J Sci Cycling. Vol. 5(3), 35-41

RESEARCH ARTICLE **Open Access**

Test-retest reliability of a 16.1 km time trial in trained cyclists using the CompuTrainer ergometer

Andy Sparks[1]✉, Emily Williams[2], Hollie Jones[3], Craig Bridge[1], David Marchant[1] and Lars McNaughton[1]

Abstract

Laboratory based cycling time trials (TT) are widely used by both researchers and practitioners, as a method of assessing cycling performance in a controlled environment. Assessments of performance often use TT durations or distances between 20 min and one hour and in the UK the 10 mile (16.1 km) TT is the most frequently used race distance for trained cyclists. The 16.1 km TT has received relatively minimal, but increased attention as a performance criterion in the literature. Therefore, the aim of this study was to assess the reliability of 16.1 km TT performance in a large cohort of trained cyclists using the CompuTrainer cycle ergometer. Trained male cyclists (n = 58, mean±SD age 35±7 yr, height 179±6 cm, weight 79.1±9.4 kg, VO_{2max}. 56.6±6.6 ml.kg.min^{-1}, PPO 365±37 W) performed an initial incremental exercise test to determine PPO and VO_{2max}. The participants then performed two 16.1 km TT on a CompuTrainer cycle ergometer separated by 3-7 days. Differences in time, power output and speed were determined using a Wilcoxon signed ranks or paired t-tests. Reproducibility of the TT performance measures was performed using the coefficient of variation (CV), intraclass correlations, and typical error (TE). There were no differences between any of the performance criteria for the whole cohort (Mean difference = 0.06 min, 0.09 km.h^{-1}, 1.5 W, for time, mean speed and power respectively) between TT1 and TT2. All TT performance data were very reproducible (CV range = 1.1-2.7%) and demonstrated trivial or small TE. The slower cyclists demonstrated marginally lower reliability (CV range = 1.3-3.2%) compared to the fastest group (CV range = 0.7-2.0%). The 16.1 km TT on the CompuTrainer represents a very reliable performance criterion for trained cyclists. Interpretation of test-retest performance outcomes should be performed in the context of the TE of each performance indicator.

Keywords: cycling, power output, typical error, reproducibility, ecological validity.

✉ **Contact email:** andy.sparks@edgehill.ac.uk (A. Sparks)

[1] *Department of Sport and Physical Activity, Edge Hill University, Lacanshire, UK*
[2] *Centre for Sports Perfomance, Leeds Beckett University, Leeds, Yorkshire, UK*
[3] *School of Psychology, University of Central Lancashire, Preston, Lancashire, UK*

Received: 25 November 2016. Accepted. 19 December 2016.

Introduction

Laboratory based cycling time trials (TTs) have been extensively used in the literature as exercise performance criteria. Distances of between 4 (Altareki et al. 2009; Stone et al. 2012) and 40 km (Jones et al. 2015a; Maunder et al. 2016) are most frequently used in cycling based research. Performance of these TTs typically represent exercise times of approximately 5-65 minutes in trained cyclists, and it is this wide spectrum which elicits diverse physiological responses, allowing the scrutiny of a broad range of experimental designs. Whilst the appropriateness of different performance test criteria has been debated (Currell and Jeukendrup 2008), it is the TT which has been extensively used, because it represents the most directly related assessment of actual cycling performance. Researchers, coaches and athletes that use such measurements of performance to monitor training, or to evaluate the efficacy of an intervention, are only able to determine meaningful changes if the sensitivity of the performance criteria is known.

The sensitivity of a test to determine a performance change is affected by the reliability, validity and accuracy of the equipment on which the exercise is being undertaken (Hopkins 2000). Consequently this research topic has received considerable attention in the literature, and there have been several attempts to provide reliability and validity data for a wide variety of cycling ergometry equipment (Astorino and Cottrell 2012; Earnest et al. 2005; Hopker et al. 2010; Kirkland et al. 2008) and exercise protocols (Che Jusoh et al. 2015; Noreen et al. 2010; Peveler 2013; Zavorsky et al. 2007). These studies suggest that the reliability, expressed as a coefficient of variation (CV) for power output, tends to be between 1.4% for steady rate cycling on the SRM (Kirkland et al 2008) and 17% for peak power output during incremental exercise on the Lode cycle ergometer (Earnest et al. 2005). Other factors such as training status and the time between trials (Clark et al. 2014) have also been shown to affect reliability. Studies that have examined the effect of training status on reliability, have demonstrated that power output may be lower in less well trained groups, by between 3.6% (Zavorsky et al. 2007) and 4.1% (Hopker et al. 2010). In studies that have investigated the reliability of cycling protocols, CV's for power are usually reported between 3.4 and 4.9% for a 20 km TT (Zavorsky et al. 2007) and a 12.9 km uphill TT (Noreen et al. 2010), respectively.

In well trained cyclists, performance changes of as little as 1% are likely to be meaningful (Lamberts et al., 2009), so the determination of the reliability of measurement criteria, to allow the analysis of performance, is extremely important. A further factor which may also affect the reliability of performance tests, is athlete or research participant familiarisation (Altareki et al. 2009). Familiarisation with endurance events in particular, plays an important role in the development of pacing schemas (Micklewright et al. 2010), and it is a vital consideration in the assessment and interpretation of the meaningful worthwhile changes, that may be elicited in performance, between experimental TT's (Currell and Jeukendrup 2008; Lamberts et al. 2009). Furthermore, familiarisation is an important experimental methodological consideration for sport and exercise scientists, since it may influence both the number of experimental trials needed in a study, and the sensitivity of the exercise performance criteria. It is therefore potentially very useful to know the test-retest reliability of performance measurements that are used in this manner. In the UK the most frequently used distance in road based TT competitions is 16.1 km (10 miles); a distance which traditionally has been less frequently selected as a laboratory performance criteria, despite its ecological validity, and the propensity of cyclists to be highly familiarised. More recently, this distance has received renewed attention because of its validity, and the familiarisation of cyclists with it (Jones et al. 2015a; Jones et al. 2015b and Williams et al. 2015).

The CompuTrainer Pro (RacerMate, Seattle, USA) represents an ergometer that allows such TT's to be easily performed, whilst also allowing athletes to use their own bicycles. This ergometer has received some attention in the cycling literature with regards to assessing both its reliability and validity (Davison et al. 2009; Clark et al. 2016). Davison et al. (2009) used a variety of ambient temperature conditions to show that initial constant power cycling, of at least 2 min, is required to off-set calibration pressure declines that may reduce the reliability and validity of the performance data. These authors also showed that at lower ambient temperatures (15 and 20ºC) the CompuTrainer underestimated power output compared to an SRM powermeter by 3.7%. Additionally, Clark et al. (2016) used the CompuTrainer to determine the validity of a 3 min all out exercise test to calculate critical power. They observed that the CV's for critical power measurements ranged between 0.5 and 3.1%, depending on the model estimate used, which represents acceptable measurement consistency. Indeed, it is consistency or reliability of ergometry equipment, which is arguably more important than validity. This is especially true, when repeated measurements are being taken, and when the data are not being directly compared to either other ergometers, or field based measurements (Davison et al. 2009). There are currently no studies that have attempted to determine the reliability of a 16.1 km TT on the CompuTrainer, despite the high level of ecological validity that this TT distance has as a cycling performance criteria. Therefore, the aim of this study was to determine the test-retest reliability of a laboratory simulated 16.1 km TT in a large cohort of trained cyclists using the CompuTrainer and RacerMateOne software.

Materials and methods

Participants

Fifty eight experienced, trained (De Pauw 2013) male cyclists (age, 34.8 ± 7.3 yr; height, 178.5 ± 5.9 cm; weight, 79.1 ± 9.4 kg; relative VO_{2peak}, 56.6 ± 6.6 $ml.kg.min^{-1}$; PPO, 365.8 ± 37.2 W) were recruited for this study. Each participant had a minimum of two year's prior experience of riding competitive 16.1 km TTs on the road and/or in a laboratory. Participants were instructed to attend the laboratory in a well hydrated state and follow the same 24 h pre-exercise dietary intake before each trial as if preparing for a competitive race. In this 24 h period, participants were also required to abstain from intense exercise and replicate any activity before future laboratory visits. Prior to data collection, participants underwent pre-exercise health screening and gave signed informed consent in accordance with the recommendations of Harriss and Atkinson (2013). The study was approved by the Departmental Ethics Committee.

Incremental Exercise Protocol

The cyclists were required to attend the laboratory at the same time of day (Atkinson and Reilly 1996) on three occasions, separated by between three and seven days. The initial visit was used to determine PPO and VO_{2peak} of participants using a graded exercise test on a laboratory cycle ergometer (Excalibur Sport, Lode, Netherlands). The participants were required to undertake a 5 min warm up at 100 W, and initial individualised workloads were determined using recognised British Cycling guidelines (Wooles et al. 2003), followed by an incremental protocol consisting of 1 min stages with 20 W increments, until volitional exhaustion had been achieved (Midgley et al. 2007). Throughout the test, breath-by-breath pulmonary ventilation and gas exchange data were recorded using a metabolic cart (Oxycon Pro, Jaeger, GmbH Hoechburg, Germany), which has previously been shown to be valid and reliable (Foss and Hallén 2005). Pulmonary oxygen uptake data were analysed using 20 s averages. The highest VO_2 measurement recorded over a 20 s period was used to classify VO_{2peak} (Dwyer 2004).

Time Trial Protocols

Following the initial laboratory visit, participants were required to bring their own TT bicycles to the final two TT sessions. Participant's bicycles were then mounted onto the ergometer (CompuTrainer, RacerMate, Seattle, USA) and rear tyre pressure was then set to 100 psi prior to the start of the warm up period. The warm up period was 10 min in duration and was performed at 70% of the individual HRmax observed in the maximal exercise test. This protocol was used to habituate the participants and to warm the ergometer and tyre, prior to calibration in accordance with the manufacturer's guidelines. Subsequent calibration of the ergometer took place immediately after the 10 min warm up, and required the participants to increase their speed to 40 $km.h^{-1}$ a minimum of three, but not more than six times. This was in order for the resistance of the tyre on the flywheel to be adjusted to the appropriate level. Following the calibration, participants were then allowed a two minute rest period, prior to the start of their TT's. During this pre TT period, the participants remained on their bicycles and were reminded that the aim of the task was to complete the TT distance in the fastest time possible. During the TT's a fan (CAM5002, Clarke, Essex, UK) was positioned diagonally to the front right of the ergometer, 2 m away from the participants face. Airflow was determined by the cyclist and replicated for the subsequent trial. Participants were allowed to consume only water during the trials, but this was in an *ad libitum* manner (which later demonstrated conformity between trials). Participants were blinded to all performance feedback data apart from distance completed. Time, speed and power output data were then recorded using the specific ergometer software (Mate One, Racer Mate, Seattle, USA) at a frequency of 34 Hz, stored, and later downloaded and exported into Excel 2013 (Microsoft Corp., Redmond, Washington, USA).

Statistical analysis

Data were analysed for the entire cohort of participants as well as the fastest and slowest half of the cyclists. All data were assessed for normality using standard graphical procedures (Grafen and Hails 2002). Thereafter, differences in time, speed, and power between TT1 and TT2 were determined using paired t-tests or Wilcoxon's Signed Rank test in accordance with the recommendations of Atkinson and Nevill (1998). Differences between groups were assessed using independent samples t-tests. Effect sizes (ES) were calculated using Cohen's d for t-tests, interpreted as 0.2 = small effect, 0.5 = medium effect and ≥0.8 = large effect. For Wilcoxon's signed rank tests, ES was calculated using r-values with the equation: $r = z/\sqrt{n}$ (where n = the total number of observations). Effect sizes represented by r-values were then interpreted using with the criteria of 0.1 = small effect, 0.3 = medium effect and ≥0.5 = large effect (Cohen 1988). Heteroscedasticity was visually assessed using a scatter plot of the mean for the performance criteria data from each TT and the difference between TT1 and TT2 (Figure 1). Raw typical error of measurement (rTE) was determined using the equation: $SDdiff/\sqrt{2}$, where SDdiff represents the standard deviation of the differences between TT1 and TT2. Standardised TE (sTE) was calculated using the method of Hopkins (2012) and interpreted using a modified Cohen scale where < 0.2 = trivial, 0.2–0.6 = small, 0.6–1.2 = moderate, 1.2 – 2.0 = large and > 2.0 = very large error). This method of TE calculation has previously been recommended in reliability study designs such as this (Hopkins 2000). Ninety-five percent limits of agreement (LoA) were used to further assess reliability of time, speed and power between TT1 and TT2. The strength of relationships of these variables between time trials was analysed using intraclass correlation coefficients (ICC) presented with CIs as recommended by (Atkinson and Nevill 1998). Calculations of variations in reproducibility of the three performance variables were also assessed using coefficients of variation (CV) to allow simple comparisons between the present study and the existing literature on cycling TT and ergometry reliability. The CV was expressed as a percentage using the equation: CV% = (SD/mean) x 100. All statistical procedures were performed using SPSS 22 for Windows (IBM UK Ltd, Portsmouth, UK) and calculations were completed using Excel 2013 (Microsoft Corp., Redmond, Washington, USA).

Results

There were no significant differences for the whole cohort of cyclists in time (mean difference = 4.0 s, z = -0.879, p = 0.379, ES = 0.12), speed (mean difference = 0.085 km.h-1, z = 0.906, p = 0.365, ES = 0.12), or power data (mean difference = 1.5 W, t = 1.171, p = 0.246, ES = 0.22), between TT1 and TT2 (Table 1). When the large data set was sub-divided into faster and slower participant performances, time (mean difference

Table 1. Mean time trial (TT) performance and difference (Δ) data (± SD) for all cyclists and subgroups.

		Total Time (s)	Mean Speed ($km.h^{-1}$)	Mean Power (W)
All	TT1	1621.1 ± 101.0	35.9 ± 2.1	249.7 ± 35.7
(n = 58)	TT2	1625.1 ± 103.3	35.8 ± 2.2	248.2 ± 37.0
	Mean	1623.1 ± 101.0	35.8 ± 2.1	248.9 ± 36.0
	ΔTT1-2	4.0 ± 28.4	0.09 ± 0.6	1.5 ± 9.6
Faster	TT1	1542.0 ± 35.6	37.6 ± 0.9	279.1 ± 18.1
(n = 29)	TT2	1544.7 ± 37.6	37.5 ± 0.9	277.4 ± 19.3
	Mean	1543.3 ± 35.6	37.6 ± 0.9	278.3 ± 18.4
	ΔTT1-2	2.7 ± 17.4	0.08 ± 0.4	1.6 ± 7.2
Slower	TT1	1700.3 ± 79.6	34.2 ± 1.5	220.3 ± 21.9
(n = 29)	TT2	1705.5 ± 83.3	34.1 ± 1.6	219.0 ± 25.2
	Mean	1702.9 ± 79.4*	34.1 ± 1.5*	220.0 ± 22.9*
	ΔTT1-2	5.2 ± 36.5	0.09 ± 0.7	1.3 ± 11.7

= 162.0 s, t = -9.9, p <0.001, CI = 126.0 – 192.0 s), speed (mean difference = 3.5 km.h^{-1}, t = 10.6, p <0.001, CI = 2.8 – 4.1 km.h^{-1}), and power (mean difference = 58.6 W, t = 10.8, p <0.001, CI = 47.7 – 69.6 W), were poorer in the slower group (Table 1).

The LoA analysis showed that there were no heteroscedastic responses for time (Figure 1a), speed (Figure 1b), or the power data (Figure 1c) with small test-retest bias between TT1 and TT2. More detailed scrutiny of the test-retest reliability (Table 2) showed that for all cyclists, there was either small (time) or trivial raw TE (speed and power). The faster and slower groups also exhibited small raw TE for all performance variables. Test-retest reliability was superior in the faster group, with a range of CV's between 0.7-2.0%, compared to the slower group that demonstrated CV's of between 1.3-3.2% (Table 2).

Typically, the ICC data also reflect good reliability between TT1 and TT2 for all cyclists (p < 0.001, r = 0.96-0.97). The faster and slower cyclist groups displayed highly significant ICC's (p < 0.001 in all cases) with r-values ranging between 0.89 – 0.93 and 0.88 – 0.90 respectively. These data represent reliable performance parameters with slightly better reliability in the faster cyclist group (Table 2).

Table 2. Reliability of performance variables for all cyclists and subgroups between TT1 and TT2. Data represent raw typical error (rTE), standardised typical error (sTE), coefficient of variation (CV), intraclass correlation coefficient (ICC) and 90% confidence intervals (CI).

			Total Time (s) (s)	Mean Speed (km.h^{-1})	Mean Power (W)
All (n = 58)					
	rTE		3.97	0.42	6.8
	sTE		0.20	0.19	0.19
	CV (%)		1.1	1.1	2.7
	ICC	r	0.96	0.97	096
		p	<0.001	<0.001	<0.001
		CI	0.94 – 0.98	0.94 – 0.98	0.94 – 0.98
Faster (n = 29)					
	rTE		2.75	0.29	5.10
	sTE		0.34	0.33	0.27
	CV (%)		0.7	0.7	2.0
	ICC	r	0.89	0.89	0.93
		p	<0.001	<0.001	<0.001
		CI	0.77 – 0.95	0.78 – 0.95	0.85 – 0.96
Slower (n = 29)					
	rTE		5.20	0.52	8.25
	sTE		0.32	0.33	0.35
	CV (%)		1.3	1.3	3.2
	ICC	r	0.90	0.89	0.88
		p	<0.001	<0.001	<0.001
		CI	0.80 – 0.95	0.78 – 0.95	0.76 – 0.94

Discussion

This is the first study to establish that the CompuTrainer cycle ergometer produces very reliable performance time, mean speed and mean power output data for 16.1 km TT's. The test-retest reliability was determined by CV's of 1.1%, 1.1% and 2.7%, and rTE of 3.97 s, 0.42 km.h-1, and 6.8 W for these performance criteria, respectively. This suggests that assessment of 16.1 km TT performance is very reproducible using this ergometer in a cohort of trained cyclists. A particular strength of the present study was the large cohort of cyclists used for the reliability analysis. Previous studies that have assessed comparable exercise protocol durations, using similar ergometers (Clark et al 2014; Driller 2012; Sporer and McKenzie 2007; Zavorsky et al. 2007), have observed reliability CV's ranging from 1.5-3.4% for mean power output. The present study demonstrates that during a 16.1 km TT the test-retest reliability of this measurement was superior on the Computrainer in this group of cyclists. This could in part be attributed to the larger number of observations in this study, compared to many cycling TT reliability studies (Atkinson and Nevill 1998).

Previous studies that have investigated the reliability of cycling TT performance using similar durations to the present study, have typically used 20 km TT's (Clark et al. 2014; Sporer and McKenzie 2007; Zavorsky et al. 2007) or 30 min fixed duration TT's for the determination of total work done (Driller 2012; Rivera and McGregor 2005). The studies that used a traditional TT format, (Clark et al. 2014; Sporer and McKenzie 2007; Zavorsky et al. 2007), all used a Velotron cycle ergometer and observed CV's of 2.0% (hilly TT), 2.1% (flat TT) and 3.4% (flat TT) for mean power output, respectively. When the duration of time between trials was increased to 28 days, reliability decreased to 3.2% (Clark et al. 2014). These data are comparible to the overall power output and performance time data in the present study, for a similar time duration between trials. However, the mean power output for the CompuTrainer was slightly worse in present study than that observed by Clark et al. (2014) for the slower group, which is likely the result of a slightly better training status of their slower cyclists. Only Driller (2012), who used a Lode, Excalibur Sport ergometer, has shown better reliability for the mean power output than the present study. Driller (2012) however, used a 30 min fixed duration TT to determine performance and physiological reliability. Between TT1 and TT2, the post exercise blood lactate response displayed the worst reliability (CV = 8.8%) and mean power output was the most reliable variable (CV = 0.7%), but this was in a group of athletes that were considerably better trained that those in the present study.

The majority of studies that have previously specifically evaluated the CompuTrainer, have done so using incremental exercise (Earnest et al. 2005; Guiraud et al. 2010), a variety of tests (Rivera and McGregor 2005) or on short duration high intensity protocols (Clark et al. 2016). All of these studies also suggest that the CompuTrainer produces reliable data in the respective protocols, but only Rivera and McGregor (2005) have assessed this ergometer using a simulated TT of a similar duration (30 minute TT) to the present study. They too found that the test-retest CV for time, speed and power ranged between 0.6-1.2% in a trained group of cyclists. This is despite fixed-time, distance capacity performance tasks, being associated with different pacing strategies (Abbiss et al. 2016) and therefore reduced task-specific familiarisation. Such time duration based performance tests, have previously been reported to have lower test-retest reliability than distance TT protocols (Hopkins et al. 2001), but the training status of the cyclists in the Rivera and McGregor (2005) and Driller (2012) studies was superior to those in the present study.

Hopkins et al. (2001), have previously suggested that fitness status, experience, and familiarity are directly related to the reliability of performance measurements. It is likely that these factors were responsible for the slightly worse reliability data in the present study for the slower compared to the faster group of participants. The reliability data for these two subgroups ranged between 0.7-2.0% and 1.3-3.2% (CV) and 0.27-0.34 and 0.32-0.35 (sTE) for the faster and slower cyclists respectively. Several other studies have also demonstrated that the reliability of cycling performance data tends to be lower in less well trained individuals (Clark et al. 2014; Hopker et al. 2010 Zavorsky et al. 2007). In a comparison of the reliability between trained cyclists and untrained individuals, Hopker et al. (2010) observed mean power output CVs to increase by 1.1% and 4.1% in the untrained participants using the SRM and Wattbike respectively. Even though the performance times were reasonably wide ranging in the present study and there was slightly lower reliability for the slower cyclists, there was no evidence of heteroscedasticity for any of the variables, suggesting no obvious pattern for worsening reliability with slower performance times.

Close scrutiny of the performance times in the present study do however highlight a discrepancy between the description of the participants and their TT ability. The mean performance times were 25.72 and 28.38 min for the faster and slower groups respectively, which is considerably slower than might be expected for trained cyclists. This is likely the result of the under-estimation of power, which is associated with the CompuTrainer (Earnest et al. 2005). Guiraud et al. (2010), reported that the CompuTrainer produced lower power output data compared to other ergometers and that this was exacerbated as power increased between 200 and 300 W. This range of power output represents the mean power for the cyclists in the present study and this under-estimation is likely responsible for the slower than expected times. Indeed, the CompuTrainer has previously been reported to produce performance times which are significantly slower than the same TT performance on the road (Peveler 2013) and lower power output compared to an SRM powermeter (Davison et al. 2009). The present study design is unable to determine the validity of the CompuTrainer, but it seems reasonable to assume that the performances would have been considerably slower on this ergometer compared to the same TT's on the road. Despite this limitation, the CompuTrainer produces test-retest power data with similar degrees of reliability to that of the SRM laboratory ergometer (Sparks et al. 2015).

This is the first study to investigate the test-retest reliability of 16.1 km TT on a CompuTrainer cycle ergometer in a trained group of cyclists, despite an increased interest in this distance as performance criteria. Indeed, several studies have previously used the 16.1 km TT for manipulating feedback (Jones et al. 2015b; Williams et al. 2015; Wilson et al. 2012) and for the assessment of nutritional ergogenic substances (Folland et al. 2008). In one such study Jeukendrup et al. (2008) used ~16 km TT performance to determine the efficacy of carbohydrate as an ergogenic, but whilst they reported a test-retest CV of 1.1%, their TT was not distance based, but total time to expend 450 kJ. This type of TT is slightly different in nature to an actual 16 km TT, and is usually used due to the capabilities of the available ergometry equipment. The performance times and test-retest data reported by Jeukendrup et al. (2008) were however similar to those of the present study. The present data show that the TE for 16.1 km TT performances on the CompuTrainer are larger than the smallest worthwhile change of 0.6% for road TT (Paton and Hopkins 2001), but it still represents a reliable ergometer on which this test can be performed in trained cyclists.

Practical application

The 16.1 km TT represents a very reliable performance criterion for trained cyclists, but researchers and practitioners should be aware that the reliability is different for time, speed and power data. These performance criteria have CV's of 1.1, 1.1 and 2.7% respectively. Poorer (but still good reliability) is observed with less well trained cyclists. For the faster cyclists reliability ranged between 0.7-2.0% and 1.3-3.2% for the slower cyclists across all performance measurements. Researchers, coaches and athletes should be aware that they are unlikely to obtain data from the CompuTrainer, that is directly comparable to road cycling performances, but it represents a reliable ergometer for assessing the performance of 16.1 km TT's in trained cyclists.

Conflict of interest

No potential conflicts of interest exist.

References

1. Abbiss CR, Thompson KG, Lipski M, Meyer T, Skorski S (2016) Pacing Differs Between Time- and Distance-Based Time Trials in Trained Cyclists. International Journal of Sports Physiology and Performance: Epub ahead of print.
2. Altareki N, Drust B, Atkinson G, Cable T, Gregson W (2009) Effects of environmental heat stress (35 degrees C) with simulated air movement on the thermoregulatory responses during a 4-km cycling time trial. International Journal of Sports Medicine 30: 9-15.
3. Astorino TA, Cottrell T (2012) Reliability and validity of the velotron racermate cycle ergometer to measure anaerobic power. International Journal of Sports Medicine 33: 205-210.
4. Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine 26: 217-238.
5. Atkinson G, Reilly T (1996) Circadian variation in sports performance. Sports Medicine 21: 292-312.
6. Che Jusoh MR, Morton RH, Stannard SR, Mundel T (2015) A reliable preloaded cycling time trial for use in conditions of significant thermal stress. Scandinavian Journal of Medicine & Science in Sports 25 (S1): 296-301.
7. Clark B, Paton CD, O'Brien, BJ (2014) The reliability of performance during computer-simulated varying gradient cycling time trials. Journal of Science and Cycling 3(3):29-33.
8. Clark IE, Gartner HE, Williams JL, Pettitt RW (2016) Validity of the 3-min all-out Exercise Test on the CompuTrainer. Journal of Strength and Conditioning Research 30: 825-829.
9. Cohen J (1988) Statistical power analysis for the behavioural sciences. Erlbaum, Hillsdale, New Jersey, USA.
10. Currell K, Jeukendrup AE (2008) Validity, reliability and sensitivity of measures of sporting performance. Sports Medicine 38: 297-316.
11. Davison RR, Corbett J, Ansley L (2009) Influence of temperature and protocol on the calibration of the Computrainer electromagnetically-braked cycling ergometer: original research article. International SportMed Journal 10: 66-76.
12. De Pauw K, Roelands, B., Cheung, S. S., De Geus, B., Rietjens, G., & Meeusen, R (2013) Guidelines to classify subject groups in sport-science research. International Journal of Sports Physiology and Performance 8: 111-122.
13. Driller MW (2012) The reliability of a 30-minute performance test on a Lode cycle ergometer. Journal of Science and Cycling 1(2): 21-27.
14. Dwyer DB (2004) A standard method for the determination of maximal aerobic power from breath-by-breath VO2 data obtained during a continuous ramp test on a bicycle ergometer. Journal of Exercise Physiology 7(5): 1-9.
15. Earnest CP, Wharton RP, Church TS, Lucia A (2005) Reliability of the Lode Excalibur Sport Ergometer and applicability to Computrainer electromagnetically braked cycling training device. Journal of Strength and Conditioning Research 19: 344-348.
16. Folland JP, Stern R, Brickley G (2008) Sodium phosphate loading improves laboratory cycling time-trial performance in trained cyclists. Journal of Science and Medicine in Sport 11: 464-468.
17. Foss Ø, Hallén J (2005) Validity and stability of a computerized metabolic system with mixing chamber. International Journal of Sports Medicine. 26(7):569-75.
18. Grafen A, Hails R (2002) Modern Statistics for the Life Sciences. Oxford University Press.
19. Guiraud T, Leger L, Long A, Thebault N, Tremblay J, Passelergue P (2010) VO2 requirement at different displayed power outputs on five cycle ergometer models: a preliminary study. British Journal of Sports Medicine 44: 449-454.
20. Harriss D, Atkinson G (2013) Ethical Standards in Sport and Exercise Science Research: 2014 update. International Journal of Sports Medicine 34(12): 1025-1028.
21. Hopker J, Myers S, Jobson SA, Bruce W, Passfield L (2010) Validity and reliability of the Wattbike cycle ergometer. International Journal of Sports Medicine 31: 731-736.
22. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Medicine 30: 1-15.
23. Hopkins WG (2012) Analysis of reliability with a spreadsheet (beta version). In: A new view of statistics. Internet Society for Sport Science, sportsci.org/resource/stats/xrely.xls
24. Hopkins WG, Schabort EJ, Hawley JA (2001) Reliability of power in physical performance tests. Sports Medicine 31: 211-234
25. Jeukendrup AE, Hopkins S, Aragon-Vargas LF, Hulston C (2008) No effect of carbohydrate feeding on 16 km cycling time trial performance. European Journal of Applied Physiology 104: 831-837.
26. Jones HS, Williams EL, Marchant D, Sparks SA, Midgley AW, Bridge CA, McNaughton L (2015a) Distance-dependent Association of Affect with Pacing Strategy in Cycling Time Trials. Medicine and Science in Sports and Exercise 47: 825-832.
27. Jones HS, Williams EL, Marchant DC, Sparks SA, Bridge CA, Midgley AW, McNaughton LR (2015b) Deception has no acute or residual effect on cycling time trial performance but negatively effects perceptual responses. Journal of Science and Medicine in Sport: pii: S1440-2440(1415)00239-X.
28. Kirkland A, Coleman D, Wiles JD, Hopker J (2008) Validity and reliability of the Ergomopro powermeter. International Journal of Sports Medicine 29: 913-916.
29. Lamberts RP, Swart J, Woolrich RW, Noakes TD, Lambert MI (2009) Measurement error associated with performance testing in well-trained cyclists: Application to the precision of monitoring changes in training status. International SportMed Journal 10: 33.
30. Maunder E, Laursen PB, Kilding AE (2016) Effect of ad libitum Ice Slurry and Cold Fluid Ingestion on Cycling Time-Trial Performance in the Heat. International Journal of Sports Physiology and Performance Aug 24:1-19. [Epub ahead of print].
31. Micklewright D, Papadopoulou E, Swart J, Noakes T (2010) Previous experience influences pacing during 20 km time trial cycling. British Journal of Sports Medicine 44: 952-960.
32. Midgley AW, McNaughton LR, Polman R, Marchant D (2007) Criteria for determination of maximal oxygen uptake. Sports Medicine 37: 1019-1028.
33. Noreen EE, Yamamoto K, Clair K (2010) The reliability of a simulated uphill time trial using the Velotron electronic bicycle ergometer. European Journal of Applied Physiology 110: 499-506.
34. Paton CD, Hopkins WG (2001) Tests of cycling performance. Sports Med 31: 489-496.
35. Peveler WW (2013) The accuracy of simulated indoor time trials utilizing a CompuTrainer and GPS data. Journal of Strength and Conditioning Research 27: 2823-2827
36. Rivera PD, McGregor SJ (2005) Validity and reliability of Computrainer TM and Powertap TM power meters during incremental and steady state exercise. Medicine and Science in Sports and Exercise 5: S22.
37. Sparks SA, Dove B, Bridge CA, Midgely AW, McNaughton LR (2015) Validity and reliability of the look Keo power pedal system for measuring power output during incremental and repeated sprint cycling. International Journal of Sports Physiology and Performance 10: 39-45.
38. Sporer BC, McKenzie DC (2007) Reproducibility of a laboratory based 20-km time trial evaluation in competitive cyclists using the velotron pro ergometer. International Journal of Sports Medicine 28: 940-944.
39. Stone MR, Thomas K, Wilkinson M, Jones AM, St Clair Gibson A, Thompson KG (2012) Effects of deception on exercise performance: implications for determinants of fatigue in humans. Medicine and Science in Sports and Exercise 44: 534-541.
40. Williams EL, Jones HS, Sparks SA, Marchant DC, Midgley AW, McNaughton LR (2015) Competitor presence reduces internal attentional focus and improves 16.1km cycling time trial performance. Journal of Science and Medicine in Sport 18: 486-491.
41. Wilson MG, Lane AM, Beedie CJ, Farooq A (2012) Influence of accurate and inaccurate 'split-time' feedback upon 10-mile time trial cycling performance. European Journal of Applied Physiology 112: 231-236.
42. Wooles A, Keen P, Palfreeman R (2003) Protocols used for the physiological testing of the Great Britain Cycling Team. Test Methods Manual.
43. Zavorsky GS, Murlas JM, Gow J, Kim DJ, Poulin-Harnois C, Kubow S, Lands LC (2007) Laboratory 20-km cycle time trial

reproducibility. International Journal of Sports Medicine 28: 743-748.