Reproducibility of cardiorespiratory and performance responses to exercise in hypoxia

Ben J Lee, Charles D Thake

Background. Knowledge of the variance in physiological performance measures between repeated trials can inform whether familiarization sessions are necessary prior to intervention studies. The purpose of this study was to assess the reliability of cardiorespiratory and exercise performance measures during both steady state hypoxic exercise and a preloaded 16.1 km hypoxic time trial. **Methods.** Eighteen male participants (age, 22 ± 4 years; height, 1.77 ± 0.04 meters; body mass, 76.8 kg; estimated body fat and VO_2peak = 3.50 ± 0.60 $L.min^{-1}$) were divided into three groups. Reliability of responses (HR, S_PO_2, VO_2, VCO_2, VE and respiratory exchange ratio; RER) to the HST (F_IO_2 0.14; 15 minutes rest, 60 minutes cycling at 50% normoxic VO_2peak) was assessed across 3 repeat trials (HST 1, 2 and 3, n = 6). Reliability of the preloaded time trial (pTT; 15 min rest, 40 minutes cycling at 50% normoxic VO_2peak, 16.1km time trial) was assessed across 3 repeat normoxic (N; $F_IO_2 \approx$ 0.21; n=6) and 3 repeat hypoxic ($F_IO_2 \approx$ 0.14; n = 6) trials. All exercise trials were undertaken at the same time of day, following exercise and dietary controls, 7 days apart. **Results**. Intra-class correlation coefficients (ICC's) for mean and peak HR, SpO_2, V_E, VO_2, VCO_2 and blood lactate within each trial were improved from HST1 to HST2 (mean data: 0.99, 0.95, 0.75, 0.62, 0.70, 0.90; peak data: 0.98, 0.96, 0.64, 0.69, 0.74, 0.75) to HST2 and HST3 (ICC = 0.99, 0.97, 0.82, 0.85, 0.87 and 0.96 respectively). The reliability for time to pTT completion was improved following one trial, and the CV (test 2 vs. 3) was similar under normoxic (CV = 0.62) and hypoxic conditions (CV = 0.63). **Conclusion.** Cardiorespiratory responses to the HST were reproducible and the pTT performance time reliable in both normoxia and hypoxia. Since the reproducibility of the measurements in HST trials and reliability of pTT improved between the second and third trials, two familiarization visit are recommended prior to employing these protocols in future studies.

1 **Title: Reproducibility of cardiorespiratory and performance responses to exercise in**

2 **hypoxia**

3

4

5 **Authors:** Lee BJ[1,2] and Thake CD[2]

6 1:   Department of Health,

7   University of Bath,

8   Bath,

9   United Kingdom.

10

11 2:   Centre for Applied Biological and Exercise Sciences,

12   Coventry University,

13   Coventry,

14   United Kingdom.

15

16 **Correspondence**.

17 Ben Lee, Ph.D

18 The University of Bath

19 Department of Health

20 Bath

21 BA2 7AY

22 **Telephone:** +44 (0)1225 383029

23 **Email:** B.J.Lee@bath.ac.uk

24

25

26

27

28

29

Abstract

Background. Knowledge of the variance in physiological performance measures between repeated trials can inform whether familiarization sessions are necessary prior to intervention studies. The purpose of this study was to assess the reliability of cardiorespiratory and exercise performance measures during both steady state hypoxic exercise and a preloaded 16.1 km hypoxic time trial. **Methods.** Eighteen male participants (age, 22 ± 4 years; height, 1.77 ± 0.04 meters; body mass, 76.8 kg; estimated body fat and VO_2peak = 3.50 ± 0.60 L.min^{-1}) were divided into three groups. Reliability of responses (HR, S_PO_2, VO_2, VCO_2, V_E and respiratory exchange ratio; RER) to the HST (F_IO_2 0.14; 15 minutes rest, 60 minutes cycling at 50% normoxic VO_2peak) was assessed across 3 repeat trials (HST 1, 2 and 3, n = 6). Reliability of the preloaded time trial (pTT; 15 min rest, 40 minutes cycling at 50% normoxic VO_2peak, 16.1km time trial) was assessed across 3 repeat normoxic (N; $F_IO_2 \approx 0.21$; n=6) and 3 repeat hypoxic ($F_IO_2 \approx 0.14$; n = 6) trials. All exercise trials were undertaken at the same time of day, following exercise and dietary controls, 7 days apart. **Results**. Intra-class correlation coefficients (ICC's) for mean and peak HR, SpO_2, V_E, VO_2, VCO_2 and blood lactate within each trial were improved from HST1 to HST2 (mean data: 0.99, 0.95, 0.75, 0.62, 0.70, 0.90; peak data: 0.98, 0.96, 0.64, 0.69, 0.74, 0.75) to HST2 and HST3 (ICC = 0.99, 0.97, 0.82, 0.85, 0.87 and 0.96 respectively). The reliability for time to pTT completion was improved following one trial, and the CV (test 2 vs. 3) was similar under normoxic (CV = 0.62) and hypoxic conditions (CV = 0.63). **Conclusion.** Cardiorespiratory responses to the HST were reproducible and the pTT performance time reliable in both normoxia and hypoxia. Since the reproducibility of the measurements in HST trials and reliability of pTT improved between the second and third trials, two familiarization visit are recommended prior to employing these protocols in future studies.

53

Introduction

Many modern sporting scenarios and military operations involve exposure to extreme environmental factors (e.g. altitude or heat) that increase physiological strain and reduce physical

57   work capacity and performance (Gore et al., 1996; 1997). Knowledge of the extent that specific

58   environmental conditions impact on these response variables can inform the preparation and

59   management of exposed individuals and groups. Therefore it is important to consider the utility

60   of tests used to measure the magnitude of response to defined environmental stressors.

61   Accordingly, a reasonable degree of confidence that any changes in outcome measures are due

62   an imposed intervention (e.g. acclimation) rather than measurement error (Atkinson and Nevill,

63   1998) or inherent high-test variability (Che Jusoh et al., 2015) is important. Fixed workload

64   protocols are often used to provide 'steady-state' data in response to a stimulus e.g. the hypoxic

65   stress test used by Lee and colleagues (2014, 2014a) to evaluate the impact of short term heat

66   acclimation on responses to exercise in hypoxia. It is generally assumed that such protocols yield

67   reproducible physiological data under defined conditions although the test-retest coefficient of

68   variation in these measures is seldom reported.

69   In order to provide an indicator of performance potential, steady-state fixed workload protocols

70   may be extended by continuing the exercise bout until exhaustion. However this approach often

71   suffers from high test – retest variability and low ecological validity. For example, trained

72   cyclists had a test-retest coefficient of variation (CV) of ~27% when cycling to exhaustion at

73   75% of maximal workload (W_{max}; Jeukendrup et al., 1996). Whereas, a performance test (a self-

74   paced 60-minute bout of cycling), conducted with the same participants, returned a test-retest CV

75   of 3.4% (Jeukendrup et al., 1996). This is typical of self-paced tests with a defined end-point (i.e.

76   time trials) that by definition provide more ecologically valid data (Mee et al., 2015; Che Jusoh

77   et al., 2015). Furthermore, under heat stress a comparable CV (3.6%) was reported by Che Jusoh

78   and colleagues (2015) following a 45 minute cycling preload (cycling at 55%$\dot{V}O_2$peak). A

79   similar performance test has been used in hypoxic conditions (Beidleman et al., 2008) although

80   no reliability data was presented for the hypoxic performance test. As hypoxic tolerance before

81   and after an intervention is now commonly reported (Lee et al., 2014; Lee 2014a; Gibson 2015),

82   knowledge of the reproducibility of physiological steady-state conditions during a fixed-work

83   protocol hypoxic stress test (HST), as well as the reliability of a preloaded self-paced cycling

84   time trial cycling test (pTT) could assist data interpretation and study design. Therefore, the

85   purpose of this study was to assess the reproducibility of physiological measures collected during

86   a steady state HST (60 minute fixed load period of cycling) and the reliability of a preloaded (40

87   minute bout of fixed-intensity cycling) 16.1km TT conducted in moderate normobaric hypoxia.

88

Methods

Participants

The study received ethical approval from the Coventry University local Ethics Committee

(Reference number P2566/P6420) and was conducted in accordance with the declaration of

Helsinki. Eighteen healthy and regularly active males (playing team sports at least 3 times per

week for a period of over 1 year) accustomed to cycling exercise provided their informed

consent to participate. Three groups of 6 participants were formed. Reproducibility of

physiological measurements to a fixed relative work-rate hypoxic stress test (HST) where made

on the HST group (age, 23 ± 3 years; height, 1.77 ± 0.05 m; body mass, 74.7 ± 6.9 kg; $\dot{V}O_2$peak,

3.50 ± 0.70 L.min^{-1}); whereas the reliability of a 16.1 km preloaded time trial (pTT) performance

test was evaluated using separate groups under normoxic ($F_IO_2 \approx 0.21$) conditions (NORM; age,

21 ± 2 years; height, 1.77 ± 0.04 m; body mass, 84.0 ± 17.0; $\dot{V}O_2$peak, 3.17 ± 0.40 L.min^{-1}), and

hypoxic ($F_IO_2 \approx 0.14$) conditions (HYP; age; 27 ± 5 years; height, 1.76 ± 0.02 m; body mass,

72.0 ± 7.6 kg; $\dot{V}O_2$peak, 3.82 ± 0.60 L.min^{-1}).

Experimental design

Each participant visited the laboratory on 4 occasions (1 preliminary and 3 experimental) each

separated by at least 7 days (range 7 – 11 days). During the preliminary visit participants gave

informed consent and measures of height, body mass, skin folds (4 sites), normoxic lactate

threshold and $\dot{V}O_2$peak were made. Thereafter, on separate visits, participants in the HST group

completed 3 fixed work-rate (50% normoxic $\dot{V}O_2$peak, 60 minutes) exercise trials whilst

breathing hypoxia ($F_IO_2 \approx 0.14$), 6 participants completed 3 preloaded time-trials whilst

breathing $F_IO_2 \approx 0.21$, and a further 6 participants completed 3 preloaded time-trials under

hypoxic conditions ($F_IO_2 \approx 0.14$).

Preliminary testing

Anthropometric data were collected in accordance with the International Society for the

Advancement of Kinathroprometry (ISAK) guidelines (Marfell-Jones et al., 2006). Lactate

threshold and peak oxygen consumption ($\dot{V}O_2$peak) were determined via an incremental exercise

116 test to volitional exhaustion on calibrated SRM cycle ergometer (Schoberer Rad Meßtechnik,

117 Welldorf, Germany). Resting blood lactate (BLa; Biosen C-Line analyser, EKF Diagnostics,

118 Germany) was determined from a finger capillary whole blood sample following a 10 minute

119 seated rest period. The test began at a workload of 70W for 4-minutes and was then increased by

120 35W every 4 minutes until a blood lactate value of $> 4 mmol.L^{-1}$ was reached. Thereafter,

121 workload increased 35W every 2 minutes until volitional exhaustion. A cadence of 70

122 revolutions per minute (RPM) was maintained throughout. Expired gas was collected during the

123 last minute of each exercise stage via a two-way nonrebreathable valve (Harvard Ltd,

124 Eldenbridge, UK) and 1.5 m of 30mm diameter polyvinyl tubing supplying a 200L Douglas bag

125 (Cranlea & Co, Birmingham, UK). Heart rate (Polar FT1, Polar Electro OY, Kempele, Finland)

126 was recorded and perceived exertion (Borg, 1976) sought within the last 10 seconds of each gas

127 collection. Respiratory gas analysis for determination of expired oxygen and carbon dioxide

128 fractions was completed using a Servomex infrared and paramagnetic gas analyzer (model 1400,

129 Servomex, Crowthorne, UK) respectively, and gas volume via a Harvard dry gas meter (Cranlea

130 and Company, Birmingham, UK). Lactate threshold was calculated according to the Dmax

131 method (Cheng et al. 1992). $\dot{V}O_2$peak was considered to be achieved if two of the following

132 criteria were met: i) a respiratory exchange ratio of >1.1, ii) a heart rate greater than 95% of age

133 predicted maximum (220−age) and iii) a final blood lactate value in excess of 8 $mmol.mL^{-1}$

134 (Winter et al. 2006).

135 *Experimental trials*

136 Laboratory attendance time was consistent for each participant in order to minimize the effects of

137 circadian variation on performance (Drust et al., 2005). Participants were requested to abstain

138 from caffeine (Luo et al., 2008) and alcohol consumption for 72 hours prior to each laboratory

139 visit and required to maintain a food and activity diary as accurately as possible for 3 days prior

140 to each experimental visit (Morton et al., 2006). Participants refrained from all supplementation

141 (i.e. vitamins, ergogenic aids) throughout the study period and were requested to abstain from

142 prolonged thermal exposures (baths, saunas, steam rooms, and tanning devices) and vigorous

143 physical activity for seven days prior to each laboratory visit, and moderate physical activity for

144 72 hours prior to each visit.

145   Participants adhered to an overnight fast (Febbraio et al., 2002) and consumed 500 ml of plain

146   water one hour before visiting the laboratory in accordance with the American College of Sports

147   Medicine position stance on hydration (Sawka *et al.*, 2007). Upon arrival (0630 – 0730 hrs),

148   participants voided their bladder to provide a sample for hydration assessment via urine specific

149   gravity (USG; Atago Refractomer, Jencons Pls, Leighton Buzzard, UK) and urine osmolarity

150   (U_{osmo}; Advanced 3300 Micro-Osmometer, Advanced Inc, Massachusetts, USA). Euhydration

151   was assumed for urine specific gravity values of ≤ 1.020 g·ml^{-1} and osmolarity values of ≤ 700

152   mOsm·kg^{-1} (Armstrong et al., 1994). This control was not violated by any participant during any

153   trial. Following this, participants measured their own nude body mass (Seca 880, Seca,

154   Hamburg, Germany), and fitted a telemetric heart rate monitor around their chest (Polar FT1,

155   Polar Electro OY, Kempele, Finland).

156   *Fixed work-rate hypoxic stress test (HST)*

157   Participants in the HST group completed 15 minutes of seated rest in normoxic conditions prior

158   to baseline data collection for all physiological variables. Following normoxic rest, a 15 minute

159   resting 'wash-in' period of breathing hypoxic gas ($F_1O_2 = 0.14$) was then completed. The gas was

160   delivered to the participant via a mouthpiece and 30 mm diameter connector (Harvard Ltd,

161   Eldenbridge, UK) connected to a two-way non-rebreathable valve (Harvard Ltd, Eldenbridge,

162   UK). Ethylene clear vinyl tubing was used to connect the inspiratory side of the valve to a series

163   of 1000L Douglas bags which were filled via a hypoxicator (Hypoxico HYP123 Hypoxicator,

164   New York, USA). At the end of the wash-in period participants began 60 minutes of cycling

165   exercise at an intensity corresponding to 50% normoxic $\dot{V}O_2$peak (145 ± 16 Watts) on an SRM

166   ergometer (Schoberer Rad Meßtechnik, Welldorf, Germany) at a cadence of 70 RPM. Measures

167   of heart rate (HR), arterial heamoglobin oxygen saturation (SpO_2), respiratory variables ($\dot{V}_E$, $\dot{V}$

168   O_2 , $\dot{V}CO_2$) via Douglas bag collections, and fingertip capillary blood samples for determination

169   of glucose and lactate concentrations were taken at the start and end of each 15 minute resting

170   period and every 10 minutes throughout exercise. Perceptual ratings of perceived exertion (RPE;

171   Borg 1976) and thermal sensation (TS) were also noted at 10 minute intervals during the exercise

172   period.

173   *Preloaded time trial (pTT)*

174   Participants undertook 30 minutes of seated rest (normoxia $F_IO_2 = 0.21$ throughout; hypoxia 15

175   min $F_IO_2 = 0.21$ followed by 15 minutes $F_IO_2 = 0.14$ 'wash in' period) prior to undertaking a 40

176   minute period of cycling exercise at an intensity corresponding to 50% normoxic $\dot{V}O_2$peak (144

177   $\pm$ 18 Watts) whilst breathing either $F_IO_2 = 0.21$ or $F_IO_2 = 0.14$. Physiological and perceptual

178   variables were collected every 10 minutes as previously described. At the end of the 40 minute

179   fixed load exercise bout and after 5 minutes of passive recovery the self-paced 16.1km cycling

180   pTT was commenced. The pTT was controlled within the SRM ergometers open-ended mode,

181   which creates a braking force that has a cubic relationship with speed, mimicking the effect of air

182   resistance on a moving bicycle. Participants were instructed to complete distance as quickly as

183   possible. The only feedback available to participants was the distance completed at any given

184   time. During the time trial only measures of HR and power output were collected in order to

185   avoid providing any external time cues.

186   *Statistical analysis*

187   Data are reported as mean $\pm$ SD unless otherwise stated. Physiological data were analysed using

188   trial x time ANOVA, with repeated measures over time. In order to assess test-retest

189   reproducibility in the mean physiological measures obtained at rest and during exercise

190   throughout HST, normoxic pTT and hypoxic pTT, the mean difference between trials (change in

191   mean), intra-class correlation coefficient (ICC) and coefficient of variation (CV) was calculated

192   between the 3 repeat conditions in each group (Hopkins, 2000). The technical error of

193   measurement (TEM) between trials was calculated for all participants by calculating the

194   difference scores and standard deviation of difference scores between trials and dividing the

195   square root of the standard deviation of the difference by 2 (Hopkins, 2000). In order to calculate

196   the smallest worthwhile change and therefore the smallest detectable change in a measured

197   variable, the TEM was multiplied by 1.5 (Hopkins, 2000).

198   **Results**

199   All participants adhered here to the activity and nutrition controls specified in the participant

200   information sheet prior to each testing session.

201   *Physiological and perceptual responses to the HST*

202　The cardiovascular responses to each HST are shown in Figure 1. Based on participants mean

203　oxygen consumption, participants were exercising at an intensity eliciting 62 ± 11, 58 ± 9 and 61

204　$\pm 16\%$ $\dot{V}O_2$peak at the end HST1, HST2 and HST3 respectively. No main effect for trial (p >

205　0.05) was observed for heart rate, S_PO_2, $\dot{V}_E$, $\dot{V}O_2$, $\dot{V}CO_2$, and RER over the course of the 3 trials.

206　The ICC, technical error of measurement, and coefficient of variation for mean and peak exercise

207　measurements are shown in Table 1. The data indicate that acceptable reliability is achieved after

208　1 familiarization for cardiovascular responses (mean exercise and peak HR and SpO_2).

209　Respiratory variables in hypoxic conditions are more variable, with improved reliability seen

210　after 2 familiarization sessions for both the mean exercise and peak measurements (Figure 2).

211　Perceptual ratings of perceived exertion averaged 14 ± 1, 13 ± 1, and 13 ± 1, and Thermal

212　sensation averaged during 4.8 ± 0.4, 4.4 ± 0.6 and 4.5 ± 0.6 during HST1, HST2 and HST3

213　respectively

214

215　*Physiological and perceptual responses to the preload period*

216　Based on participants mean $\dot{V}O_2$ and HR participants in the normoxic group were exercising at

217　an intensity eliciting $51.1 \pm 8\%$, 51.3 ± 11 and $52 \pm 10\%$ $\dot{V}O_2$peak, and participants in the

218　hypoxic group at $67 \pm$, 65 ± 9 and $65 \pm \% \pm$ for pTT1, pTT2, and pTT3 respectively. Main

219　effects for time (p < 0.0001) but not trial (p > 0.05) or time x trial interaction (p > 0.05) were

220　observed for HR, SpO_2, $\dot{V}O_2$, $\dot{V}_E$ and RER during the 40 minute preload, with reliability indices

221　mirroring those observed during the prolonged 60 minute HST (data not shown). Mean and peak

222　physiological and perceptual data during the preload period are presented in Table 2.

223　*Time trial performance*

224　Physiological data from the time trial are shown in Figure 3. Based on mean power output during

225　each TT, the normoxic group completed the distance at $63 \pm 5\%$, $63 \pm 6\%$ and $66 \pm 9\%$ and the

226　hypoxic group at $60 \pm 6\%$, $62 \pm 6\%$ and $61 \pm 6\%\%$ of normoxic W_{max} for TT1, TT2 and TT3

227　respectively. Time to complete the 16.1km course was longer in HYP compared to NORM

228　during TT1 (42.3 ± 5.5 versus 39.4 ± 1.1 minutes), TT2 (41.4 ± 5.5 versus 39.1 ± 1.5 minutes)

229　and TT3 (41.8 ± 5.3 versus 38.8 ± 0.6 minutes), although no TT x condition effect was observed

230　(p = 0.250). From Table 3 it can be seen that the reliability of the performance time, typically the

231 main outcome measure obtained from a time trial, was improved when 2 familiarization trials are

232 performed. Data regarding the smallest worthwhile change for all variables measured during the

233 HST and pTT are shown in Table 4.

234

Discussion

235

236 This study reports that physiological responses to a 60 minute steady-state HST are reproducible

237 and that preloaded (40 minutes of steady state cycling) 16.1km TT performance, conducted

238 under both normoxic and acute normobaric hypoxic conditions ($FiO_2 \approx 0.14$), is reliable.

239 Importantly these data demonstrate that the reproducibility of the measurements in HST trials

240 and reliability of pTT performance (under both normoxic and hypoxic conditions) improved

241 between the second and third trials. Accordingly when using such protocols in future

242 investigations we would advise conducting at least one familiarisation trial where participants

243 carryout the full protocol under the proposed experimental conditions, with 2 familiarization

244 protocols reducing the technical error of measurement further. Furthermore an awareness of the

245 typical variance of each of these tests under the conditions studied can help provide context to

246 the effect of an intervention on defined outcome measures.

247 During each HST trial classic markers of hypoxic physiology (HR, SpO_2, $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER

248 and blood lactate) were measured. Our data indicate that this simple to conduct fixed-workload

249 test had good agreement for all the physiological variables studied (Table 1) when tests are

250 conducted one week apart, supporting its use in future studies investigating hypoxic tolerance.

251 Typically, a correlation coefficient of over 0.90 is considered high, 0.70-0.80 moderate, and

252 below 0.70 considered too low for a reliable and sensitive physiological test (Vincent, 1995). In

253 the present investigation the ICC for each physiological variable was over 0.80 when HST2 and

254 HST3 were considered. Based on this observation, the whole body physiological measurements

255 of HR, SpO_2, $\dot{V}O_2$, $\dot{V}CO_2$, RER and $\dot{V}_E$ all have an acceptable level of reproducibility between

256 repeated trials in the conditions studied if two familiarisation visits are conducted. With

257 reference to the smallest worthwhile change, calculated from the TEM, we suggest that changes

258 in mean exercise HR, SpO_2, $\dot{V}_E$ and $\dot{V}O_2$ of 2 - 3 beats.min^{-1}, 1%, 3.6 - 4.7 L.min^{-1}, 0.24 – 0.32

259 L.min^{-1} respectively, would constitute a statistically meaningful change in these variables

260   between HST tests conducted before and after an intervention (e.g. an acclimation period)

261   provided 2 familiarization sessions are completed.

262   In addition to the fixed workload HST we also examined the reliability of a 16.1km TT

263   conducted immediately following a 40-minute, submaximal preload. This type of experimental

264   design allows researchers to assess both steady state responses, and incorporate a measure of

265   physical performance. As the TEM and CV for normoxia and hypoxia are similar, this indicates

266   that any familiarisation sessions can be completed in normoxic conditions and reduce the

267   potential for conferred acclamatory effect prior to subsequent hypoxic exposure. It should be

268   noted that, while not trained cyclists, all participants in the present investigation were not naive

269   to laboratory testing protocols. To avoid further confounds care was taken to remove the

270   influence of diurnal variation, previous activity, and diet, as these may all effect endurance

271   performance (Atkinson et al., 2005; Rauch et al., 2005). Like other time trial procedures,

272   feedback in the form of distance completed was provided (Jeukendrup et al., 1996; Tyler and

273   Sutherland et al., 2008; Biedleman et al., 2014). This increases the external validity of the test

274   and allows participants to use a pacing strategy to a known end-point (Che Jusoh et al., 2015).

275   One criticism levelled at the use of fixed intensity steady-state preload periods is that by the time

276   participants begin the self-paced performance period they are already near to their physiological

277   limit (Che Jusoh et al., 2015). To minimise this potential confound a workrate of 50% normoxic

278    $\dot{V}O_2$ max was selected as pilot data indicated this allowed the participants to perform exercise

279   below both the normoxic and hypoxic lactate threshold. At the end of the preload period, HR was

280   140, 142, and 140 for the normoxic group and 156, 156, and 158 for the hypoxic group, which

281   would suggest that while this 40 minute period presents a moderate physiological challenge,

282   participants began the time trial with considerable capacity to increase physiological strain, i.e.,

283   from 156 beats.min^{-1} to 180 beats.min^{-1} at the end of the TT. The smallest worthwhile change for

284   performance time following one experimental visit (TT1 and TT2) was between 0:59 and 1:19

285   minutes and 1:09 and 1:26 minutes for the normoxic and hypoxic time trials. However if two

286   familiarization trials are conducted prior to an experimental trial the smallest worthwhile change

287   is reduced to 0:35 – 0:47 minutes (normoxia) and 0:32 – 0:43 minutes (hypoxia), indicating that

288   a significant change in performance time induced by an intervention would need to be > 43

289   seconds to be meaningful.

290

Conclusion

Performance is a key primary outcome variable commonly used when assessing the efficacy of interventions. Thus separating real treatment effects from the typical variance associated with a test and/or participants is important yet seldom reported. Additionally, knowledge of the variance in physiological and performance measures between repeated trials is essential to inform the number of familiarization trials required prior to exposing participants to an intervention. The results of this study show that for moderately trained healthy males, both a steady state HST and a preloaded 16.1km time trial conducted on an SRM ergometer are highly reproducible when performed under laboratory conditions. Using these protocols, practitioners and researchers can determine whether common interventions (e.g. acclimation) reduce physiological strain (HST/preload) during steady state work conditions and improve physical performance (TT time) within a test. In order to reduce the known error and remove confounding learning effects associated with these protocols we recommend 2 familiarization sessions be performed prior to employing these protocols in future studies.

305

Acknowledgements

References

Armstrong LE, Maresh CM, Castellani JW, Bergeron M, Kenefick R, LaGasse K, and Riebe D. 1994. Urinary indices of hydration status. *International journal of sport nutrition* 4:265-279.

Atkinson G, and Nevill AM. 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine* 26:217-238.

316 Beidleman BA, Muza SR, Fulco CS, Cymerman A, Sawka MN, Lewis SF, and Skrinar GS.
317   2008. Seven intermittent exposures to altitude improves exercise performance at 4300 m.
318   *Med Sci Sports Exerc* 40:141-148.
319 Borg G. 1976. Simple rating methods for estimation of perceived exertion. *Physical work and*
320   *effort*:39-46.
321 Che Jusoh M, Morton R, Stannard S, and Mündel T. 2015. A reliable preloaded cycling time trial
322   for use in conditions of significant thermal stress. *Scandinavian journal of medicine &*
323   *science in sports* 25:296-301.
324 Cheng B, Kuipers H, Snyder A, Keizer H, Jeukendrup A, and Hesselink M. 1992. A new
325   approach for the determination of ventilatory and lactate thresholds. *Int J Sports Med*
326   13:518-522.
327 Drust B, Waterhouse J, Atkinson G, Edwards B, and Reilly T. 2005. Circadian rhythms in sports
328   performance-an update. *Chronobiology international* 22:21-44.
329 Febbraio MA, Steensberg A, Walsh R, Koukoulas I, Hall Gv, Saltin B, and Pedersen BK. 2002.
330   Reduced glycogen availability is associated with an elevation in HSP72 in contracting
331   human skeletal muscle. *J Physiol* 538:911-917.
332 Gibson OR, Turner G, Tuttle JA, Taylor L, Watt PW, and Maxwell NS. 2015. Heat Acclimation
333   attenuates physiological strain and the Hsp72, but not Hsp90α mRNA response to acute
334   normobaric hypoxia. *J Appl Physiol (1985)*:jap. 00332.02015.
335 Gore CJ, Hahn AG, Scroop GC, Watson D, Norton KI, Wood R, Campbell D, and Emonson D.
336   1996. Increased arterial desaturation in trained cyclists during maximal exercise at 580 m
337   altitude. *J Appl Physiol (1985)* 80:2204-2210.
338 Gore CJ, Little S, Hahn AG, Scroop GC, Norton KI, Bourdon PC, Woolford S, Buckley JD,
339   Stanef T, and Campbell D. 1997. Reduced performance of male and female athletes at
340   580 m altitude. *European journal of applied physiology and occupational physiology*
341   75:136-143.
342 Heled Y, Peled A, Yanovich R, Shargal E, Pilz-Burstein R, Epstein Y, and Moran DS. 2012.
343   Heat acclimation and performance in hypoxic conditions. *Aviation, space, and*
344   *environmental medicine* 83:649-653.
345 Hopkins WG. 2000. Measures of reliability in sports medicine and science. *Sports Medicine*
346   30:1-15.

347 Jeukendrup A, Saris W, Brouns F, and Kester AD. 1996. A new validated endurance
348   performance test. *Med Sci Sports Exerc* 28:266-270.

349 Lee BJ, Emery-Sinclair EL, Mackenzie RW, Hussain A, Taylor L, James RS, and Thake CD.
350   2014. The impact of submaximal exercise during heat and/or hypoxia on the
351   cardiovascular and monocyte HSP72 responses to subsequent (post 24 h) exercise in
352   hypoxia. *Extreme physiology & medicine* 3:15.

353 Lee BJ, Mackenzie RW, Cox V, James RS, and Thake CD. 2014. Human Monocyte Heat Shock
354   Protein 72 Responses to Acute Hypoxic Exercise after 3 Days of Exercise Heat
355   Acclimation. 2014a. *BioMed Research International*. vol. 2015, Article ID 849809, 15
356   pages, 2015. doi:10.1155/2015/849809

357 Marfell-Jones M, Olds T, Stewart A, and CARTER J. 2006. ISAK: Potchefstroom: International
358   Standards for Anthropometric Assessment.

359 Mee JA, Doust J, and Maxwell NS. 2015. Repeatability of a running heat tolerance test. *Journal*
360   *of Thermal Biology* 49:91-97.

361 Morton JP, MacLaren DP, Cable NT, Bongers T, Griffiths RD, Campbell IT, Evans L, Kayani
362   A, McArdle A, and Drust B. 2006. Time course and differential responses of the major
363   heat shock protein families in human skeletal muscle following acute nondamaging
364   treadmill exercise. *J Appl Physiol (1985)* 101:176-182.

365 Rauch H, Gibson ASC, Lambert E, and Noakes T. 2005. A signalling role for muscle glycogen
366   in the regulation of pace during prolonged exercise. *British Journal of Sports Medicine*
367   39:34-38.

368 Sawka MN, Burke LM, Eichner ER, Maughan RJ, Montain SJ, and Stachenfeld NS. 2007.
369   American College of Sports Medicine position stand. Exercise and fluid replacement.
370   *Med Sci Sports Exerc* 39:377-390.

371 Tyler C, and Sunderland C. 2008. The effect of ambient temperature on the reliability of a
372   preloaded treadmill time-trial. *Int J Sports Med* 29:812-816.

373 Vincent W, and Weir J. 1994. *Statistics in kinesiology*: Human Kinetics.

374 Winter EM, Jones AM, Davison RR, Bromley PD, and Mercer TH. 2006. *Sport and Exercise*
375   *Physiology Testing Guidelines: Volume I–Sport Testing: The British Association of Sport*
376   *and Exercise Sciences Guide*: Routledge.

377

Table 1(on next page)

Measures of reliability for mean exercise and peak physiological variables during the HST ($n = 6$).

1 **Table 1.** Measures of test - retest reliability for mean exercise and peak physiological variables during the HST ($n = 6$).

	HST1 – HST 2			HST2 – HST 3		
	ICC	TEM	CV	ICC	TEM	CV
Mean HR (bts·min^{-1})	0.99 (0.93 – 0.99)	1.6	0.91	0.99 (0.96 – 0.99)	1.7	0.90
Peak HR (bts·min^{-1})	0.98 (0.89 – 0.99)	2.44	1.15	0.97 (0.81 – 0.99)	3.92	1.84
Mean SpO$_2$ (%)	0.95 (0.65 – 0.99)	0.91	0.81	0.97 (0.73 – 0.99)	0.50	0.61
Peak SpO$_2$ (%)	0.96 (0.69 – 0.99)	0.83	1.15	0.83 (0.50 – 0.97)	1.44	1.37
Mean $\dot{V}_E$BTPS (L·min^{-1})	0.75 (0.74 – 0.97)	3.92	4.84	0.82 (0.59 – 0.97)	2.37	2.90
Peak $\dot{V}_E$BTPS (L·min^{-1})	0.64 (-2.7 – 0.95)	6.20	2.23	0.97 (0.83 – 0.99)	1.25	1.53
Mean $\dot{V}O_2$ (L·min^{-1})	0.62 (-3.5 – 0.95)	0.11	4.10	0.85 (-19 – 0.98)	0.16	5.64
Peak $\dot{V}O_2$ (L·min^{-1})	0.69 (-0.61 – 0.95)	0.17	5.75	0.74 (0.72 – 0.95)	0.22	8.06
Mean $\dot{V}CO_2$ (L·min^{-1})	0.70 (-2.2 – 0.96)	0.12	5.65	0.87 (0.15 – 0.98)	0.10	2.41
Peak $\dot{V}CO_2$ (L·min^{-1})	0.74 (-0.41 – 0.96)	0.09	8.61	0.85 (0.33 – 0.97)	0.19	12.04
Mean BLa (mmol^{-1})	0.90 (0.30 – 0.99)	0.62	15.0	0.96 (0.72 – 0.99)	0.39	8.7
Peak BLa (mmol^{-1})	0.75 (-0.29 – 0.96)	0.65	19.2	0.98 (0.85 – 0.99)	0.26	3.04

2 ICC = Intraclass correlation coefficient (lower 95% and upper 95% confidence interval. TEM = technical error of measurement. CV = coefficient of variation.

3 HR = heart rate. BTPS = Body temperature pressure saturated.

4

Table 2(on next page)

Measures of reliability in normoxic (n = 6) and hypoxic (n = 6) conditions during the time trial. Data are calculated from the mean TT power output (Watts) and heart rate.

1 **Table 2.** Mean ± SD for mean peak exercise physiological, respiratory and perceptual responses

2 during the 40 minute preload period in normoxia (n = 6) and hypoxia (n = 6).

	Normoxia			Hypoxia		
	pTT1	pTT2	pTT3	pTT1	pTT2	pTT3
Mean HR (bts·min^{-1})	138 ± 10	137 ± 7	136 ± 8	153 ± 8	154 ± 9	155 ± 6
Peak HR (bts·min^{-1})	140 ± 12	142 ± 7	140 ± 10	156 ± 6	156 ± 9	158 ± 4
Mean SpO$_2$ (%)	96 ± 1	97 ± 1	97 ± 1	80 ± 4	79 ± 4	80 ± 3
Peak SpO$_2$ (%)	96 ± 1	97 ± 1	97 ± 1	79 ± 4	79 ± 3	80 ± 2
Mean $\dot{V}_E$BTPS (L·min^{-1})	52.5 ± 4.4	54.5 ± 5.1	54.3 ± 4.8	61.2 ± 12.9	61.8 ± 9.8	64.0 ± 6.3
Peak $\dot{V}_E$BTPS (L·min^{-1})	53.4 ± 3.7	56.7 ± 5.1	53.4 ± 6.8	63.8 ± 14.7	62.4 ± 8.9	63.1 ± 8.6
Mean $\dot{V}O_2$ (L·min^{-1})	2.02 ± 0.11	2.03 ± 0.19	2.04 ± 0.18	1.94 ± 0.36	1.93 ± 0.33	1.96 ± 0.31
Peak $\dot{V}O_2$ (L·min^{-1})	2.08 ± 0.10	2.07 ± 0.25	2.09 ± 0.18	2.07 ± 0.42	1.96 ± 0.33	1.96 ± 0.35
Mean $\dot{V}CO_2$ (L·min^{-1})	1.88 ± 0.21	1.85 ± 0.16	1.88 ± 0.20	1.92 ± 0.34	1.94 ± 0.30	1.92 ± 0.26
Peak $\dot{V}CO_2$ (L·min^{-1})	2.01 ± 0.23	1.88 ± 0.19	1.98 ± 0.22	1.96 ± 0.35	1.96 ± 0.26	1.95 ± 0.28
Mean RPE	13 ± 1	13 ± 1	13 ± 1	13 ± 2	12 ± 1	12 ± 1
Peak RPE	13 ± 1	13 ± 1	13 ± 1	14 ± 1	14 ± 1	14 ± 1
Mean TS	4.9 ± 0.4	4.8 ± 0.4	4.8 ± 0.1	4.6 ± 0.2	4.9 ± 0.4	4.8 ± 0.4
Peak TS	5.0 ± 0.0	4.8 ± 0.5	4.8 ± 0.5	5.0 ± 0.8	4.8 ± 0.5	4.8 ± 0.5

3 HR = heart rate. BTPS = Body temperature pressure saturated. RPE = Rating of perceived exertion. TS = Thermal
4 sensation.

5

Table 3(on next page)

Measures of reliability in normoxic (n = 6) and hypoxic (n = 6) conditions during the time trial. Data are calculated from the mean TT power output (Watts) and heart rate.

1   **Table 3.** Measures of reliability in normoxic (n = 6) and hypoxic (n = 6) conditions during the

2   time trial. Data are calculated from the mean TT power output (Watts) and heart rate.

	TT1 – TT2			TT2 – TT3		
	ICC	TEM	CV	ICC	TEM	CV
Normoxia						
Time	0.79 (-0.52 – 0.97)	0.66	0.46	0.82 (0.05 – 0.99)	0.39	0.62
Watts	0.91 (0.35 – 0.99)	5.01	0.45	0.92 (0.41 – 0.99)	5.70	3.01
HR	0.94 (0.59 – 0.99)	3.0	1.56	0.97 (0.79 – 0.99)	2.0	1.06
Hypoxia						
Time	0.98 (0.86 – 0.99)	0.77	1.53	0.99 (0.97 – 0.99)	0.36	0.63
Watts	0.86 (-0.19 – 0.98)	2.30	2.30	0.79 (-0.65 – 0.97)	6.30	0.71
HR	0.92 (0.47 – 0.99)	3.00	0.43	0.98 (0.86 – 0.99)	3.00	0.71

3   ICC = Intraclass correlation coefficient. TEM = technical error of measurement. CV =

4   coefficient of variation. HR = heart rate.

Table 4 (on next page)

Reliability measures during the hypoxic tests.

Smallest worthwhile changes for mean exercise data during the HST, pTT preload period, and time trial conducted in hypoxic conditions following 2 familiarization sessions. Data are calculated as 1.5 x the TEM, and 2.0 x the TEM (Hopkins 2000).

1 **Table 4.** Smallest worthwhile changes for mean exercise data during the HST, pTT preload

2 period, and time trial conducted in hypoxic conditions following 2 familiarization sessions. Data

3 are calculated as 1.5 x the TEM, and 2.0 x the TEM (Hopkins 2000).

|	Smallest worthwhile change	
	1.5	**2.0**
HST/pTT preload		
Mean HR (bts·min^{-1})	2	3
Mean SpO$_2$ (%)	1.0	1.0
Mean $\dot{V}_E$BTPS (L·min^{-1})	3.6	4.7
Mean $\dot{V}O_2$ (L·min^{-1})	0.24	0.32
Mean $\dot{V}CO_2$ (L·min^{-1})	0.15	0.20
Mean BLa (mmol^{-1})	0.58	0.78
Time trial		
Performance time (seconds)	30	43
Mean power output (Watts)	9	12
Mean HR (bts·min^{-1})	5	6

4

Figure 1 (on next page)

Heart rate and arterial oxygen saturation throughout the 60 minute HST.

Resting and exercise heart rate (A) and arterial oxygen saturation (B) were no different throughout the 3 HSTs. Line graphs show mean ± SD throughout rest and exercise (n = 6) and box plots display the mean exercise values, and highest and lowest value for HST1, HST2 and HST3.
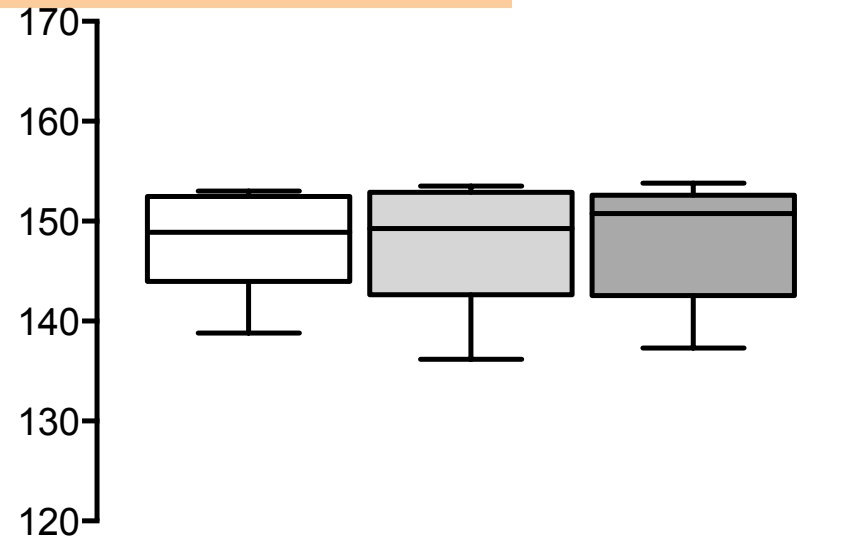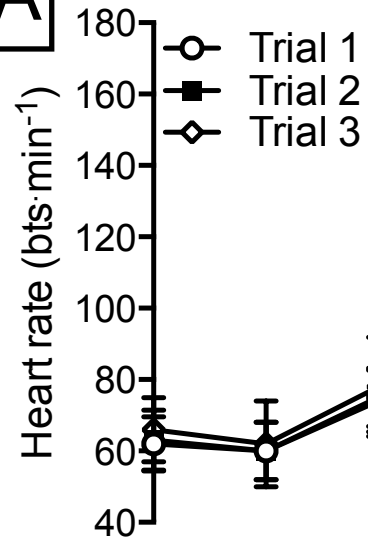
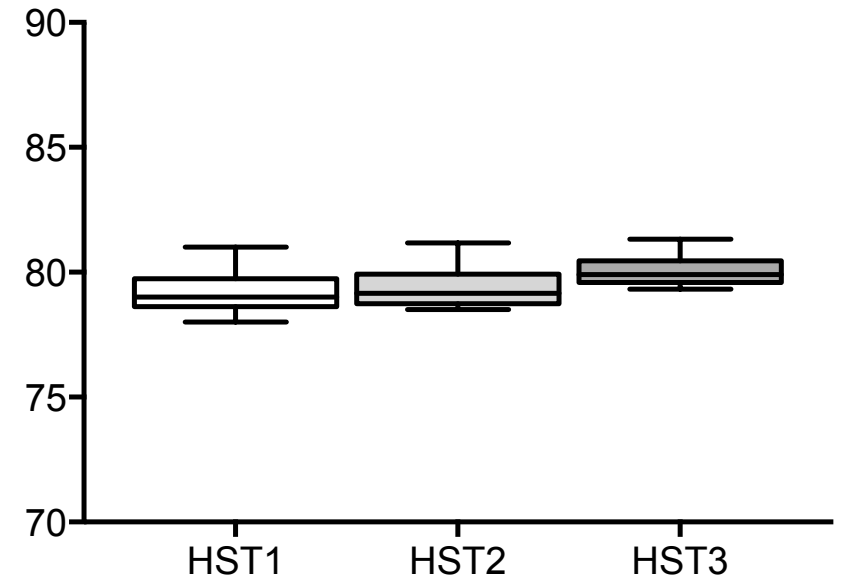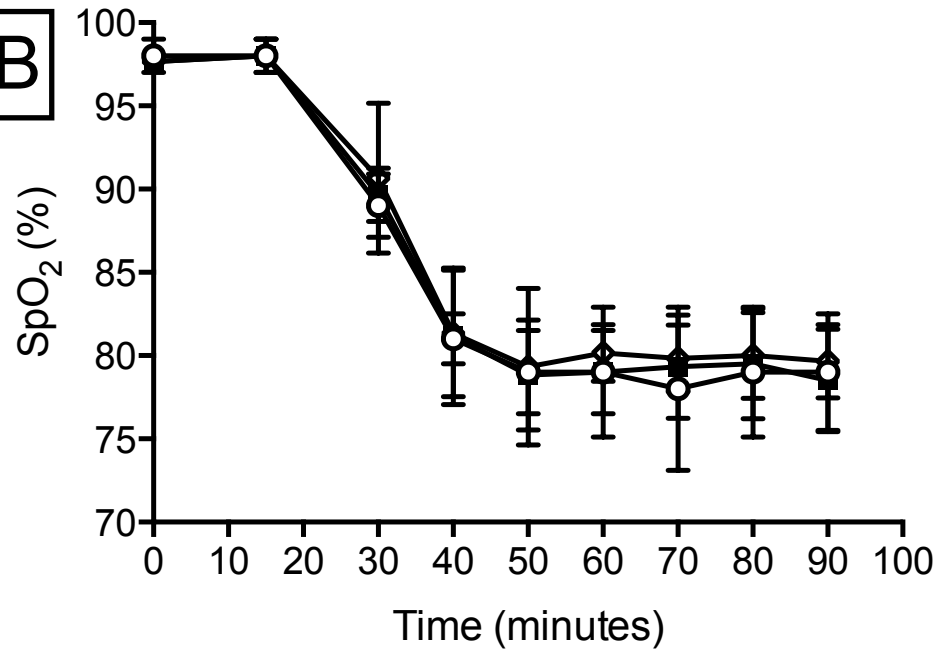# Figure 2 (on next page)

Respiratory responses to the HST

Resting and exercise ventilation (A), oxygen consumption (B), and carbon dioxide production (C) were no different between the 3 HSTs. Line graphs show the mean ± SD values and box plots represent the mean exercise values with error bars showing the highest and lowest values recorded.
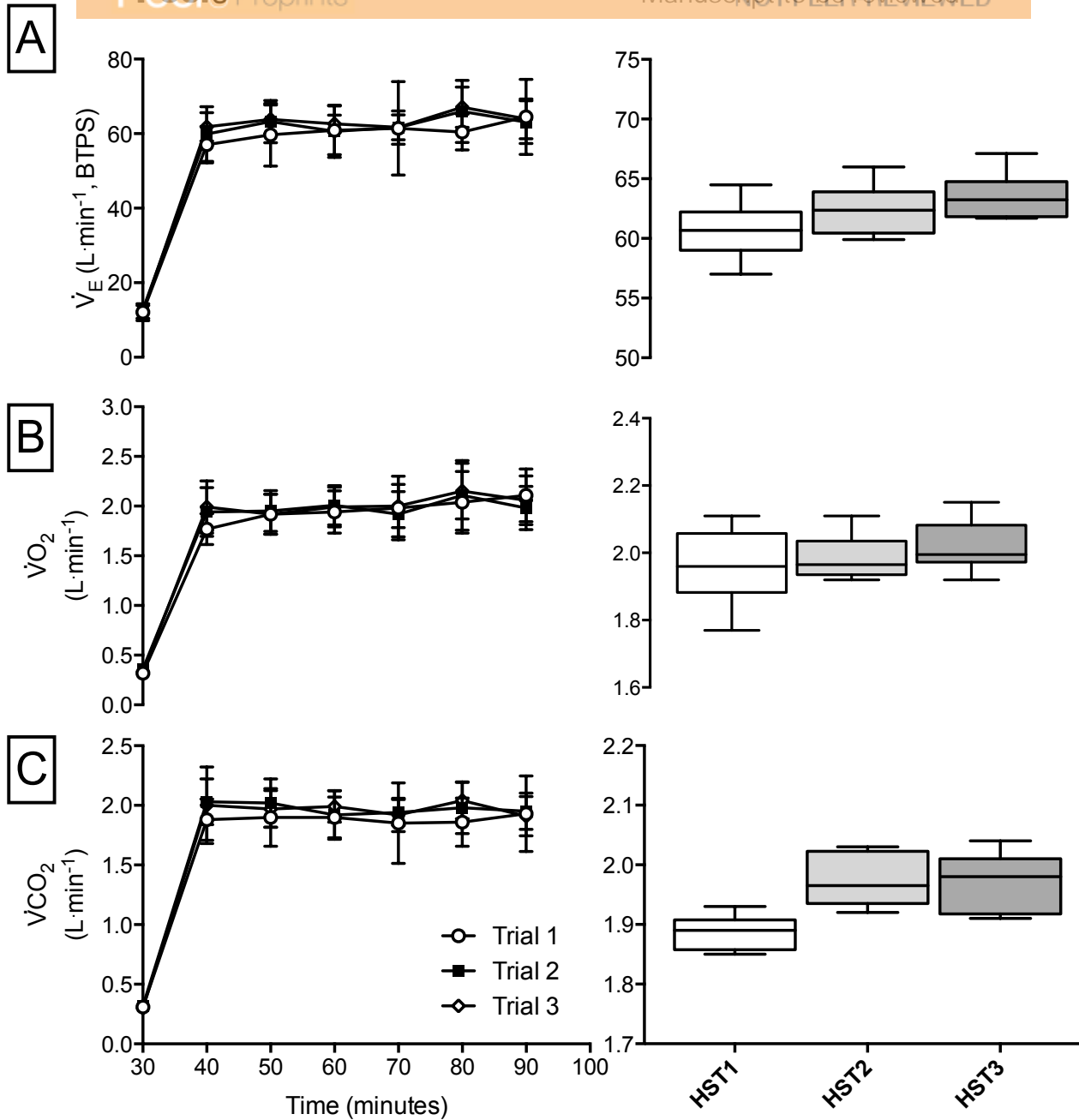
Figure 3(on next page)

Physiological responses to the preloaded time trial in normoxia and hypoxia.

Time trial completion time, mean power output throughout each kilometer, and mean heart rate during each kilometer for the normoxic (A, B, C) and hypoxic (D, E, F) time trials. Box plots display the individual performance times (dots), and highest and lowest times recorded. Line graphs display the mean ± SD responses throughout (n = 6).