Journal of Human Performance in Extreme Environments

Volume 17 | Issue 1 Article 5

Published online: 5-5-2022

Errors of Measurement for Blood Parameters and Physiological and Performance Measures After the Decay of Short-Term Heat Acclimation

Andrew T. Garrett
University of Hull, a.garrett@hull.ac.uk

Nancy J. Rehrer
University of Otago

Mark J. Patterson
Defence Science Technology Organisation (DSTO), Australia

Andrew J. Simpson
University of Hull

James D. Cotter
University of Otago

Recommended Citation

Garrett, Andrew T.; Rehrer, Nancy J.; Patterson, Mark J.; Simpson, Andrew J.; and Cotter, James D. (2022) "Errors of Measurement for Blood Parameters and Physiological and Performance Measures After the Decay of Short-Term Heat Acclimation," *Journal of Human Performance in Extreme Environments*: Vol. 17 : Iss. 1, Article 5.
DOI: 10.7771/2327-2937.1140
Available at: https://docs.lib.purdue.edu/jhpee/vol17/iss1/5

Errors of Measurement for Blood Parameters and Physiological and Performance Measures After the Decay of Short-Term Heat Acclimation

Cover Page Footnote

Special thanks are given to the participants in this study and the technical expertise provided by Mrs R. McKay and Miss D. Wilson. This work was supported by grants from the Australian Defence Science Technology Organisation (DSTO) and the School of Physical Education, Sport and Exercise Science, University of Otago, New Zealand.

Errors of Measurement for Blood Parameters and Physiological and Performance Measures After the Decay of Short-Term Heat Acclimation

Andrew T. Garrett[1,2], Nancy J. Rehrer[2], Mark J. Patterson[3], Andrew J. Simpson[1], and James D. Cotter[2]

[1]*University of Hull, UK*
[2]*University of Otago, New Zealand*
[3]*Defence Science Technology Organisation (DSTO), Australia*

Abstract

Introduction: It is important to determine the accuracy of measurements relative to potential treatment effects, with time intervals between tests. Purpose: The aim of this study was to assess the error of measurement for blood parameters and physiological and performance measures after the decay of short-term heat acclimation. Methods: Ten trained males (mean ± SD: age 28 ± 7 years; body mass 74.6 ± 4.4 kg; 4.26 ± 0.37 L min^{-1}; peak power output (PPO) 329 ± 42 W) completed an exercising heat stress test at baseline, the second day after acclimation, and then during decay at 1, 2, 3, and 5–6 weeks. CoV (95% CI), SE (95% CI), and Pearson's *r* were used for analysis of blood volume (blood, plasma, red cell volume, mean hemoglobin mass); plasma (aldosterone, arginine vasopressin, total protein, albumin, sodium); physiological measures (rectal temperature, cardiac frequency); and performance (exercise performance capacity, PPO). Results: The CoV (95% CI), SE (95% CI), and *r* with a 1-week interval for blood volume was 2.3% (1.6 to 4.3; 1.9 [1.3 to 3.4 mL kg^{-1}]; $r = 0.93$; $n = 10$). After 2 weeks and 5–6 weeks this had increased to 4.9% (3.4 to 9.3; 3.8 [2.6 to 7.0 mL kg^{-1}]; $r = 0.76$; $n = 9$) and 5.5% (3.6 to 12.8; 4.5 [2.9 to 10.0 mL kg^{-1}]; $r = 0.65$; $n = 7$) respectively. Conclusions: Blood volume and physiological measures demonstrated the least error one week apart but error increased thereafter. Plasma concentrations and performance markers had the greatest error with repeat measures after one week. Therefore, for greater reliability and low measurement error, measures should be taken no more than one week apart in repeated experimentation.

Keywords: blood volume, standard error, reliability

Introduction

Physiological and biological outcome measures are a cornerstone of sport and exercise science research. Humans often exhibit a wide range of responses to environmental stressors (Hecksteden et al., 2015, 2018; Hopkins, 2015). This variability is a composite of individual physiological differences and measurement errors. A comprehensive understanding of the variability and reproducibility of outcome measures is important for research design (Hayden et al., 2004) and to better understand confidence in the data generated. Providing details regarding the range of reliability and validity statistics in combination with descriptive statistics has been proposed (Bland & Altman, 1986, 2003; Brunton et al., 2000; Currell & Jeukendrup, 2008).

Measurement error can be determined from the typical (standard) error of measurement expressed as a coefficient of variation (CoV) (Petersen et al., 2001; Poulsen et al., 1998) and is typically used for reliability of dependent measures (Hopkins et al., 1999, 2009, 2001). It is equivalent to the standard deviation of an individual's repeated measurements, expressed as a percentage of the mean test score. It is also informative to report an absolute measure of variability, such as typical error or standard error (SE) (Hopkins et al., 2001).

The CoV has been reported as a useful measure of reliability in sport and exercise (Hopkins, 2000; Hopkins et al., 2001). A reliable method will have small measurement error and is a prerequisite of validity (Johnstone et al., 2012a, 2012b, 2012c), the extent to which a test actually measures what it intends to measure (Gore et al., 2005). The within-subject variance and the measurement reproducibility used for repeated measures are useful for establishing the internal quality control of a study (Hopkins, 2000). We have previously reported the reliability of blood, plasma, red cell volume, and mean hemoglobin mass as CoV (Garrett et al., 2009).

The measurement of blood volume is an important marker of adaptation to heat stress conditions (Sawka et al., 2000), yet the reliability of this and other physiological outcomes over repeated measures has received little interest in the literature. The reliability of the carbon monoxide (CO) rebreathing technique for the measurement of blood volume (a technique used in the current study) is ~2.5%, from findings in a meta-analysis of 346 estimates of error measurement for several blood parameters (Gore et al., 2005). Adjusted for 1 day between trials and expressed as CoV, they reported mean errors for mean hemoglobin mass (2.2%; 90% confidence interval 1.4 to 3.5%) and red cell volume (6.7%; 3.4 to 14%).

http://dx.doi.org/10.7771/2327-2937.1140

Furthermore, Gore and colleagues (2005) reported that blood measures even with the smallest error, e.g., the CO rebreathing technique, showed some increase in error with increasing time between trials. For example, Eckfeldt and colleagues (1994) reported within-subject variability expressed as CoV, 14 days apart, in serum for sodium (0.6%), albumin (2.8%), and total protein (2.5%). They recorded the CoV for the measurement techniques used for the same samples analyzed at least one week apart for sodium (1.0%), albumin (2.6%), and total protein (2.4%). The within-subject variability may strongly influence clinical interpretation of repeated laboratory tests and increase internal quality control (Eckfeldt et al., 1994).

An acceptable reliability boundary for CoV ($<$10%) for exercise performance capacity has been cited in some papers, though this is not accepted unanimously in the literature (Atkinson & Nevill, 1998; Currell & Jeukendrup, 2008). Therefore, a reliability study must be designed to obtain an acceptable precision of within-subject variation and should take into account the possibility of greater variability in a sample of participants with a wider range of ability (Hopkins et al., 1999). Furthermore, the difference between within-subject variation and repeated measures should be shown as confidence intervals, as they are useful for evaluating the limits for internal quality control of a study (Hopkins, 2000; Hopkins et al., 2009).

There is often measurement error in repeated exercise performance tests associated with biological test–retest variation and the type of assessment undertaken. Therefore, reliability of power in tests of physical performance affects the precision for assessment of study participants and random error in a performance test should be minimized by the choice of test (Hopkins et al., 2001). For example, performance tests that are based on physiological measures (e.g., $\dot{V}O_{2\,peak}$) have been shown to produce a random error of ~2–3% in the measure of power output (Paton & Hopkins, 2001) and this error is a mixture of ergometer error and biological test–retest variation.

In summary, the determination of the error of measurement over time, within one longitudinal study, for blood volume, plasma, physiological measures, and performance measures has received limited attention in the literature. Therefore, the aim of this work was to calculate the error of measurement for blood volume, plasma, physiological measures, and performance measures after the decay of short-term (5-day) heat acclimation, at intervals of 7–9, 14–16, 22–23, and 41–49 days.

Method

Participants

Participants were 10 volunteers from the University of Otago. They were male, moderately well trained, and in the age range of 18–37 years (mean $\pm$ SD: age 28 $\pm$ 7 years; body mass 74.6 $\pm$ 4.4 kg; $\dot{V}O_{2\,peak}$ 4.26 $\pm$ 0.37 L min^{-1}; and peak power output (PPO) 329 $\pm$ 42 W). This determination of moderately well-trained status was based on the $\dot{V}O_{2\,peak}$ data and all participants actively trained 3–4 days per week. All participants provided written informed consent that detailed the purposes of the study and were in good health and asymptomatic from cardiovascular dysfunction. The University of Otago Human Ethics Committee granted ethical approval (Approval #02/035).

Experimental Design and Overview

Ten participants repeated the same exercising heat stress test (HST) administered one week before and the second day after acclimation and then on days 9, 16, 23, and 42–49. This ensured that the reliability of dependent measures calculated in this study was free from the effectiveness of the heat acclimation protocol undertaken (Garrett et al., 2011). The data from this study have previously been reported by the authors in studies to investigate the induction and decay of heat acclimation for moderately trained athletes (Garrett et al., 2009, 2014) and highly trained athletes (Garrett et al., 2012). Therefore, this work analyses secondary data. In brief, the HST involved participants cycling at 40% PPO for a duration of 90 minutes (35℃, 60% RH, with a wind speed $<$ 0.5 m s^{-1}). They then rested for 10 minutes before commencing a ramp protocol (2% PPO each 30 s) to volitional fatigue, or a rectal temperature (T_{re}) $\geqslant$39.5℃. Participants' plasma, physiological, and performance parameters were measured at rest and in response to the HST. Heat acclimation consisted of 90-minute exposure on 5 consecutive days (40℃, 60% RH), using controlled hyperthermia (T_{re} of 38.5℃), with no fluid intake allowed during acclimation bouts.

Participants were fully informed (verbally and in writing) of the pre-experimental and daily procedures, and asked to refrain from strenuous exercise for 24 hours prior to each HST. They were advised to consume a meal high in carbohydrate and remain fully hydrated the evening before and immediately after the HST. The detail for standardized nourishment was based on the feedback from a 3-day nutritional diary, carried out as part of the familiarization procedures. This involved the participants reporting in a diary their typical timing and food consumption over a 3-day period. The participants were asked to use this for each HST throughout the study to ensure consistency in nutritional and fluid intake. They were asked to refrain from alcohol and caffeine consumption on the day of the HSTs and days 1 to 5 of the acclimation sessions. Information from the physical activity and nutrition diaries was not analyzed in the study and was used to provide consistency for each of the subjects prior to each visit to the laboratory.

On arrival at the environmental chamber before the HSTs participants consumed 250 mL of a 4% carbohydrate

(CHO) fluid to increase the likelihood that they would begin in a euhydrated state (Armstrong et al., 1994, 1997; Popowksi et al., 2001). Fluid osmolality was between 240 and 270 mmol kg^{-1}, as measured using a vapor pressure osmometer (Model 5520 Vapro, Wescor, USA). At approximately 15-minute intervals during HSTs, 150 mL of 4% CHO solution was consumed, for a total consumption of 900 mL in 90 minutes. To monitor hydration status, whole body sweat rate (mL h^{-1}) was calculated using measurements of pre- and post-exercise nude body mass using scales ($\pm$20 g; Wedderburn Scales, Teraka Seiko, Tokyo, Japan), calibrated to an accuracy of 0.1 kg with calibrated masses. Sweat was toweled off before post-exercise body mass was recorded. Furthermore, urine samples were obtained before and after HST and acclimation bouts. Using fresh urine samples, urine specific gravity and urine color were measured using a calibrated refractometer (Uricon-N, urine specific gravity refractometer, Atago Co., Tokyo, Japan) and urine color chart (Armstrong et al., 1998, 1994) respectively. Urine volume was recorded and urine osmolality was analyzed after the experiment. Nominal fluid replacement (100 mL) was given immediately before the acclimation bouts to limit participants' perception of fluid deprivation and no fluid was consumed during the bout itself (Garrett et al., 2009, 2012, 2014).

Blood volume was measured one week before the pre-acclimation HST and the first day after acclimation, followed by days 8, 15, 22, and 41–48 using the CO rebreathing technique (Burge & Skinner, 1995). This technique uses mean values for total hemoglobin [Hb], pre- and post-CO-rebreathing [HbCO], and hematocrit [Hct] to calculate absolute blood and plasma volumes. Fluid regulatory measures (aldosterone and arginine vasopressin [AVP]) were recorded on day one and day five of the acclimation regime.

Data Analysis

Reliability of dependent measures was calculated using the CoV (95% CI), SE (95% CI), and Pearson product moment correlation (r) of the corresponding dependent variable measures. Resting blood volume (BV), plasma volume (PV), red cell volume (RCV), and mean hemoglobin mass (M_{Hb}) were measured at baseline followed by days 8, 15, 22, and 41–48. Plasma, physiological, and performance measures at rest and end-exercise were calculated at intervals from 9 to 16 days, 9 to 23 days, and baseline to 42–49 days respectively. Plasma measures were aldosterone $[aldo]_p$, AVP $[AVP]_p$, total protein $[TP]_p$, albumin $[alb]_p$, and sodium $[Na^+]_p$. The physiological measures were rectal temperature (T_{re}) and cardiac frequency (f_c). Performance measures were exercise performance capacity and PPO. The reliability of dependent measures protocol is shown in Figure 1.

Results

All 10 participants completed the baseline and 2- and 9-day HSTs. Seven participants completed the post 16, 23, and 42–49 days. Blood and plasma volume was measured from all ten participants for baseline, 1, 8, and 14 days. Nine participants were measured for 22 days and seven at 41–48 days.

Reliability of Dependent Measures

In the present study, it had been established that by 9 days and thereafter post-acclimation, the measured dependent variables under investigation had returned to baseline pre-acclimation values, apart from the f_c and exercise performance variables. Therefore, the reliability of dependent measures after 7, 14, and 42–49 days was calculated using the CoV (95% CI), SE (95% CI), and r of the corresponding dependent variable measures from baseline versus 9 to 16, 9 to 23, and 42–49 days post-acclimation respectively. The SE reflects the change in two variables. Because the f_c and exercise performance variables had a later return to baseline after 16 days post-acclimation, the reliability of dependent measures for these variables, after a 7-day interval, was calculated on the corresponding dependent variable measures from baseline versus 16 to 23 days post-acclimation. As a consequence there are no 14-day reliability data available for f_c and exercise performance, but baseline to 42–49 days interval was calculated the same as for the other dependent measures.

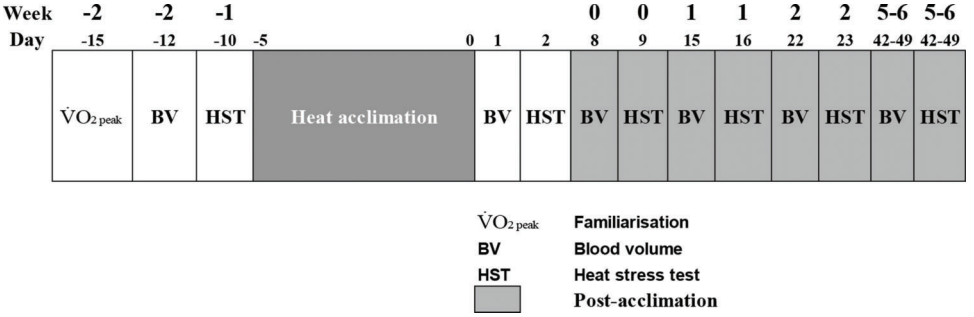

Figure 1. Experimental design for the reliability of dependent measures at 1, 2, and 5–6 weeks relative to 0-week baseline post-acclimation.

Blood Volume

Reliability of BV, PV, RCV, and M_{Hb} at intervals of 7, 14, and 42–49 days relative to baseline is shown in Table 1.

Low CoV and SE and strong relationships between trials were recorded for BV, PV, RCV, and M_{Hb} when measured 7 days apart. However, CoVs and SEs for BV, PV, RCV, and M_{Hb} were substantially higher when measured after 14 days and especially 42–49 days.

Plasma Measures

Reliability of plasma measures post-acclimation was determined at intervals of 7 (Table 2), 14 (Table 3), and 42–49 (Table 4) days relative to baseline. It is important to take into account the intra-assay CoVs for $[aldo]_p$ (9.9%), $[AVP]_p$ (5.6%), and $[cortisol]_p$ (12.1%) when determining the reliability of the fluid-regulatory and stress hormones.

Physiological Measures

At rest and after 90 minutes of exercise, both T_{re} and f_c recorded low CoV and SE, after a 7-day interval. Furthermore, there was moderate to strong relationships within all of these parameters, between repeated reliability trials. Similar findings were reported for CoVs and SEs of T_{re} and f_c at rest and after 90 minutes of exercise, for 14 days. However, greater CoV and SE were evident, relative to baseline, for T_{re} and f_c after 90 minutes of exercise, for 42–49 days (Table 5).

Performance Measures

Reliability of exercise performance and power output was measured at intervals of 7 and 42–49 days relative to baseline. Results are shown in Table 6.

Table 1
Reliability of blood volume, plasma volume, red cell volume and mean hemoglobin mass at intervals of 7, 14, and 42–49 days relative to baseline.

Dependent variable and time interval	*n*	CoV (95% CI) (%)	SE (95% CI)	*r*
7 days				
BV	10	2.3 (1.6 to 4.3)	1.9 (1.3 to 3.4 mL kg^{-1})	0.93
PV	10	2.9 (2.1 to 5.7)	1.3 (0.9 to 2.4 mL kg^{-1})	0.92
RCV	10	2.8 (2.0 to 5.3)	1.1 (0.7 to 2.0 mL kg^{-1})	0.89
M_{Hb}	10	2.8 (2.0 to 5.3)	0.4 (0.3 to 0.7 mmol)	0.86
14 days				
BV	9	4.9 (3.4 to 9.3)	3.8 (2.6 to 7.0 mL kg^{-1})	0.76
PV	9	5.2 (3.9 to 10.8)	2.4 (1.7 to 4.4 mL kg^{-1})	0.74
RCV	9	4.3 (3.0 to 8.2)	1.5 (1.1 to 2.8 mL kg^{-1})	0.80
M_{Hb}	9	4.6 (3.2 to 9.3)	0.6 (0.4 to 1.1 mmol)	0.75
42–49 days				
BV	7	5.5 (3.6 to 12.8)	4.5 (2.9 to 10.0 mL kg^{-1})	0.65
PV	7	7.2 (4.7 to 17.1)	3.2 (2.1 to 7.0 mL kg^{-1})	0.57
RCV	7	4.3 (2.8 to 9.9)	1.6 (1.1 to 3.6 mL kg^{-1})	0.77
M_{Hb}	7	5.0 (3.2 to 13.1)	0.7 (0.4 to 1.6 mmol)	0.61

Note. CoV (95% CI), coefficient of variation and 95% confidence interval; SE (95% CI), standard error and 95% confidence interval; *r*, Pearson product moment correlation.

Table 2
Reliability of plasma measures at 7 days relative to baseline.

Dependent variable	*n*	CoV (95% CI) (%)	SE (95% CI)	*r*
$[aldo]_p$ rest	7	76.4 (63.7 to 438.4)	95.3 (61.4 to 209.7 pg mL^{-1})	0.23
$[aldo]_p$ 90 minutes	7	20.5 (14.1 to 57.1)	77.7 (50.1 to 171.1 pg mL^{-1})	0.84
$[AVP]_p$ rest	7	24.9 (17.4 to 72.9)	1.2 (0.7 to 2.5 pmol L^{-1})	0.92
$[AVP]_p$ 90 minutes	7	42.7(31.7 to 155.9)	2.9 (1.9 to 6.5 pmol L^{-1})	0.69
$[cortisol]_p$ rest	7	26.6 (18.7 to 79.8)	2.5 (1.6 to 5.5 μg dL^{-1})	0.40
$[cortisol]_p$ 90 minutes	7	31.9 (22.9 to 102.1)	4.6 (2.9 to 10.0 μg dL^{-1})	0.45
$[TP]_p$ rest	7	3.9 (2.5 to 8.9)	0.3 (0.2 to 0.6 mg mL^{-1})	0.81
$[TP]_p$ 90 minutes	7	4.6 (3.0 to 10.7)	0.4 (0.2 to 0.8 mg mL^{-1})	0.54
$[alb]_p$ rest	7	5.4 (3.6 to 12.7)	0.5 (0.3 to 0.8 mg mL^{-1})	0.54
$[alb]_p$ 90 minutes	7	2.8 (1.8 to 6.3)	0.1 (0.1 to 0.3 mg mL^{-1})	0.56
$[Na^+]_p$ rest	7	2.1 (1.3 to 5.3)	2.9 (1.8 to 7.1 mmol L^{-1})	0.18
$[Na^+]_p$ 90 minutes	7	0.4 (0.2 to 1.0)	0.6 (0.4 to 1.4 mmol L^{-1})	0.95

Note. CoV, coefficient of variation; 95% CI of CoV, 95% confidence interval of coefficient of variation; *r*, Pearson product moment correlation.

Table 3
Reliability of plasma measures at 14 days relative to baseline.

Dependent variable	*n*	CoV (95% CI) (%)	SE (95% CI)	*r*
$[aldo]_p$ rest	7	49.8 (37.8 to 199.4)	99.4(64.1 to 219.0 pg mL^{-1})	0.67
$[aldo]_p$ 90 minutes	7	23.5 (16.4 to 67.9)	80.2 (51.7 to 176.5 pg mL^{-1})	0.87
$[AVP]_p$ rest	7	59.1 (46.4 to 267.7)	3.7 (2.4 to 8.2 pmol L^{-1})	0.67
$[AVP]_p$ 90 minutes	7	60.7 (47.9 to 260.8)	4.3 (2.8 to 9.4 pmol L^{-1})	0.69
$[cortisol]_p$ rest	7	22.4 (15.5 to 63.8)	2.4 (1.5 to 5.2 μg dL^{-1})	0.65
$[cortisol]_p$ 90 minutes	7	20.0 (13.8 to 55.4)	2.0 (1.3 to 4.4 μg dL^{-1})	0.87
$[TP]_p$ rest	7	4.8 (3.2 to 11.2)	0.3 (0.2 to 0.7 mg mL^{-1})	0.57
$[TP]_p$ 90 minutes	7	1.9 (1.2 to 4.2)	0.1 (0.1 to 0.3 mg mL^{-1})	0.92
$[alb]_p$ rest	7	5.7 (3.7 to 13.3)	0.2 (0.2 to 0.5 mg mL^{-1})	0.47
$[alb]_p$ 90 minutes	7	3.5 (2.3 to 8.1)	0.2 (0.1 to 0.4 mg mL^{-1})	0.52
$[Na^+]_p$ rest	7	1.9 (1.2 to 4.7)	2.6 (1.6 to 6.3 mmol L^{-1})	0.42
$[Na^+]_p$ 90 minutes	7	0.8 (0.5 to 1.9)	1.1 (0.7 to 2.7 mmol L^{-1})	0.81

Note. CoV, coefficient of variation; 95% CI of CoV, 95% confidence interval of coefficient of variation; *r,* Pearson product moment correlation.

Table 4
Reliability of plasma measures at 42–49 days relative to baseline.

Dependent variable	*n*	CoV (95% CI) (%)	SE (95% CI)	*r*
$[aldo]_p$ rest	5	39.0 (24.7 to 327.5)	72.1 (40.1 to 269.0 pg mL^{-1})	0.94
$[aldo]_p$ 90 minutes	5	22.2 (14.2 to 89.4)	82.7 (49.6 to 237.8 pg mL^{-1})	0.83
$[AVP]_p$ rest	5	5.9 (3.1 to 44.4)	0.1 (0.1 to 0.6 pmol L^{-1})	0.99
$[AVP]_p$ 90 minutes	5	51.7 (34.1 to 588.4)	2.7 (1.5 to 10.0 pmol L^{-1})	0.46
$[cortisol]_p$ rest	5	25.0 (16.2 to 105.3)	1.7 (-2.3 to 5.6 μg dL^{-1})	0.32
$[cortisol]_p$ 90 minutes	5	15.8 (9.9 to 57.4)	2.7 (1.6 to 7.6 μg dL^{-1})	0.55
$[TP]_p$ rest	5	4.1 (2.5 to 12.3)	2.9 (1.7 to 8.3 mg mL^{-1})	0.55
$[TP]_p$ 90 minutes	5	1.1 (0.7 to 3.2)	0.9 (0.5 to 2.4 mg mL^{-1})	0.89
$[alb]_p$ rest	5	8.3 (5.1 to 26.9)	0.4 (0.2 to 1.1 mg mL^{-1})	0.05
$[alb]_p$ 90 minutes	5	5.7 (3.5 to 17.8)	0.3 (0.2 to 0.7 mg mL^{-1})	0.60
$[Na^+]_p$ rest	5	2.3 (1.4 to 6.9)	3.3 (2.0 to 9.4 mmol L^{-1})	-0.49
$[Na^+]_p$ 90 minutes	5	0.4 (0.3 to 1.3)	0.6 (0.4 to 1.8 mmol L^{-1})	0.91

Note. CoV, coefficient of variation; 95% CI of CoV, 95% confidence interval of coefficient of variation; *r*, Pearson product moment correlation.

Table 5
Reliability of rectal temperature and cardiac frequency measurements at intervals of 7, 14, and 42–49 days relative to baseline.

Dependent variable and time interval	*n*	CoV (95% CI) (%)	SE (95% CI)	*r*
7 days				
T_{re} rest	7	0.7 (0.4 to 1.5)	0.3 (0.2 to 0.6℃)	0.36
T_{re} 90 minutes	7	0.4 (0.2 to 0.8)	0.1 (0.1 to 0.3℃)	0.89
f_c 90 minutes	7	5.6 (3.7 to 13.1)	7 (5 to 13 b min^{-1})	0.91
14 days				
T_{re} rest	7	0.4 (0.3 to 0.9)	0.2 (0.1 to 0.3℃)	0.80
T_{re} 90 minutes	7	0.4 (0.3 to 0.9)	0.2 (0.1 to 0.4℃)	0.90
42–49 days				
T_{re} rest	6	0.4 (0.3 to 1.1)	0.2 (0.1 to 0.4℃)	0.82
T_{re} 90 minutes	6	1.4 (0.9 to 2.7)	0.5 (0.4 to 1.0℃)	0.20
f_c 90 minutes	6	9.1 (5.9 to 25.1)	13 (8 to 33 b min^{-1})	0.46

Note. CoV (95% CI), coefficient of variation and 95% confidence interval; SE (95% CI), standard error and 95% confidence interval; *r*, Pearson product moment correlation.

Discussion

In this study we calculated the variation of measurement for blood volume, plasma, physiological, and performance measures following short-term heat acclimation. The CoVs and SEs tended to be higher for the plasma (Tables 2–4) and exercise performance measures (Table 6), especially the plasma hormone concentrations of aldosterone, AVP, and cortisol. There was considerably lower variability in the blood volume parameters (Table 1) and physiological measures of T_{re} and f_c (Table 5). Generally, measures even with the smallest error showed some increase in error with increasing time between trials. For example, the CoV for BV was 2.3% with a 7-day interval between repeated tests.

Table 6
Reliability of performance measurements at intervals of 7 and 42–49 days relative to baseline.

Dependent variable and time interval	*n*	CoV (95% CI) (%)	SE (95% CI)	*r*
7 days				
Time to exhaustion	7	13.7 (9.2 to 35.3)	93.0 (59.9 to 204.7 s)	0.73
Power output	7	7.9 (5.2 to 19.1)	22.8 (14.7 to 50.3 W)	0.58
42–49 days				
Time to exhaustion	5	17.1 (10.8 to 63.3)	99.3 (59.5 to 285.3 s)	−0.40
Power output	5	5.3 (3.2 to 16.3)	13.8 (8.3 to 39.8 W)	0.92

Note. CoV (95% CI), coefficient of variation and 95% confidence interval; SE (95% CI), standard error and 95% confidence interval; *r*, Pearson product moment correlation.

However, after 14 and 42–49 days, this had increased to 4.9% and 5.5% respectively (Table 1).

Low 7-day measurement errors (CoVs and SEs) were recorded for BV, PV, RCV, and M_{Hb} indicating that the CO rebreathing technique used in this study (Burge & Skinner, 1995) was reliable. This is supported by data presented in the meta-analysis on blood volume by Gore and colleagues (2005). Furthermore, the CoV for the measurement of BV in the 7-day reliability trial (2.3%; $n = 10$) was lower than the CoV reported in pilot testing (2.9%; $n = 6$).

In the present study, the low measurement error recorded for RCV (2.8%; 95% CI 2.0 to 5.3%) and M_{Hb} (2.8; 2.0 to 5.3%) compares favorably with similar findings reported by Gore and colleagues (2005). Indeed, Gore and colleagues (2005) found that blood measures even with the smallest error, such as the CO rebreathing technique, showed some increase in error with increasing time between trials. This is further illustrated in the present study, with progressive increases in CoV and SEs for BV, PV, RCV, and M_{Hb} after 7, 14, and 42–49 days. For example, the treatment effect for PV that would need to be evident to reveal a difference in experimentation is after 7 (3.0%), 14 (5.3%) and 42–49 (7.3%) days apart. This has implications for monitoring and interpreting change in blood parameters. For example, a PV measure with a large error of 7.3% (42–49 days apart) is only useful for determining a change of that magnitude and not accurate enough in a clinical setting (Gore et al., 2005). Therefore, an important application of the current research is the accuracy of measurements relative to potential treatment effects with various time intervals between tests.

There tended to be higher CoVs in the plasma and exercise performance measures. This may reflect the large biological within-person and methodological variability of plasma measures (Eckfeldt et al., 1994). Furthermore, to the authors' knowledge, information on the reproducibility in the heat of the fluid-electrolyte hormones (aldosterone and AVP), under investigation in the present study, appears incomplete in the literature and requires further attention. The performance test used in this study, based on physiological measures (e.g., $\dot{V}O_{2\,peak}$), has been shown to produce a random error that includes a mixture of ergometer error and biological test–retest variation (Hopkins et al., 1999; Paton & Hopkins, 2001). Therefore, the random error in a performance test should be minimized by the choice of test reported by Hopkins and colleagues (2001).

This work demonstrated low between-trial variation of exercise f_c and T_{re} and this is similar to results of previous reliability investigations in hot (Hayden et al., 2004) and cool (Becque et al., 1993; Wilmore et al., 1998) conditions. Furthermore, the expected low variation in T_{re} was similar to findings in previous reports (Barnett & Maughan, 1993; Hayden et al., 2004). However, to the authors' knowledge, quantifying the reproducibility of exercise f_c and T_{re} in the heat has surprisingly received limited attention in the literature. Of the information available on the reliability of repeated measures in heat stress, the reproducibility of f_c, T_{re}, and $\bar{T}_{sk}$ has been determined by Hayden and colleagues (2004). Seven male participants cycled for 60 minutes at 29.5% peak work rate, on three separate occasions, one week apart, in hot environmental conditions (36℃, 60% RH). They found that there were no between-trial differences evident for f_c, T_{re}, and $\bar{T}_{sk}$ reporting mean CoVs of $0.3 \pm 0.2\%$, $0.7 \pm 0.6\%$ and $3.9 \pm 1.7\%$ respectively. However, they reported that f_c tended to be greater in the first trial ($P = 0.08$) and suggested that although this may imply a heat acclimation effect, they concluded that it was more likely that the reduced f_c reflected greater familiarity of the participants with the experimental procedures. They further reported that the observed CoV for f_c reliability during exercise in a hot environment (Hayden et al., 2004) was similar to those reported in cooler conditions (Becque et al., 1993; Wilmore et al., 1998). Furthermore, Wilmore and colleagues (1998) found high reproducibility of cardiovascular variables (e.g., f_c), from four different laboratories, using the same techniques and equipment, as a CoV below 9% and interclass correlations over 0.80. These variables were measured with the same participants, on three separate occasions, separated by at least 4–10 days, using a submaximal protocol (Bouchard et al., 1995). Under heat stress conditions (35℃, 60% RH), Barnett and Maughan (1993) recorded an unaltered between-trial T_{re} response, separated at weekly intervals. They further suggested that no PV expansion had occurred between trials and there was no significant difference in

serum Na^+ concentration, although no CoVs were reported for all these variables.

There are several limitations worthy of discussion here. Firstly, participants were male and moderately trained; therefore, generalizations to females and populations who are either highly trained or not trained should be made with caution. Further, whilst the sample size included herewith was modest, it is in line with our previous published research in the field (Garrett et al., 2009, 2012, 2014) making these data a valid reference sample. Due to the small sample size, however, a degree of caution should be made when interpreting the inferential statistics (i.e., Pearson correlation coefficients). To begin to address the issue of small sample size in this field of research, the authors have opted to make their data open (supplementary material). We propose future research is required for both males and females, across a range of athletic abilities and conducted in a larger cohort.

Conclusion

An important application of this work is for researchers interested in the accuracy of measurements relative to potential treatment effects with various time intervals between tests. For example, it supports the use of the CO rebreathing technique, for monitoring changes in blood volume parameters of moderately trained participants, taken ideally no more than seven days apart. This is further supported in the present study by the technique's relative ease of handling and the safety for participants with repeated measures. Therefore, the CO dilution technique (Burge & Skinner, 1995) used in this study gave a reliable measure of blood volume repeated after seven days with moderately trained participants but showed some increase in measurement error with increasing time between trials thereafter.

Acknowledgments

Special thanks are given to the participants in this study and the technical expertise provided by Mrs R. McKay and Miss D. Wilson. This work was supported by grants from the Australian Defence Science Technology Organisation (DSTO) and the School of Physical Education, Sport and Exercise Science, University of Otago, New Zealand.

References

Armstrong, L. E., Herrera Soto, J. A., Hacker, F. T., Casa, D. J., Kavouras, S. A., & Maresh, C. M. (1998). Urinary indices during dehydration, exercise, and rehydration. *International Journal of Sport Nutrition*, *8*, 345–355. https://doi.org/10.1123/ijsn.8.4.345

Armstrong, L. E., Maresh, C. M., Castellani, J. W., Bergeron, M. F., Kenefick, R. W., LaGasse, K. E., & Riebe, D. (1994). Urinary indices of hydration status. *International Journal of Sport Nutrition*, *4*, 26–279. https://doi.org/10.1123/ijsn.4.3.265

Armstrong, L. E., Maresh, C. M., Gabaree, C. V., Hoffman, J. R., Kavouras, S. A., Kenefik, R. W., … Ahlquist, L. E. (1997). Thermal and circulatory responses during exercise: Effects of hypohydration, dehydration, and water intake. *Journal of Applied Physiology*, *86*(6), 2028–2035. https://doi.org/10.1152/jappl.1997.82.6.2028

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

Barnett, A., & Maughan, R. (1993). Response of unacclimatized males to repeated bouts of exercise in the heat. *British Journal of Sports Medicine*, *27*(1), 39–44. https://doi.org/10.1136/bjsm.27.1.39

Becque, M. D., Katch, V. L., Marks, C., & Dyer, R. (1993). Reliability and within-subject variability of VE, VO_2, heart rate, and blood pressure during submaximum cycle ergometery. *International Journal of Sports Medicine*, *14*, 220–223. https://doi.org/10.1055/s-2007-1021167

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*, 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analysis of measurement studies. *Ultrasound in Obstetrics & Gynecology*, *22*, 85–93. https://doi.org/10.1002/uog.122

Bouchard, C., Leon, A. S., Rao, D. C., Skinner, J. S., Wilmore, J. H., & Gagnon, J. (1995). The HERITAGE family study: Aims, design and measurement protocol. *Medicine and Science in Sports and Exercise*, *27*, 721–729.

Brunton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, *86*, 94–99. https://doi.org/10.1016/S0031-9406(05)61211-4

Burge, C. M., & Skinner, S. L. (1995). Determination of hemoglobin mass and blood volume with CO: Evaluation and application of a method. *Journal of Applied Physiology*, *79*(2), 623–631. https://doi.org/10.1152/jappl.1995.79.2.623

Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance *Sports Medicine*, *38*, 297–316. https://doi.org/10.2165/00007256-200838040-00003

Eckfeldt, J. H., Chambless, A. E., & Shen, Y. L. (1994). Short-term, within-person variability in clinical chemistry test results. *Archives of Pathology & Laboratory Medicine*, *118*, 496–500.

Garrett, A. T., Creasy, R., Rehrer, N. J., Patterson, M. J., & Cotter, J. D. (2012). Effectiveness of short-term heat acclimation for highly trained athletes *European Journal of Applied Physiology*, *112*(5), 1827–1837. https://doi.org/10.1007/s00421-011-2153-3

Garrett, A. T., Goossens, N. G., Rehrer, N. J., Patterson, M. J., & Cotter, J. D. (2009). Induction and decay of short-term heat acclimation. *European Journal of Applied Physiology*, *107*(6), 659–671. https://doi.org/10.1007/s00421-009-1182-7

Garrett, A. T., Goosens, N. G., Rehrer, N. J., Patterson, M. J., & Cotter, J. D. (2014). Short-term heat acclimation is effective and is not impaired by dehydration during acclimation. *American Journal of Human Biology*, *26*(3), 311–320. https://doi.org/10.1002/ajhb.22509.

Garrett, A. T., Rehrer, N. J., & Patterson, M. J. (2011). Induction and decay of short-term heat acclimation in moderately and highly trained athletes. *Sports Medicine*, *41*(9), 757–771. https://doi.org/10.2165/11587320-000000000-00000

Gore, C. J., Hopkins, W. G., & Burge, C. M. (2005). Errors of measurement for blood volume parameters: A meta analysis. *Journal of Applied Physiology*, *99*, 1745–1758. https://doi.org/10.1152/japplphysiol.00505.2005.

Hayden, G., Milne, H. C., Patterson, M. J., & Nimmo, M. A. (2004). The reproducibility of closed-pouch sweat collection and thermoregulatory responses to exercise-heat stress. *European Journal of Applied Physiology*, *91*(5–6), 748–751. https://doi.org/10.1007/s00421-004-1057-x

Hecksteden, A., Kraushaar, J., Scharhag-Rosenberger, F., Theisen, D., Senn, S., & Meyer, T. (2015). Individual response to exercise training:

A statistical perspective. *Journal of Applied Physiology*, *118*(12), 1450–1459. https://doi.org/10.1152/japplphysiol.00714.2014

Hecksteden, A., Werner, P., Rosenberger, F., & Meyer, T. (2018). Repeated testing for the assessment of individual response to exercise training. *Journal of Applied Physiology*, *124*(6), 1567–1579. http://doi.org/10.1152/japplphysiol.00896.2017

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*(1), 1–15. https://doi.org/10.2165/00007256-200030010-00001

Hopkins, W. G. (2015). Individual responses made easy. *Journal of Applied Physiology*, *118*(12), 1444–1446. https://doi.org/10.1152/japplphysiol.00098.2015

Hopkins, W. G., Hawley, J. A., & Burke, L. M. (1999). Design and analysis of research on sport performance enhancement. *Medicine and Science in Sports and Exercise*, *31*(3), 472–485. https://doi.org/10.1097/00005768-199903000-00018

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*, 3–12. https://doi.org/10.1249/MSS.0b013e31818cb278

Hopkins, W. G., Schabort, E. J., & Hawley, J. A. (2001). Reliability of power in physical performance tests. *Sports Medicine*, *31*(3), 211–234. https://doi.org/10.2165/00007256-200131030-00005

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., & Garrett, A. T. (2012a). Bioharness multivariable monitoring device. Part 1: Validity. *Journal of Sports Science & Medicine*, *11*(3), 400–408.

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., & Garrett, A. T. (2012b). Bioharness multivariable monitoring device. Part 2: Reliability. *Journal of Sports Science & Medicine*, *11*(3), 409–417.

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., Mitchell, A. C. S., & Garrett, A. T. (2012c). Field based reliability and validity of the Bioharness™ multivariable monitoring device. *Journal of Sports Science & Medicine*, *11*(4), 643–652.

Paton, C. D., & Hopkins, W. G. (2001). Tests of cycling performance. *Sports Medicine*, *31*(7), 489–496. https://doi.org/10.2165/00007256-200131070-00004

Petersen, P. H., Stockl, D., Westgard, J. O., Sandberg, S., Linnet, K., & Thienpont, L. (2001). Models of combining random and systematic errors, assumptions and consequences for different models. *Clinical Chemistry and Laboratory Medicine*, *39*, 589–595. https://doi.org/10.1515/CCLM.2001.094

Popowksi, L. A., Oppliger, R. A., Patrick Lambert, G., Johnson, R. F., Johnson, A. K., & Gisolfi, C. V. (2001). Blood and urinary measures of hydration status during progressive acute dehydration. *Medicine and Science in Sports and Exercise*, *33*(5), 747–753. https://doi.org/10.1097/00005768-200105000-00011

Poulsen, T. D., Klausen, T., Richalet, J. P., Kanstrup, I.-L., Fogh-Andersen, N., & Olsen, N. V. (1998). Plasma volume in acute hypoxia: Comparison of a carbon monoxide rebreathing method and dye dilution with Evans' blue. *European Journal of Applied Physiology*, *77*, 457–461. https://doi.org/10.1007/s004210050360

Sawka, M. N., Convertino, V. A., Eichner, R. E., Schnieder, S. M., & Young, A. J. (2000). Blood volume: Importance and adaptations to exercise training, environmental stresses, and trauma/sickness. *Medicine and Science in Sports and Exercise*, *32*(2), 332–348. https://doi.org/10.1097/00005768-200002000-00012

Wilmore, J. H., Stanforth, P. R., Turley, K. R., Gagnon, J., Daw, E. W., Leon, A. S., … Bouchard, C. (1998). Reproducibility of cardiovascular, respiratory, and metabolic responses to submaximal exercise: The HERITAGE family study. *Medicine and Science in Sports and Exercise*, *30*, 259–265. https://doi.org/10.1097/00005768-199802000-00014