Reliability of time to exhaustion treadmill running as a measure of human endurance capacity

Abdullah F. Alghannam [1*]

*Corresponding author

E-mail: A.F.Alghannam@bath.ac.uk

Dawid Jedrzejewski[1]

Mark G. Tweddle[1]

Hannah Gribble [1]

James Bilzon[1]

James A. Betts [1]

[1] Human Physiology Research Group, Department for Health, University of Bath, Bath, BA2 7AY, UK

Abstract:

Little if any research has examined the variability in time to exhaustion (TTE) during submaximal treadmill running. This study investigated the test-retest reliability of submaximal treadmill TTE as a measure of endurance capacity. Sixteen endurance-trained males (n=14) and females (n=2) completed a run to exhaustion at 70% $\dot{V}O_{2max}$ (T1) and repeated the same run three weeks later (T2). At 30 min intervals during each run, expired gas, heart rate (HR) and ratings of perceived exertion (RPE) were collected. Mean ± SD TTE was 96 ± 20 min in T1 versus 101 ± 29 min in T2 (P=0.3). The mean ± 95% confidence intervals (CI) of the coefficient of variance (CV) was 5.4% (1.4 – 9.6). The average intraclass correlation coefficient (±95% CI) was 0.88 (0.67 – 0.96) between trials. The respiratory-exchange ratio was not different between trials, T1: 0.87±0.1 and T2: 0.89±0.1 (P>0.05) and neither was total whole-body carbohydrate oxidation (2.1±0.4 $g \cdot min^{-1}$ and 2.3±0.6 $g \cdot min^{-1}$), fat oxidation (0.6±0.2 $g \cdot min^{-1}$), HR (178±8 and 175±7 $beats \cdot min^{-1}$) or RPE (17±3 and 16±3). These results suggest that use of prolonged treadmill-based TTE can be a reliable research tool to assess human endurance capacity in aerobically-trained men and women.

Introduction:

To examine the effects of different nutritional interventions on performance and fatigue, exercise protocols often require individuals to exercise until the point of volitional exhaustion (time to exhaustion; TTE) or complete a set distance or amount of work as quickly as possible (time trial; TT). A meaningful physiological performance test requires reliability, such that reproducible results are obtained when a test is performed repeatedly [16]. The internal validity of TTE is well established to measure fatigue [5,23,33], with external validity also apparent for various exercise scenarios and occupational tasks [1]. However, a focus of much attention and debate is the reliability of TTE in study protocols, particularly relative to TT tests [11,16], evidenced by some investigations reporting insufficient reliability of TTE with a coefficient of variation (CV) of 13-27 % [18,21,24]. Nevertheless, it can be postulated that the reliability of any exercise protocol will depend on the specific measures/protocol employed and thus no single CV can be assigned to a given exercise test, which can be expected to vary within and between laboratories.

The mode and intensity of a TTE test appears to influence the degree of variability [24]. With longer duration fixed intensity, TTE protocols may carry greater variability than those with shorter duration [11,16]. Nonetheless, it has been argued that TTE may not be a reliable measure when the exercise intensity is increased above 80 % $\dot{V}O_{2max}$ [19], which may be related to differences in fatigue mechanisms and inter-individual differences once test intensity approximates the aerobic/anaerobic 'threshold' [6]. To support this notion in TTE exercise protocols, it has been shown that exercise intensities < 80 % $\dot{V}O_{2max}$ appear to have lower CV (≈9 %) [13,23] than those utilising maximal short-duration protocols [7], but this is not universal

[10]. The reliability of different treadmill running protocols has been investigated in the context of TT performance [26-28]. To our knowledge, however, there is no information pertaining to the reliability of TTE during submaximal treadmill running in endurance-trained individuals. In concordance, the aim of this study was to investigate the test-retest reliability of prolonged moderated- to high-intensity run time to exhaustion as a measure of human endurance capacity.

Materials and Methods

Participants

Sixteen healthy active males (n=14) and females (n=2) participated in this study (Mean ± SD; age 23 ± 7 years, body mass (BM), 70.4 ± 8.6 kg, and $\dot{V}O_{2max}$ 62 ± 5.4 $mL \cdot kg^{-1} \cdot min^{-1}$). The chosen target population were healthy non-smoking recreationally active men and women who included endurance training in the form of running as a central component (≥2 hours/week) of their training, and are aged between 18-48 years old. Each individual was fully briefed regarding the nature of the study and provided informed consent prior to participation. The study was approved by National Health Services Research Ethics Committee (Ref: 09/H0101/82) in accordance with international standards [14] and was part of a wider project on the effects of nutrition on post-exercise recovery and repeated exercise [1].

Preliminary measurements

Participants underwent one preliminary visit to determine submaximal ($\dot{V}O_2$) and maximal ($\dot{V}O_{2max}$) oxygen uptakes [31] on a motorised treadmill (Ergo ELG70, Woodway, Weil am Rhein, Germany) and a second familiarisation visit during which participants underwent the main experimental procedures exactly as per the main experimental arms described below.

Experimental design

Participants completed two main trials (T1 and T2) in a repeated measures experimental design separated by a space of three weeks (95% CI; 2-3 weeks), an interval which was consistent with a previous assessment of reliability of endurance performance using a treadmill [27]. Each trial involved a run to exhaustion at 70 % $\dot{V}O_{2max}$ until the point of volitional exhaustion. A weighed dietary record was completed 48 hours preceding the familiarisation trial, and was subsequently repeated prior to the commencement of the main trials (2665 ± 601 $kcal.d^{-1}$; 53 ± 5 % CHO; 22± 10 % fat; 25 ± 7 % protein). Participants were provided with a standardised meal (760 kcal; 57 % CHO; 24 % Protein; 19 % fat) to be consumed as their final caloric intake before testing, consumed in the evening (12 ± 1 h) before the familiarisation trial and replicated prior to each main trial. Participants also abstained from alcohol consumption and refrained from strenuous physical activity (with any light exercise recorded and matched) during this 48 h period of dietary and lifestyle standardisation between trials.

Experimental protocol

The experimental protocol is described in detail elsewhere [1]. Briefly, each participant arrived to the laboratory in the morning (≈08:00) in a post-absorptive state following ≥ 10 h overnight fast. After confirming their informed consent to take part in the study, each participant provided a urine sample to assess hydration via cryoscopic osmometer (Advanced Instruments, Inc, Norwood, MA, USA) and adequate hydration was assumed for osmolality values ≤ 900 mOsm·kg^{-1} [29] and then nude BM was recorded (Weylux 424, Fereday & Sons Ltd., UK) to the nearest 0.1 kg. A standardised 5-min warm-up at 60 % $\dot{V}O_{2max}$ was used before running at a speed equivalent to 70 % $\dot{V}O_{2max}$ until the point of volitional exhaustion. Once participants indicated that they were unable to sustain the exercise intensity, the prescribed running speed was reduced to walking (walk-1; 4.4 km·h^{-1}) for 2-min intervals. This procedure was repeated for a second time (walk-2) and only on the third occasion when participants indicated they could no longer sustain the prescribed exercise intensity was fatigue accepted and time to exhaustion (TTE) recorded. The time during walking was excluded from the total TTE. Average one-minute hearts rates (Polar FT2, Kempele, Finland), and ratings of perceived exertion (RPE; [8]) were taken at 30 min intervals during the runs. Water intake was permitted *ad libitum* during the familiarisation trial and then matched for subsequent trials (0.4 ± 0.3 liters). Nude BM was then recorded to assess hydration status through percentage change in mass. Ambient temperature and humidity were recorded at 60 min intervals throughout the trials using a portable weather station (WS 6730; Technoline, Berlin, Germany). Participants exhibited adequate degrees of pre-exercise hydration status among experimental conditions (*P*= 0.7), with mean urine osmolality values of 533 ± 298 and 506 ± 266 mOsm.kg^{-1} in T1 and T2, respectively. Similarly, the sweat loss during the exhaustive run was not significantly different (*P*= 0.7) across T1 and T2, as reflected by -1.8 ± 0.8 % and -1.7 ± 0.8 %, respectively. Environmental conditions were standardised between the experimental trials, with barometric pressure (741 ± 9 and 741 ± 9 mmHg; *P*= 0.7), ambient temperature (19.8 ± 0.9 and 19.8 ± 0.9 °C; *P*= 0.9) and humidity not statistically different (47 ± 7 and 47 ± 8 %; *P*= 0.9) between T1 and T2, respectively. Background music and verbal encouragement was standardised between trials [4] and participants were unaware of the time elapsed during the exercise capacity test.

Statistical analysis

Paired differences were tested for normality using the Shapiro-Wilk test and single comparisons between two means were analysed by using a paired sample *t*-tests when normally distributed. Where data were deemed in violation of normality, a non-parametric equivalent (i.e Wilcoxon signed rank test) was employed to compare medians. A visual inspection using quantile-quantile (*Q*-*Q*) plots showed that run times to exhaustion data were within the quantiles of a standard normal distribution, with a skeweness of 0.029 (standard error; SE= 0.564) and kurtosis of -0.902 (SE= 1.091) in T1 and a skewness of 0.923 (SE= 0.564) and kurtosis of 0.010 (SE= 1.091). Coupled with the Shapiro-Wilk test ($P > 0.05$) the data can be assumed as normally distributed [15,30]. The mean coefficient of variation (standard deviation/mean x 100) with the associated 95 % confidence intervals (CI) was calculated to establish error in measurement between TTE protocols (i.e. absolute reliability). Intra-class correlation coefficients (ICC) with a two-way mixed effects was used to determine relative test-retest reliability, with ≤ 0.50 indicating moderate reliability, 0.70-0.89 high reliability, and ≥ 0.90 very high reliability [25]. Relative reliability was supplemented by including Pearson product moment correlation coefficient (*r*) and coefficient of determination (R^2) to assess the association between the performed exercise protocols. Bland and Altman plots were used to determine absolute bias using 95 % limits of agreement (LoA). A two-way linear mixed model (trial x time) was used to identify differences between trials over time, with participants inserted as random effects and trial and time as fixed effects (covariates). This statistical approach allows to model for nonlinear changes in a dependent variable across time, which is often associated with a number of physiological variables (e.g. heart rate), while also acknowledging both group and individual changes over time [20]. A post hoc analysis (G*power version 3.1.7; University Düsseldorf, Düsseldorf, Germany) based on the sample of the current experiment (n=16) provided 80% power to detect a correlation of ≥ 0.5 between repeated bouts of exhaustive treadmill running at alpha level 0.05 with a coefficient of determination of 0.73. Statistical procedures were performed using commercially available software (IBM SPSS version 21.0, SPSS Inc., Chicago, IL) and significance was set at an alpha level ≤ 0.05. Unless otherwise stated, all results were reported as the mean ± standard deviation (SD).

Results

There was no systematic bias in time to exhaustion between T1 and T2, and neither were any trial order differences apparent at the relative time points of walk-1 (84 ± 19 versus 90 ± 29 min), walk-2 (91 ± 19 versus 96 ± 29 min) or TTE (96 ± 19 versus 101 ± 29 min), respectively (Figure 1). The mean ± 95 % CI of the typical error of measurement expressed as CV was 5.4 % (1.4 – 9.6) at the point of exhaustion, hence was marginally lower than walk-1 (6.3 %; 1.7 – 11.0) and walk-2 (5.7 %; 1.4 – 10). Percent change in mean TTE ± 95 % CI across T1 and T2 trials was 3.7 % (-4 – 11.3), which was also lower than percent change in mean to reach walk-1 (5.7 %; -3 – 15) and walk-2 (4.4 %; -4 – 13). In relation to differences from walk-1 to the point of volitional exhaustion (TTE), no trial x time interactions were identified (P= 0.8). Participants were able to run 11 ± 3 min from walk-1 until the point of volitional exhaustion, although this was not statistically significant (P= 0.07). Participants were able to run for 6 ± 2 min from walk-1 to walk-2 and 5 ± 2 min from walk-2 to TTE.

Relative exercise intensities were successfully matched between trials and averaged 69.6 ± 4.1 % $\dot{V}O_{2max}$ in T1 and 69.2 ± 3.8 % $\dot{V}O_{2max}$ in T2. This was additionally verified by the overall heart rate (178 ± 8 and 175 ± 7 beats·min^{-1}) and RPE (17 ± 3 and 16 ± 3) measurements in T1 and T2, respectively (Table 1). The respiratory exchange ratio using indirect calorimetry was similar between T1 (0.87 ± 0.1) and T2 (0.89 ± 0.1). Whole-body carbohydrate oxidation was 2.1 ± 0.4 g·min^{-1} and 2.3 ± 0.6 g·min^{-1}, while fat oxidation was 0.6 ± 0.2 g·min^{-1} and 0.6 ± 0.2 g·min^{-1} in T1 and T2, respectively. Thus, no systematic bias was observed in exercise intensity, HR, RPE or substrate metabolism across trials ($P > 0.05$).

The average intra-class correlation coefficient (± 95 % CI) revealed high reliability of 0.88 (0.67 – 0.96) at the time of exhaustion between the trials. A similar pattern was shown at walk-1 and walk-2 with ICC of 0.86 (0.62 - 0.95) and 0.87 (0.65 – 0.96), respectively. A high correlation coefficient (r= 0.86; 0.62 – 0.95) was observed between the two TTE exercise protocols ($P < 0.001$; Figure 2), as evidenced by the coefficient of determination (R^2= 0.73). The correlation coefficient and coefficient of determination were similarly high during walk-1 (r=0.84 (0.59 – 0.94); R^2=0.70) and walk-2 (r= 0.85 (0.62 – 0.95); R^2=0.73). Figure 3 illustrates the difference between trials through the use of Bland and Altman plots displaying the absolute bias and 95 % LoA. The absolute bias ± 95 % LoA was 4 ± 31 minutes.

Discussion

Relative reliability is concerned with measurement error relative to between-subject variability, whereas absolute reliability is the degree to which repeated measurements vary for individuals [3]. This study investigated the test-retest reliability of prolonged moderate- to high-intensity run time to exhaustion as a measure of human endurance capacity. Here we demonstrate high relative and absolute test-retest reliability of time to exhaustion as measure of endurance capacity during prolonged (> 60 min) treadmill running and thus may be acceptable to detect small but meaningful effects upon endurance capacity. To our knowledge, this is the first study to assess the reliability of a treadmill based prolonged exhaustive exercise protocol.

Previous comparisons between TT and TTE have examined the reliability of these measures, and notable variability in the reliability of TTE exercise protocols can be observed relative to TT measures [11,16]. However, numerous extraneous variables may affect reliability such as the mode of exercise employed, individuals being tested, laboratory environment and, importantly, familiarisation. The low measurement error presented here is in agreement with a study using prolonged (> 60 min) cycling [23] but is in contrast to other prolonged cycling-based investigations [13,18]. Moreover, the present findings are not consistent with other shorter duration running-based TTE protocols of only 6-18 min [7,21]. Gleser and Vogel (1971) have reported the presence of a systematic order effect and an associated higher degree of variability (CV= 13 %) when repeated TTE exercise bouts were performed in untrained participants unaccustomed to exercise tests to exhaustion wand were not familiarised to the exercise protocol [13]. This was further corroborated by McLellan et al. (1995), who reported a CV of 17 % in participants who were not aerobically trained [24]. Indeed, TT may carry greater reliability from the outset, particularly if using relevant athletes and/or if repetitive familiarisation with TTE is to be avoided. However, the current findings further support the notion that the reliability of an exercise protocol may vary substantially between laboratories and the degree of variability in TTE tests can be influenced by familiarisation, participant characteristics and the specific protocol utilised. In the current study, young, healthy and aerobically trained participants underwent a familiarisation trial identical to the experimental procedures. These considerations may have a profound impact in the sensitivity of an exercise protocol [13,24]. Critically, no training effect would therefore be expected nor was detected in our cohort of participants, as evidenced by the similar relative exercise intensity, RPE and HR response between trials. In addition to the aforementioned steps, we also employed a number of measures to increase the absolute reliability of the TTE protocol. We standardised the nutritional status of participants prior to the commencement of each trial to diminish any carry over effect to metabolism and/performance [9]. Indeed, it has been postulated that longer duration protocols may present higher variation than short protocols [11], mainly associated with the known role of prior nutrition and the inherent link between pre-exercise glycogen availability and the endurance capacity [5].

With regards to the prolonged nature of the adopted TTE protocol, it is also conceivable that environmental conditions and hydration status may influence the reliability of the tests. In the present study, the laboratory environment was successfully standardised as shown in the similar ambient conditions (temperature, humidity and barometric pressure), pre-trial hydration status and changes in BM across the experimental procedures. Collectively, the control measures conducted in the present study may have contributed to the minimised absolute measurement error (CV = 5.4 %) associated with previously reported TTE protocols [7,21].

An important quality of a valid test of prolonged endurance capacity is that participants reach a 'metabolic endpoint' (i.e. volitional exhaustion coincident with metabolic substrate disturbance/depletion as opposed to injury, stomach discomfort or mere boredom). To help participants gauge their fatigue more accurately, we incorporated two 2-min walks in the current protocol before terminating the TTE test. Prior to each trial, participants were informed not to use any of the walks for tactical reasons but rather to indicate the need to reduce the running speed only once they felt unable to sustain the exercise intensity (i.e. where they would stop if there were to be no further opportunity to continue). These individuals were able to run for 11 ± 3 min from walk-1 to TTE, which represents ≈11 % of the total time to exhaustion during the endurance capacity test. Muscle glycogen resynthsis would be restored at very low rates in the absence of carbohydrate ingestion (0.5 mmol.kg dry mass^{-1}.min^{-1}; [17]). Thus, given that muscle glycogen utilisation would be 3.2 mmol.kg dry mass^{-1}.min^{-1} during treadmill running the intensity employed here [32], any resynthesis of muscle glycogen during the 4-min of walking could only account for < 1 min of extended exercise time. It is therefore reasonable to suggest that including walking intervals is important to help participants better gauge their level of fatigue and more completely deplete their muscle glycogen levels.

There were several important limitations in the current study. The experiment was not designed to assess the specific role of individual extraneous variables in improving reliability of TTE exercise protocol. As a consequence, it is not possible to determine the precise factor(s) contributing to the lower variability reported here than previously. It is also noteworthy that expired gas and blood sampling measurements were obtained during the TTE protocol (blood data not shown), which may have had a negative influence on the mental concentration of participants [18], albeit that is the context in which TTE tests are normally used. It is also worth stating that only two female participants took part in the current study and thus sex-specific difference on the variability of TTE exercise protocols remain largely unknown. Notwithstanding that, the variability between the two trials in these females (CV=1.5 %) reflects a similar pattern to their male counterparts; a larger sample of female participants is required to establish the presence of any sex-specific differences on variability of TTE measures.

Reliability could be considered as the amount of measurement error that has been deemed acceptable for the effective practical use of a measurement tool [3]. In this respect, it may be considered that TT and TTE are inherently different measurement tools, each encompassing their own level of acceptable 'noise'. Indeed, it has been postulated that the signal-to-noise ratio may be greater in TTE when compared to TT [16]. Accordingly, a reliable test can be judged on the basis of its sensitivity for detecting worthwhile changes [21]. The bias present between the two TTE tests was shown to be 4 % and in accordance can be deemed as acceptable measurement tool under the commonly accepted measurement error of $\leq$ 5% [2]. This is further supported when considering a greater signal-to-noise ratio (i.e. relative higher variability than TT but larger changes in the outcome measure) as demonstrated by > 20 % change typically seen in TTE in response to various experimental interventions [12,22,23], despite the fact that the magnitude of change is likely to be dependent on the specific intervention. Therefore, when considering the culmination of the relative and absolute reliability measurements observed in the current study, it would appear that the current TTE protocol can be a reliable tool for research purposes.

In conclusion, this study demonstrates high absolute and relative reliability between two exhaustive bouts of prolonged moderate- to high-intensity treadmill running in endurance trained participants. The use of aerobically trained individuals who are familiarised to TTE exercise protocol may have contributed in obtaining more reliable measurement outcomes [16]. Moreover, the inclusion of short intervals of reduced intensity during the exercise protocol may be an important consideration to reach a 'metabolic endpoint' and therefore a more accurate reflection of endurance capacity.

References

1. Alghannam AF, Tsintzas K, Thompson D, Bilzon J, Betts JA. Exploring mechanisms of fatigue during repeated exercise and the dose dependent effects of carbohydrate and protein ingestion: study protocol for a randomised controlled trial. Trials 2014; 15: 95

2. Atkinson G. What is this thing called measurement error? In: Reilly T, Marfell-Jones M eds, Kinanthropometry VIII: Proceedings of the 8th International Conference of the International Society for the Advancement of Kinanthropometry (ISAK). London: Routledge; 2003: 3-13

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

4. Atkinson G, Wilson D, Eubank M. Effects of music on work-rate distribution during a cycling time trial. Int J Sports Med 2004; 25: 611-615

5. Bergstrom J, Hermansen L, Hultman E, Saltin B. Diet, muscle glycogen and physical performance. Acta Physiol Scand 1967; 71: 140-150

6. Betts JA, Stevenson E, Williams C, Sheppard C, Grey E, Griffin J. Recovery of endurance running capacity: effect of carbohydrate-protein mixtures. Int J Sport Nutr Exerc Metab 2005; 15: 590-609

7. Billat V, Renoux JC, Pinoteau J, Petit B, Koralsztein JP. Reproducibility of running time to exhaustion at VO2max in subelite runners. Medicine and Science in Sports and Exercise 1994; 26: 254-257

8. Borg GA. Perceived exertion: a note on "history" and methods. Med Sci Sports 1973; 5: 90-93

9. Braun B, Brooks GA. Critical importance of controlling energy status to understand the effects of "exercise" on metabolism. Exercise and sport sciences reviews 2008; 36: 2-4

10. Coggan AR, Costill DL. Biological and technological variability of three anaerobic ergometer tests. Int J Sports Med 1984; 5: 142-145

11. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. Sports Medicine 2008; 38: 297-316

12. Galloway SD, Maughan RJ. The effects of substrate and fluid provision on thermoregulatory and metabolic responses to prolonged exercise in a hot environment. J Sports Sci 2000; 18: 339-351

13. Gleser MA, Vogel JA. Endurance exercise: effect of work-rest schedules and repeated testing. J Appl Physiol 1971; 31: 735-739

14. Harriss DJ, Atkinson G. Ethical standards in sport and exercise science research: 2014 update. Int J Sports Med 2013; 34: 1025-1028

15. Henderson AR. Testing experimental data for univariate normality. Clinica chimica acta; international journal of clinical chemistry 2006; 366: 112-129

16. Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Medicine 2001; 31: 211-234

17. Ivy JL, Katz AL, Cutler CL, Sherman WM, Coyle EF. Muscle glycogen synthesis after exercise: effect of time of carbohydrate ingestion. Journal of Applied Physiology 1988; 64: 1480-1485

18. Jeukendrup A, Saris WH, Brouns F, Kester AD. A new validated endurance performance test. Medicine and Science in Sports and Exercise 1996; 28: 266-270

19. Krebs PS, Powers SK. Reliability of a laboratory endurance test. Medicine and Science in Sports and Exercise 1989; 21: S10.

20. Krueger C, Tian L. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biological research for nursing 2004; 6: 151-157

21. Laursen PB, Francis GT, Abbiss CR, Newton MJ, Nosaka K. Reliability of time-to-exhaustion versus time-trial running tests in runners. Med Sci Sports Exerc 2007; 39: 1374-1379

22. Maughan RJ, Bethell LR, Leiper JB. Effects of ingested fluids on exercise capacity and on cardiovascular and metabolic responses to prolonged exercise in man. Exp Physiol 1996; 81: 847-859

23. Maughan RJ, Fenn CE, Leiper JB. Effects of fluid, electrolyte and substrate ingestion on endurance capacity. Eur J Appl Physiol Occup Physiol 1989; 58: 481-486

24. McLellan TM, Cheung SS, Jacobs I. Variability of time to exhaustion during submaximal exercise. Can J Appl Physiol 1995; 20: 39-51

25. Munro BH. Statistical methods for health care research 3rd edition New York Lippincott; 1997

26. Rollo I, Williams C, Nevill A. Repeatability of scores on a novel test of endurance running performance. J Sports Sci 2008; 26: 1379-1386

27. Russell RD, Redmann SM, Ravussin E, Hunter GR, Larson-Meyer DE. Reproducibility of endurance performance on a treadmill using a preloaded time trial. Med Sci Sports Exerc 2004; 36: 717-724

28. Schabort EJ, Hopkins WG, Hawley JA. Reproducibility of self-paced treadmill performance of trained endurance runners. Int J Sports Med 1998; 19: 48-51

29. Shirreffs SM, Maughan RJ. Urine osmolality and conductivity as indices of hydration status in athletes in the heat. Medicine and Science in Sports and Exercise 1998; 30: 1598-1602

30. Tabachnick BG, Fidell LS. Using multivariate statistics 6th edition. Essex, UK: Pearson Education Limited 2014

31. Taylor H, Buskirk E, Henschel A. Maximal oxygen uptake as an objective measure of cardio-respiratory performance. Journal of Applied Physiology 1955; 8: 73-80

32. Tsintzas OK, Williams C, Boobis L, Greenhaff P. Carbohydrate ingestion and single muscle fiber glycogen metabolism during prolonged running in men. Journal of Applied Physiology 1996; 81: 801-809

33. Wilson WM, Maughan RJ. Evidence for a possible role of 5-hydroxytryptamine in the genesis of fatigue in man: administration of paroxetine, a 5-HT re-uptake inhibitor, reduces the capacity to perform prolonged exercise. Exp Physiol 1992; 77: 921-924

Table 1. % $\dot{V}O_{2max}$, Ratings of perceived exertion (RPE) and heart rate responses to the first and second run times to exhaustion. TTE, time to exhaustion; T1, first trial; T2, second trial. Data are means ± SD

	Pre	30 min	60 min	90 min	TTE
% $\dot{V}O_{2max}$					
T1	-	67.7 ± 5.6	69.1 ± 4.6	71.5 ± 1.7	70.0 ± 4.4
T2	-	67.5 ± 3.4	68.3 ± 3.9	70.8 ± 1.9	70.3 ± 5.9
RPE (6-20)					
T1	-	13 ± 1	16 ± 2	18 ± 1	20 ± 0
T2	-	13 ± 2	15 ± 2	17 ± 2	20 ± 0
Heart rate (beats·min^{-1})					
T1	72 ± 13	172 ± 8	179 ± 9	181 ± 10	181 ± 9
T2	72 ± 12	169 ± 6	175 ± 7	178 ± 9	178 ± 9

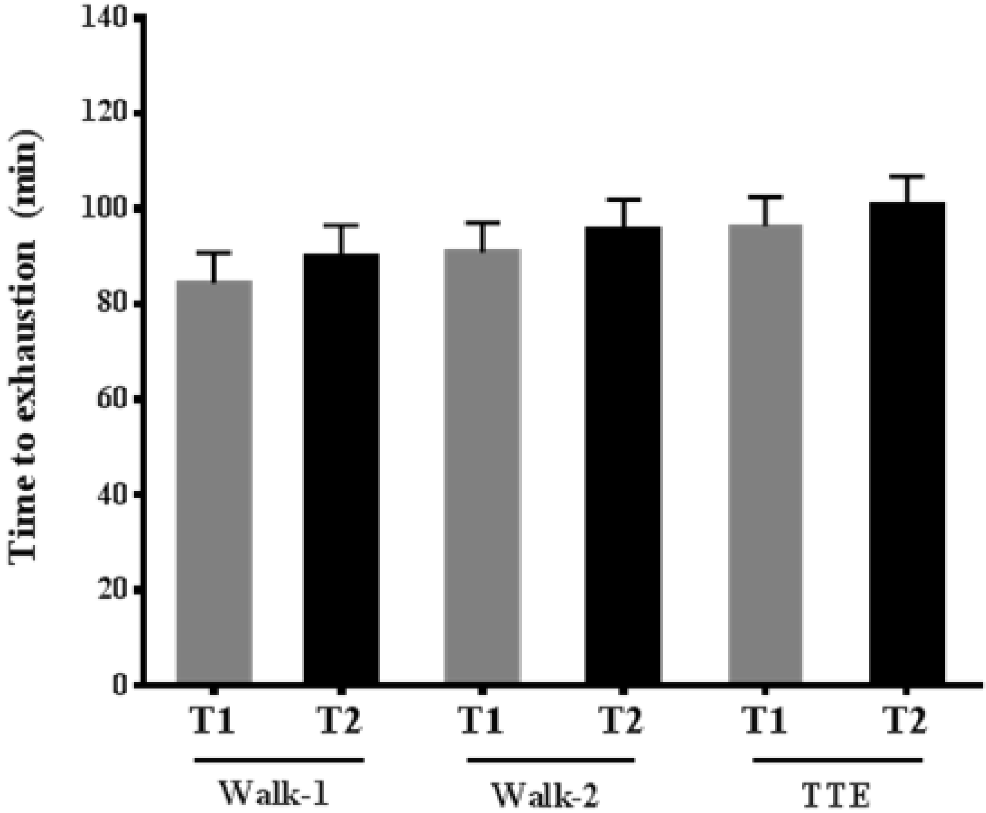

Figure 1. Run times compared between trials during walk-1, walk-2 and time to exhaustion. Data are means ± nCI

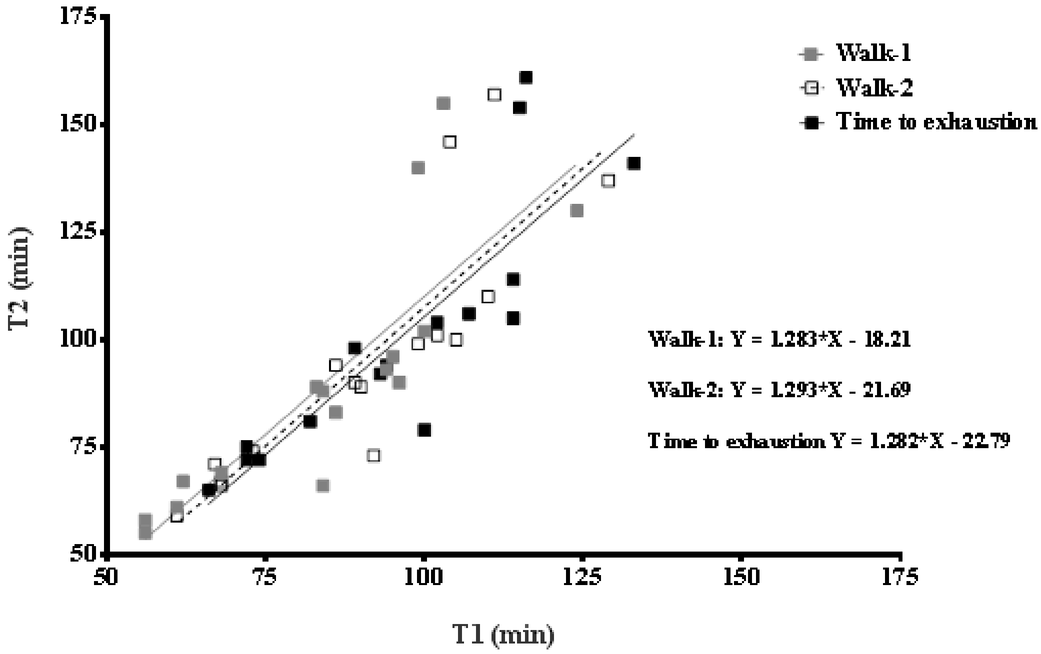

Figure 2. Scatterplot showing the relationships between the two run times to exhaustion. Straight lines represent best fit for walk-1, walk-2 and time to exhaustion (*P*< 0.001). T1, first trial; T2, second trial

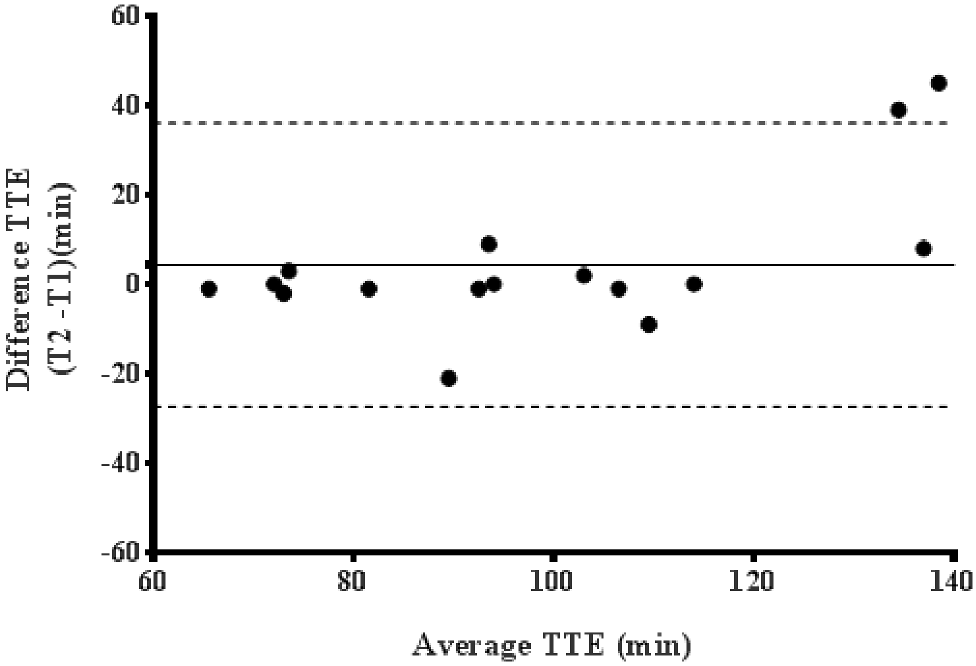

Figure 3. Bland-Altman plot of absolute agreement between two run times to exhaustion TTE, time to exhaustion; T1, first trial; T2, second trial