Road cycle TT performance: Relationship to the power-duration model and association with FTP

Original Investigation

Paul T. Morgan[1], Matthew I. Black[1], Stephen J. Bailey[1, 2], Andrew M. Jones[1] and Anni Vanhatalo[1]

[1]Sport and Health Sciences, College of Life and Environmental Sciences, University of Exeter, St. Luke's Campus, Heavitree Road, Exeter, EX1 2LU, UK.

[2]School of Sport, Exercise and Health Sciences, Loughborough University, Ashby Road, Loughborough, Leicestershire LE11 3TU

Address for Correspondence:
Anni Vanhatalo, PhD.
Department of Sport and Health Sciences, College of Life and Environmental Sciences, University of Exeter, St. Luke's Campus, Heavitree Road, Exeter, EX1 2LU, UK.
Tel: 01392 722 815; **E-mail**: A.Vanhatalo@exeter.ac.uk

Running head: Practical application of CP and FTP

Abstract word count: 221 words

Text-only word count: 4620 words

Number of figures and tables: 4 Figures, 1 Table, 1 Supplementary Table

ABSTRACT

Purpose: To determine the accuracy of critical power (CP) and *W′* (the curvature constant of the power-duration relationship) derived from self-paced time-trial (TT) prediction trials using mobile power meters to predict 16.1-km road cycling TT performance. This study also aimed to test the agreement between functional threshold power (FTP) and CP. **Methods:** Twelve competitive male cyclists completed an incremental test to exhaustion, a 16.1-km road TT, an FTP test, and 4–5 self-paced TT bouts on a stationary bike within the lab, using mobile power meters. **Results:** CP and *W′* derived from the power-duration relationship closely predicted TT performance. The 16.1-km road TT completion time (26.7 ± 2.2 min) was significantly correlated with the predicted time-to-completion (27.5 ± 3.3 min, r= 0.89, P<0.01). CP and FTP were not significantly different (275 ± 40 W vs. 278 ± 42 W, P>0.05); however, the limits of agreement between CP and FTP were 30 to -36 W. **Discussion:** The findings of this study indicate that CP and *W′* determined using mobile power meters during maximal, self-paced TT prediction trials can be used to accurately predict 16.1-km cycling performance, supporting the application of the CP and *W′* for performance prediction. However, whilst we demonstrated that the FTP was not significantly different from CP, the limits of agreement were too large to consider FTP and CP interchangeable.

Key words: Critical power; functional threshold power; power meter; power-duration relationship; time-trial

MAIN TEXT INTRODUCTION

The ability to perform high-intensity (i.e., within the severe-intensity domain) exercise is described by the hyperbolic relationship between power output (PO) and time to the limit of tolerance (T_{lim}) (Jones, Vanhatalo, Burnley, Morton, & Poole, 2010; Morton, 2006). The power asymptote of this relationship is termed the critical power (CP), which reflects the highest work rate that can be sustained without a progressive loss of intramuscular and systemic homeostasis (Black, Jones, Kelly, Bailey, & Vanhatalo, 2016; Poole, Ward, Gardner, & Whipp, 1988; Poole, Ward, & Whipp, 1990; Vanhatalo et al., 2016). The curvature constant of this relationship, *W′*, represents a fixed amount of work that can be completed above CP before T_{lim} (Moritani, Nagata, deVries, & Muro, 1981; Poole et al., 1988; Poole et al., 1990; Vanhatalo et al., 2016). During exercise above CP, the tolerable duration of exercise is predictable according to the following equation (derived from the power-duration relationship):

$$T_{lim} = W' / (P - CP) \qquad \text{[Eqn. 1]}$$

where P is a given severe-intensity PO. Determination of the power-duration relationship is, therefore, of considerable value for understanding high-intensity exercise tolerance and for predicting athletic performance (Jones et al., 2010; Morton, 2006).

Previous research has demonstrated that determination of the power-duration relationship permits accurate estimation of laboratory-based exercise performance (Chidnok et al., 2012; Chidnok et al., 2013; Hill, Poole, & Smith, 2002; Murgatroyd, Ferguson, Ward, Whipp, & Rossiter, 2011) and predicts field-based cycling performance (Black, Durant, Jones, & Vanhatalo, 2014; Smith, Dangelmaier, & Hill, 1999). However, it should be noted that these field-based investigations predicted performance using regression equations derived from time-trial (TT) performance and CP rather than incorporating both CP and *W′* in a prediction equation:

$$T_{lim} = (W - W')/CP \qquad \text{[Eqn. 2]}$$

When exercising within the severe-intensity domain, knowledge of the 2-parameter CP model (eqn. 2) permits a more accurate determination of exercise performance (Jones et al., 2010; Morton, 2006).

Despite evidence supporting CP as a powerful determinant of endurance performance (Black et al., 2014; Smith et al., 1999), its application has been hindered by the need for specialist equipment, and the arduous and time consuming protocol, which requires the performance of several (~3–5) maximal trials spanning ~2–15 minutes in duration (Jones et al., 2010; Hill et al., 2002; Poole et al., 1988; Vanhatalo, Doust, & Burnley, 2007). However, the advent of commercially available power meters, that are both valid and reliable (Bertucci, Duc, Villerius, Pernin, & Grappe, 2005; Gardner et al., 2004), provide the opportunity to assess the power-duration relationship using equipment widely available to cyclists. The use of a cycle-mounted power meter in combination with a static trainer enables the power-duration relationship to be derived using a series of self-paced maximal TTs where the cyclist is able to control the gear, cadence, and pacing strategy, more accurately replicating conditions in the field relative to the conventionally used constant work rate time-to-exhaustion trials (Hopkins, Schabort, & Hawley, 2001; Jeukendrup, Saris, Brouns, & Kester, 1996; Laursen, Francis, Abbiss, Newton, & Nosaka, 2007). Previous research has demonstrated the utility of cycle-mounted power meters to estimate CP in the field (Karsten, Jobson, Hopker, Jimenez, & Beedie, 2014; Karsten, Jobson, Hopker, Stevens, & Beedie, 2015). However, it remains unclear whether the power-duration relationship derived from TTs using cycle-mounted power meter equipment commonly used by cyclists can accurately predict performance in the field.

The functional threshold power (FTP) is a popular index of fitness used among cyclists to provide an estimate of the maximal sustainable (~1 h) PO (Gavin et al., 2012). Equivalent to 95% of the mean PO sustained during a maximal self-paced 20-min TT (Allen & Coggan, 2006), determination of the FTP can be incorporated into training rides with relative ease. However, despite recent research reporting correlations between FTP (derived from the 20-min TT) and $\dot{V}O_2$ max (Denham, Scott-Hamilton, Hagstrom, & Gray, 2017) and the so-called individual ‘anaerobic threshold’ (Borszcz, Tramontin, Bossi, Carminatti, & Costa, 2018), there is little evidence to support the physiological underpinnings of the FTP and no previous work has established its ability to predict performance.

The purpose of this study was: 1) to determine the accuracy with which 16.1-km road cycling TT performance may be predicted by CP and *W′* derived using a road-bike, static trainer and a cycle-mounted power meter; and 2) to assess the agreement between CP and FTP. We hypothesised that the CP and *W′* can be used to accurately predict 16.1-km road cycling TT performance; that FTP would be correlated with 16.1-km road cycling TT performance; and that CP and FTP would be positively correlated and not significantly different from one another.

MATERIALS AND METHODS

Subjects

Twelve healthy, club-level cyclists (mean ± SD: age, 25 ± 7 years, height 1.80 ± 0.05 m, body mass 75.6 ± 5.9 kg) volunteered and gave written informed consent to participate in this study, which had been approved by the University of Exeter Research Ethics Committee. This study conformed to the principles of the World Medical Association Declaration of Helsinki. Subjects reported to all testing sessions well-hydrated, having avoided strenuous exercise and caffeine ingestion for 24 h and 3 h prior to testing, respectively. All subjects were provided with general recommendations on maintaining adequate hydration prior to arrival. Testing was performed at the same time of day (± 90 min) for each subject and separated by at least 24 h.

Design

Subjects visited the laboratory on 9–10 occasions, and completed a 16.1-km road-based TT over a 6-week period during pre-season and a minimum of 72 h between testing sessions. All subjects completed: (i) an incremental test to determine peak oxygen uptake ($\dot{V}O_{2peak}$), the gas exchange threshold (GET) and peak aerobic PO; (ii) 4–5 TT prediction trials for determination of the power-duration relationship; (iii) a 20-min TT for determination of FTP; and (iv) a 16.1-km road TT (performed mid-way through the testing protocol). Following the initial incremental test, all tests were randomised (excluding the 16.1-km road TT), and performed on the subjects own road-bike with PO and work done measured via a mobile power meter integrated into the rear wheel (PowerTap G3 Hub, CycleOps, Madison, USA) connected wirelessly to a data logger (Edge 500, Garmin, Chicago, USA). The PowerTap G3 device was calibrated according to manufacturer's instructions prior to each test. The road-bike was loaded onto a static trainer (Elite Volare Trainer Mag Alu, Fontaniva, Italy) for the prediction trials (test ii) and 20-min TT (test iii), during which maximal resistance was placed upon the rear wheel. The trainer resistance (set at arbitrary units of '5') and tyre pressure (110 psi) was checked prior to each test. All laboratory-based tests were performed in similar environmental conditions (temperature, 18–20ºC; relative humidity, 45–55%). Subjects were provided with visual and verbal feedback regarding the elapsed distance completed, distance remaining as well as the elapsed work done and work remaining during the laboratory TTs to replicate feedback typically received during TT efforts performed in the field.

Incremental test

On the first laboratory visit, subjects performed a ramp-incremental cycling test for the determination of the GET, peak aerobic PO and the peak oxygen uptake on an electronically braked cycle ergometer (Lode Excalibur Sport, Groningen, the Netherlands). The $\dot{V}O_{2peak}$ was required as a validation criterion of a maximal test during the experimental trials (as described below).

Determination of the power-duration relationship

CP and *W′* were estimated via 4–5 (4 trials, n = 6; 5 trials, n = 6) self-paced, maximal TTs to obtain a range of times between ~2 and 15 min. This range of work rates was selected to ensure participants were exercising within the severe-intensity domain, which was verified by the measurement of $\dot{V}O_2$ during each trial. Subjects were instructed to complete a target total work as quickly as possible with the shortest trial completed within ~2 min and the longest lasting <15 min, with two trials spaced equally in between, with a minimum 5 min separating the shortest and longest trials (Bishop, Jenkins & Howard, 1998). Prior to each trial, each subject performed a standardised warm-up at a work rate below GET, followed by 5 min of passive rest. Subjects then completed a further 3 min of pedalling at their preferred cadence but with no resistance on the rear wheel. Subjects were familiarised to the maximal self-paced TTs and specific conditions of the task. Subjects performed a minimum of 2 TTs of a set amount of work to result in completion in ~5–7 min until the difference in repeated TT duration was <1.3% (Sparks et al., 2016). These trials were not included in the subsequent data analysis. As a quality control measure of the mathematical modelling of the power-duration parameters, *a priori* criteria were set for the standard errors associated with the CP and *W′*, such that if the standard errors exceeded 5% and 10%, respectively, after 4 prediction trials, additional trials were completed until the standard error of estimate (SEE) was considered acceptable. Any prediction trials where the end-exercise $\dot{V}O_2$ was <95 % of the subject's ramp test determined $\dot{V}O_{2peak}$ were excluded from the modelling of the power-duration relationship. In the one instance where this occurred, the participant was willing to revisit the laboratory to re-perform this trial.

Determination of the functional threshold power

Following the same experimental setup as described for the prediction trials, the FTP test started with 3 min of baseline pedalling at <90% GET at preferred cadence, followed by a 20-min maximal, self-paced TT. The FTP was defined as 95% of the mean PO achieved during the 20-min TT (Allen & Coggan, 2006). All subjects reported completing the FTP test as part of their regular training regime.

16.1-km road TT

The TT was performed in Exeter (Devon, UK), on a dry day, with minimal wind, and an ambient air temperature of ~14ºC. All subjects were familiar with the 16.1-km road TT route which they used regularly as part of club training sessions. The course initially directed participants out by ~7 km before making a U-turn and covering the same course in the opposite direction. There was minimal elevation, or variation in gradient or terrain throughout the trial (Figure 1). Subjects followed their normal pre-competition warm-up, and were instructed to perform maximally during the 16.1-km TT and not to draft. All subjects performed the TT on the same day, within the same hour. To reduce the possibility of drafting, start times were separated by a 1-min interval and assigned based on previous TT performance, so that the fastest cyclist started first. Time-to-completion was recorded to the nearest second and PO was used to calculate total work done (time integral x difference in PO). All subjects completed the 16.1-km TT on their own individual road bike, which were all of a similar high-standard, fitted with the same power meter device that was used for laboratory assessments such that 12 power meters of the same model were used throughout the study.

Breath-by-breath gas analysis

During all laboratory tests, pulmonary gas exchange was measured breath-by-breath using an online gas analyser (Mobile Jaegar Oxygen Pro, Hoechberg, Germany). The analyser was calibrated before each test with gases of known concentration, and a calibration syringe of known volume (3-L; Hans Rudolph, KS).

Data Analysis

The CP and W' parameters were estimated using 3 models: the hyperbolic power-time ($P\text{-}T_{lim}$) model (eqn. 1); the linear work-time ($W\text{-}T_{lim}$) model, where total work done is plotted against time (eqn. 3); and the linear inverse-of-time ($1/T_{lim}$) model, where PO is plotted against the inverse of time (eqn. 4):

$$W = CP \cdot T_{lim} + W' \qquad \text{(Eqn. 3)}$$

$$P = W' \cdot (1/T_{lim}) + CP \qquad \text{(Eqn. 4)}$$

The power (P) during the TTs was defined as the mean PO measured across the duration of the trial. The SEE associated with the CP and W' was expressed as coefficient of variation (CV%). The "total error" associated with the modelling of the power-duration parameters was calculated as the sum of the CV% associated with the CP and W'. The sum of the CV% was optimised for each individual by selecting the model with the smallest and highest total error (eqn. 1, 3 or 4) to produce the "best individual fit" (BIF) and "worst individual fit" (WIF) parameter estimates (Black, Jones, Bailey, & Vanhatalo, 2015; Black et al., 2016). The BIF and WIF parameter estimates were then used to predict 16.1-km road TT performance retrospectively by using eqn. 2 and the individual total work done which was measured for each subject.

Statistical analysis

One-way analysis of variance was used to assess differences in: (i) power-duration parameters between models (eqn. 1–3), and; (ii) $\dot{V}O_{2peak}$ achieved in the ramp incremental test, prediction trials, and the FTP test. Paired samples *t*-tests and Bland-Altman analyses were used to assess differences and agreement between the actual and predicted 16.1-km road TT performance times, between CP and FTP, and between BIF and WIF models. Similarity between actual and predicted time-trial performance as well as between CP and FTP was also assessed via the mean bias. Mean bias was calculated as the difference between this estimated value and the 'true' value of the parameter being estimated as expressed as a percentage of the true variable. The prediction of TT performance using FTP was assessed via regression analysis. Paired samples *t*-tests and Bland-Altman analyses were used to assess differences and agreement between the actual and predicted 16.1-km road TT performance times using the FTP. Pearson's product moment correlation coefficients were used to assess relationships between 16.1-km road TT performance and the GET, $\dot{V}O_{2peak}$, CP, and FTP. For calculation of effect size, Cohen's *d* was used for paired *t*-tests. Statistical significance was accepted when $P<0.05$ and data are presented as mean ± SD.

RESULTS

16.1-km road TT performance

Subjects completed the 16.1-km road TT in 26.7 ± 2.2 min. The mean PO and cadence were 296 ± 38 W and 94 ± 6 rpm, respectively. Total work done in the road TT was 467 ± 39 kJ. The group mean pacing strategy is displayed in Figure 1.

Ramp incremental test

During the ramp incremental test, subjects attained a peak aerobic PO of 427 ± 34 W and a $\dot{V}O_{2peak}$ of 4.73 ± 0.49 $L{\cdot}min^{-1}$ (63.5 ± 6.5 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$). The GET occurred at 2.08 ± 0.28 $L.min^{-1}$ and 157 ± 32 W. The $\dot{V}O_{2peak}$ measured during the ramp incremental test was not different from the end-exercise $\dot{V}O_2$ measured during the prediction trials (4.83 ± 0.46 $L.min^{-1}$) and the FTP test (4.81 ± 0.48 $L.min^{-1}$; P=0.59). The $\dot{V}O_{2peak}$ (r= -0.70, $P<0.01$) and peak PO (r= -0.84, $P<0.01$) during the ramp incremental test were significantly correlated with 16.1-km road TT performance (i.e. completion time), but no relationship was observed between TT performance and the GET (*r*= -0.31, *P*=0.33).

Power-duration parameters

During the maximal, self-paced TT prediction trials, mean PO ranged from 246 to 528 W, which resulted in completion times ranging from 84 to 789 s. Importantly, $\dot{V}O_{2peak}$ was attained during all trials, confirming that each trial was performed within the severe-intensity domain and, thus could be used to establish the power-duration relationship (Burnley & Jones, 2007; Hill et al. 2002). The mean cadence during the TT prediction trials was 94 ± 5 rpm, and was not significantly different to the preferred cadence selected during the ramp incremental test or the 16.1-km road TT (*P*=0.77). Group mean CP and *W′* estimated using the BIF were 275 ± 42 W and 20.0 ± 7.0 kJ, respectively. When using the WIF, CP (273 ± 42 W) and *W′* (20.6 ± 8.2 kJ) were not significantly different from BIF estimates (*P*=0.69, *d*=0.04). In addition, there were no differences between equations 1, 3 and 4 in CP or *W′* estimates (P=0.79; Table 1), which indicated low levels of random error within the prediction trial data (Hill & Smith, 1994). There was no difference between predicted TT performance using the BIF or WIF (BIF: 27.5 ± 3.3 min vs. WIF: 27.8 ± 2.9min, *P*=0.45, *d*=0.12). The BIF parameter estimates were obtained from the P-T_{lim} model in 2 subjects and from the 1/T_{lim} and W-T_{lim} in 6 and 4 subjects, respectively. The WIF parameter estimates were obtained from the P-T_{lim} model in 7 subjects and from the 1/T_{lim} and W-T_{lim} in 1 and 4 subjects, respectively. However, whilst the actual TT performance (26.7 ± 2.2 min) was not different to that predicted from the BIF model (*P*=0.13, *d*=0.29), there was a significant difference between actual performance and the prediction derived from the WIF model (*P*=0.02, *d*=0.47). Therefore, BIF model parameters were used in further analyses. The predicted time-to-completion underestimated actual TT performance with a mean bias of -49

± 104 s which corresponded to 2.9% of actual TT performance (Figure 2b). The CP was inversely correlated with 16.1-km road TT performance (*r*= -0.89, *P*<0.01; Figure 3a). No significant relationship was observed between *W′* and TT performance (*r*= 0.43, *P*=0.16). The 16.1-km road TT completion time was significantly correlated with the predicted time (27.5 ± 3.3min) to completion (*r*= 0.88, *P*<0.01; Figure 2a).

Functional threshold power

The group mean PO during the 20-min FTP test was 292 ± 44 W, which corresponded to an FTP of 278 ± 42 W (Figure 4). The FTP was not significantly different to CP (275 ± 42 W, *P*=0.57, *d*=0.07; Figure 2c; 2d). The FTP predicted TT performance with a mean difference of 60 s (3.8% difference, *P*=0.99). The FTP was positively correlated with the CP (*r*= 0.92, *P*<0.01; Figure 2c) with a mean bias of -3 ± 17 W (1.3%; Figure 2d). The deducted 5% of work output during the 20-min FTP test was not significantly correlated with the *W′* (*r*= 0.42, *P*=0.26). The FTP was inversely correlated with 16.1-km TT completion time (*r*= -0.87, *P*<0.01; Figure 3b).

DISCUSSION

The current study assessed the accuracy with which road cycling TT performance may be predicted by the parameters of the power-duration relationship (i.e., CP and *W′*) derived using equipment commonly used by cyclists for training. The main finding of this study was that consistent with our primary hypothesis, the power-duration relationship established in the laboratory provided a prediction that was not statistically different from, and was strongly correlated with, actual 16.1-km road TT performance. In agreement with our second hypothesis, the FTP was correlated with, and not different from, the CP, but the limits of agreement between CP and FTP were relatively large (+10.9 to -13.1%). Collectively, these findings substantiate the scientific foundations for the translation of laboratory-based assessments of the power-duration relationship to predict athletic performance in the field. However, the FTP should not be considered interchangeable with CP and the physiological justification for the FTP protocol remains questionable.

Previous research has shown that the CP alone is a strong predictor of field based cycling performance (Black et al., 2014; Smith et al., 1999). However, knowledge of the power-duration parameters (i.e., CP and *W′*) derived from the 2-parameter CP model (eqn. 2) should permit a more accurate determination of exercise performance >CP (i.e. within the severe intensity domain, Jones et al., 2010; Morton, 2006). To test this assumption, we used the 2-parameter CP model to predict 16.1-km road TT performance, which is typically performed at a mean PO slightly above CP (Brickley, Dekerle, Hammond, Pringle, & Carter, 2007). Indeed, the PO sustained during the 16.1-km TT (~296 W) was 7.8% greater than CP and 6.7% greater than FTP. This is in agreement with previous research showing that a power output at or close to CP can be maintained for ~20-60 min (Brickley et al., 2002; Bull et al., 2000; Housh et al., 1989; McLellan and Cheung, 1992). In addition, we showed that the 2-parameter CP model provided a prediction that was not statistically different from, and was in close agreement with, actual TT performance. The prediction underestimated the 16.1-km TT duration by ~2.9% (Figure 3), which is similar to TT test-retest reliability for trained cyclists performing in a laboratory setting (1.3–3.2%, Sparks et al., 2016).

Despite the 2-parameter CP model (r= 0.88) having a superior predictive capability compared to traditional performance parameters including $\dot{V}O_{2peak}$ (r= 0.70), ramp test peak PO (r= 0.84) and GET (r= -0.31), a similarly accurate performance prediction was provided based on

CP alone (r= 0.89). It should, however, be noted that knowledge of the power-duration relationship and the total work to be completed permits the estimation of performance according to equation 2. In contrast, knowledge of the CP without knowledge of either, or both, *W′* or total work to be completed necessitates the use of a regression equation to predict performance which is specific to a particular course and distance (Black et al., 2014). It may, therefore, be prudent to predict TT performance using the CP and *W′* rather than the CP or FTP alone.

In this study, we replicated the current procedures that are recommended to and adopted by athletes and coaches for measuring FTP (Allen & Coggan, 2006). The FTP, which is equivalent to 95% of the PO achieved during a maximal self-paced 20-min TT, is popular among cyclists to determine "sustainable PO for 1 hour", to distinguish between performance capabilities and to track changes in fitness (Gavin et al., 2012). We observed no significant difference and a statistically high level of agreement, between CP and FTP suggesting that the FTP may provide a practical means of estimating CP outside the laboratory. Given the strong relationship between CP and FTP (r= 0.92), and the similar trial duration of the FTP protocol and the 16.1-km TT, it is unsurprising that the FTP (~278 W) was in close agreement with the mean PO sustained during the 16.1-km TT and correlated with TT completion time (r= -0.87). However, it is important to note that the limits of agreement between CP and FTP in this study (+30 to -36 W) may be considered too large to be practically meaningful for athletes and coaches, and that the agreement between the two variables is likely to be coincidental rather than mechanistically linked.

The *W′* (~20 kJ) accounted for a very small proportion (~4.3%) of the total work performed during the 16.1-km TT (~466 kJ) and therefore, unsurprisingly, no significant relationship was found bet´een the *W′* and the amount of work performed above FTP during the 20-min TT. Whilst the physiological underpinnings of *W′* remain to be fully elucidated, it is indicative of a finite amount of work that can be performed above CP, and is associated with the $\dot{V}O_2$ slow component, depletion of muscle phosphocreatine, and the accumulation of fatigue-related metabolites (Burnley & Jones, 2007; Poole, Burnley, Vanhatalo, Rossiter, & Jones, 2016). Accordingly, during exercise above CP, exercise tolerance and/or performance is defined by the magnitude of the *W′* and its rate of utilisation (Burnley & Jones, 2016; Chidnok et al., 2013; de Souza et al., 2016; Skiba, Chidnok, Vanhatalo, & Jones, 2012).

Importantly, during exercise below the CP, the *W′* can be recovered at a rate dependent on the intensity and duration of the recovery interval, such that a greater reconstitution occurs at lower intensities and during longer duration recovery intervals (Burnley & Jones, 2016; Chidnok et al., 2013; de Souza et al., 2016; Skiba et al., 2012). During the self-paced 20-min TT, subjects cycled for periods below their CP (Figure 4), which would have permitted some reconstitution of the *W′*. However, we found mean *W′* (~20 kJ) to be higher than the group mean '5% power-time integral' (~ 16.8 kJ) from the FTP test. This is likely explained by 'lost time' spent below CP and the setting of a metabolic limit on the utilization of *W′* (Fukuba & Whipp, 1999). Future research is warranted to compare the accuracy of performance prediction by the 2-parameter CP model for continuous exercise (present study) against the intermittent CP model (Morton & Billet, 2004) and the '*W′* balance' model (Skiba et al., 2012) which account for fluctuations in PO including periods <CP. Furthermore, it should be noted that the determination of the FTP is based on the arbitrary subtraction of 5% of the

mean power output during a 20-min TT. Indeed, the work performed above FTP was not equivalent to W' thus questioning the physiological bases of the FTP protocol.

As mentioned above, the predictive accuracy of the power-duration relationship is influenced by a number of factors. It is also important to note that the power-duration relationship is only applicable in the severe intensity domain where a number of additional factors have an increasing important influence on performance and exercise tolerance (Black et al. 2017; Burnley & Jones, 2007), such that attainment of the $\dot{V}O_{2max}$ must be considered an obligatory criterion for a prediction trial to be included in mathematical modelling (Burnley & Jones, 2007; Hill et al. 2002). Furthermore, TTs have been reported to be more reliable than constant work rate trials, and better reflect the demands of field-based competition (Hopkins et al., 2001; Jeukendrup et al., 1996; Laursen et al., 2007). Therefore, in the current study, to minimise the influence of performance variability on the power-duration relationship, all exercise trials were self-paced, maximal TTs, where $\dot{V}O_{2peak}$ was attained, and to which subjects were familiarised. Moreover, all exercise trials were performed on subjects' personal road bikes allowing each subject to freely control their gear, cadence, and pacing strategy, thus accurately replicating conditions in the field. These considerations (i.e., matching of trials, familiarisation, good reliability and consistent attainment of $\dot{V}O_{2peak}$) ensured a good fit of the experimental data to the model.

Previous research investigating the power-duration relationship has typically adopted a single model (i.e., P-T_{lim}, 1/T_{lim} or W-T_{lim}) for the estimation of the power-duration parameters. To examine the importance of (in)accuracies in the modelled fit to the experimental data we compared the 2-parameter models associated with the least (BIF) and most (WIF) total error for each subject. Interestingly, and despite a good fit of each 2-parameter model to the experimental data, the WIF significantly underestimated actual TT performance whereas no significant difference was observed in actual and predicted performance using the BIF. The findings of the current study, therefore, highlight the importance of model selection, and support the adoption of the BIF model to ensure accurate prediction of performance (Black et al., 2015).

Practical Applications

This study demonstrates that the laboratory-based estimate of the power-duration relationship, determined using equipment readily available to cyclists, enables the accurate prediction of field-based cycling performance. It is recommended that athletes and practitioners consider incorporating assessment of the power-duration relationship into their normal training routine to monitor fitness and predict performance. The predictive accuracy can be improved by selecting the 2-parameter model that is associated with the lowest modelling error (i.e. BIF). However, whilst we demonstrate, statistically, that FTP provides a close approximate to CP, the limits of agreement are too large to consider FTP and CP interchangeable. Therefore, considering the arbitrary definition of the FTP as 95% of mean PO during a 20-min TT, we recommend that FTP is used with caution and instead encourage the use of the power-duration relationship for performance prediction.

Conclusion

The results of the present study demonstrate that the parameters of the power-duration relationship, CP and W', provide a performance prediction that is not statistically significant from, and is closely correlated with, 16.1-km road cycling TT performance. The CP was more strongly associated with performance than the ramp test peak PO, peak oxygen uptake, *W′*, GET, and FTP, providing further evidence to support the predictive validity and performance relevance of the CP. However, despite observing a close agreement between CP and FTP, the limits of agreement were too large to consider these variables as equivalent. Furthermore, due to its arbitrary definition, and that no relationship was observed between the work performed above FTP and *W′*, the physiological relevance of the FTP is questioned. Given the superior predictive capability of the power-duration relationship parameters compared to other traditional physiological variables, we encourage applied practitioners and athletes to incorporate CP testing into their training and testing routine as an aid to monitor fitness and predict performance.

Acknowledgements

This research was not supported by external funding. The authors are grateful to Jacob Durant and the Mid-Devon Cycling Club for their assistance during data collection.

Declaration of interest statement

The authors report to conflict of interest.

REFERENCES

Allen, H., & Coggan, A. *Training and Racing with a Power Meter.* Boulder (CO): VeloPress: p. 5–20, 41–65; 2006.

Bertucci, W., Duc, S., Villerius, V., Pernin, J. N., & Grappe, F. (2005). Validity and reliability of the PowerTap mobile cycling powermeter when compared with the SRM device. *International Journal of Sports Medicine*, *26*(10), 868–873. https://doi.org/10.1055/s-2005-837463

Bishop, D., Jenkins, D.G., & Howard, A. (1998). The critical power function is dependent on the duration of the predictive exercise tests chosen. *International Journal of Sports Medicine, 19*, 125–129. https://doi.org/10.1055/s-2007-971894

Black, M. I., Durant, J., Jones, A. M., & Vanhatalo, A. (2014). Critical power derived from a 3-min all-out test predicts 16.1-km road time-trial performance. *European Journal of Sport Science*, *14*(3), 217–223. https://doi.org/10.1080/17461391.2013.810306

Black, M. I., Jones, A. M., Bailey, S. J., & Vanhatalo, A. (2015). Self-pacing increases critical power and improves performance during severe-intensity exercise. *Applied Physiology, Nutrition, and Metabolism*, *40*(7), 662–670. https://doi.org/10.1139/apnm-2014-0442

Black, M. I., Jones, A. M., Kelly, J. A., Bailey, S. J., & Vanhatalo, A. (2016). The constant work rate critical power protocol overestimates ramp incremental exercise performance. *European Journal of Applied Physiology*, *116*(11–12), 2415–2422. https://doi.org/10.1007/s00421-016-3491-y

Black, M. I., Jones, A. M., Blackwell, J. R., Bailey, S. J., Wylie, L. J., McDonagh, S. T. J., … Vanhatalo, A. (2017). Muscle metabolic and neuromuscular determinants of fatigue during cycling in different exercise intensity domains. *Journal of Applied Physiology*, *122*(3), 446–459. https://doi.org/10.1152/japplphysiol.00942.2016

Brickley, G., Dekerle, J., Hammond, A.J., Pringle, J., & Carter, H. (2007). Assessment of maximal aerobic power and critical power in a single 90-s isokinetic all-out cycling test. *International Journal of Sports Medicine*, 28(5), 414-419.

Borszcz, F.K., Tramontin, A.F., Bossi, A.H., Carminatti, L.J., & Costa, V.P. (2018). Functional Threshold Power in Cyclists: Validity of the Concept and Physiological Responses. *International Journal of Sports Medicine, 25*. doi: 10.1055/s-0044-101546. [Epub ahead of print]

Burnley, M., & Jones, A. M. (2007). Oxygen uptake kinetics as a determinant of sports performance. *European Journal of Sport Science*, *7*(2), 63–79. https://doi.org/10.1080/17461390701456148

Burnley, M., & Jones, A. M. (2018). Power–duration relationship: Physiology, fatigue, and the limits of human performance. *European Journal of Sport Science*, *18*(1), 1–12. https://doi.org/10.1080/17461391.2016.1249524

Chidnok, W., DiMenna, F.J., Bailey, S.J., Wilkerson, D.P., Vanhatalo, A., & Jones, A.M. (2012). All-out Critical Power Test Predicts Time-to-exhaustion During Ramp Incremental and Constant-work-rate Exercise. *Medicine and Science in Sports and Exercise*, 44, 613-613.

Chidnok, W., Fulford, J., Bailey, S. J., DiMenna, F. J., Skiba, P. F., Vanhatalo, A., & Jones, A. M. (2013). Muscle metabolic determinants of exercise tolerance following exhaustion: relationship to the "critical power." *Journal of Applied Physiology, 115*(2), 243–250. https://doi.org/10.1152/japplphysiol.00334.2013

Denham, J., Scott-Hamilton, J., Hagstrom, A.D., & Gray, A.J. (2017). Cycling Power Outputs Predict Functional Threshold Power And Maximum Oxygen Uptake. *Journal of Strength and Conditioning Research, 1*. https://doi.org/10.1519/JSC.0000000000002253

de Souza, K. M., Dekerle, J., Salvador, P. C. do N., de Lucas, R. D., Guglielmo, L. G. A., Greco, C. C., & Denadai, B. S. (2016). Rate of utilization of a given fraction of W′ (the curvature constant of the power-duration relationship) does not affect fatigue during severe-intensity exercise. *Experimental Physiology*, *101*(4), 540–548. https://doi.org/10.1113/EP085451

Gardner, A. S., Stephens, S., Martin, D. T., Lawton, E., Lee, H., & Jenkins, D. (2004). Accuracy of SRM and Power Tap power monitoring systems for bicycling. *Medicine and Science in Sports and Exercise*, *36*(7), 1252–1258. https://doi.org/10.1249/01.MSS.0000132380.21785.03

Gavin, T. P., Van Meter, J. B., Brophy, P. M., Dubis, G. S., Potts, K. N., & Hickner, R. C. (2012). Comparison of a field-based test to estimate functional threshold power and power output at lactate threshold. *Journal of Strength and Conditioning Research*. https://doi.org/10.1519/JSC.0b013e318220b4eb

Hill, D.W., Poole, D.C., & Smith, J.C. (2002). The relationship between power and the time to achieve .VO(2max). *Medicine & Science in Sports & Exercise*, 34(4), 709-714. https://doi.org/10.1097/00005768-200204000-00023

Hill, D.W., & Smith, J.C. (1994). A method to ensure the accuracy of estimates of anaerobic capacity derived using the critical power concept. *Journal of Sports Medicine and Physical Fitness*, 34, 23–37.

Hopkins, W. G., Schabort, E. J., & Hawley, J. A. (2001). Reliability of power in physical performance tests. *Sports Medicine (Auckland, N.Z.)*, *31*(3), 211–234. https://doi.org/10.2165/00007256-200131030-00005

Karsten, B., Jobson, S.A., Hopker, J., Jimenez, A., & Beedie, C. (2014). High agreement between laboratory and field estimates of critical power in cycling. *International Journal of Sports Medicine*, 35, 298–303. doi:10.1055/s-0033-1349844.

Karsten, B., Jobson, S.A., Hopker, J., Stevens, L., & Beedie, C. (2015). Validity and reliability of critical power field testing. *European Journal of Applied Physiology,* 115, 197–204. doi:10.1007/s00421-014-3001-z.

Jeukendrup, A., Saris, W. H. M., Brouns, F., & Kester, A. D. M. (1996). A new validated endurance performance test. *Medicine and Science in Sports and Exercise*, *28*(2), 266–270. https://doi.org/10.1097/00005768-199602000-00017

Jones, A.M., & Vanhatalo, A. (2017). The 'Critical Power' Concept: Applications to Sports Performance with a Focus on Intermittent High-Intensity Exercise. *Sports Medicine, 47*, 65–78. https://doi.org/10.1007/s40279-017-0688-0

Jones, A. M., Vanhatalo, A., Burnley, M., Morton, R. H., & Poole, D. C. (2010). Critical power: Implications for determination of VO2max and exercise tolerance. *Medicine and Science in Sports and Exercise*, *42*(10), 1876–1890. https://doi.org/10.1249/MSS.0b013e3181d9cf7f

Laursen, P. B., Francis, G. T., Abbiss, C. R., Newton, M. J., & Nosaka, K. (2007). Reliability of time-to-exhaustion versus time-trial running tests in runners. *Medicine and Science in Sports and Exercise*, *39*(8), 1374–1379. https://doi.org/10.1249/mss.0b013e31806010f5

Moritani, T., Nagata, A., Devries, H. A., & Muro, M. (1981). Critical power as a measure of physical work capacity and anaerobic threshold. *Ergonomics*, *24*(5), 339–350. https://doi.org/10.1080/00140138108924856

Morton, R. H. (2006). The critical power and related whole-body bioenergetic models. *European Journal of Applied Physiology*. https://doi.org/10.1007/s00421-005-0088-2

Morton, R. H., & Veronique Billat, L. (2004). The critical power model for intermittent exercise. *European Journal of Applied Physiology*, *91*(2–3), 303–307. https://doi.org/10.1007/s00421-003-0987-z

Murgatroyd, S. R., Ferguson, C., Ward, S. A., Whipp, B. J., & Rossiter, H. B. (2011). Pulmonary O2 uptake kinetics as a determinant of high-intensity exercise tolerance in humans. *Journal of Applied Physiology*, *110*(6), 1598–1606. https://doi.org/10.1152/japplphysiol.01092.2010

Poole, D. C., Burnley, M., Vanhatalo, A., Rossiter, H. B., & Jones, A. M. (2016). Critical power: An important fatigue threshold in exercise physiology. *Medicine and Science in Sports and Exercise*, *48*(11), 2320–2334. https://doi.org/10.1249/MSS.0000000000000939

Poole, D. C., Ward, S. A., & Whipp, B. J. (1990). The effects of training on the metabolic and respiratory profile of high-intensity cycle ergometer exercise. *European Journal of Applied Physiology and Occupational Physiology*, *59*(6), 421–429. https://doi.org/10.1007/BF02388623

Poole, D. C., Ward, S. A., Gardner, G. W., & Whipp, B. J. (1988). Metabolic and respiratory profile of the upper limit for prolonged exercise in man. *Ergonomics*, *31*(9), 1265–1279. https://doi.org/10.1080/00140138808966766

Skiba, P. F., Chidnok, W., Vanhatalo, A., & Jones, A. M. (2012). Modeling the expenditure and reconstitution of work capacity above critical power. *Medicine and Science in Sports and Exercise*, *44*(8), 1526–1532. https://doi.org/10.1249/MSS.0b013e3182517a80

Smith, J. C., Dangelmaier, B. S., & Hill, D. W. (1999). Critical power is related to cycling time trial performance. *International Journal of Sports Medicine*, *20*(6), 374–378. https://doi.org/10.1055/s-2007-971147

Sparks, A.S., Williams, E., Jones, H., Bridge, C., Marchant, D., & McNaughton, L. (2016). Test-retest reliability of a 16.1 km time trial in trained cyclists using the CompuTrainer ergometer. *Journal of Science and Cycling*, *5*(3), 35–41.

Vanhatalo, A., Black, M. I., DiMenna, F. J., Blackwell, J. R., Schmidt, J. F., Thompson, C., … Jones, A. M. (2016). The mechanistic bases of the power–time relationship: muscle metabolic responses and relationships to muscle fibre type. *Journal of Physiology*, *594*(15), 4407–4423. https://doi.org/10.1113/JP271879

Vanhatalo, A., Doust, J. H., & Burnley, M. (2007). Determination of critical power using a 3-min all-out cycling test. *Medicine and Science in Sports and Exercise*, *39*(3), 548–555. https://doi.org/10.1249/mss.0b013e31802dd3e6

Figure and table captions

Figure 1

Group mean pacing strategy (white circles) during road 16.1-km time-trial (TT). The dashed black line represents the mean critical power (CP) and the dashed grey line represents the mean functional threshold power (FTP).

Figure 2

Correlation (A and C) and Bland–Altman analyses (B and D) for predicted and actual time-trial (TT) performance (s) and for critical power (CP) and functional threshold power (FTP). In panels A and C, the solid line is the best-fit linear regression, and the dashed line is the line of identity. Time trial performance was predicted using the 2-parameter power-duration model (i.e. CP and *W′*). In panels B and D (Bland–Altman plots), the dashed horizontal lines represent the 95% limits of agreement (LOA) and the solid black line represents the mean difference (MD) between the two measures.

Figure 3

Correlation between critical power (CP) and time-trial (TT) performance (panel A) and functional threshold power (FTP) and TT performance (panel B). Panel C illustrates Bland-Altman analysis for predicted and actual time-trial (TT) performance (s). Time trial performance was predicted using the FTP during linear regression. The solid black line represents the mean difference (MD) and the dashed horizontal lines represent the 95% limits of agreement (LOA) between the two measures.

Figure 4

Group mean power output (PO) during 20 min functional threshold power (FTP) test. The PO equivalent to critical power has also been included (dashed line).

Table 1

The parameter estimates derived from eqs. 1-3 and the best and worst individual fits for the time trials (TT). Total error indicates the sum of coefficients of variation (CV %) associated with critical power (CP) and the curvature constant (W').