviation of long-term HRV from the Poincaré plot.

Table 4.1. Comparison between log transformed HRV values derived from ultra-short and criterion short recordings from an ECG in the seated position (n = 48).

					Limits of agreement				
Value	Mean (*SD*)	*p*	*ES*	*ICC (95% CI)*	*CE* (1.96 SD_d)	Ratio	Lower	Upper	*r*
RMSSD									
Short	3.57 (0.73)								
Ultra-short	3.47 (0.78)	0.098	-0.14	0.91 (0.84-0.95)	-0.10 (0.83)	0.12	-0.93	0.72	0.12
SDNN									
Short	3.84 (0.55)								
Ultra-short	3.72 (0.61)	0.066	-0.21	0.82 (0.68-0.90)	-0.12 (0.86)	0.11	-0.99	0.74	0.15
LF									
Short	7.28 (1.07)								
Ultra-short	6.76 (1.40)	0.007	-0.42	0.61 (0.31-0.78)	-0.52 (2.47)	0.18	-2.99	1.95	0.29*
LFnu									
Short	4.32 (0.19)								
Ultra-short	4.15 (0.45)	0.004	-0.53	0.52 (0.17-0.73)	-0.17 (0.75)	0.09	-0.91	0.58	0.73**
HF									
Short	5.95 (1.63)								
Ultra-short	5.83 (1.78)	0.447	-0.07	0.89 (0.81-0.94)	-0.15 (2.09)	0.18	-2.24	1.94	0.15
HFnu									
Short	2.94 (0.66)								
Ultra-short	3.15 (0.84)	0.025	0.27	0.79 (0.62-0.88)	0.21 (1.22)	0.20	-1.01	1.43	0.31*
SD1									
Short	3.23 (0.73)								
Ultra-short	3.15 (0.79)	0.180	-0.11	0.91 (0.84-0.95)	-0.08 (0.83)	0.13	-0.91	0.75	0.14
SD2									
Short	4.12 (0.53)								
Ultra-short	3.99 (0.60)	0.054	-0.23	0.78 (0.61-0.88)	-0.13 (0.89)	0.11	-1.01	0.76	0.16

* significant trend in error <.05; ** significant trend in error <.01; *CE*, constant error; *ES*, effect size; *ICC*, intra-class correlation; LF, low frequency power; LFnu, low frequency power in normalized units; *p*, probability of two-tailed dependent t-test; *r*, trend in error between the observed difference and the average of each measure as displayed by the Pearson correlation coefficient; SD_d, standard deviation of the difference; RMSSD, root mean square of successive R-R differences; SDNN, standard division of all normal-to-normal R-R intervals; HF, high frequency power; HFnu, high frequency power in normalized units; SD1, standard deviation of short-term HRV from the Poincaré plot; SD2, standard deviation of long-term HRV from the Poincaré plot.

CHAPTER 5

CONCLUSION

Technological developments have provided less obtrusive avenues for acquiring measures of HRV which have provided increased practicality for athletic monitoring. However, the degree of accuracy in HRV measures derived from a multitude of portable devices has yet to be quantified. Additionally, the use of HRV for athletic monitoring has typically been performed longitudinally within athletes who are primarily involved in bouts of aerobic training. Less is known about the potential use of HRV monitoring in the confines of resistance training and the relationship of cardio-autonomic perturbations to muscular performance recovery. Furthermore, the accuracy of ultra-short measures of HRV surrounding a strenuous bout of resistance training has yet to be established. Therefore, this series of studies provided useful insight into the accuracy of HRV metrics acquired from portable devices and the relationship between the time course in HRV and acute muscular performance recovery was assessed. Furthermore, the use of ultra-short measures surrounding a strenuous bout of resistance training was also addressed.

The first study involved a systematic review and meta-analysis on the accuracy of portable devices for acquiring HRV. Twenty-three studies yielded 301 effects that were included in the analysis. Results indicated that HRV measures acquired from portable devices significantly differed from ECG (*ES*=0.23, 95% *CI*: 0.05, 0.42), although this effect was small and highly heterogeneous (I^2=78.6%, 95% *CI:* 76.2%, 80.7%). Moderator analysis revealed that HRV metric (p<0.001), position (p=0.033), and biological sex (β=0.45, 95% *CI:* 0.30, 0.61; p<0.001), but not portable device, modulated the degree of absolute error. Absolute error was

significantly higher within metric when expressed as SDNN (*ES*=0.44) compared to any other metric. However, SDNN was no longer significantly different after a sensitivity analysis removed outliers. Similarly, the error associated with the tilt/recovery position was significantly higher than any other position and remained significantly different without outliers in the model. These findings suggest that HRV measures acquired from portable devices demonstrate a small amount of absolute error when compared to ECG. However, this small error is acceptable when considering the improved practicality and compliance of HRV measures acquired through portable devices in the field setting. Practitioners and researchers should consider the cost-benefit along with the simplicity of the measure when attempting to increase compliance in acquiring HRV measures.

The second study examined the relationship between the time course in recovery of criterion short-term HRV measures and muscular performance recovery over a 72-hr period following an exhaustive bout of resistance training. Results showed that all HRV metrics had a significant interaction with muscular performance *"performance"* over time ($p < .01$) indicating change scores in *performance* and HRV following the physiological stressor were not parallel. Furthermore, the mean change scores in all HRV metrics significantly differed from *performance* ($p < .05$) across time, except SDNN, LF, and SD2 at the 0.5-hr mark, and HF at the 24-hr time point. The repeated measures correlation analysis indicated a lack of intra-individual association between the change in *performance* and HRV over time (all < .45). The findings indicated that recovery in HRV measures following an exhaustive bout of resistance training precede the recovery in acute muscular performance recovery.

The third study evaluated the agreement between ultra-short and criterion short-term HRV measures surrounding a bout of exhaustive resistance training. The highest levels of

agreement were displayed from lnRMSSD [LOA = -0.91– 0.69, *ICC* = .91, *p* = .082, *ES* = 0.15] and lnSD1 [LOA = -0.90 – 0.72, *ICC* = .91, *p* = .156, *ES* = 0.13] compared to all other metrics. These results support the use of ultra-short lnRMSSD and lnSD1 as they provide accurate measures compared to the criterion short-term recordings.

The collective findings of this dissertation support the use of portable devices and ultra-short recordings to derive accurate measures of HRV. Furthermore, it was also identified that recoveries in criterion short-term HRV measures were not associated with muscular performance recovery over the 72-hr period following a strenuous bout of resistance training. However, as HRV metrics which are believed to represent a combination of parasympathetic and sympathetic activity (e.g., LF) did not show good agreement with criterion measures when extrapolated around the strenuous resistance training protocol, future research should investigate if ultra-short measures of these indices provide a better association with muscular performance recovery.

REFERENCES

1. Malik M, Camm AJ, Bigger Jr JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17(3):354-81.

2. Prinsloo GE, Rauch HG, Derman WE. A brief review and clinical application of heart rate variability biofeedback in sports, exercise, and rehabilitation medicine. Phys Sportsmed. 2014;42(2):88-99.

3. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol. 2010;141(2):122-31.

4. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol. 2014;5:73.

5. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol. 2014;5:1040.

6. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. Sports Med. 2013;43(9):773-81.

7. Shaffer F, Shearman S, Meehan ZM. The promise of ultra-short-term (UST) heart rate variability measurements. Biofeedback. 2016;44(4):229-33.

8. Flatt AA, Esco MR. Heart rate variability stabilization in athletes: towards more convenient data acquisition. Clin Physiol Funct Imaging. 2016;36(5):331-6.

9. Thompson WR. Worldwide survey of fitness trends for 2018: the CREP edition. ACSMs Health Fit J. 2017;21(6):10-9.

10. Thompson WR. Worldwide survey of fitness trends for 2017. ACSMs Health Fit J. 2016;20(6):8-17.

11. Radespiel-Tröger M, Rauh R, Mahlke C, Gottschalk T, Muck-Weymann M. Agreement of two different methods for measurement of heart rate variability. Clin Auton Res. 2003;13(2):99-102.

12. Wallen MB, Hasson D, Theorell T, Canlon B, Osika W. Possibilities and limitations of the Polar RS800 in measuring heart rate variability at rest. Eur J Appl Physiol. 2012;112(3):1153-65.

13. Esco MR, Flatt AA, Nakamura FY. Agreement between a smartphone pulse sensor application and electrocardiography for determining lnRMSSD. J Strength Cond Res. 2017;31(2):380-5.

14. Bellenger CR, Fuller JT, Thomson RL, Davison K, Robertson EY, Buckley JD. Monitoring athletic training status through autonomic heart rate regulation: a systematic review and meta-analysis. Sports Med. 2016;46(10):1461-86.

15. Flatt AA, Esco MR. Evaluating individual training adaptation with smartphone-derived heart rate variability in a collegiate female soccer team. J Strength Cond Res. 2016;30(2):378-85.

16. Plews DJ, Laursen PB, Buchheit M. Day-to-day heart-rate variability recordings in world-champion rowers: appreciating unique athlete characteristics. Int J Sports Physiol Perform. 2017;12(5):697-703.

17. Chen JL, Yeh DP, Lee JP, Chen CY, Huang CY, Lee SD, et al. Parasympathetic nervous activity mirrors recovery status in weightlifting performance after training. J Strength Cond Res. 2011;25(6):1546-52.

APPENDIX

THE UNIVERSITY OF ALABAMA | Office of the Vice President for Research & Economic Development
Office for Research Compliance

June 20, 2018

Michael Esco, Ph.D.
Assistant Professor
Department of Kinesiology
College of Education
The University of Alabama
Box 870312

Re: IRB Protocol # 16-021-ME-R1-A
"Heart Rate Variability for Reflecting Psychophysiology Recovery Following Physically and Mentally Stressful Events"

Dr. Esco:

The University of Alabama Medical Institutional Review Board has reviewed the revision to your previously approved full board protocol. The board has approved the change in your protocol.

Please remember that your protocol will expire on June 13, 2019.

Should you need to submit any further correspondence regarding this proposal, please include the assigned IRB application number. Changes in this study cannot be initiated without IRB approval, except when necessary to eliminate apparent immediate hazards to participants.

Good luck with your research.

Sincerely,

J. Grier Stewart, MD, FACP
Medical IRB Chair

358 Rose Administration Building | Box 870127 | Tuscaloosa, AL 35487-0127
205-348-8461 | Fax 205-348-7189 | Toll Free 1-877-820-3066