Autonomic Neuroscience: Basic and Clinical 221 (2019) 102582

Contents lists available at ScienceDirect

Autonomic Neuroscience: Basic and Clinical

journal homepage: www.elsevier.com/locate/autneu

Reproducibility of post-exercise heart rate recovery indices: A systematic review

Rafael Y. Fecchio[a,*], Leandro Brito[a], Anthony S. Leicht[b], Cláudia L.M. Forjaz[a], Tiago Peçanha[a]

[a] School of Physical Education and Sport, Exercise Hemodynamic Laboratory, University of São Paulo, São Paulo, Brazil
[b] Sport and Exercise Science, James Cook University, Townsville, Qld, Australia

ABSTRACT

Heart rate recovery (HRR) has been widely used to evaluate the integrity of the autonomic nervous system with a slower HRR being associated with greater cardiovascular risk. Different HRR indices have been proposed. Some evaluate HR changes from the end of exercise to a specific recovery moment (e.g. 60s – HRR60s; 120s – HRR120s; 300s – HRR300s) and others calculate time-constant decays of HR for different recovery intervals (e.g. first 30s – T30; the entire period – HRRt). Several studies have examined the reproducibility of these commonly-used HRR indices, but reported discordant findings. Thus, this systematic review was designed to synthesize the reproducibility of HRR. We included studies that evaluated short-term (< 1 year) reproducibility of HRR after dynamic exercise by employing typical measures of reliability (intraclass correlation coefficient, ICC) and agreement (coefficient of variation, CV). The electronic database PubMed/Medline was searched for relevant studies published up to July 2018. From the initial 120 records identified, 15 studies were retained for the qualitative synthesis of 24 experimental conditions. During most experimental conditions, high ICC and desirable CV were reported for HRR60s (62.5 and 76.2%, respectively), HRR120s (55.6 and 71.4%) and HRR300s (50.0 and 100.0%). While, it were reported during the minority of conditions for HRRt (37.5 and 42.9%) and in none condition for T30 (0.0 and 0.0%). In conclusion, HRR60s, HRR120s and HRR300s exhibited good reproducibility for evaluating HRR in predominantly healthy males within research and clinical settings. In contrast, caution should be taken when employing other HRR indices (T30, HRRt) due to their poorer reproducibility.

1. Introduction

Immediately after exercise, heart rate (HR) exponentially decreases towards its pre-exercise levels, which is known as “heart rate recovery” (HRR) (Coote, 2010; Imai et al., 1994; Peçanha et al., 2014a, 2014b; Perini et al., 1989a). This response is mainly determined by cardiac autonomic restoration and, therefore, HRR has been widely used as a non-invasive tool for evaluating the integrity of the autonomic nervous system (Coote, 2010; Imai et al., 1994; Peçanha et al., 2016; Peçanha et al., 2014a, 2014b; Perini et al., 1989b). In healthy humans, post-exercise HR usually takes 5-–10 min to reach resting values (Peçanha et al., 2014a, 2014b), but a longer period might be necessary after high-intensity exercises (Seiler et al., 2007). In various clinical populations (Bienias et al., 2018; Kanegusuku et al., 2016; Peçanha et al., 2018; Racine et al., 2003), a slower HRR has been reported and identified as a marker of cardiac dysautonomia. In addition, longitudinal studies have shown that a slower HRR after maximal and submaximal dynamic exercise is an independent and powerful predictor of overall and cardiovascular mortalities in patients with (Lachman et al., 2018) and without cardiovascular disease (Qiu et al., 2017).

Regarding the kinetics of HRR, two distinct phases can be identified (Peçanha et al., 2014a, 2014b). The first phase consists of a rapid decline of post-exercise HR and is primarily modulated by parasympathetic reactivation (Imai et al., 1994). The second phase is characterized by a slower reduction of post-exercise HR resulting from sympathetic withdrawal and parasympathetic reactivation (Perini et al., 1989b). Different indices have been reported as markers of HRR with their physiological meanings dependent upon the HRR phase they evaluate (Peçanha et al., 2016). For example, the HR reduction after 1 min of recovery (HRR60s) and the short-term time constant of HRR (T30) represent the fast phase of HRR and parasympathetic reactivation (Peçanha et al., 2014a, 2014b). On the other hand, the HR reduction after 2 (HRR120s) and 5 min (HRR300s) of recovery, as well as the time constant of HRR after exponential fitting (HRRt) mainly reflect the slow phase (Peçanha et al., 2014a, 2014b) of HRR and the actions of both autonomic nerve branches (sympathetic withdrawal and parasympathetic restoration). Therefore, the use of specific HRR indices should be carefully considered with selection based on the HRR phase and the mechanism of interest. Further, the selection of HRR indices should take into account their reproducibility as a key methodological element for assuring consistent findings (Atkinson and Nevill, 1998; Hopkins, 2000).

Numerous studies have examined the reproducibility of several HRR indices (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008, 2010; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009;

* Corresponding author at: School of Physical Education and Sport, University of São Paulo, São Paulo, São Paulo, Brazil.
E-mail address: rafael.fecchio@usp.br (R.Y. Fecchio).

https://doi.org/10.1016/j.autneu.2019.102582
Received 18 February 2019; Received in revised form 3 July 2019; Accepted 21 August 2019

Lamberts et al., 2011; Mellis et al., 2014; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016; Yawn et al., 2003) and a wide range of results has been reported. For example, the intraclass correlation coefficient (ICC), an index of reliability that defines the capacity of a measure to distinguish subjects in a given population, was reported to range from 0.12 (T30) to 0.99 (HRR60s) (Dupuy et al., 2012; Lamberts et al., 2009). Similarly, for the coefficient of variation (CV), an index of agreement that quantifies the within-individual variability between repeated measurements, values were reported to range from 3.4 (HRR60s) to 90.1% (T30) (Arduini et al., 2011; Lamberts et al., 2011). These large variations in results raise serious concerns about the reproducibility of HRR indices and require further clarification.

Given the widespread use of HRR in assessing healthy and clinical populations (Peçanha et al., 2014a, 2014b), the prognostic significance of HRR (Lachman et al., 2018; Qiu et al., 2017), and the inconsistencies amongst prior studies of HRR reproducibility, the current study aimed to systematically review the reproducibility of commonly employed HRR indices. It was expected that the review outcomes would clarify the reproducibility of HRR indices and assist clinicians and researchers with the selection of the most appropriate indices for their intended purposes.

2. Methods

2.1. Literature search

The US National Library of Medicine and National Institutes of Health (i.e., PubMed) database was independently searched by two authors using the following terms and strategy: ("heart rate recovery" or "post-exercise heart rate" or "parasympathetic reactivation") and ("reproducibility" or "reliability" or "agreement" or "validity" or "test-retest" or "accuracy" or "typical error" or "intraclass correlation coefficient" or "standard error of measurement" or "limits of agreement" or "coefficient of variation"). The search included all sources up to July 31, 2018. To complement the database, additional manual searches of the reference lists of identified articles were also performed.

2.2. Selection criteria

A flow diagram depicting the search and screening process is shown in Fig. 1. The literature search initially identified 118 records and 2 other records were identified by manual search. In a first analysis, records were reduced to 61 by removing duplicate records. Afterwards, studies were excluded based upon not meeting the following criteria after abstract review: 1) being an original study; 2) conducted with humans; 3) conducted using dynamic exercise (e.g. walking, running, cycling, rowing, etc.), given that HRR association with non-fatal and fatal cardiovascular outcomes has only been demonstrated with this type of exercise (Lachman et al., 2018; Qiu et al., 2017); and 4) including repeated tests. The full texts of the remaining studies were assessed for eligibility, and studies were excluded if they did not utilize typical measures of reliability (i.e. ICC) (Atkinson and Nevill, 1998; de Vet et al., 2006) and/or agreement (i.e. CV) (Atkinson and Nevill, 1998; de Vet et al., 2006; Hopkins, 2000). The combination of ICC and CV was chosen to provide a comprehensive reproducibility evaluation by assessing both reliability and agreement. In addition, both indices have proposed cutoffs limits for their interpretations, allowing posterior descriptive analyses, and their results are expressed in relative terms (ICC varies within a 0.0 to 1.0 scale and CV is a percentage of measure error in relation to variable mean value), allowing the comparison of different HRR indices and studies results. Lastly, one study was excluded because it evaluated longer-term HRR reproducibility, employing an interval of 1 year between the repeated tests. After these steps, 15 studies were retained and included in the qualitative synthesis. From these, some studies evaluated more than one experimental condition by comparing different populations or exercise protocols. For example, the Dupuy et al.'s study (Dupuy et al., 2012) evaluated, in a cross-over design, the HRR reproducibility after both submaximal and maximal exercises, and reported reproducibility results for each exercise intensity. In such cases, 2 experimental conditions were considered. In the cited example, the result for the maximal exercise and the result for the submaximal exercise were considered as two different experimental conditions. Given that, the 15 retained studies resulted in 24 different experimental conditions.

2.3. Data extraction

Data were extracted by one author (R.Y.F) and checked by another (T.P.) with discrepancies resolved by critical discussion. The following descriptive characteristics of each study were extracted: 1) sample size; 2) sample characteristics (i.e. gender, age, physical training status and health condition); 3) exercise protocol characteristics (i.e. exercise mode, exercise intensity and recovery protocol) (Table 1). Subsequently, the reproducibility outcomes of the following HRR indices were extracted: 1) HRR60s (i.e. Decay of Heart Rate in 60s after exercise); 2) T30 (i.e. Short Term Time Constant); 3) HRR120s (Decay of Heart Rate in 120 s after exercise); 4) HRR300s (Decay of HR in 300s after exercise); and 5) HRRt (Time Constant of Heart Rate Recovery after Exponential Fitting) (Table 2).

2.4. Data analysis

Reproducibility was considered by measures of reliability based on ICC and agreement based on CV (Atkinson and Nevill, 1998; de Vet et al., 2006). Reliability was considered high when the ICC was ≥ 0.75, moderate when the ICC was between 0.75, and 0.40 and low when the ICC was < 0.40 (Szklo and Nieto, 2000). Agreement was considered desirable when the CV $< 20.0\%$, intermediate when the CV was between 20.0 and 30.0% and undesirable when the CV was $> 30.0\%$ (Rosner, 2011).

3. Results

3.1. Study characteristics

Table 1 lists the characteristics of the reviewed studies. Considering the fifteen studies (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016), sample size ranged from 12 to 52 subjects and the mean age varied from 16 to 64 years. Ten studies (66.7%) evaluated only males (Al Haddad et al., 2011; Bonato et al., 2018; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Otter et al., 2015; Rabbani et al., 2018; Veugelers et al., 2016), one (6.7%) evaluated only females (Tonello et al., 2016) and four studies (26.7%) included both males and females (Arduini et al., 2011; Bosquet et al., 2008; Lamberts et al., 2011; Tulumen et al., 2011). Regarding health status, fourteen studies (93.3%) examined healthy subjects, (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016) while only one study (6.7%) examined a clinical population composed of patients with intermittent claudication (Fecchio et al., 2018). Most of the studies employed an exercise protocol involving walking or running exercise (73.3%) (Al Haddad et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2011; Rabbani et al., 2018; Tulumen et al., 2011; Veugelers et al., 2016),

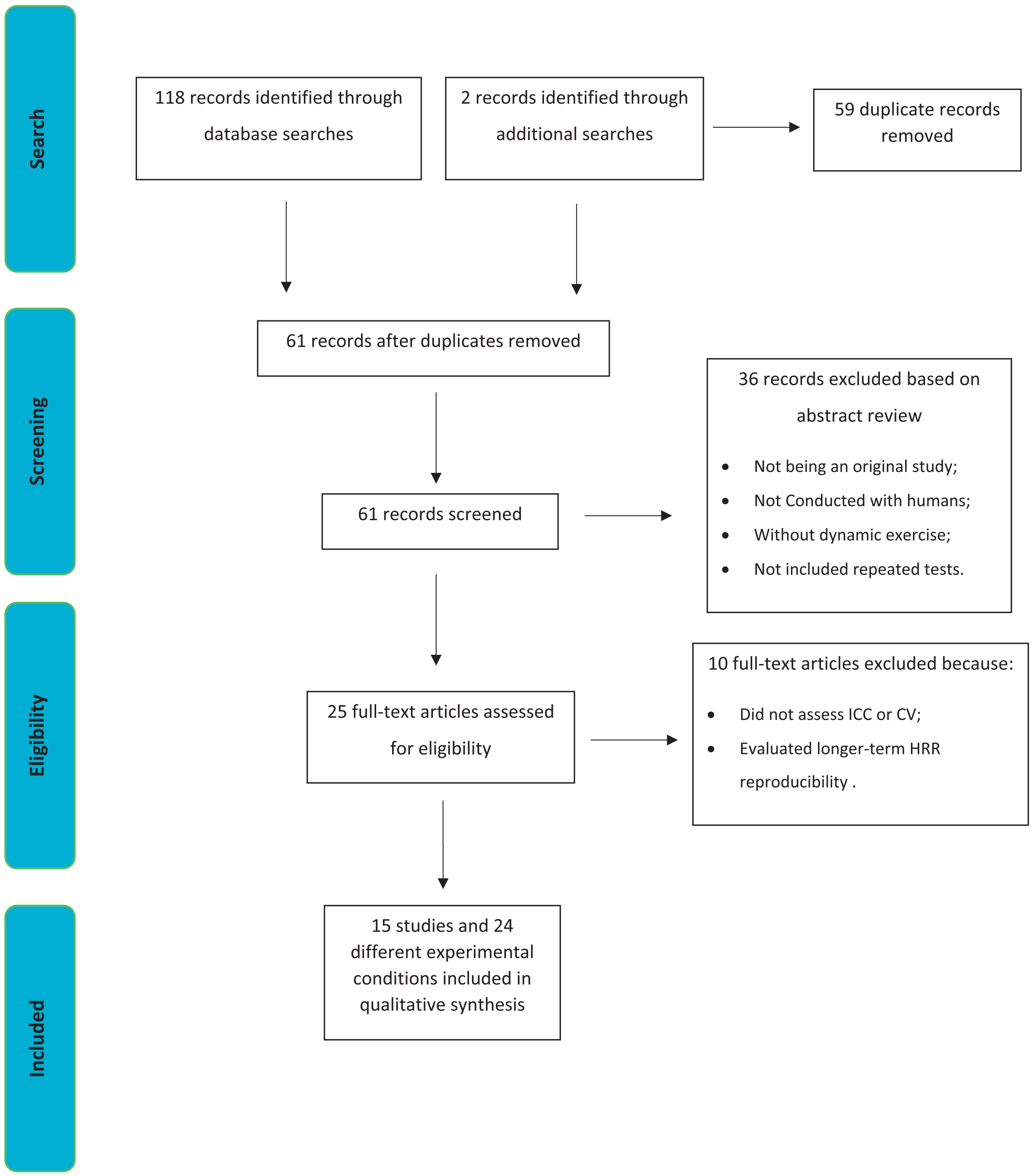

Fig. 1. Flow diagram of the study. CV = coefficient of variation; HRR = heart rate recovery; ICC = intraclass correlation coefficient.

submaximal intensity workload (80.0%) (Al Haddad et al., 2011; Arduini et al., 2011; Bosquet et al., 2008; Buchheit et al., 2008; Dupuy et al., 2012; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016), and inactive recovery (80.0%) (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Veugelers et al., 2016). Lastly, thirteen studies (86.7%) (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Veugelers et al., 2016) obtained HR through telemetric breast belt devices (e.g. Polar HR monitor) and the other two studies (13.3%) (Fecchio et al., 2018; Tulumen et al., 2011) by electrocardiography.

3.2. *Reproducibility of HRR indices – fast phase*

Reproducibility of HRR60 (Fig. 2) was examined by all fifteen studies (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016). Reliability analysis was conducted in sixteen different experimental conditions. The HRR60s was reported as highly reliable (ICC = 0.77–0.99) for ten conditions (62.5%) (Arduini et al., 2011; Bonato et al., 2018; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Otter et al., 2015; Rabbani et al., 2018; Tulumen et al., 2011; Veugelers et al., 2016) and moderately reliable (ICC = 0.43–0.70) for the remaining six conditions (37.5%) (Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Tonello et al., 2016). HRR60s agreement was evaluated during twenty one different experimental conditions with desirable agreement (CV = 3.4–14.7%) reported for sixteen of them (76.2%) (Al Haddad et al., 2011; Bonato et al., 2018; Dupuy et al.,

Table 1
Characteristics of sample and exercise protocols employed in the studies examining heart rate recovery reproducibility.

Study	Participants /gender	Age (years)	Physical training status	Health condition	Exercise mode	Exercise intensity	Recovery protocol
Al Haddad et al. (2011)	15 M	22 ± 1	Trained	Healthy	Running and cycling	Submaximal (60% VO_{2max}) and Supramaximal (30-s Wingate)	Inactive
Arduini et al. (2011)	17 M, 4F	26 ± 5	Untrained and trained	Healthy	Cycling	Lower submaximal (65% HR_{max}) and Higher submaximal (80% HR_{max})	Inactive
Bonato et al. (2018)	14 M	24 ± 4	Trained	Healthy	Running	Supramaximal (30-m sprints, 5 times)	Inactive
Bosquet et al. (2008)	22 M, 8F	33 ± 10	Trained	Healthy	Running	Submaximal (80% PVR) and Maximal	Inactive
Boullosa et al. (2014)	12 M	30 ± 5	Trained	Healthy	Running	Maximal	Active
Buchheit et al. (2008)	15 M	16 ± 1	Trained	Healthy	Running	Submaximal (60% PVR)	Inactive
Dupuy et al. (2012)	30 M	27 ± 8	–	Healthy	Running	Submaximal (75% PVR) and Maximal	Inactive
Fecchio et al. (2018)	19 M	64 ± 8	Untrained	Patients with IC	Walking	Maximal	Inactive
Lamberts et al. (2009)	17 M	31 ± 4	Trained	Healthy	Cycling	Submaximal (60–90% HR_{max})	Inactive
Lamberts et al., 2011	21 M, 10F	LFL: 24 ± 4; MFL: 25 ± 4; HFL: 24 ± 3.	Trained	Healthy	Running	Lower submaximal (6.0–9.6 km/h), Standard submaximal (7.2–10.8 km/h) and Higher submaximal (8.4–12.0 km/h)	Inactive
Otter et al. (2015)	12 M	20 ± 1	Trained	Healthy	Rowing	Submaximal (70–90% HR_{max})	Inactive
Rabbani et al. (2018)	12 M	25 ± 4	Trained	Healthy	Running	Submaximal (12 km/h)	Inactive
Tonello et al. (2016)	21F	35 ± 6	Untrained	Healthy	Cycling	Submaximal (86% HR_{max})	–
Tulumen et al. (2011)	22 M, 30F	30 ± 10	–	Healthy	Running	Submaximal (85% HR_{max})	Active in the 1st minute and inactive in the remaining 4 min
Veugelers et al. (2016)	25 M	23 ± 4	Trained	Healthy	Running	Submaximal (75–85% HR_{max})	Inactive

F = females; HFL = high fitness level group; HR_{max} = maximal heart rate; IC = intermittent claudication; LFL = low fitness level group; M = males; MFL = medium fitness level group; PVR = peak velocity reached; km/h = kilometers per hour; VO_{2max} = maximal oxygen consumption.

Table 2
Summary of reproducibility measures reported in eligible studies of the systematic review.

Study	HRR60s	T30	HRR120s	HRR300s	HRRt
Al Haddad et al. (2011)	**Submaximal** CV = 25.7% **Supramaximal** CV = 14.7%#				**Submaximal** CV = 32.1% **Supramaximal** CV = 24.3%
Arduini et al. (2011)	**Lower submaximal** ICC = 0.81* CV = 34.0% **Higher submaximal** ICC = 0.78* CV = 34.9%	**Lower submaximal** ICC = 0.56 CV = 90.1% **Higher submaximal** ICC = 0.55 CV = 67.7%	**Lower submaximal** ICC = 0.68 CV = 41.3% **Higher submaximal** ICC = 0.76 * CV = 29.7%		
Bonato et al. (2018)	**Supramaximal** ICC = 0.91* CV = 10.9%#				**Supraximal** ICC = 0.96* CV = 13.5%#
Bosquet et al. (2008)	**Submaximal** ICC = 0.43 **Maximal** ICC = 0.58		**Submaximal** ICC = 0.60 **Maximal** ICC = 0.62	**Submaximal** ICC = 0.57 **Maximal** ICC = 0.71	**Submaximal** ICC = 0.43 **Maximal** ICC = 0.71
Boullosa et al. (2014)	**Maximal** ICC = 0.64 CV = 23.3%		**Maximal** ICC = 0.89* CV = 8.4%#	**Maximal** ICC = 0.78 CV = 7.0%#	**Maximal** ICC = 0.84* CV = 13.2%#
Buchheit et al. (2008)	**Submaximal** ICC = 0.70	**Submaximal** ICC = 0.73			**Submaximal** ICC = 0.86*
Dupuy et al. (2012)	**Submaximal** ICC = 0.69 CV = 12.5%# **Maximal** ICC = 0.77* CV = 10.8%#	**Submaximal** ICC = 0.23 CV = 50.0% **Maximal** ICC = 0.12 CV = 72.6%			**Submaximal** ICC = 0.47 CV = 29.8% **Maximal** ICC = 0.74 CV = 11.5%#
Fecchio et al. (2018)	**Maximal** ICC = 0.92* CV = 13.2%#		**Maximal** ICC = 0.88* CV = 12.2%#	**Maximal** ICC = 0.93 CV = 8.6%#	
Lamberts et al. (2009)	**Submaximal** ICC = 0.99* CV = 4.2%#				
Lamberts et al., 2011	**Submaximal standard protocol** *Lower fitness level* CV = 10.6%# *Medium fitness level* CV = 3.4% # *Higher fitness level* CV = 10.5%# **Submaximal adapted protocol** *Lower fitness level* CV = 4.7%# *Medium fitness level* CV = 5.2%# *Higher fitness level* CV = 4.4%#				
Otter et al. (2015)	**Submaximal** ICC = 0.93* CV = 8.1%#				
Rabbani et al. (2018)	**Submaximal** ICC = 0.84* CV = 7.0%#				
Tonello et al. (2016)	**Submaximal** ICC = 0.51 CV = 6.3%#	**Submaximal** ICC = 0.20 CV = 75.3%	**Submaximal** ICC = 0.39 CV = 6.5%#	**Submaximal** ICC = 0.37 CV = 6.9%#	**Submaximal** ICC = 0.36 CV = 31.3%
Tulumen et al. (2011)	**Submaximal** ICC = 0.88 CV = 23.2%		**Submaximal** ICC = 0.88* CV = 12.4%#	**Submaximal** ICC = 0.82* CV = 10.1%#	
Veugelers et al. (2016)	**Submaximal** ICC = 0.94* CV = 9.2%#		**Maximal** ICC = 0.90* CV = 6.0%#		

CV = coefficient of variation; HRR60s = heart rate reduction after 60s of recovery; HRR120s = heart rate reduction after 120 s of recovery; HRR300s = heart rate reduction after 300 s of recovery; HRRt = long-term time constant of HRR after exponential fitting ICC = intraclass correlation coefficient; T30 = short-term time constant of HRR after linear fitting. ICC ≥ 0.75 was considered as high reliability (*), ICC between 0.75 and 0.40 as moderate reliability and ICC < 0.40 as low reliability. CV < 20% was considered as desirable agreement (#), CV between 20% and 30% was considered as intermediate agreement and CV > 30% was considered as undesirable agreement.

2012; Fecchio et al., 2018; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Tonello et al., 2016; Veugelers et al., 2016), intermediate agreement (CV = 23.2–25.7%) for three (14.3%) (Al Haddad et al., 2011; Boullosa et al., 2014; Tulumen et al., 2011) and undesirable agreement (CV = 34.0–34.9%) for the last two conditions (9.5%) (Arduini et al., 2011).

The reproducibility of T30 (Fig. 2) was examined by four studies (Arduini et al., 2011; Buchheit et al., 2008; Dupuy et al., 2012; Tonello

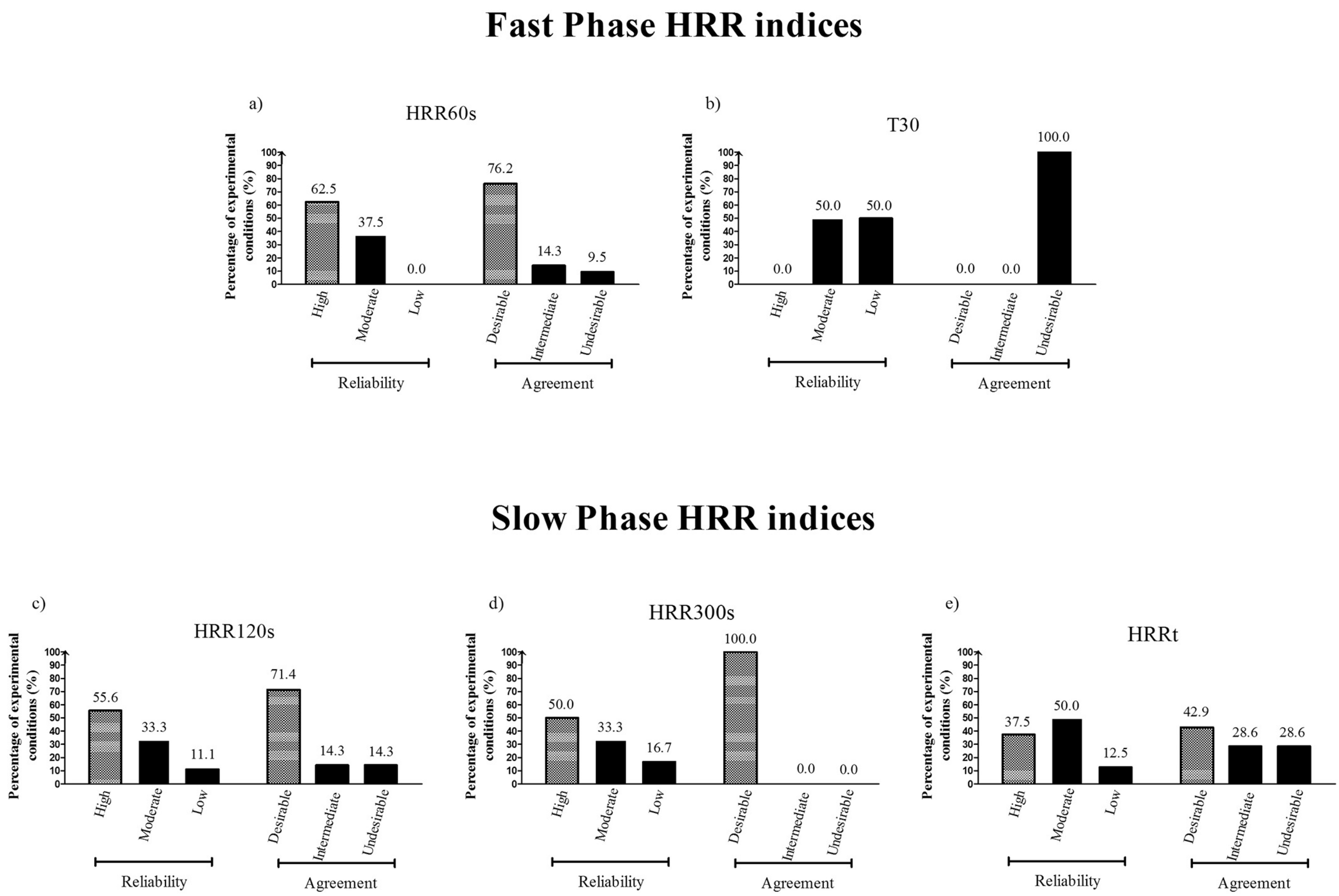

Fig. 2. Synthesis of the reproducibility (reliability and agreement) results from included studies that evaluated the fast (HRR60 – panel a; T30 – panel b) and slow phase (HRR120s – panel c; HRR300 – panel d; HRRt – panel e) HRR indices.

et al., 2016). Reliability was examined during six different experimental conditions, with moderate reliability (ICC = 0.55–0.73) being reported in three conditions (50.0%) (Arduini et al., 2011; Buchheit et al., 2008) and low (ICC = 0.12–0.23) reliability in the other three conditions (50.0%) (Dupuy et al., 2012; Tonello et al., 2016). T30 agreement was analyzed during five experimental conditions, and undesirable agreement (CV = 50.0–90.1%) was reported for all of them (100.0%) (Arduini et al., 2011; Dupuy et al., 2012; Tonello et al., 2016).

3.3. Reproducibility of HRR indices – slow phase

The reproducibility of HRR120s (Fig. 2) was examined by seven studies (Arduini et al., 2011; Bosquet et al., 2008; Boullosa et al., 2014; Fecchio et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016). Reliability was analyzed during nine experimental conditions. High reliability (ICC = 0.76–0.90) was reported during five conditions (55.6%) (Arduini et al., 2011; Boullosa et al., 2014; Fecchio et al., 2018; Tulumen et al., 2011; Veugelers et al., 2016), moderate reliability (ICC = 0.60–0.68) during three conditions (33.3%) (Arduini et al., 2011; Bosquet et al., 2008) and low reliability (ICC = 0.39) for the last condition (11.1%) (Tonello et al., 2016). HRR120s agreement was examined during seven experimental conditions, and desirable agreement (CV = 6.0–12.4%) was reported for five conditions (71.4%) (Boullosa et al., 2014; Fecchio et al., 2018; Tonello et al., 2016; Tulumen et al., 2011; Veugelers et al., 2016), intermediate agreement (CV = 29.7%) for one condition (14.3%) (Arduini et al., 2011) and undesirable agreement (CV = 41.3%) for the remaining condition (14.3%) (Arduini et al., 2011).

HRR300s reproducibility (Fig. 2) was evaluated by five studies (Bosquet et al., 2008; Boullosa et al., 2014; Fecchio et al., 2018; Tonello et al., 2016; Tulumen et al., 2011). Its reliability was examined during six experimental conditions. High reliability (ICC = 0.78–0.93) was reported for three conditions (50.0%) (Boullosa et al., 2014; Fecchio et al., 2018; Tulumen et al., 2011), moderate (ICC = 0.57–0.71) for two conditions (33.3%) (Bosquet et al., 2008) and low (ICC = 0.37) for the remaining conditions (16.7%) (Tonello et al., 2016). HRR300s agreement was reported as desirable (CV = 6.9–10.1%) for all four conditions in which it was evaluated (100.0%) (Boullosa et al., 2014; Fecchio et al., 2018; Tonello et al., 2016; Tulumen et al., 2011).

HRRt reproducibility (Fig. 2) was examined by seven studies (Al Haddad et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Tonello et al., 2016). Reliability analysis was performed in eight experimental conditions. It was reported as highly reliable (ICC = 0.84–0.96) for three conditions (37.5%) (Bonato et al., 2018; Boullosa et al., 2014; Buchheit et al., 2008), moderately reliable (ICC = 0.43–0.74) for four conditions (50.0%) (Bosquet et al., 2008; Dupuy et al., 2012) and with low reliability (ICC = 0.36) for the last condition (12.5%) (Tonello et al., 2016). HRRt agreement was analyzed during seven experimental conditions with desirable agreement (CV = 11.5–13.5%) identified for three conditions (42.9%) (Bonato et al., 2018; Boullosa et al., 2014; Dupuy et al., 2012), intermediate (CV = 24.3–29.8%) for two (28.6%) (Al Haddad et al., 2011; Dupuy et al., 2012) and undesirable (CV = 31.3–32.1%) for the last two conditions (28.6%) (Al Haddad et al., 2011; Tonello et al., 2016).

4. Discussion

The current systematic review demonstrated that HRR60s, HRR120s and HRR300s exhibited high reliability and desirable agreement predominantly while T30 and HRRt mostly exhibited moderate to low reliability and intermediate to undesirable agreement. Therefore, defining all indices of HRR as a reproducible measure is a generalization that clinicians and researchers should consider with caution.

Our review revealed that the commonly-used HRR indices, calculated by the absolute difference between HR at the end of exercise and at a specific moment of recovery (i.e. HRR60s, HRR120s and HRR300s), exhibited good reproducibility as defined by high ICC and low CV. These results might be explained by the simplicity of the calculation for these indices and by the inherent good reproducibility of HR measures obtained both during exercise and recovery (Bonato et al., 2018; Bosquet et al., 2008; Boullosa et al., 2014; Fecchio et al., 2018; Lamberts et al., 2009; Rabbani et al., 2018; Tulumen et al., 2011). On the other hand, HRR indices calculated based on mathematical models, such as T30 and HRRt, demonstrated medium to poor reproducibility (i.e. most conditions exhibited moderate to low ICC and intermediate to high CV). This poorer outcome might be related to the adequacy of the curve fitting process for both T30 and HRRt. Therefore, subtle variations in the response of heart rate after exercise may lead to slight errors between real and modelled heart rates that results in significantly diminished reproducibility of both indices (Peçanha et al., 2016). For instance, T30 is calculated by a semi-logarithmic analysis of the HR decay during the initial 30 s of recovery (Imai et al., 1994; Peçanha et al., 2016). During this immediate period, heart rate responses may include a brief plateau or slight increase that substantially alters the model fit and calculation of T30. Similarly, HRRt is also calculated via modeling and expressed as a single time-constant that estimates the HRR return over the entire recovery period (Peçanha et al., 2016). This modeling typically involves the use of a mono-exponential function, which may not provide the best fit for HRR modeling in some circumstances (e.g. maximal exercise) (Pierpont et al., 2000). Errors associated with modeling HRR indices may lead to greater variability in outcome measures and poorer reliability, as shown in the current review. Identification of these methodological limitations is essential to help guide implementation of HRR indices.

The current results will help clinicians and researchers with the selection of the most appropriate HRR indices for clinical and/or research use. Reliability defines the capacity of a measure to distinguish the individuals in a given population. It evaluates the degree by which the subjects maintain their positions (i.e. rankings) between repeated tests (Atkinson and Nevill, 1998; de Vet et al., 2006). Thus, given that the current review has clearly indicated that HRR60s, HRR120s and HRR300s were the most reliable measures, these indices should be preferentially employed to classify patients in clinical settings. In practical terms, this would allow the clinicians and researches to identify with greater accuracy the subjects who present greater autonomic dysfunction, which would increase cardiovascular risk prediction and improve clinical screening.

On the other hand, agreement quantifies the error of a measurement, which has important implications, since a higher error of measurement will impair its sensitivity to detect effects of treatments (Atkinson and Nevill, 1998; Hopkins, 2000). In this sense, the current results indicated that HRR60s, HRR120s and HRR300s have greater agreement when compared to T30 and HRRt. In practical terms, clinicians and researches would have a greater assurance, when tracking their patients over time, that an observed change on HRR reflects a real effect of treatment instead a merely random variation by employing HRR60s, HRR120s and HRR300s. Moreover, the current agreement results also have applications for study designs, since the error of measurement is commonly employed as a parameter for sample size estimations. Given that, the current results indicated the HRR indices (HRR60s, HRR120s and HRR300s) that will demand smaller sample sizes to detect group responses with statistical significance. These results might explain why some studies found divergent results by employing different HRR indices, such T30 and HRRt would demand clinical trials with a higher number of subjects to detect treatment effects.

The current review has provided novel insight to the reproducibility of HRR indices across a range of populations including the degree of variation in reproducibility results for each index. For example, ICCs values ranged from 0.43 to 0.99 and CVs varied between 3.4 and 34.9% for the most widespread HRR index, HRR60s. Differences between studies, such as population characteristics and/or the exercise, and recovery protocols may contribute to this variation on HRR reproducibility that requires further investigation.

Regarding population characteristics, only one study, included in this current review, investigated HRR reproducibility in females (Tonello et al., 2016) and reported lower HRR60s reliability (ICC = 0.51) than all the studies conducted with men (ICCs = 0.64–0.99) (Bonato et al., 2018; Boullosa et al., 2014; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Otter et al., 2015; Rabbani et al., 2018; Veugelers et al., 2016). This poorer reproducibility for women may reflect a true gender difference that remains to be confirmed. This gender difference may be related to the potential effects of the menstrual cycle on cardiovascular autonomic modulation stability (Minson et al., 2000), which was not controlled in the mentioned study (Tonello et al., 2016). Differences in health status have been reported to impact on cardiac autonomic activity (Dogdu et al., 2010; Gupta et al., 2013; Kanegusuku et al., 2016; Racine et al., 2003) and should be expected to impact HRR reproducibility. However, in the current review only one study (Fecchio et al., 2018) examined a non-healthy population (i.e. peripheral artery disease patients) that reported good reproducibility with results (ICC: 0.92; CV = 13.2%) within the range observed for healthy individuals (ICC = 0.43–0.99; CV = 3.4–34.9%). This was an interesting finding and suggests that health status may not affect HRR reproducibility, although other diseases might distinctly affect cardiac autonomic regulation (Haapaniemi et al., 2001) and thus HRR reproducibility, which deserves further investigation.

Regarding the possible influence of exercise protocol, overall the current review identified higher agreement after maximal (CV ≤ 23.3%) than submaximal exercise (CV ≤ 34.9%). Similarly, the two studies that directly compared different exercise intensities reported better reliability (ICC = 0.77 vs 0.69) (Dupuy et al., 2012) and agreement (CV = 14.7 vs 25.7%) (Al Haddad et al., 2011) after higher exercise intensities. Greater reproducibility after higher-intensity exercise may result from the greater suppression of HR variability after maximal than submaximal exercise (de Oliveira et al., 2013) with the use of higher intensity exercises recommended for reliable HRR results. Finally, most of the studies included in this review examined HRR60s during inactive recovery and reported moderate to good reliability and agreement (ICCs = 0.43–0.99; CVs = 3.4 to 34.9%) (Al Haddad et al., 2011; Arduini et al., 2011; Bonato et al., 2018; Bosquet et al., 2008; Buchheit et al., 2008; Dupuy et al., 2012; Fecchio et al., 2018; Lamberts et al., 2009; Lamberts et al., 2011; Otter et al., 2015; Rabbani et al., 2018; Veugelers et al., 2016). Only one study examined active recovery (Boullosa et al., 2014) and reported similar HRR60s reproducibility (ICC = 0.64; CV = 23.3%) to that of passive recovery studies suggesting that recovery protocol may not impact reproducibility of certain HRR indices, although more studies are needed to confirm the impact of recovery protocol on HRR indices.

The present systematic review has several strengths. To the best of our knowledge, it is the first attempt to synthetize literature regarding HRR reproducibility, and its results have important clinical and research applications by supporting the use of HRR60s, HRR120s and HRR300s, and recommending some caution when using T30 and HRRt. To achieve that, a rigorous process of literature review was employed with the search and selection of studies performed in duplicate by two independent investigators without any restriction on the dates of publications. Moreover, the current review analyzed data from a large sample of 326 subjects and the reproducibility results were consistent within each HRR index, excepted for HRRt. Lastly, the review demonstrated that most studies were conducted with similar populations (healthy males) and exercise protocols (walking or running and inactive recovery), and that HRR reproducibility remains poorly explored in other experimental conditions (e.g., women, elderly, patients with chronic diseases, or following water-based exercises), which gives directions for future research.

While the present review evaluated HRR reproducibility with consistent conclusions for the most-commonly used HRR indices, it is important to consider some limitations. First, the systematic literature search was performed only in PubMed, and the inclusion of other databases might have increased the number of studies. However, PubMed database is the most comprehensive and adopted database in health/medical sciences. In addition, the execution of additional manual searches based on papers references lists might have attenuated the potential database bias. Second, it might exist some divergences between studies in ICC and CV formulas used, as well as, in the manner to express HRR values (i.e. absolute HR values, relative values to peak HR or log-transformed). Third, it is important to mention that the selection of HRR index to be employed by clinicians/researchers should also consider other methodological aspects beyond its reproducibility, such as clinical validity, physiological meaning and technical availability.

5. Conclusion

The current systematic review demonstrated that amongst the most-used fast-phase HRR indices, HRR60s had better reproducibility than T30, while amongst the low-phases HRR indices, HRR120s and HRR300s were more reproducible than HRRt. Based upon the current extensive review, HRR60s, HRR120s and HRR300s exhibited good reproducibility for assessing HRR and cardiac autonomic restoration after exercise in research and clinical settings. Clinicians and researchers should employ a degree of caution when using T30 and HRRt as post-exercise indicators of cardiac autonomic restoration due to their reported poorer reproducibility.

Declaration of competing interest

None.

References

Al Haddad, H., Laursen, P.B., Chollet, D., Ahmaidi, S., Buchheit, M., 2011. Reliability of resting and postexercise heart rate measures. Int. J. Sports Med. 32, 598–605. https://doi.org/10.1055/s-0031-1275356.

Arduini, A., Gomez-Cabrera, M.C., Romagnoli, M., 2011. Reliability of different models to assess heart rate recovery after submaximal bicycle exercise. J. Sci. Med. Sport 14, 352–357. https://doi.org/10.1016/j.jsams.2011.02.012.

Atkinson, G., Nevill, A., 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 26, 217–238. https://doi.org/10.2165/00007256-199826040-00002.

Bienias, P., Ciurzyński, M., Chrzanowska, A., Dudzik-Niewiadomska, I., Irzyk, K., Oleszek, K., Kalińska-Bienias, A., Kisiel, B., Tłustochowicz, W., Pruszczyk, P., 2018. Attenuated post-exercise heart rate recovery in patients with systemic lupus erythematosus: the role of disease severity and beta-blocker treatment. Lupus 27, 217–224. https://doi.org/10.1177/0961203317716318.

Bonato, M., Meloni, A., Merati, G., Torre, A.L., Agnello, L., Vernillo, G., 2018. Effect of repeated-sprints on the reliability of short-term parasympathetic reactivation. PLoS One 13, 1–10. https://doi.org/10.1371/journal.pone.0192231.

Bosquet, L., Camelin, F.X., Berthoin, S., 2008. Reliability of postexercise heart rate recovery. Int. J. Sports Med. 29, 238–243. https://doi.org/10.1055/s-2007-965162.

Boullosa, D.A., Barros, E.S., Del Rosso, S., Nakamura, F.Y., Leicht, A.S., 2014. Reliability of heart rate measures during walking before and after running maximal efforts. Int. J. Sports Med. 35, 999–1005. https://doi.org/10.1055/s-0034-1372637.

Buchheit, M., Millet, G.P., Parisy, A., Pourchez, S., Laursen, P.B., Ahmaidi, S., 2008. Supramaximal training and postexercise parasympathetic reactivation in adolescents. Med. Sci. Sports Exerc. 40, 362–371. https://doi.org/10.1249/mss.0b013e31815aa2ee.

Buchheit, M., Mendez-Villanueva, A., Quod, M.J., Poulos, N., Bourdon, P., 2010. Determinants of the variability of heart rate measures during a competitive period in young soccer players. Eur. J. Appl. Physiol. 109, 869–878. https://doi.org/10.1007/s00421-010-1422-x.

Coote, J.H., 2010. Recovery of heart rate following intense dynamic exercise. Exp. Physiol. 95, 431–440. https://doi.org/10.1113/expphysiol.2009.047548.

de Oliveira, T.P., de Alvarenga Mattos, R., da Silva, R.B.F., Rezende, R.A., de Lima, J.R.P., 2013. Absence of parasympathetic reactivation after maximal exercise. Clin. Physiol. Funct. Imaging 33, 143–149. https://doi.org/10.1111/cpf.12009.

de Vet, H.C.W., Terwee, C.B., Knol, D.L., Bouter, L.M., 2006. When to use agreement versus reliability measures. J. Clin. Epidemiol. 59, 1033–1039. https://doi.org/10.1016/j.jclinepi.2005.10.015.

Dogdu, O., Yarlioglues, M., Kaya, M.G., Ardic, I., Oguzhan, N., Akpek, M., Sahin, O., Akyol, L., Kelesoglu, S., Koc, F., Ozdogru, I., Oguzhan, A., 2010. Deterioration of heart rate recovery index in patients with systemic lupus erythematosus. J. Rheumatol. 37, 2511–2515. https://doi.org/10.3899/jrheum.100163.

Dupuy, O., Mekary, S., Berryman, N., Bherer, L., Audiffren, M., Bosquet, L., 2012. Reliability of heart rate measures used to assess post-exercise parasympathetic reactivation. Clin. Physiol. Funct. Imaging 32, 296–304. https://doi.org/10.1111/j.1475-097X.2012.01125.x.

Fecchio, R.Y., Chehuen, M., Oliveira, T.P., Cucato, G., Costa, L., Leicht, A., Ritti-Dias, R.M., Wolosker, N., Forjaz, C.L.D.M., 2018. Reproducibility of heart rate recovery in patients with intermittent claudication. Clin. Physiol. Funct. Imaging 38, 603–609. https://doi.org/10.1111/cpf.12452.

Gupta, M., Bansal, V., Chhabra, S.K., 2013. Abnormal heart rate recovery and chronotropic incompetence on exercise in chronic obstructive pulmonary disease. Chron. Respir. Dis. 10, 117–126. https://doi.org/10.1177/1479972313493097.

Haapaniemi, T.H., Pursiainen, V., Korpelainen, J.T., Huikuri, H.V., Sotaniemi, K.A., Myllylä, V.V., 2001. Ambulatory ECG and analysis of heart rate variability in Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 70, 305–310.

Hopkins, W.G., 2000. Measures of reliability in sports medicine and science. Sports Med. 30, 1–15. https://doi.org/10.2165/00007256-200030050-00006.

Imai, K., Sato, H., Hori, M., Kusuoka, H., Ozaki, H., Yokoyama, H., Takeda, H., Inoue, M., Kamada, T., 1994. Vagally mediated heart rate recovery after exercise is accelerated in athletes but blunted in patients with chronic heart failure. J. Am. Coll. Cardiol. 24, 1529–1535. https://doi.org/10.1016/0735-1097(94)90150-3.

Kanegusuku, H., Silva-Batista, C., Peçanha, T., Nieuwboer, A., Silva Jr., N.D., Costa, L.A., de Mello, M.T., Piemonte, M.E., Ugrinowitsch, C., Forjaz, C.L., 2016. Blunted maximal and submaximal responses to cardiopulmonary exercise tests in patients with Parkinson disea. Arch. Phys. Med. Rehabil. 97, 720–725. https://doi.org/10.1016/j.apmr.2015.12.020.

Lachman, S., Terbraak, M.S., Limpens, J., Jorstad, H., Lucas, C., Scholte op Reimer, W., Boekholdt, S.M., ter Riet, G., Peters, R.J.G., 2018. The prognostic value of heart rate recovery in patients with coronary artery disease: a systematic review and meta-analysis. Am. Heart J. https://doi.org/10.1016/j.ahj.2018.02.008.

Lamberts, R.P., Maskell, S., Borresen, J., Lambert, M.I., 2011. Adapting workload improves the measurement of heart rate recovery adapting workload improves the measurement of heart rate recovery. Int. J. Sports Med. 32, 698–702. https://doi.org/10.1055/s-0031-1275357.

Lamberts, R.P., Swart, J., Noakes, T.D., Lambert, M.I., 2009. A novel submaximal cycle test to monitor fatigue and predict cycling performance. Br. J. Sports Med. 45, 797–804. https://doi.org/10.1136/bjsm.2009.061325.

Mellis, M.G., Ingle, L., Carroll, S., 2014. Variability in heart rate recovery measurements over 1 year in healthy, middle-aged adults. Int. J. Sport 135–138.

Minson, C.T., Halliwill, J.R., Young, T.M., Joyner, M.J., 2000. Influence of the menstrual cycle on sympathetic activity, baroreflex sensitivity, and vascular transduction in young women. Circulation 101, 862–868. https://doi.org/10.1161/01.CIR.101.8.862.

Otter, R., Brink, M., Lamberts, R., Lemmink, K., 2015. A new submaximal rowing test to predict 2,000-m rowing ergometer performance. J. Strength Cond. Res. 29, 2426–2433.

Peçanha, T., Prodel, E., Bartels, R., Nasario-Junior, O., Paula, R.B., Silva, L.P., Laterza, M.C., Lima, J.R.P., 2014a. 24-h cardiac autonomic profile after exercise in sedentary subjects. Int. J. Sports Med. 35, 245–252. https://doi.org/10.1055/s-0033-1349873.

Peçanha, T., Silva-Júnior, N.D., Forjaz, C.L. de M., 2014b. Heart rate recovery: autonomic determinants, methods of assessment and association with mortality and cardiovascular diseases. Clin. Physiol. Funct. Imaging 34, 327–339. https://doi.org/10.1111/cpf.12102.

Peçanha, T., Bartels, R., Brito, L.C., Paula-Ribeiro, M., Oliveira, R.S., Goldberger, J.J., 2016. Methods of assessment of the post-exercise cardiac autonomic recovery: a methodological review. Int. J. Cardiol. https://doi.org/10.1016/j.ijcard.2016.10.057.

Peçanha, T., Rodrigues, R., Pinto, A.J., Sá-Pinto, A.L., Guedes, L., Bonfiglioli, K., Gualano, B., Roschel, H., 2018. Chronotropic incompetence and reduced heart rate recovery in rheumatoid arthritis. J. Clin. Rheumatol. 24, 375–380. https://doi.org/10.1097/RHU.0000000000000745.

Perini, R., Orizio, C., Comandè, A., Castellano, M., Beschi, M., Veicsteinas, A., 1989a. Plasma norepinephrine and heart rate dynamics during recovery from submaximal exercise in man. Eur. J. Appl. Physiol. Occup. Physiol. 58, 879–883.

Perini, R., Orizio, C., Comande, A., Castellano, M., Beschi, M., Veicsteinas, A., Comandè, A., Castellano, M., Beschi, M., Veicsteinas, A., 1989b. Plasma norepinephrine and heart rate dynamics during recovery from submaximal exercise in man. Eur. J. Appl. Physiol. Occup. Physiol. 58, 879–883.

Pierpont, G.L., Stolpman, D.R., Gornick, C.C., 2000. Heart rate recovery post-exercise as an index of parasympathetic activity. J. Auton. Nerv. Syst. 80, 169–174. https://doi.org/10.1016/S0165-1838(00)00090-4.

Qiu, S., Cai, X., Sun, Z., Li, L., Zuegel, M., Steinacker, J.M., Schumann, U., 2017. Heart

rate recovery and risk of cardiovascular events and all-cause mortality: a meta-analysis of prospective cohort studies. J. Am. Heart Assoc. https://doi.org/10.1161/JAHA.117.005505.
Rabbani, A., Kargarfard, M., Twist, C., 2018. Reliability and validity of a submaximal warm-up test for monitoring training status in professional soccer players. J. Strength Cond. Res. 32, 326–333. https://doi.org/10.1519/JSC.0000000000002335.
Racine, N., Blanchet, M., Ducharme, A., Marquis, J., Boucher, J.-M., Juneau, M., White, M., 2003. Decreased heart rate recovery after exercise in patients with congestive heart failure: effect of beta-blocker therapy. J. Card. Fail. 9, 296–302. https://doi.org/10.1054/jcaf.2003.47.
Rosner, B., 2011. Fundamentals of Biostatistics, 7th ed. Brooks/Cole, Cengage Learning, Boston. https://doi.org/10.1002/bimj.4710350205.
Seiler, S., Haugen, O., Kuffel, E., 2007. Autonomic recovery after exercise in trained athletes: intensity and duration effects. Med. Sci. Sports Exerc. 39, 1366–1373. https://doi.org/10.1249/mss.0b013e318060f17d.
Szklo, R., Nieto, F., 2000. Epidemiology: Beyond the Basis. Aspen Publications, Gaithersburg, MD. https://doi.org/10.1093/oxfordjournals.epirev.a017974.
Tonello, L., Reichert, F.F., Oliveira-Silva, I., Del Rosso, S., Leicht, A.S., Boullosa, D.A., 2016. Correlates of heart rate measures with incidental physical activity and cardiorespiratory fitness in overweight female workers. Front. Physiol. 6, 1–11. https://doi.org/10.3389/fphys.2015.00405.
Tulumen, E., Khalilayeva, I., Aytemir, K., Ergun Baris Kaya, F.E.S.C., Sinan Deveci, O., Aksoy, H., Kocabas, U., Okutucu, S., Tokgozoglu, L., Kabakci, G., Ozkutlu, H., Oto, A., 2011. The reproducibility of heart rate recovery after treadmill exercise test. Ann. Noninvasive Electrocardiol. 16, 365–372. https://doi.org/10.1111/j.1542-474X.2011.00464.x.
Veugelers, K., Naughton, G., Duncan, C., Burgess, D., Graham, S., 2016. Validity and reliability of a submaximal intermittent running test in elite Australian football players. J. Strength Cond. Res. 30, 3347–3353. https://doi.org/10.1519/JSC.0000000000001441.
Yawn, B.P., Ammar, K.A., Thomas, R., Wollan, P.C., 2003. Test-retest reproducibility of heart rate recovery after treadmill exercise. Ann. Fam. Med. 1, 236–241. https://doi.org/10.1370/afm.37.