RESEARCH ARTICLE

Reproducibility of a Laboratory Based 1-km Wattbike Cycling Time Trial in Competitive Cyclists

Phillip M Bellinger [1] ✉ and Clare Minahan [1]

Abstract

The purpose of this study was to evaluate the reliability of a 1-km cycling time trial using a Wattbike cycling ergometer in competitive cyclists. Ten competitive male cyclists (mean ± SD; 63.9 ± 5.1 mL•kg^{-1}•min^{-1}) performed a VO2max test and a familiarization of the 1-km time trial (TT) on the same day, and 3 separate 1-km TT (TT1, TT2 and TT3) separated by >24 hours. Mean performance time was not significantly different between TT (70.27 ± 5.95 vs. 70.06 ± 5.75 vs. 69.79 ± 6.0; P = 0.47) and displayed a low coefficient of variation (CV%; 90% CI: 0.6%; 0.4 – 1.1%). The mean reliability, expressed as the coefficient of variation and typical error of measurement over the three TT was performance time 0.6%, 0.42 s (95% CL 0.28-0.77); mean power 1.2%, 6.3 W (95% CL 4.7-12.0); peak power 7%, 64.8 W (95% CL 45.3-124.9); blood lactate concentration 3.2%, 0.5 mmol•L^{-1} (95% CL 0.9-0.3); heart rate 0.8%, 1.4 bpm (95% CL 1.1-2.5). Such high reliability makes the combination of the Wattbike, athlete and test protocol very suitable for detecting "real" changes in performance where improvements may be small but still considered worthwhile in a competitive sport setting.

Keywords: reliability, highly-trained, performance testing, power output

✉**Contact email:** _p.bellinger@griffith.edu.au_ (PM. Bellinger)

[1] School of Allied Health Sciences Gold Coast Campus, Griffith University QLD 4222, Australia

Introduction

In order to monitor changes in athletic performance, the laboratory-based performance test used must be reliable, valid, and sensitive to small changes in response to training, illness, and interventions (Hopkins et al. 1999). Indeed, it is now considered important to report the reliability of an exercise test because it determines how well a test can monitor changes in performance of athletes in an applied research setting (Hopkins 2000). Furthermore, the estimation of the error in an exercise protocol is a key consideration for the determination of sample sizes in research and for making pragmatic assumptions on the effectiveness of an intervention (Hopkins 2000).

Reliability refers to the reproducibility of the observed value when the measurement is repeated (Hopkins 2000). There are three components of reliability that are considered important when athletes undergo repeated testing: changes in the mean value, retest correlation, and within-participant variation (Hopkins 2000). Changes in the mean value indicate the extent to which an athlete's performance improves or deteriorates with respect to the original performance trial. This may be indicative of a possible learning effect during subsequent trials or the presence of

residual fatigue attenuating successive performance. Retest correlation represents how the rank order of participants in one trial is replicated in the subsequent trials. Within-participant variation is considered the most important type of reliability measure for researchers, because it affects the precision of estimates of change in the dependant variable of an experimental study. It is also the most important type of reliability measure for coaches, sport scientists and other professionals using tests to monitor the performance of their athletes (Hopkins 2000). In these situations, the smaller the within-participant variation, the easier it is to detect a "real" change in performance where improvements may be small but still considered worthwhile in a competitive sport setting (Hopkins 2004). The knowledge of test–retest reliability may also help in the calculation of sample size and determining precision of measurement between different ergometers.

The Wattbike is an air-braked ergometer that is used by competitive cyclists at sporting institutes to assess and monitor cycling performance and for sports performance research at Universities worldwide (Hopker et al. 2010). The reliability of the Wattbike cycle ergometer has previously been assessed in trained cyclists (Driller et al. 2014; Driller et al. 2013; Hopker et al. 2010). Hopker et al. (Hopker et al. 2010) reported a CV of 2.6% (95% CI: 1.8-5.1%) under four constant load work rates ranging from 50-300 W. Two studies from the Australian Institute of Sport (Driller et al. 2014; Driller et al. 2013) have reported a CV for mean power output of 2.4% (90% CI: 2.0-3.1%) and 2.7%

(95% CI: 2.3-3.5%) in a 30-s sprint and 4-min performance trial, respectively. However, it is unknown whether similar reliability would exist for a short distance based (1-km) TT on the Wattbike cycle ergometer.

Previous research indicates the laboratory-based time trials (TT), such as constant-distance cycling tests, are considered to have the highest performance validity in cyclists (Paton and Hopkins 2001). As far as we are aware, there are no published reports of the reliability of performance and physiological variables during a short, distance-based cycling TT in trained cyclists, especially when using cycling-specific ergometry such as with the Wattbike ergometer. Determining the variation in performance of a short distance-based TT on a Wattbike will provide insight into monitoring performance and has important implications on research design and the sample size required to detect the smallest worthwhile change in performance (Hopkins et al. 1999). Furthermore, a 1-km Wattbike cycling TT may simulate real world performance, such as the individual 1000 m TT that features at the UCI Track Cycling World Championships. Therefore, the purpose of this study was to assess the reproducibility of performance and physiological variables during a laboratory based 1-km cycling TT performance test in competitive cyclists using the Wattbike cycling ergometer.

Materials and methods

Ten highly-trained cyclists (mean ± SD; age = 25.3 ± 5.4 yr; mass = 73.1 ± 4.0 kg; VO_{2max} = 63.9 ± 5.1 $mL·kg^{-1}·min^{-1}$) were recruited for the current study. All cyclists were considered highly-trained and consistently accumulated >10 $h·wk^{-1}$ of training while regularly competing in local A grade criterion and TT (>10 km) cycling races. All cyclists were informed verbally and in writing as to the requirements of the study and all gave their written informed consent. Cyclists had not taken any nutritional supplements in the 3 months prior to the study with the exception of three cyclists who were consuming a multi-vitamin supplement and two cyclists who were consuming a fish-oil supplement. The study was conducted in the Griffith University Sport Science laboratory and was approved by the Griffith University Human Research Ethics Committee and meets the ethical standards of the Journal of Science and Cycling (Harriss and Atkinson 2011).

Cyclists attended the laboratory on four separate occasions at the same time of day (± 2 h) and standard laboratory conditions were maintained during testing (a temperature of 22.5-23°C and 50-55% relative humidity). Cyclists were provided with a minimum of 24 h between each visit and all testing was completed within a period of 3 weeks. The initial visit to the laboratory consisted of a graded-exercise test (GXT) to determine VO_{2max} and a familiarisation of the maximal

1-km cycling TT. The remaining three visits to the laboratory were for the completion of the 1-km TTs.

The GXT was performed on the Velotron Pro cycle ergometer (RacerMate Inc., Seattle, WA, USA) modified with clip-in pedals and low-profile racing handlebars. The saddle and handlebar position of the cycle ergometer were adjusted to accurately match the dimensions of each of the cyclists' road bicycle measurements and each cyclist warmed up at a self-selected power output for 5 min. The test consisted of an initial workload of 100 W, followed by a ramped protocol in which power output was increased by 15 $W·30-s^{-1}$ until volitional fatigue (Stöggl and Sperlich 2014). Respiratory variables were measured and recorded every 30 s (Parvomedics Trueone 2400, Utah, USA).

Following completion of the GXT, cyclists undertook a 20-min recovery interval which incorporated 10 min of cycling at a self-selected power output and 10 min of complete rest. Following the 20-min recovery interval the cyclists performed a practice trial of the 1-km cycling TT on the Wattbike cycle ergometer (Wattbike Pro, Nottingham, UK).

Prior to reporting to the laboratory for the TTs, each cyclist was asked to abstain from caffeine and alcohol for a period of 12 h and avoid strenuous exercise for the 24 h preceding each TT. Cyclists were required to record a 24 h diet diary leading up to the first TT which was then replicated in the 24 h preceding each subsequent TT. Adherence to these requests was confirmed by each cyclist verbally prior to each trial.

When the cyclists reported to the laboratory on the day of each TT, they were weighed using SECA balance beam scales (Birmingham, UK). Each 1-km TT was preceded by a standardized 20-min warm-up protocol which consisted of 5 min of cycling at 2.5 $W·kg^{-1}$, an 8-min progressive build up to each cyclists peak power output achieved during the GXT, 2 min of cycling at a self-selected power output, 2 min of cycling to include three 6-s maximal sprints and finished with 3 min of cycling at a self-selected power output. Following completion of the warm-up there was a 60-s period where cyclists were instructed to sit passively before a standardized countdown initiated the TT.

The cyclists then performed an all-out maximal cycling time-trial for 1 km. During each TT, cyclists had access to the distance remaining and were required to complete the TT in the quickest time possible. With the exception of verbal encouragement, no other information was provided. The gearing was self-selected by the cyclists on the Wattbike during the practice trial and then replicated during each TT.

Saddle and handlebar height and position were replicated for each trial. Cyclists wore their own cycling shoes and their own pedals were attached. All cyclists were instructed to perform the sprint as a maximal all-out effort, and, as such, they began each TT out of the saddle and were typically seated for the final seconds of the sprint.

The computer attached to the Wattbike cycle ergometer was used to record performance time, mean power, peak power and mean cadence during the 1-km TT. Blood was sampled via a capillary fingertip sample and subsequently analysed for lactate concentration with a Lactate-Pro analyzer (Arkray, Japan). Measurements were taken immediately following the completion of the TT. Heart rate was measured continuously (RS800cx, Polar Electro Oy, Finland) during the TTs. The 1-km cycling TT protocol implemented in the current study was designed to closely mimic the track-cycling event of the identical distance (1000 m) that is contested at the UCI Track Cycling World Championships.

Mean values for all performance and physiological variables were compared using a one-way repeated measures ANOVA. Data were log transformed and analysed using an Excel spreadsheet for reliability (Hopkins 2009) as described by Hopkins and colleagues (Hopkins et al. 1999). Data were analysed using intra-class correlations. An individual's CV for a specific trial was calculated as the SD of an individual's repeated measurement, expressed as a percentage of their individual mean test score (Hopkins 2000). Typical error is presented as the CV% and as an absolute value along with upper and lower 90% CI. Statistical significance was accepted at the $P < 0.05$ level.

Results

The performance and physiological variables for each trial are shown in Table 1. The mean difference in performance time between TT3 (69.79 ± 6.0 s) and TT1 (70.27 ± 5.95 s) and TT2 (70.06 ± 5.75 s) was equal to approximately 0.48 s (TT1) and 0.21 s (TT2), however, these differences were not statistically significant ($F_{2,18} = 4.9$, $P = 0.47$). Similarly, there was no significant variation among TT for mean power ($F_{2,18} = 0.5$, $P = 0.40$), peak power ($F_{2,18} = 0.8$, $P = 0.20$), relative power ($F_{2,18} = 1.3$, $P = 0.50$) and cadence $F_{2,18} = 1.0$, $P = 0.40$) and there was no obvious effect of fatigue (i.e., decrease in performance in subsequent trials) or learning (i.e., an increase in performance)

Table 1. Mean performance and physiological variables (± SD) determined during and after three 1-km time trials (TT) performed on a Wattbike.

	TT1	TT2	TT3	Average
Performance time (s)	70.27 ± 5.95	70.06 ± 5.75	69.79 ± 6.0	70.04 ± 5.95
Mean power (W)	520 ± 101	523 ± 100	524 ± 102	522 ± 101
Peak power (W)	909 ± 144	945 ± 181	921 ± 174	925 ± 166
Relative power (W·kg^{-1})	7.2 ± 1.1	7.3 ± 1.1	7.3 ± 1.8	7.3 ± 1.3
Mean cadence (rev·min^{-1})	122 ± 10	124 ± 9	124 ± 9	124 ± 9
Blood lactate (mmol·L^{-1})	14.3 ± 1.0	14.6 ± 0.9	14.2 ± 1.0	14.4 ± 1.0
Heart rate (beats·min^{-1})	176 ± 9	175 ± 9	177 ± 9	176 ± 9

between TT1 and TT3.

In agreement with the athlete-dependant variables, the athlete-independent variables such as blood lactate concentration ($F_{2,18} = 0.5$, $P = 0.27$) and peak heart rate values ($F_{2,18} = 2.12$, $P = 0.37$) did not change across trials.

All reliability measures are shown in Table 2. The mean reliability, expressed as the typical error of measurement (TEM; Watts) and coefficient of variation (CV; %) over the three TTs were performance time 0.6%, 0.42 s (95% CL 0.28-0.77); mean power 1.2%, 6.3 W (95% CL 4.7-12.0); peak power 7.0%, 64.8 W (95% CL 45.3-124.9); relative power 1.3%, 0.1 W (95% CL 0.07-0.17); mean cadence 1.0%, 1.2 rpm (95% CL 0.9-2.2).

The physiological variables also displayed acceptable reliability across trials with peak blood lactate concentration 3.2%, 0.5 mmol/L (95% CL 0.9-0.3) and peak heart rate 0.8%, 1.4 bpm (95% CL 1.1-2.5) displaying values that are considered within the commonly reported reliability criteria in athletic testing (CV: <10%). Differences from the mean of the three 1-km TTs for performance time, mean power, peak power, relative power, peak heart rate and peak blood lactate are shown in Figure 1.

Discussion

The present study is the first to determine the reproducibility of a short, distance-based (1-km) TT in highly-trained cyclists on the Wattbike cycle ergometer. Data from the present study showed that the performance time in three 1-km TT performed a minimum of 24 h apart on the Wattbike cycle ergometer are not significantly different from each other. It was also demonstrated that performance was

Table 2. Mean within-subject intraclass correlation (ICC) and typical error as a coefficient of variation (CV%) of the between-tests change.

	Performance time (s)	Mean power (W)	Peak power (W)	Relative power (W·kg^{-1})	Mean cadence (rev·min^{-1})	Blood lactate (mmol·L^{-1})	Heart rate (beats·min^{-1})
ICC 2-1	.99 (.97-.99)	.99 (.97-.99)	.87 (.53-.97)	.99 (.96-.99)	.98 (.95-.99)	.84 (.44-.96)	.98 (.95-.99)
ICC 3-2	.99 (.98-.99)	.99 (.98-.99)	.90 (.63-.98)	.99 (.98-.99)	.98 (.93-.99)	.80 (.35-.95)	.96 (.84-.99)
ICC mean	.99 (.98-.99)	.99 (.98-.99)	.89 (.58-.98)	.99 (.97-.99)	.98 (.94-.99)	.80 (.35-.95)	.97 (.90-.99)
CV 2-1	.7 (.4-1.2)	1.5 (1.1-2.9)	7.2 (5.1-14.0)	1.6 (1.1-3.0)	0.9 (.6-1.6)	3.3 (2.3-6.2)	.5 (.4-1.0)
CV 3-2	.5 (.3-.9)	.9 (.6-1.7)	6.7 (4.7-13.0)	0.9 (0.6-1.6)	1.0 (.7-1.8)	3.0 (2.1-5.7)	1.0 (.7-1.8)
CV mean	.6 (.4-1.1)	1.2 (.9-2.3)	7.0 (4.9-13.5)	1.3 (0.9-2.3)	1.0 (.7-1.8)	3.2 (2.2-6.0)	.8 (.6-1.4)
Absolute TEM	.42 (.28-.77)	6.3 (4.7-12.0)	64.8 (45.3-124.9)	0.1 (.07-.17)	1.2 (.9-2.2)	.5 (.3-.9)	1.4 (1.1-2.5)
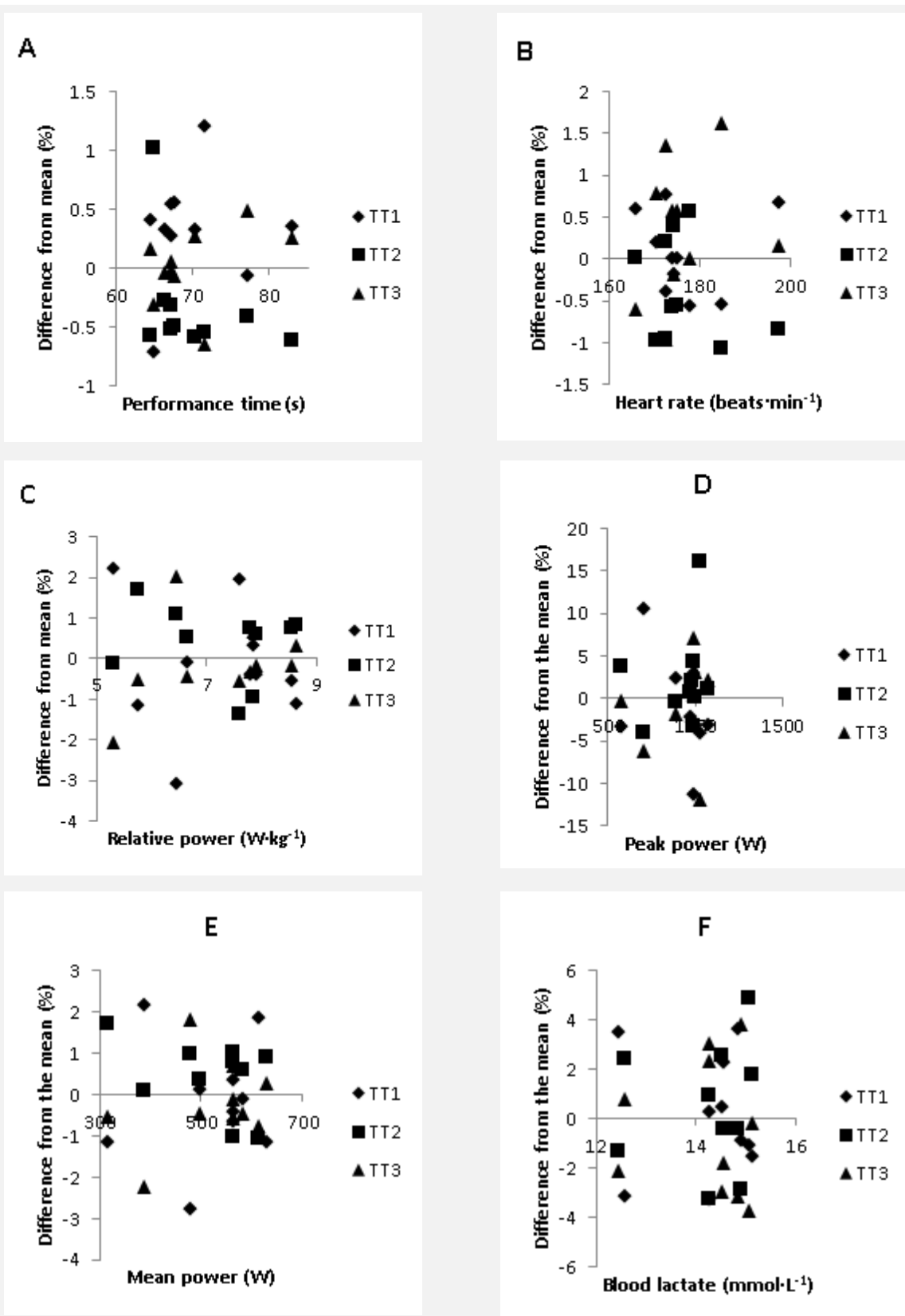
Figure 1. Difference (%) of each individual trial from the mean of three trials for each athlete for (A) performance time, (B) mean power, (C) peak power, (D) relative power, (E) peak heart rate and (F) blood lactate concentration.

highly reproducible with respect to mean, peak, and relative power and mean cadence. The physiological variables also displayed values that are considered within the commonly reported reliability criteria in athletic testing (CV: <10%; ICC >0.8) (Atkinson et al.

1999). Performance time demonstrated the lowest CV between trials (0.7% or less) with peak power and peak blood lactate displaying a CV of <7.0% and <3.2%, respectively, which are considered within the

commonly reported reliability criteria in athletic testing.

The mean CV of measurement for performance time and mean power was 0.6% and 1.2%, respectively, with a low CV being observed between TT1 and TT2 (mean power 1.5%; performance time 0.7%) and TT2 and TT3 (mean power 0.9%; performance time 0.5%). The high reproducibility of performance across trials is comparable to that found in other reliability studies employing air braked ergometers and duration or distance-based cycling TT protocols (Driller et al. 2014; Driller et al. 2013; Laursen et al. 2003; Smith et al. 2001; Sporer and McKenzie 2007). Previous research has demonstrated a low mean CV for performance time (1.0-1.9%) and mean power (2.0%) across trials over the 40-km distance (Laursen et al. 2003; Smith et al. 2001). Similarly, Sporer et al. (Sporer and McKenzie 2007) reported a CV for performance time (0.8%) and mean power (2%) across trials in a laboratory based 20-km TT. Two previous studies have determined the reliability of high-intensity fixed duration cycling protocols utilizing the Wattbike (Driller et al. 2014; Driller et al. 2013). The CV for mean power was 2.7% in a 4-min performance trial (Driller et al. 2014) and 2.4% in a 30-s sprint test (Driller et al. 2013), respectively. Given the low CV across trials in these studies involving trained cyclists, we support previous findings that suggest recruiting highly trained individuals is a valid approach to lower the chance of random error or between test variation (Zavorsky et al. 2007).

It has been suggested that at least one practice trial that fully familiarizes participants with the testing procedures should precede formal testing, especially with athletes, to limit the potential learning effect in subsequent trials (Hopkins et al. 2001). Previous research suggests that this may not always be necessary in trained cyclists (Paton and Hopkins 2001; Sporer and McKenzie 2007). In the current study, the CV for performance time between the TT1 and TT2 (0.7%) is quite similar to the CV between TT2 and TT3 (0.5%). Although none of the cyclists had used the Wattbike ergometer before, all had previously used other cycle ergometers, completed short TTs, and regularly undertook repeated high intensity bouts of interval training. In addition, cyclists completed a practice 1-km TT following the incremental exercise test to determine VO_{2max}. The absence of a substantial learning effect between TT1 and TT2 suggests that a separate practice trial may not be necessary for trained cyclists provided they are given an opportunity to become familiar with the Wattbike ergometer (i.e., following a VO_{2max} test).

An important aspect for evaluation of a laboratory performance test is its reliability compared to trials performed in the field or indeed, the event itself (Hopkins et al. 1999). When compared to performance time across trials for a 40-km field TT (mean CV: 1.7%), laboratory trials appear to demonstrate higher reproducibility (mean CV: 1.0%) (Smith et al. 2001). A similar comparison can be made between cyclists performing a "real world" 1-km track cycling TT (Paton and Hopkins 2006) and the Wattbike cycling TT in the current study. Indeed, when compared between the two, the typical variation in performance time (and its 95% likely limits) of a cyclist performing a 1-km track cycling TT is 1.2% (0.8-2.2%) compared to a 1-km Wattbike TT of 0.6% (0.4-1.1%). The higher reliability of laboratory trials is likely due to the control of several factors that can affect speed (wind, topography, temperature, rolling resistance, and aerodynamics). Furthermore, the effects of individual aerodynamics and environmental conditions may limit the comparison of performance time for a simulated laboratory 1-km TT and a 1-km track cycling TT. The measurement of power output may provide a more valid measure of exercise performance and exercise intensity than performance time and is therefore considered to be a more direct method of monitoring exercise performance (Jeukendrup and Diemen 1998). Further research is necessary to determine the validity of performance variables using the Wattbike compared to an outdoor TT.

In the current study, the cyclists were considered highly-trained and consistently accumulated >10 h·wk^{-1} of training while regularly competing in local A grade criterion and TT (>10 km) cycling races. Despite these competitions being classified as endurance events, the cyclists must be able to maintain a relatively high velocity over the course of these races. This emphasizes the role of anaerobic characteristics in highly-trained endurance athletes. Furthermore, recent evidence has suggested that endurance athletes can optimize their training by devoting ~15-20% of their training volume to maximal and supramaximal training intensities (Stöggl and Sperlich 2014). This emphasizes the importance of reporting the reliability of short, maximal TT cycling performance in endurance athletes as this test may provide an appropriate measure for monitoring the performance of highly-trained endurance cyclists during training and leading into competition.

In addition to providing reliable performance data, it is important that physiological responses are similar between trials. Although muscle pH and lactate concentration was not directly measured in this study, blood lactate concentration was measured as it is a parameter that is routinely recorded to monitor training intensity (Billat 1996). Peak blood lactate concentration displayed a low CV of 3.2% (2.2-6.0%) suggesting that if peak blood lactate concentration is a key variable of interest, the 1-km Wattbike TT would be sensitive to detect changes in this variable. The low CV of peak HR (0.8%) also makes this variable sensitive to small changes in between trials.

The 1-km Wattbike cycling TT is a highly reliable sport-relevant test designed to replicate real world performance, such as the individual TT (1000 m) that features at the UCI Track Cycling World Championships. The current study has shown that when using cycling-specific ergometry such as with the Wattbike ergometer, high reliability exists across 1-km Wattbike cycling TTs in highly-trained cyclists. Due to

the small CV across tests the 1-km Wattbike cycling TT is sensitive to small changes in performance that would be meaningful to a coach and athlete.

Practical applications

The high reliability of the 1-km Wattbike cycling TT allows the protocol to be suitable for monitoring the performance of highly-trained cyclists. The high reliability of the test will allow researchers to use realistic sample sizes to detect small but potentially worthwhile/harmful changes in performance that are important to athletes. Furthermore, the low error of measurement across all trials suggests that only a single practice trial is required for this test protocol when using highly-trained cyclists.

Acknowledgment

References
1. Atkinson G, Nevill A, Edwards B (1999) What is an acceptable amount of measurement error? The application of meaningful 'analytical goals' to the reliability of sports science measurements made on a ratio scale. Journal of Sports Science 17: 18
2. Billat LV (1996) Use of blood lactate measurements for prediction of exercise performance and for control of training. Sports Medicine 22: 157-175
3. Driller M, Argus C, Bartram J, Boniventura J, Martin D, West N, Halson S (2014) The reliability of a two-bout exercise test on a Wattbike cycle ergometer. International Journal of Sports Physiology and Performance 9: 340-345
4. Driller MW, Argus CK, Shing CM (2013) The reliability of a 30-s sprint test on the Wattbike cycle ergometer. International Journal of Sports Physiology and Performance 8: 379-383
5. Harriss D, Atkinson G (2011) Update–ethical standards in sport and exercise science research. International Journal of Sports Medicine 32: 819-821
6. Hopker J, Myers S, Jobson S, Bruce W, Passfield L (2010) Validity and reliability of the Wattbike cycle ergometer. International Journal of Sports Physiology and Performance 31: 731-736
7. Hopkins W (2000) Measures of Reliability in Sports Medicine and Science. Sports Medicine 30: 1-15
8. Hopkins WG (2004) How to interpret changes in an athletic performance test. Sportscience 8: 1-7
9. Hopkins WG (2009) Calculating the reliability intraclass correlation coefficient and its confidence limits (Excel spreadsheet) newstats.org/xICC.xls.
10. Hopkins WG, Hawley JA, Burke LM (1999) Design and analysis of research on sport performance enhancement. Medicine Science and Sports Exercercise 31: 472-485
11. Hopkins WG, Schabort EJ, Hawley JA (2001) Reliability of power in physical performance tests. Sports Medicine 31: 211-234
12. Jeukendrup A, Diemen AV (1998) Heart rate monitoring during training and competition in cyclists. Journal of Sports Sciences 16: 91-99
13. Laursen P, Shing C, Jenkins D (2003) Reproducibility of a laboratory-based 40-km cycle time-trial on a stationary wind-trainer in highly trained cyclists. International Journal of Sports Medicine 24: 481-485
14. Paton C, Hopkins W (2001) Tests of cycling performance. Sports Medicine 31: 489-496
15. Paton CD, Hopkins WG (2006) Variation in performance of elite cyclists from race to race. European Journal of Sport Science 6: 25-31
16. Smith M, Davison R, Balmer J, Bird S (2001) Reliability of mean power recorded during indoor and outdoor self-paced 40 km cycling time-trials. International Journal of Sports Medicine 22: 270-274
17. Sporer B, McKenzie D (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
18. Stöggl T, Sperlich B (2014) Polarized training has greater impact on key endurance variables than threshold, high intensity or high volume training. Name: Frontiers in Physiology 5: 33
19. Zavorsky G, Murias J, Gow J, Kim D, Poulin-Harnois C, Kubow S, Lands L (2007) Laboratory 20-km cycle time trial reproducibility. International Journal of Sports Medicine 28: 743-748