J Sci Cycling. Vol. 5(3), 23-27

RESEARCH ARTICLE

Open Access

The Reliability of a Laboratory-based 4 km Cycle Time Trial on a Wahoo KICKR Power Trainer

Emma K Zadow[1]✉, James W Fell[1], Cecilia M Kitic[1]

Abstract

The purpose of the present study was to evaluate the reliability of a laboratory-based 4 km cycling time trial using a Wahoo KICKR Power Trainer. Twelve trained male cyclists (age: 34.0 ± 6.5 years; height: 1.78 ± 0.62 m; training per week: 11.9 ± 2.6 hours) completed three 4 km time trials on the Wahoo KICKR Power Trainer, with each time trial separated by a minimum of two days. During all time trials, mean power (W), cadence (rpm), speed (km.h-1), heart rate (bpm) and total time (s) were recorded with rating of perceived exertion (6-20) collected immediately post time trial. Average Intraclass Correlation Coefficients (ICC) between time trials (2v1, 3v2, 3v1) for power was 0.94 (95%CI: 0.85-0.98), cadence 0.73 (95%CI: 0.46-0.90), speed 0.54 (95%CI: 0.22-0.82), heart rate 0.93 (95%CI: 0.84-0.98) and total time 0.64 (95%CI: 0.34-0.86). Mean reliability expressed as the coefficient of variation (CV) and typical error of measurement over the three time trials was 3.4%, 5.2%, 4.2%, 1.6% and 4.3% for power, cadence, speed, heart rate and total time, respectively. Average power measured during a laboratory-based 4 km cycling time trial is highly reliable in trained cyclists making it a reliable method for monitoring cycling performance, however, caution should be applied when assessing cadence, speed and total time due to the larger typical errors when completed on the Wahoo KICKR Power Trainer..

Keywords: reproducibility, power output, cycling, athletic performance, performance test

✉ **Contact email:** *emma.zadow@utas.edu.au* (EK. Zadow)

[1] *School of Health Sciences, University of Tasmania, Australia.*

Received: 20 May 2016. Accepted. 14 October 2016.

Introduction

In professional cycling, small differences in performance can often determine the difference between a finish on the podium and a finish within the peloton, therefore, the ability to monitor training and competitive performance changes in highly-trained cyclists is of high importance (Currell and Jeukendrup 2008; Lamberts et al. 2009). To detect these meaningful changes within power output, laboratory based performance tests which replicate competitive performances have been shown to possess good test-retest reliability and have the precision to detect changes as small as 1% (Currell and Jeukendrup 2008; Hopkins et al. 2001; Paton and Hopkins 2006). In addition, the knowledge of test-retest reliability may determine how sensitive a test is to monitor changes in performance, inform sample size calculations for research studies and enable the comparison of ergometer precision (Hopkins et al. 2001).

Time trials (TTs), in which athletes complete a set amount of work in as short a time as possible, have been shown to provide coaches and sports scientists with the ability to monitor responses and detect changes associated with training and fatigue whilst providing accurate representations of the bioenergetics required in competitive cycling (Hopkins et al. 2001; Meeusen et al. 2010). In a review of the validity, reliability and sensitivity of measures of sporting performance, TTs of distances ranging from 5 to 40 km, have been shown to be highly reliable in well-trained cyclists with coefficient of variations (CVs) of less than 5% (Currell and Jeukendrup 2008), with smaller CVs of 1.9-2.4% observed when completed on cyclists' own bicycles using a Kingcycle™ and SRM™ ergometer (Smith et al. 2001).

When accustomed to the exercise protocol and cyclists' own bicycles are used, the ability to replicate the physiological demands of cycling (Abbiss et al. 2008) and movement economy (Driller et al. 2013) are improved, enhancing the ecological validity in measures of performance (Abbiss et al. 2009; Driller et al. 2013; Hopkins et al. 2001). Indeed, Lamberts et al. (Lamberts et al. 2009) reported low CVs of 0.7% and 1.7% in both 40 km performance time and 40 km mean power in cyclists who rode their own bicycle on an electromagnetically-braked cycle ergometer. The use of a cyclist's own bicycle in performance assessments as suggested by Paton and Hopkins (Paton and Hopkins 2001), is critical for producing reliable results predictive of competitive performance.

The Wahoo KICKR Power Trainer (KICKR) is an electromagnetically braked portable ergometer, which allows cyclists to use their own bicycles. Indeed, the

KICKR has been shown to provide valid measures of power (Zadow et al. 2016), falling within the recommended range of ergometer error of <2% (Hopkins et al. 1999; Hopkins et al. 2001). However, the reliability of a laboratory-based 4 km cycle time trial (TT) with cyclists using their own bicycle on the KICKR has yet to be reported. A 4 km TT may simulate real world performance such as the 4 000m individual and team pursuit that features at the UCI Track Cycling World Championships. Therefore, the aim of the present study was to determine the reliability of a 4 km cycling TT when completed on a Wahoo KICKR Power Trainer in trained cyclists.

Materials and methods

Participants

Twelve trained male cyclists (age: 34.0 ± 6.5 years, height: 1.78 ± 0.62 m, body mass: 76.8 ± 9.6 kg) with minimum weekly cycling duration of 10 h and previous TT experience volunteered to participate in the study. Participants were provided with written description of the risks and benefits of this study and provided signed informed consent. Ethics was obtained from the Institutional Human Research Ethics Committee. Furthermore, this study conforms to the ethical standards of the Journal of Science and Cycling (Harriss and Atkinson 2009).

Study Design

In a repeated measures study design, participants performed three 4 km cycling TTs in standard laboratory conditions (19 ± 1.2^{o}C and 45 ± 7.5% relative humidity) over three separate occasions within a two-week period. Testing sessions were separated by a minimum of 48 and a maximum of 72 h with testing performed at the same time of day (± 1 h) to minimise the effects of diurnal variation. Participants were required to avoid strenuous activity (<24 h) and caffeine (<12 h) before and on the day of testing. Participants completed a 24 h food diary prior to their first visit and were required to replicate their diet as closely as possible before each subsequent visit. All sessions were performed using the same equipment.

All TTs were performed on participants' own bicycles fitted to the Wahoo KICKR Power Trainer (KICKR: Wahoo Fitness, Atlanta, GA). The KICKR was calibrated using the Wahoo Fitness Utility Application (Wahoo Fitness, 2014, version 2.5) prior to and immediately post each participant's warm up before each TT. For a successful calibration, participants were required to reach a speed of 35.4 km.h-1 (22 mph) and cease pedalling until 16.0 km.h-1 (10 mph) had been reached.

A ten-minute self-selected intensity warm-up immediately followed the calibration process with participants free to alter their pedalling cadence and gear ratio as required. Immediately post warm-up, participants re-calibrated the KICKR with a standardised 60 s period of passive recovery provided following the re-calibration period, with a 10 s non-verbal countdown beginning the TT. Participants were instructed to perform the 4 km TTs as fast as possible, commencing from a standing start position with no rolling resistance. Only feedback on distance elapsed was provided throughout each TT. A ten-minute cool-down of self-selected intensity immediately commenced upon completion of the 4 km TT. During each TT, heart rate was recorded at a beat by beat frequency (Wahoo Fitness Blue HR, Atlanta, GA) with speed, cadence and power output recorded at a frequency of 1 Hz via the Wahoo Fitness Application for the KICKR (Wahoo Fitness, 2014, version 5.1.1) and Wahoo Fitness Blue Speed and Cadence (Wahoo Fitness, Atlanta, GA), respectively. Rating of perceived exertion (RPE) for the TT was determined upon immediate completion of the TT using a 6-20 Borg Scale (Borg 1970).

Statistical analysis

Average power, cadence, speed, heart rate and total time between TTs were logarithmically transformed and evaluated using Intraclass Correlation Coefficient (ICC) in combination with 95% confidence intervals (CI), analysed using an Excel spreadsheet for reliability (Hopkins 1997). Thresholds for assigning qualitative terms to the strength of within participant intraclass correlations were as follows: 0.5-0.69 low; 0.7-0.79 moderate; 0.8-0.89 high; 0.9-1.0 nearly perfect (Vincent 2005). Typical error expressed as a CV% of an absolute value with upper and lower 95% CI were examined between TTs using the Excel spreadsheet of Hopkins (Hopkins 1997). Based on previous research (Hopkins et al. 2001), a coefficient of variation lower than 3.5% was regarded as having high test-retest reliability. A one-way ANOVA with repeated measures was used to determine any differences in power output and was analysed using GraphPad Prism version 5.03 for windows (GraphPad software, La Jolla California, USA). Due to a loss of signal of the cadence sensor as a result of operator error on one occasion, cadence data was obtained from 11 of the 12 cyclists.

Results

Mean (±SD) data for measures of power, total time, cadence, speed, heart rate and post-exercise ratings of perceived exertion for TTs 1, 2 and 3 are presented in Table 1. The ICC for mean power and heart rate were nearly perfect at 0.97 (95%CI: 0.92-0.99) with CVs < 2.4% between TT 1 and 2 (Table 2 and 3). Moderate to high ICCs of 0.87 (95%CI: 0.58-0.96), 0.70 (95%CI: 0.23-0.90) and 0.77 (95%CI: 0.36-0.93) were reported for cadence, speed and total time between TT 2 and 3 (Table 2). The average ICC and CV with 95%CI for measurements between TTs (2v1, 3v2, and 3v1) are presented in Table 2. Power output during the three time trials was not significantly different (P=0.198).

Table 1. Mean and standard deviation for total time (s), power (W), cadence (rpm), speed (km-h^{-1}) and heart rate (bpm) measured during each 4 km cycling time trial (TT). Ratings of perceived exertion (RPE) were measured upon completion of the 4 km TT.

	TT 1	TT 2	TT 3	Average
Total Time (s)	416.4 ± 22.8	421.0 ± 29.9	410.8 ± 27.6	416.1 ± 26.8
Power (W)	342 ± 42	341 ± 45	349 ± 37	344 ± 41
Cadence (rpm)	92 ± 10	93 ± 7	91 ± 8	92 ± 8
Speed (km.h-1)	33.7 ± 2.0	34.7 ± 2.4	35.4 ± 2.6	34.6 ± 2.3
Heart Rate (bpm)	174 ± 9	172 ± 8	171 ± 9	172 ± 9
RPE	18.4 ± 1.6	18.1 ± 1.7	18.5 ± 1.5	18.3 ± 1.6

Note that cadence n = 11.

Table 2. Mean within–participant intraclass correlation (ICC) and coefficient of variation (CV) between time trials. Data are presented as mean (95%CI).

	Power (W)	Cadence (rpm)	Speed (km.h-1)	Heart Rate (bpm)	Total Time (s)
ICC (2 to 1)	0.97 (0.91- 0.99)	0.78 (0.36- 0.93)	0.36 (-0.24- 0.76)	0.98 (0.94- 1.00)	0.51 (-0.07- 0.84)
ICC (3 to 2)	0.92 (0.75- 0.98)	0.87 (0.58- 0.96)	0.70 (0.23-0.90)	0.91 (0.70- 0.97)	0.77 (0.36- 0.93)
ICC (3 to 1)	0.80 (0.45- 0.94)	0.34 (-0.29- 0.77)	0.49 (-0.08- 0.82)	0.84 (0.34- 0.29)	0.52 (-0.02- 0.84)
Mean	0.94 (0.85- 0.98)	0.73 (0.46- 0.90)	0.54 (0.22- 0.81)	0.93 (0.84- 0.98)	0.64 (0.34- 0.86)
CV (2 to 1)	2.4 (1.7- 4.0)	4.9 (3.4- 8.8)	4.5 (3.1- 7.7)	0.8 (0.6- 1.4)	5.0 (3.4- 8.9)
CV (3 to 2)	3.8 (2.7- 6.5)	3.5 (2.4- 6.2)	3.9 (2.7- 6.7)	1.8 (1.3- 3.3)	3.7 (2.6- 6.7)
CV (3 to 1)	3.8 (2.7- 6.5)	6.8 (4.7- 12.2)	4.4 (3.1- 7.5)	1.8 (1.3- 3.2)	4.1 (2.9- 7.4)
Mean	3.4 (2.7- 4.7)	5.2 (4.1- 7.4)	4.2 (3.3- 5.9)	1.6 (1.2- 2.2)	4.3 (3.4- 6.1)

Note that cadence n = 11.

Discussion

The present study is the first to determine the reliability of a 4 km cycle TT in trained cyclists on the Wahoo KICKR Power Trainer. The results of this investigation show that power within a 4 km cycle TT when completed on the KICKR is highly reliable with a CV of 3.4% whereas measures of cadence, speed and total time were shown to be unreliable measures of performance with CV's of 5.2%, 4.2% and 4.3%, respectively.

The importance of detecting performance changes in athletes has been previously emphasized in a review by Hopkins et al. (Hopkins et al. 2001) with measures of performance required to show a strong relationship with competitive cycling performance (Hopkins et al. 1999; Smith et al. 2001; Sporer and McKenzie 2007). Laboratory based TTs are frequently used to detect these changes due to the low coefficients of variation (<5%) observed when using various laboratory ergometers (Currell and Jeukendrup 2008; Hopkins et al. 2001). When power output is the key performance variable in repeated TTs, lower CVs of 1.9-3.6% (Laursen et al. 2003; Smith et al. 2001; Sporer and McKenzie 2007) are reported, indicating better test-retest reliability and a more direct method for monitoring of exercise performance (Jeukendrup et al. 1996; Jeukendrup and VanDiemen 1998). For a 4 km TT completed on the KICKR, our findings indicate this cycling test to be highly reliable for power output. We observed a mean CV of 3.4% (CI: 2.4-4.7%) and a within-subject ICC of 0.94 (CI: 0.89-0.98) across three TTs, with the lowest CV (2.4%; CI: 1.7-4.0%) observed between TTs one and two (Table 2). While highly reliable, further research is required to quantify how changes in 4 km TT on the KICKR reflect changes in on road performance.

In the investigation of both short and longer duration TTs, mean power output for competitive and well-trained cyclists have been observed to consistently exceed 250W, falling within the acceptable range of ergometer error for the Wahoo KICKR Power Trainer (250-700W and 80-120rpm) (Zadow et al. 2016).

Indeed, on average, powers of 346± 38W, 323± 35W and 303± 35W have been reported for a 4, 20 and 40 km cycling TT, respectively (Smith et al. 2001; Sporer and McKenzie 2007; Zadow et al. 2015). The average power of 344± 41W (Table 1) observed within the present study is consistent with previously reported power outputs for TTs of varying distances.

To determine the reliability between TTs (within-subject variation), coefficient of variations for power output (2v1, 3v2, 3v1) was evaluated (Table 2). Based upon previous studies investigating multiple performance TTs, learning effects have been proposed to contribute to lower CVs observed between TTs two and three (0.9-2.1%) with larger CVs observed between TTs one and two (2.1-3.0%) (Hopkins 2000; Laursen et al. 2003; Sporer and McKenzie 2007; Thomas et al. 2012). To ensure the changes observed within repeated performances are not the result of a learning effect (Moir et al. 2004; Moir et al. 2005), a single familiarisation session has been shown to establish a high degree of reliability and should precede experimental TTs (Abbiss et al. 2008; Laursen et al. 2003). In contrast to previous findings, there appeared to be no learning effect within the present study in the absence of a familiarisation session with a larger CV of 3.8% (CI: 2.7-6.5%) observed between TTs two to three when compared to TTs one to two (CV: 2.4%, CI: 1.7-4.0%) (Table 2). It has been previously suggested that a familiarisation session may not always be necessary in trained cyclist (Paton and Hopkins 2001; Sporer and McKenzie 2007) with the lack of any learning effect from TT one to TT two observed due to the training status and previous TT experience of the cyclists recruited (Jeukendrup et al. 1996). The larger CV observed between TTs two to three can be attributable to an increase in average power output observed within the third and final TT (Table 1). With our findings similar to those observed by Jeukendrup et al. (Jeukendrup et al. 1996) in which performance was greater in the final TT, we propose the improved performance may be due to the participants knowledge that this was their final TT and thus may have been more motivated to complete this TT (Hopkins 2000).

With performance time popular in the assessment of reliability within cycling tests due to the ease of measurement, Sporer et al.(Sporer and McKenzie 2007), Palmer et al.(Palmer et al. 1996) and Schabort et al.(Schabort et al. 1998) have demonstrated time to be a reliable method for assessing laboratory based cycling TTs, with CVs of 0.8-1.7% for TTs ranging from 20 to 100 km in trained cyclists. The CV for total time to complete a 4 km cycling TT in our study contrasts previous findings with a CV of 4.3% observed. In combination with the large typical errors observed within speed and cadence (Table 2), this suggests total time, speed and cadence should not be used as key performance outcomes when assessing cycling performance when using the KICKR.

In conclusion, the mean power for a 4 km TT performed on the Wahoo KICKR Power Trainer is highly reliable in trained cyclists. With a mean CV of 3.4% for power a 4 km TT on the KICKR is able to detect cycling performance changes of 1.7% in trained cyclists.

Practical application

A 4 km TT performed on the Wahoo KICKR Power Trainer is a reliable and therefore sensitive test for coaches and sports scientists to monitor responses associated with performance, training, fatigue and ergogenic aid use within trained cyclists when monitoring and measuring power output only. When assessing 4 km TT performance on the Wahoo KICKR Power Trainer, cadence, speed and total time should not be used due to greater variations in reliability observed within this study. The ability for cyclists to use their own bicycles when attached to the Wahoo KICKR Power Trainer is highly advantageous and critical for producing reliable results predictive of competitive performance.

Acknowledgements

We would like to thank the cyclists for volunteering their time to complete this study. We would also like to thank Mr Stephen Stone, Mr Taylor Wilczynski and Mr Torben Partridge-Madsen for their assistance with the collection of data. At the time of this study, the corresponding author was the recipient of an Australian government funded Postgraduate Research Scholarship (Australian Postgraduate Award).

Conflict of interest

None declared

References

1. Abbiss CR, Levin G, McGuigan MR, Laursen PB (2008) Reliability of power output during dynamic cycling. Int J Sports Med 29: 574-578
2. Abbiss CR, Quod MJ, Levin G, Martin DT, Laursen PB (2009) Accuracy of the Velotron ergometer and SRM power meter. Int J Sports Med 30: 107-112
3. Borg G (1970) Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 3: 92-98
4. Currell K, Jeukendrup AE (2008) Validity, reliability and sensitivity of measures of sporting performance. Sports Med 38: 297-316
5. Driller MW, Argus CK, Shing CM (2013) The reliability of a 30-s sprint test on the Wattbike cycle ergometer. Int J Sports Physiol Perform 8: 379-383
6. Harriss DJ, Atkinson G (2009) International Journal of Sports Medicine - ethical standards in sport and exercise science research. Int J Sports Med 30: 701-702
7. Hopkins WG (1997) A new view of statistics. Will G Hopkins, Sportscience [online], p. Available from: URL: http://sportsci.org/resource/stats
8. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30: 1-15
9. Hopkins WG, Hawley JA, Burke LM (1999) Design and analysis of research on sport performance enhancement. Med Sci Sports Exerc 31: 472-485
10. Hopkins WG, Schabort EJ, Hawley JA (2001) Reliability of power in physical performance tests. Sports Med 31: 211-234
11. Jeukendrup A, Saris WH, Brouns F, Kester AD (1996) A new validated endurance performance test. Med Sci Sports Exerc 28: 266-270
12. Jeukendrup A, VanDiemen A (1998) Heart rate monitoring during training and competition in cyclists. J Sports Sci 16 Suppl: S91-99

13. 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: 33-44
14. Laursen PB, Shing CM, Jenkins DG (2003) Reproducibility of a laboratory-based 40-km cycle time-trial on a stationary wind-trainer in highly trained cyclists. Int J Sports Med 24: 481-485
15. Meeusen R, Nederhof E, Buyse L, Roelands B, de Schutter G, Piacentini MF (2010) Diagnosing overtraining in athletes using the two-bout exercise protocol. Br J Sports Med 44: 642-648
16. Moir G, Button C, Glaister M, Stone MH (2004) Influence of familiarization on the reliability of vertical jump and acceleration sprinting performance in physically active men. J Strength Cond Res 18: 276-280
17. Moir G, Sanders R, Button C, Glaister M (2005) The influence of familiarization on the reliability of force variables measured during unloaded and loaded vertical jumps. J Strength Cond Res 19: 140-145
18. Palmer GS, Dennis SC, Noakes TD, Hawley JA (1996) Assessment of the reproducibility of performance testing on an air-braked cycle ergometer. Int J Sports Med 17: 293-298
19. Paton CD, Hopkins WG (2001) Tests of cycling performance. Sports Med 31: 489-496
20. Paton CD, Hopkins WG (2006) Ergometer error and biological variation in power output in a performance test with three cycle ergometers. Int J Sports Med 27: 444-447
21. Schabort EJ, Hawley JA, Hopkins WG, Mujika I, Noakes TD (1998) A new reliable laboratory test of endurance performance for road cyclists. Med Sci Sports Exerc 30: 1744-1750
22. Smith MF, Davison RC, Balmer J, Bird SR (2001) Reliability of mean power recorded during indoor and outdoor self-paced 40 km cycling time-trials. Int J Sports Med 22: 270-274
23. Sporer BC, McKenzie DC (2007) Reproducibility of a laboratory based 20-km time trial evaluation in competitive cyclists using the Velotron Pro ergometer. Int J Sports Med 28: 940-944
24. Thomas K, Stone MR, Thompson KG, St Clair Gibson A, Ansley L (2012) Reproducibility of pacing strategy during simulated 20-km cycling time trials in well-trained cyclists. Eur J Appl Physiol 112: 223-229
25. Vincent JW (2005) Statistics in Kinesiology. Human Kinetics Books, Champaign, IL
26. Zadow EK, Gordon N, Abbiss CR, Peiffer JJ (2015) Pacing, the missing piece of the puzzle to high-intensity interval training. Int J Sports Med 36: 215-219
27. Zadow EK, Kitic CM, Wu SS, Smith ST, Fell JW (2016) Validity of Power Settings of the Wahoo KICKR Power Trainer. Int J Sports Physiol Perform