RESEARCH ARTICLE **Open Access**

Comprehensive reliability analysis of a work-based (~420 kJ) cycling time-trial in recreationally-trained individuals

Luana F. de Oliveira[1]✉, Guilherme Yamaguchi[1], Victor de S. Painelli[1], Rafael P da Silva[1], Lívia S. Gonçalves[1], Bruno Gualano[1] and Bryan Saunders[1]

Abstract

The aim of this study was to comprehensively determine the reliability of performance and physiological measurements during a simulated work-based cycling time-trial (TT), without (Part A) and with (Part B) a familiarisation session using a battery of statistical tests. Fifty recreationally-trained cyclists performed a work-based cycling TT test on two (Part A) or three (Part B) occasions. Mean power output, blood lactate and ratings of perceived exertion (RPE) were recorded following completion of 25, 50, 75, 90 and 100 % of the test. Overall mean power output was analysed using intraclass correlations (r), systematic bias and ratio limits of agreement, coefficient of variation (CV), t-tests and Cohen's d (d). Pacing strategy data, blood lactate and RPE were analysed using repeated measures ANOVA and Tukey tests for post hoc comparisons. Overall mean power output in Part A (P = 0.11, d = 0.08, r = 0.95, CV = 3.04 ± 2.25 %) and in Part B (P = 0.72, d = 0.05, r = 0.87, CV = 2.93 ± 2.65 %) was not different between trials. Mean power output, blood lactate and RPE were not different between trials at any time point throughout the TT in Part A and B. The simulated cycling TT was shown to be reliable using a battery of statistical tests in recreationally-trained cyclists, with and without a familiarisation session.

Keywords: cycling, work-based time-trial, learning effect, familiarisation, pacing strategy, individual variability, reliability.

✉ luanaindaia@yahoo.com.br (LF. de Oliveira)

[1] University of Sao Paulo, Sao Paulo, Brazil.

Received: 17 September 2016. Accepted: 2 February 2017

Introduction

Individual cycling time-trials (TTs) require the individual to complete a set distance within the shortest period of time possible. These stages are considered the ultimate test of an individual's athletic ability (Mujika and Padilla 2001). Simulated TTs performed on a stationary ergometer in a laboratory provide a controlled environment in which researchers can investigate the physiological profile of an individual during this type of exercise and the effects of a dietary or training intervention on performance. These laboratory tests are usually distance or work-based protocols and the exact performance task that is selected will likely impact upon the ability to detect differences to an intervention.

Prior to accepting that all variation in the exercise response to an intervention is due to inter-individual differences, it is important to determine the likely measurement error associated with the protocol employed (Shephard et al. 2004). Cycling TTs have been shown to be reliable exercise protocols with low coefficients of variation (CVs (Bellinger and Minahan 2014; Currell and Jeukendrup 2008; Driller 2012; Jeukendrup et al. 2008; Jeukendrup et al. 1996)). Previous research has used the CV of a test to determine an individual's response (or non-response (Saunders et al. 2014)) to an intervention, though the use of CVs is only one of several methods that can be employed. Studies have employed 2× the typical error of a measurement (Gurd et al. 2016) or the systematic bias and random error components of the 95% limits of agreement (LoA) on the ratio scale (Hulse et al. 2013) to determine individuals who responded and those that did not. Whatever the preferred method, it seems appropriate to perform a robust statistical analysis of the chosen test to ensure its suitability and sensitivity to detect the desired differences due to an intervention.

Blood data and perception of effort are also commonly employed measures throughout exercise that require a certain level of sensitivity to ascertain as to whether these markers have changed following an intervention. Thus, it is also important that these measured variables throughout the test are consistent between repeated trials. Pacing strategy relates to the changes in power output that occur throughout exercise (Abbiss and Laursen 2008). Since alterations in pacing strategy have been shown following nutritional interventions (Correia-Oliveira et al. 2014; Santos et al. 2013), leading to an improved overall performance, it is necessary to ensure that the variation in pacing strategy is sufficiently small to allow detection of meaningful changes.

Although CVs for TTs have been shown to be low (<5% (Currell and Jeukendrup 2008)), measurement variation will always exist due to a number of random and systematic errors (Lamberts et al. 2009) including the ability of the ergometer to accurately measure power output (Paton and Hopkins 2001), circadian variation (Fernandes et al. 2014), pre-trial diet (Jeacocke and Burke 2010) and the ability of the individual (Hopkins et al. 2001). To minimise variation, studies generally perform trials at the same time of day using the same equipment following similar nutritional intake while also including a familiarisation session in order to avoid possible learning effects contributing to any significant changes in subsequent performances. Jeacocke and Burke (2010) noted that almost 20% of studies did not sufficiently implement standardisation of nutritional intake prior to performance trials, which may have resulted in dietary differences between trials influencing performance. Although studies have demonstrated the importance of familiarisation to a protocol (Kohler et al. 2010; Mendez-Villanueva et al. 2007), the more experienced the cyclist, the lower the between-test variation (Hopkins et al. 2001). Therefore, it could be hypothesised that individuals who regularly engage in the exercise activity investigated will display a high level of consistency without the necessity of a familiarisation session.

The aim of this study was to comprehensively determine the reliability of a simulated work-based TT in recreationally-trained cyclists. Furthermore, we aimed to determine whether reliability was shown between trials with, and without, a prior familiarisation session. It was hypothesised that performance and physiological measurements would exhibit a high degree of consistency with and without prior familiarisation.

Materials and methods

Subjects

Fifty (Part A: age 37±8 y, body mass 74.3±8.8 kg, height 1.76±0.06 m, maximum oxygen uptake [VO_{2max}] 50.7±7.3 ml•kg^{-1}•min^{-1}, maximal power output [Wmax] 329±52 W, experience 13±11 y, training 11±7 h•week^{-1} and 283±137 km•week^{-1}) recreationally-trained male cyclists competing in regional, national and international competition volunteered and gave their written informed consent to participate in this study. Sixteen (age 37 ± 9 y, body mass 73.7 ± 10.4 kg, height 1.77 ± 0.05 m, VO2max 49.1 ± 7.0 ml·kg^{-1}·min^{-1}, maximal power output 339 ± 38 W) of these individuals took part in an extension of this study (Part B). All individuals were experienced in performing cycling TTs. Exclusion criteria included the use of creatine or beta-alanine in the past six months, the presence of any musculoskeletal disorder, or the use of anabolic steroids. The study was first approved by the University of São Paulo Ethics Review Committee and is in agreement with the ethical standard of the journal (Harriss and Atkinson 2011).

Design

Participants attended the laboratory on three (Part A) or four (Part B) occasions separated by a minimum of 72 h, with all trials performed at the same time of day to account for circadian variation (Fernandes et al. 2014). The first session comprised of an incremental cycling test to exhaustion to determine VO_{2max} and W_{max}. In the remaining two sessions individuals performed a simulated work-based TT (Part A); in Part B, individuals performed a familiarisation session on a separate day before performing two main sessions of the TT.

Methodology

Twenty-four hours prior to the trials, participants were required to refrain from alcohol, caffeine and strenuous exercise, while individuals were requested to maintain similar dietary intake during this 24-h. Food intake was monitored using a 24-h food diary; participants were given a book with instructions and illustrative examples on how to fill out the dietary recall. Food diaries were analysed by a nutritionist in the presence of the individual so that as much specificity relating to portion sizes and cooking methods could be determined. Energy and macronutrient intake were subsequently analyzed using Avanutri online software (Avanutri, Rio de Janeiro, Brazil).

Individual set up of the cycle ergometer (Lode Excalibur Sport, Groningen, Netherlands) was determined prior to the maximal test, recorded electronically and maintained for all subsequent trials. Participants were required to perform four submaximal 4-min stages starting at 75 W increased by 50 W until 225 W. Thereafter, intensity increased by 30 W every minute until volitional exhaustion. Ventilatory and gas exchange measurements were recorded using a portable breath-by-breath system (Cosmed K4b2, Rome, Italy); the highest mean value over a 30 s period during the test was defined as VO_{2max}. The last completed stage plus the fraction of time spent in the final non-completed stage multiplied by 30 W was defined as an individual's W_{max}.

A 5-min cycling warm up was performed at 125 W immediately followed by the individual TT. Participants were required to complete a predetermined amount of work equivalent to 25 min at 85% of their individual W_{max}; this protocol was based on Jeukendrup et al. (2008). The formula for total amount of work to be performed was as follows:

$$\text{Total amount of work} = 0.85 \times W_{max} \times 1500\text{s}$$

The cycle ergometer was set in linear mode, meaning work load was cadence dependent according to the formula:

$$W = \alpha \times (\text{rpm})2$$

The α value was based on W_{max} so that individuals were working at 85% of individual W_{max} at a cadence of 95 rpm. Participants were instructed to complete the exercise in the fastest possible time. No motivation or visual feedback (cadence or power output) was given to the individuals during the test although they were

Table 1. Absolute and relative reliability measures of mean power output during the time trial.

	Part A (N = 50)	Part B (N = 16)
Trial 1 (W)	227.2±35.4	221.9±20.3
Trial 2 (W)	224.5±34.7	223.2±28.2
CV (%)	3.04	2.93
Trial 1 (ln)	5.40 ±0.16	5.39±0.10
Trial 2 (ln)	5.40±0.16	5.38±0.13
Systematic Error (W)	8.0	9.4
Measurement Error (W)	11.2	12.6
Repeatability (W)	9.43 ± 6.72	9.39 ± 9.09
Smallest meaningful change (W)	7.0	4.8
Systematic Bias	0.989	1.002
×/÷ Ratio LoA	1.109	1.118
ICC (CI)	0.95 (0.91 – 0.97)	0.87 (0.67 – 0.95)
t-test	P=0.11	P=0.72
Variation LoA	225.8; 201.4, 247.6	222.5: 199.5, 249.3
Variation CV	225.8; 219.0, 232.7	222.5; 216.0, 229.1

W = Watts; ln = log transformed; CV = coefficient of variation; ICC = intra-class correlations; CI = confidence interval; LoA = limits of agreement.

informed when they had completed 25%, 50%, 75% and 90% of the exercise, since this coincided with data collection (see below). Participants were given no performance feedback until all trials were completed.

Finger-prick blood samples were taken at baseline; following completion of 25%, 50%, 75%, 90% and 100% of the test; and 5-min post-exercise. A 20 µL volume of blood was stored in the same volume of ice-cold 2% NaF solution, centrifuged for 5 minutes at 2g at 4°C and the resultant plasma stored at -80°C until analysis. Plasma lactate was determined spectrophotometrically using an enzymatic-colorimetric method (Katal, Interteck, São Paulo, Brazil). Ratings of perceived exertion (RPE) according to the 6-20 Borg scale (Borg 1982) were taken following 25%, 50%, 75%, 90% and 100% of the test.

Statistical Analysis

All analyses in Part A and B were performed on data from the final two sessions. Data are presented as mean±1SD, unless stated otherwise. Heteroscedasticity was examined using the correlation between the absolute differences between the two trials and their mean, on both raw and log transformed values. Overall mean power output (Part A: N=50; Part B: N=16) was analysed using intra-class correlations (ICC, 2 way fixed, repeated measures, absolute model (Weir 2005)), systematic bias ratio (Nevill and Atkinson 1997), ratio LoA (Bland and Altman 1986), CVs, t-tests and Cohen's d effect sizes (Cohen 1988). Measurement error, repeatability (Bartlett and Frost 2008) and the smallest meaningful change were also calculated (Hopkins 2004; Paton and Hopkins 2006). Exercise data throughout the test, namely mean power output (Part A: N=47; Part B: N=15), blood lactate (Part A: N=38; Part B: N=13) and RPE (Part A: N=44; Part B: N=14), were analysed using mixed-models with repeated measures in SAS (SAS 9.2, SAS Institute Inc., USA), with Trial and Time as fixed factors and Participants as a random factor. Tukey tests were used for post-hoc analyses. Statistical significance was accepted at P≤0.05.

Results

Part A: Overall mean power output was not different between trials (P=0.11, d=0.08 (Table 1)). Ratio systematic bias and LoAs, ICC and the CV (Range: 0.09 to 10.08%; Median: 2.39%) for mean power output are presented in Table 1. Time-to-completion (1861±142 and 1885±170 s, P=0.10, d=0.15, CV=3.04±2.25) and cadence (86±4 and 85±4 rev•min^{-1}, P=0.15, d=0.11, CV=1.52±1.13) were not different between trials.

There was a main effect of Time on mean power output throughout exercise (F=3.39; P=0.01), with mean power output between 75-90% significantly lower than 0-25% (P=0.007), with no Trial×Time interaction (P=0.59) indicating no differences between trials at any time point (Figure 1, Panel A). The CVs of mean power output throughout the test were 5.2±4.5, 3.0±2.6, 4.4±4.2, 5.8±5.5 and 8.3±7.3% for the 0-25, 25-50, 50-75, 75-90 and 90-100% splits.

Blood lactate was not different between trials (Trial, P=0.45), but was significantly increased from baseline throughout the test and following 5-min of recovery in all trials (for all time points, P≤0.001). There were no differences between trials at any time point (Trial×Time, P=0.26). RPE increased throughout the test (Time, P<0.0001) although there were no differences between trials at any time point (Trial×Time, P=0.89; Table 2). No differences in food intake were shown between trials (all P>0.05; Table 3).

Part B: Overall mean power output was not different between trials (P=0.72, d=0.05). Ratio systematic bias and LoAs, ICC and the CV (Range: 0.07 to 8.71%; Median: 4.5%) for mean power output are presented in Table 1. Time-to-completion (1951±161 and 1947±163 s, P=0.88, d=0.03, CV=2.93±2.65) and cadence (83±3 and 84±3 rev•min-1, P=0.88, d=0.03, CV=1.47±1.32) were not different between trials.

A main Time effect was shown on mean power output throughout the test (F=2.59; P=0.04), with a reduction in mean power output at 75-90% compared to 0-25% (P=0.03) (Figure 1, Panel B). However, there was no Trial×Time interaction (P=0.56), with similar values at

Table 2. Blood lactate and RPE in Part A and Part B throughout exercise. Data are mean ± 1SD. *P ≤ 0.001 from Pre-exercise. $^{\#}$P ≤ 0.001 from previous time point.

	Pre-exercise	25%	50%	75%	90%	100%	5-min post-exercise
Lactate (mmol·L^{-1})							
Part A							
Trial 1	0.9±0.9	4.5±3.3*	4.5±2.6*	4.4±2.7*	4.3±2.4*	4.7±2.7*	3.3±2.5*
Trial 2	1.3±1.3	4.6±2.5*	4.1±2.6*	4.0±2.4*	3.5±2.5*	3.8±2.3*	3.0±2.0*
Part B							
Trial 1	0.8±0.7	3.7±2.1*	3.8±2.2*	3.5±1.8*	2.7±1.9*	3.5±2.1*	2.4±1.4*
Trial 2	0.4±0.4	4.3±2.8*	4.8±2.5*	4.5±2.5*	4.4±1.8*	4.8±2.1*	3.3±1.7*
RPE							
Part A							
Trial 1		13±2	15±2$^{\#}$	16±2$^{\#}$	17±2$^{\#}$	17±2$^{\#}$	
Trial 2		13±2	14±2$^{\#}$	15±2$^{\#}$	17±2$^{\#}$	17±3$^{\#}$	
Part B							
Trial 1		13±1	14±2$^{\#}$	15±2$^{\#}$	16±2$^{\#}$	17±2$^{\#}$	
Trial 2		13±2	14±2$^{\#}$	15±2$^{\#}$	16±2$^{\#}$	17±2$^{\#}$	

every time point between trials (Figure 1, Panel B). The CVs of mean power output throughout the test were 3.1±3.0, 3.5±2.7, 2.9±2.8, 5.3±4.8 and 8.6±10.3% for the 0-25, 25-50, 50-75, 75-90 and 90-100% splits.

Blood lactate was significantly increased throughout the test compared to baseline (all P<0.0001; Table 2). There was no difference between trials for blood lactate (Trial, P=0.14) nor were there any differences at any time point between trials (Trial×Time, P=0.13; Table 2). RPE increased throughout the test (P<0.0001) without differences between trials at any time point (Trial×Time, P=0.97; Table 2). There were no differences in food intake between trials (all P>0.05; Table 3).

Discussion

The reliability of a work-based (~420 kJ) cycling TT was demonstrated using a battery of statistical tests; overall mean power output was shown to be highly reproducible, while pacing strategy, blood lactate and perceived exertion throughout exercise were similar between trials. The reliability of these variables was shown both with, and without, a prior familiarisation session.

The CV of the test here (~3%) is in line with the <5% variation shown for time-trials in a review (Currell and Jeukendrup 2008), although it is higher than the 1.1% previously shown using this exact test (Jeukendrup et al. 2008). The differences between these two studies may be due to the athletic level of the cyclists; our cyclists were deemed recreationally-trained according to the criteria of De Pauw et al. (2013) while those recruited to the study of Jeukendrup et al. (2008) would be characterised as well-trained. Indeed, the ~30 min completion times in the current study were far slower than the ~25 min times reported by Jeukendrup et al. (2008). Thus, despite the current results suggesting that this test is reliable in recreationally-trained cyclists, there may have been less absolute variability than if a more trained sample population were employed. That notwithstanding, Saunders et al. (2016) showed a 4.1% improvement in TT performance using the same test in similarly recreationally-trained cyclists (despite their terminology of "trained cyclists"), highlighting the sensitivity of this test in the sample population employed to detect changes following an intervention designed to

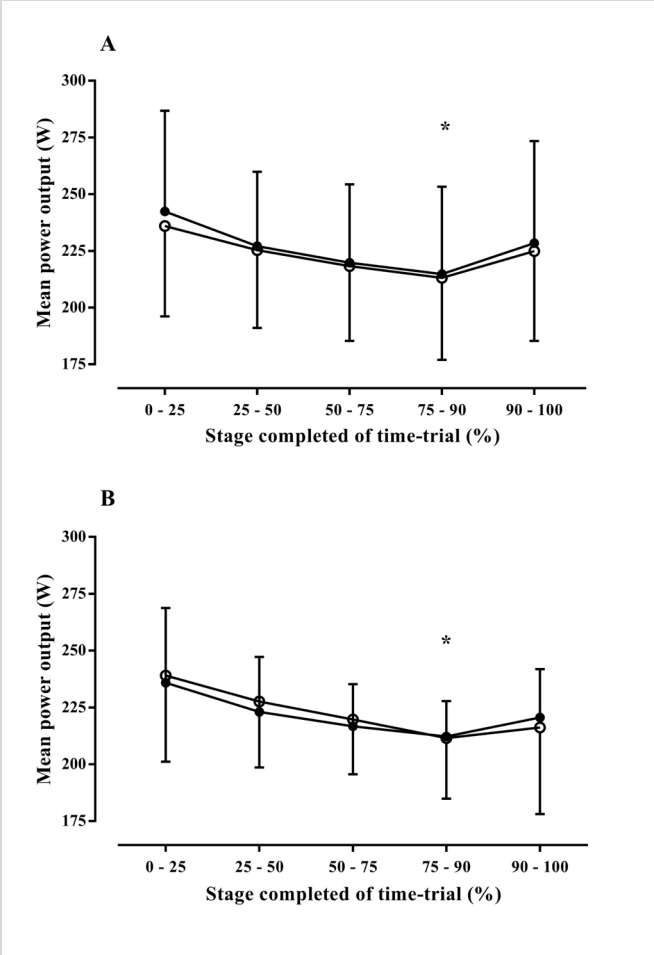

Figure 1. Mean power output (W) throughout the work-based time-trial in Part A (Panel A) and Part B (Panel B). Trial 1 is represented by black circles and Trial 2 by white circles. * P < 0.05 from 0-25%.

improve performance. Nonetheless, it would be of interest to determine the reliability of such a protocol using different sample populations ranging from untrained through to professional, according to the classifications of De Pauw et al. (2013).

Relative and absolute reliability for overall mean power output during the work-based TT, with and without a familiarisation trial, was demonstrated by non-significant ratio bias values close to 1, narrow agreement ratios and ICCs just below and above 0.9. The systematic bias and random error components of the 95% LoA on the ratio scale is an alternative method used to establish the range of variation associated with the measure (Hulse et al. 2013). The calculated ratio bias is multiplied and divided by the agreement ratio, and subsequently used to calculate the normal range within which a variable could lie. As an example, for an individual in Part A who performed the TT with an mean power output of 225.8 W, the systematic bias and agreement ratio was 0.989 and 1.109 respectively, giving subsequent values of 1.097 (0.989*1.109) and 0.892 (0.989/1.109). Therefore, given the measurement error indicated by the bias and ratio, in order to be 95% certain performance was better or worse than their previous attempt, mean power output would have to be above 247.7 W (1.096*225.8) or below 201.4 W (0.892/225.8). However, the systematic bias and ratio agreement method appears to be a highly stringent method by which to analyse changes in performance. The CV can also be used to determine the variation of the test where, if an individual maintains 225.8 W throughout a test with a CV of 3.04%, an improved performance would have to be above 232.7 W (225.8+[225.8*0.0304]) or below 219.0 W (225.8–[225.8*0.0304]).

Although reliability and reproducibility has long been considered an essential aspect of an exercise protocol, it is too often overlooked in research meaning results following an intervention cannot be contextualised. However, it is now becoming increasingly important due to interest in individual variability and characterisation of responders and non-responders to an intervention (Dias et al. 2015; Saunders et al. 2014; Shephard et al. 2004). To determine whether an intervention has had an overall effect it is possible to perform a number of analyses to confirm this statistically. However, to determine whether a solitary individual has benefitted or not, it is not possible to robustly test this with statistics. It has previously been considered that an individual improving above the CV of a test may be a suitable method to determine an individual's response (or non-response (Saunders et al. 2016; Saunders et al. 2014)). Other studies have used 2× the typical error of a measurement to determine individuals who responded and those that did not (Raleigh et al. 2016). It must be noted that the different reliability calculations used here resulted in a substantial range of variation between methods. Whichever method employed, individual inferences and comparisons require reliability data for the chosen exercise measure and, thus, here we provide a number of different measures to allow determination of the most suitable or stringent method desired.

Table 3. Food intake prior to the trials in Part A and Part B.

	Trial 1	Trial 2	P
Part A			
Protein (%)	18.3±5.8	18.7±5.8	0.73
Carbohydrate (%)	53.4±9.3	54.1±8.4	0.69
Fat (%)	28.1±7.5	27.3±6.3	0.55
Caloric intake (Kcal)	2407±961	2595±916	0.16
Part B			
Protein (%)	18.2±5.1	18.1±5.9	0.94
Carbohydrate (%)	54.0±9.7	53.3±7.8	0.95
Fat (%)	27.32±9.0	28.6±7.9	0.94
Caloric intake (Kcal)	2542±1082	2490±1034	0.84

Mean power output at the various split times throughout the test revealed no differences at any time point between trials, with or without a familiarisation session. Previous research has shown differences in pacing strategy following an intervention (Correia-Oliveira et al. 2014; Santos et al. 2013), although these studies did not determine whether changes were above the variation of the test employed. The CVs for mean power output at the various splits in the current test ranged from 2.9 to 5.8% up to 90% of the test completed, while the final 10% of the test showed the largest variability (8.3 and 8.6%). This higher degree of variability at the end of the TT is in line with previous reliability data during endurance cycling (Thomas et al. 2012), although the CVs throughout the protocol here were generally higher than the aforementioned study. This may be due to differences in the intervals at which mean power output was measured. The intervals used in the current study were chosen to coincide with the moments at which other measurements (i.e., blood lactate and RPE) were taken, since knowledge of how much of the test was remaining may have influenced pacing strategy. Indeed, mean power output during the final 10% of the test was recovered (i.e., increased) from the previous split. Thus, although mean power output at various splits throughout this work-based were not different between trials, different split times or better trained cyclists may result in less variation.

Blood lactate was not different between trials at any time point, with the maximal absolute mean difference between trials at any time point during exercise reaching 1.0 mmol•L^{-1} in study A and 1.7 mmol•L^{-1} in part B. This compared favourably with previous reliability studies on prolonged cycling (Driller 2012; Thomas et al. 2012). Data from our laboratory have shown increased lactate concentrations following caffeine supplementation with absolute increases in excess of 2.0 mmol•L-1 (unpublished data), suggesting that interventions which may exert an influence upon lactate production will likely result in differences above those shown in the current study. Furthermore, perception of effort, as measured by RPE, was not significantly different

between trials with nearly identical values at every time point measured. Therefore, it appears lactate and RPE responses during the TT may be considered sufficiently reliable and sensitive to detect changes when investigating the effect of an intervention on these outcome measures.

It has been suggested that the more experienced the cyclist, the lower the chance of between-test variation (Hopkins et al. 2001), although, to our knowledge, no study has directly investigated whether reliability data were similar with and without prior familiarisation of the protocol employed. Thus, data from the current study suggest that recreationally-trained individuals may not require a familiarisation session since all measured variables were not different between trials both with and without a prior familiarisation trial. Indeed, the absolute and relative measures of reliability were very similar between the two parts of this study. Nonetheless, caution must be taken since all individuals in this study were recreationally-trained cyclists who had competed in cycling TTs previously, thus were experienced in the particular task performed. Indeed, Mendez-Villanueva et al. (2007) showed that individuals who were unfamiliar with the performance task undertaken required at least one familiarisation to obtain reliable results, although increasing familiarisation sessions led to even more stable results. Thus, it would appear reasonable to suggest that a familiarisation may not be necessary for individuals already familiar with the type of exercise task being undertaken, which would be of benefit in time-limited situations (i.e., during the competitive season), although inclusion of a familiarisation session may lead to more robust data.

One of the strengths of the current study is that dietary intake during the 24-h preceding the exercise trials was assessed using a food recall and comprehensively analysed by a trained nutritionist. Dietary intake will likely influence metabolism and subsequent exercise performance, although we are unaware of any study directly quantifying the effect of pre-trial food intake on the reliability of exercise performance. Nonetheless, several reviews exist that discuss the importance of controlling nutritional intake due to its potential impact upon exercise performance (Currell and Jeukendrup 2008; Hopkins et al. 1999). We allowed individuals to freely choose their own 24-h pre-trial food intake, but requested them to repeat this as closely as possible prior to each subsequent trial. Analysis suggests that dietary intake was closely replicated prior to each trial, meaning that food intake did not contribute significantly to variation in performance between trials. Intervention and, in particular, reliability studies should similarly employ comprehensive analysis of dietary intake to ensure this does not influence performance.

Conclusions

This study showed that a work-based cycling time-trial is as a useful and reliable test able to measure overall performance, as well as physiological and perception parameters in recreationally-trained cyclists, with or without a familiarisation session.

Practical Application

This work-based cycling time-trial is a useful and reliable test and can be used to investigate the effects of an intervention on prolonged cycling performance. Blood lactate responses and perception of effort during the test were similar between trials, suggesting that they are sufficiently sensitive to detect any changes due to a given intervention. Furthermore, it appears that recreationally-trained cyclists may not require a familiarization session prior to performing this work-based time-trial. Coaches can safely use this work-based time-trial without prior familiarization if the athlete's schedule does not allow it in the knowledge that performance data will be reliable should the athlete have prior experience in time-trials. The comprehensive battery of statistical tests employed allows sports scientists to determine individuals who responded or did not respond to an intervention using their preferred method, although some caution must be taken since the choice of measure will change the stringency of the analysis.

Conflict of interest

The authors declare that they do not have conflict of interests.

References

1. Abbiss CR, Laursen PB (2008) Describing and understanding pacing strategies during athletic competition. Sports Med 38: 239-252
2. Bartlett JW, Frost C (2008) Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 31: 466-475
3. Bellinger PM, Minahan C (2014) Reproducibility of a laboratory based 1-km Wattbike cycling time trial in competitive cyclists. Journal of Science and Cycling 3: 5
4. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1: 307-310
5. Borg GA (1982) Psychophysical bases of perceived exertion. Med Sci Sports Exerc 14: 377-381
6. Cohen J (1988) Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale (NJ)
7. Correia-Oliveira CR, Santos RA, Silva-Cavalcante MD, Bertuzzi R, Kiss MA, Bishop DJ, Lima-Silva AE (2014) Prior low- or high-intensity exercise alters pacing strategy, energy system contribution and performance during a 4-km cycling time trial. PLoS One 9: e110320
8. Currell K, Jeukendrup AE (2008) Validity, reliability and sensitivity of measures of sporting performance. Sports Med 38: 297-316
9. De Pauw K, Roelands B, Cheung SS, de Geus B, Rietjens G, Meeusen R (2013) Guidelines to classify subject groups in sport-science research. Int J Sports Physiol Perform 8: 111-122
10. Dias GFA, Silva VE, Painelli VS, Sale C, Artioli GG, Gualano B, Saunders B (2015) (In)Consistencies in responses to sodium bicarbonate supplementation: A randomised, repeated measures, counterbalanced and double-blind study. PLoS One 10: e0143086
11. Driller MW (2012) The reliability of a 30-minute performance test on a Lode cycle ergometer. Journal of Science and Cycling 1: 6
12. Fernandes AL, Lopes-Silva JP, Bertuzzi R, Casarini DE, Arita DY, Bishop DJ, Lima-Silva AE (2014) Effect of time of day on performance, hormonal and metabolic response during a 1000-M cycling time trial. PLoS One 9: e109954
13. Gurd BJ, Giles MD, Bonafiglia JT, Raleigh JP, Boyd JC, Ma JK, Zelt JG, Scribbans TD (2016) Incidence of nonresponse and

individual patterns of response following sprint interval training. Appl Physiol Nutr Metab 41: 229-234
14. Harriss DJ, Atkinson G (2011) Update--Ethical standards in sport and exercise science research. Int J Sports Med 32: 819-821
15. Hopkins WG (2004) How to Interpret Changes in an Athletic Performance Test. Sportscience 8: 7
16. Hopkins WG, Hawley JA, Burke LM (1999) Design and analysis of research on sport performance enhancement. Med Sci Sports Exerc 31: 472-485
17. Hopkins WG, Schabort EJ, Hawley JA (2001) Reliability of power in physical performance tests. Sports Med 31: 211-234
18. Hulse MA, Morris JG, Hawkins RD, Hodson A, Nevill AM, Nevill ME (2013) A field-test battery for elite, young soccer players. Int J Sports Med 34: 302-311
19. Jeacocke NA, Burke LM (2010) Methods to standardize dietary intake before performance testing. Int J Sport Nutr Exerc Metab 20: 87-103
20. Jeukendrup A, Hopkins S, Aragon-Vargas LF, Hulston C (2008) No effect of carbohydrate feeding on 16 km cycling time trial performance. Eur J Appl Physiol 104: 831-837
21. Jeukendrup A, Saris WH, Brouns F, Kester AD (1996) A new validated endurance performance test. Med Sci Sports Exerc 28: 266-270
22. Kohler RM, Rundell KW, Evans TM, Levine AM (2010) Peak power during repeated wingate trials: implications for testing. J Strength Cond Res 24: 370-374
23. Lamberts RP, Swart J, Woolrich RW, Noake 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: 11
24. Mendez-Villanueva A, Bishop D, Hamer P (2007) Reproducibility of a 6-s maximal cycling sprint test. J Sci Med Sport 10: 323-326
25. Mujika I, Padilla S (2001) Physiological and performance characteristics of male professional road cyclists. Sports Med 31: 479-487
26. Nevill AM, Atkinson G (1997) Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. Br J Sports Med 31: 314-318
27. Paton CD, Hopkins WG (2001) Tests of cycling performance. Sports Med 31: 489-496
28. Paton CD, Hopkins WG (2006) Variation in performance of elite cyclists from race to race. Eur J Sport Sci 6: 7
29. Raleigh JP, Giles MD, Scribbans TD, Edgett BA, Sawula LJ, Bonafiglia JT, Graham RB, Gurd BJ (2016) The impact of work-matched interval training on VO2peak and VO2 kinetics: diminishing returns with increasing intensity. Appl Physiol Nutr Metab 41: 706-713
30. Santos RA, Kiss MA, Silva-Cavalcante MD, Correia-Oliveira CR, Bertuzzi R, Bishop DJ, Lima-Silva AE (2013) Caffeine alters anaerobic distribution and pacing during a 4000-m cycling time trial. PLoS One 8: e75399
31. Saunders B, de Oliveira LF, da Silva RP, de Salles Painelli V, Goncalves LS, Yamaguchi G, Mutti T, Maciel E, Roschel H, Artioli GG, Gualano B (2016) Placebo in sports nutrition: a proof-of-principle study involving caffeine supplementation. Scand J Med Sci Sports
32. Saunders B, Sale C, Harris RC, Sunderland C (2014) Sodium bicarbonate and high-intensity-cycling capacity: variability in responses. Int J Sports Physiol Perform 9: 627-632
33. Shephard RJ, Rankinen T, Bouchard C (2004) Test-retest errors and the apparent heterogeneity of training response. Eur J Appl Physiol 91: 199-203
34. 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
35. Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19: 231-240