

The reliability and construct validity of the functional threshold power test in trained cyclists

Lewis Gough ¹ , Ben Dobson ² , Jake Williams ¹, Gregory Downes ¹, Savannah Sturridge ³ , Ashley Warner ⁴ , Andy Sparks ⁵  and Daniel Brown ²

Received: 13 June 2024

Accepted: 19 March 2025

Published: 11 July 2025

¹ Human Performance and Health group, Centre for Life and Sport Sciences (CLaSS), Birmingham City University, B15 3TN, UK

² Sports Nutrition and Performance Research Group, Department of Sport and Physical Activity, Edge Hill University, Ormskirk, United Kingdom

³ School of Sport and Exercise Sciences, University of Kent, Canterbury, CT2 7PE, UK

⁴ Division of Sport and Exercise, School of Health and Life Sciences, University of the West of Scotland, Glasgow, UK

⁵ Maurten AB, Gothenburg, Sweden

⁶ Centre for Human Performance, Carnegie School of Sport, Leeds Beckett University, Leeds, UK

Correspondence

Ben Dobson

Sports Nutrition and Performance Research Group, Department of Sport and Physical Activity, Edge Hill University, Ormskirk, United Kingdom

bendobsonresearch@outlook.com

Abstract

The functional threshold power (FTP) 20-min test (FTP20) is popular amongst cyclists and coaches due to the theory it can predict the power output that can be sustained for 60-mins. However, little is known in terms of the reliability and validity of this construct, therefore the aim of this study was to assess the reliability of the FTP20 test and the construct validity of this test to predict 60-min power. Twenty-two male trained cyclists (age = 32 ± 10 years, body mass (BM) = 77.2 ± 6.8 kg, maximal oxygen uptake ($\text{VO}_{2\text{max}}$) = 59.4 ± 5.6 ml.kg⁻¹.min⁻¹ BM) completed five trials consisting of a $\text{VO}_{2\text{max}}$ test, a familiarisation trial of the FTP20, two experimental FTP20 tests, and a time to volitional exhaustion (T_{LIM}) at 95% FTP20. The repeatability for mean power output (MPO) during the FTP20 was excellent ($r = 0.94$, CI 0.82, 0.98, $p < 0.001$). Mean T_{LIM} (at 95% FTP20) was 42 ± 17 -min, with six participants within 10-min of the 60-min suggested threshold. These results suggest that the FTP20 is reliable, however it does not predict 60-min power with a high level of validity. Future research should explore adapting the calculation of FTP whereby the intensity may be lowered (i.e., 80-90% MPO of FTP20), particularly as most participants' T_{LIM} was far below the suggested 60-min time frame.

Keywords

FTP; Cycling Performance; Physiology; Exercise Performance; Wattbike Pro; Time to Exhaustion



This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



1 Introduction

In recent years, the adoption of cycling power meters has become more prevalent, with many athletes using them to evaluate both training and race performances (Passfield et al., 2017). It was first proposed by Allen and Coggan (2012) that power meters could serve as a practical method to determine a field-based equivalent to the maximal lactate steady-state (MLSS), thereby proposing the notion of a functional threshold power (FTP), which is defined as the maximal mean power output (MPO) that can be sustained for ~60 mins in a 'quasi-steady state'. Although its use is highly contentious, FTP has now become the most common method of exercise prescription and training monitoring purposes by athletes, coaches, and online platforms (TrainingPeaks, Intervals.ICU, Zwift etc.). When the construct of FTP was originally proposed, the testing method required athletes to complete a 60-min time trial as this was shown to predict actual competitive (40 km) performance more appropriately than a range of other predictors such as maximum oxygen uptake ($\text{VO}_{2\text{max}}$), skeletal muscle capillarization, fibre type, and mitochondrial content (Coyle et al., 1992). More recently, a condensed version of the FTP test has become popular that only requires cyclists to perform a 20-min time trial (FTP20) (Allen and Coggan, 2010). Consequently, this offers significant practical benefits to the traditional 60-min time trial (TT) due to the reduced time to complete, and reduction in post-exercise fatigue that may interfere with other training sessions (Allen and Coggan, 2010). Due to the shorter duration of FTP20, 95% of mean power output (MPO) achieved during the FTP20 is considered as the cyclist's estimation of FTP and this intensity should be sustainable for approximately 60-mins as demonstrated by Allen and Coggan, 2010. Whilst this may potentially be a credible method of estimation, there is a paucity of

literature to suggest this method is reliable and valid to predict the highest power that a cyclist can maintain for 60-mins. Cyclists and coaches therefore, run the risk of using FTP to determine and evaluate exercise training zones inappropriately, which could subsequently have negative consequences on training (i.e., session intensity) and competitive performance (i.e., pacing strategies).

Despite these potential negative consequences, there are a considerable number of methods to determine FTP and set threshold zones are available, including commercially available devices and software (Mackey and Horner, 2021). This is of concern if the appropriate scientific evaluation has either not been conducted or does not support the use of this method. Several studies have reported the reliability of the of FTP20 (Borszcz et al., 2018; McGrath et al., 2019; Lillo-Bevia et al., 2022; MacInnis et al., 2019). In studies using elite athletes, MacInnis et al. (2019) reported that the reliability of a 20-min cycling TT displayed excellent reliability ($\text{CV}\% = 1.4$; Intraclass correlation coefficient, $\text{ICC} = 0.99$, $p < 0.001$) with a technical error of measurement (TEM) of 4.6 W. In the same study, linear regression analysis also compared the relationship between FTP20 and 60-min TT MPO and found this method was suitable to predict FTP ($r = 0.92$, $p < 0.001$), suggesting the test is suitable to predict FTP. Nonetheless, this study was conducted in a small sample ($n = 8$) and within a highly trained/elite cohort of cyclists ($\text{VO}_{2\text{max}} = 65.1 - 71.1 \pm 4.7 \text{ ml.kg}^{-1}.\text{min}^{-1}$). Cyclists of this trained status are not only known to be able to produce more repeatable efforts compared to lesser trained individuals (Currell and Jeukendrup, 2008), but they are also likely to be able to maintain power output at a higher percentage of peak power output (PPO) and have higher absolute critical power (Leo et al., 2022). Conversely, the most likely group to perform such predictive tests are recreational

and lesser trained athletes, due to interactions with platforms such as Zwift™ which has more than 3 million subscribers (Reed, 2021) and offers FTP20 in order to prescribe exercise intensity domains. Whilst the FTP20 performed using online platforms such as Zwift™ have been shown to be repeatable (Matta et al., 2022), Borszcz et al. (2018) reported that cyclists of the sub-elite training status ($\text{VO}_{2\text{max}} = 59.4 \pm 5.9 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) were only able to sustain FTP20 for 50.9 ± 15.7 minutes, which suggests that the validity of the FTP20 to predict 60 min maximal power output is questionable. In addition, the authors used a fixed workload, which does not reflect the quasi-steady state nature of real-world cycling. The purpose of this study therefore was to assess the reliability of the condensed FTP20 test and to assess the construct validity of FTP20 to predict 60-min max power output.

2 Material and Methods

2.1 Participants

Twenty-two male trained cyclists completed the study (age = 32 ± 9 years, body mass (BM) = $76.7 \pm 6.3 \text{ kg}$, $\text{VO}_{2\text{max}} = 59.4 \pm 5.6 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ BM, PPO = $366 \pm 38 \text{ W}$) (McKay et al., 2022). All participants were frequent cyclists with a minimum of two sessions per week, and some were members of clubs registered with British Cycling ($n = 9/22$). All participants were aware of FTP, however, only five participants had used it previously to inform training. Each participant provided consent to take part in the study, which had institutional ethical approval (1680 /R(B) /2018 /Apr /HELS FAEC).

2.2 Study Overview

Participants visited the laboratory on five separate occasions at approximately the same time of day ($\pm 1\text{-h}$) and in a euhydrated state. In visit one, an incremental ramp test to assess maximal oxygen consumption ($\text{VO}_{2\text{max}}$) and

PPO was completed. Participants then completed a further four trials, whereby a full familiarisation trial of the 20 min TT was completed (FTP20), followed by two experimental trials of the same protocol. Finally, on a separate occasion a time to volitional exhaustion (T_{LIM}) test at 95% of MPO of the FTP20 test was completed. The FTP was determined by calculating the MPO achieved during the two FTP20 tests. Trials were separated by a minimum of 3 days and maximum of 7 days and completed within a five-week period. During this time, participants were permitted to train but were asked to keep this consistent for the duration of the study. Nutritional intake was also recorded 24-h prior to each trial, and participants were required to replicate this intake for each trial. Participants were encouraged to prepare as they normally would for an intense training ride or race. Other pre-testing dietary controls included abstinence from alcohol and any other acutely ingested ergogenic supplements or compound (i.e., caffeine, nitrates) for the 24-h prior to each trial.

2.3 Procedures

2.3.1 Incremental Ramp Test

Expired gas samples were determined using an online breath-by-breath gas analyser (Metalyzer 3B, Cortex, Germany) continuously for rate of oxygen consumption (VO_2), expired carbon dioxide (VCO_2), and respiratory exchange ratio (RER). Data were averaged over the highest rolling 30 s period of exercise to determine $\text{VO}_{2\text{max}}$. The protocol was conducted on an electromagnetically braked cycle ergometer (Excalibur Sport, Lode, Netherlands) as per a previously described method (Gough et al., 2017). Heart rate (HR) (Polar, FT1, Finland) and rating of perceived exertion (RPE) (6-20 scale; Borg, 1982) were monitored every 1-min, and blood lactate (BLa) was recorded pre- and post-exercise. All

participants achieved the set criteria of a valid maximal test as outlined by Bird and Davison (1997).

2.3.2 FTP20

All trials were performed on a cycle ergometer (Wattbike Pro, Nottingham, UK) in the participants preferred seat geometry, which was recorded and replicated for each trial. This cycle ergometer has previously been shown to be valid and reliable at the power outputs used in this study (Hopker et al., 2010). Participants were permitted to use their preferred cycling attire and pedals, and these were kept consistent for all trials. Participants were asked to treat each FTP20 as a race, and to complete as much work within the given time. They were also blinded from all available variables except for time elapsed and cadence. Prior to each trial, a self-selected warm-up was conducted, and this was replicated for each subsequent trial. This was selected due to pilot work revealing athletes ($n = 25$) would use their own warm-up prior to the FTP test, rather than the recommended one by Allen and Coggan (2010).

In the familiarisation trial, measures for RPE for overall body exertion (RPE-O), legs (RPE-L), and HR were measured at rest and every 5-min of the exercise. In experimental trials, procedures were identical with the addition of a fingertip capillary blood sample collected every 5 min for BL_a and expired gas samples were collected for 1 min every 5 min for VO_2 , VCO_2 , and RER. Consistent verbal encouragement was provided during each trial, and all trials were conducted in a temperature and humidity-controlled laboratory ($18 \pm 1^\circ\text{C}$; $40 \pm 5\%$).

2.3.3 Time to Volitional Exhaustion Test

Participants were required to cycle at 95% of MPO based on the FTP20 tests, which was determined as the intensity they should be able

to sustain for 60-mins. A self-selected cadence was used (range 70-100 $\text{r}\cdot\text{min}^{-1}$) and this was kept consistent throughout the trial until volition exhaustion (T_{LIM}). If MPO dropped 1% below FTP, the participants were given one warning and were required to increase their power back to within 1% of FTP. On the second occurrence the test was terminated. This small degree of freedom was given to participants due to the quasi-steady state nature of cycling in the real world (Borszcz, 2018), therefore a fixed power output was avoided. The participant was blinded to all available variables except for time elapsed, mean power output and cadence ($\text{rev}\cdot\text{min}^{-1}$). Participants were allowed to ingest fluids ad libitum that could contain carbohydrates, as per previous research (Macinnis et al., 2019). This was permitted as this would better reflect the nutritional strategy of a cyclist during a 60 min TT. During the T_{LIM} , respiratory measures for VO_2 , VCO_2 , RER were measured every 10 min, alongside lactate, HR, RPE-O, and RPE-L. All measurements were repeated immediately post-exercise upon the participant reaching T_{LIM} .

2.4 Statistical Analysis

Normality was checked using standard geographical methods and the Shapiro Wilk test. To assess differences in MPO between the FTP20 tests, a paired t test was used. To assess physiological or perceptual responses across the FTP20 tests (every 5 min) a two-way repeated measures ANOVA was used with Bonferroni post-hoc pairwise comparisons for HR, VO_2 , VCO_2 , RER, lactate, RPE-O, and RPE-L. Partial eta squared (η^2) effect size was reported and interpreted as small (0.01), medium (0.06) or large (0.14) (Cohen, 1988). To assess reliability, Intraclass Correlation Coefficient (ICC) were calculated using a two-way mixed effects model, with consistency and average measures used for interpretation. The

r value is reported, along with confidence intervals of the ICC estimate and significance. Interpretation was decided by the r value, with values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively (Koo and Li, 2016). Significance was accepted at $p < 0.05$. Mean and standard deviation (\pm) is reported unless otherwise stated. Statistical procedures were completed using Statistical Package for the Social Sciences (SPSS) (V28, IBM, Chicago, USA).

3 Results

Mean power for the two FTP20 tests displayed excellent reliability ($r = 0.94$, CI 0.82, 0.98, $p < 0.001$). The mean power difference between the first and second FTP20 was 5 W (271 ± 30 vs. 276 ± 26 W; $p = 0.020$; Figure 1). Blood lactate displayed excellent reliability ($r = 0.92$, CI 0.86, 0.96, $p < 0.001$), with no differences at any time point ($p = 0.516$, $\eta^2 = 0.038$). The repeatability of $\dot{V}O_2$ was good ($r = 0.86$, CI 0.69, 0.95, $p < 0.001$), and no significant differences at any time point were observed ($p = 0.812$, $\eta^2 = 0.035$). A similar level of reliability and difference was seen for $\dot{V}CO_2$ ($r = 0.89$, CI 0.76, 0.97, $p < 0.001$; $p = 0.757$, $\eta^2 = 0.045$), and RER displayed moderate reliability ($r = 0.52$, CI -0.03, 0.84, $p = 0.029$) with no significant differences at any time point ($p = 0.135$, $\eta^2 = 0.180$). Heart rate displayed excellent reliability ($r = 0.94$, CI 0.88, 0.97, $p < 0.001$), with no significant differences between the two FTP20's ($p = 0.686$, $\eta^2 = 0.016$). Perceptual responses for RPE-O displayed excellent reliability ($r = 0.941$, 0.89, 0.97, $p < 0.001$) with no differences between the FTP20 tests ($p = 0.918$, $\eta^2 = 0.008$). Similarly, RPE-L displayed excellent reliability ($r = 0.916$, 0.85, 0.96, $p < 0.001$) and no significant differences between the FTP20's ($p = 0.277$, $\eta^2 = 0.059$; Table 1).

The T_{LIM} was highly variable (CV: 41%), with a median of 41 min and a range between 13 to 74 minutes (Figure 2). The mean T_{LIM} to exhaustion was 42 ± 17 min. Time course mean changes in HR, $\dot{V}O_2$, $\dot{V}CO_2$, RER, lactate, RPE-L, and RPE-O are depicted in Table 2. Only 6/22 participants were within 10-minutes of the 60-min target, and many were at least 20-min below (9/22).

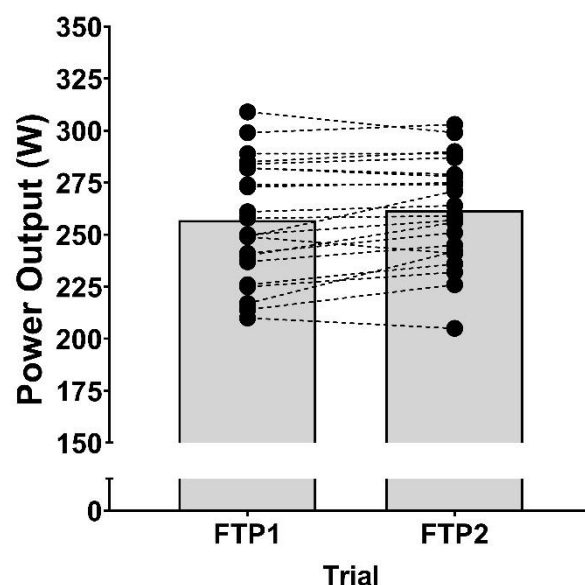


Figure 1. Power output (watts) for FTP1 and FTP2. Bars represent mean (\pm SD), and lines represent individual data.

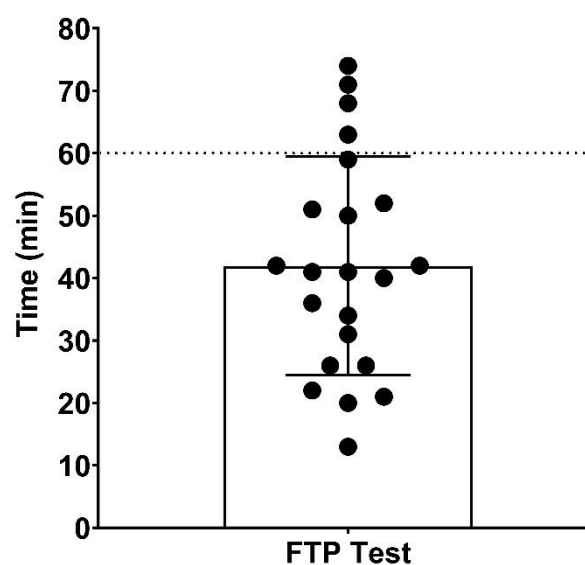


Figure 1. Time to volition tolerance (TLIM). Bar represents mean (\pm SD), and solid dots represent individual data.

Table 1. Mean (\pm SD) data for the repeated FTP20 tests.

Variable	Time Point (min)									
	Rest		5		10		15		20	
	FTP1	FTP2	FTP1	FTP2	FTP1	FTP2	FTP1	FTP2	FTP1	FTP2
VO ₂ (L·min ⁻¹)	0.49 \pm 0.12	0.49 \pm 0.12	3.63 \pm 0.65	3.59 \pm 0.66	3.85 \pm 0.54	3.78 \pm 0.46	3.96 \pm 0.55	3.82 \pm 0.52	4.25 \pm 0.86	4.06 \pm 0.50
VCO ₂ (L·min ⁻¹)	0.52 \pm 0.15	0.46 \pm 0.13	3.96 \pm 0.61	3.82 \pm 0.46	3.85 \pm 0.50	3.81 \pm 0.48	3.89 \pm 0.59	3.86 \pm 0.43	4.05 \pm 0.48	4.11 \pm 0.58
Heart rate (b·min ⁻¹)	69 \pm 2	61 \pm 7	159 \pm 12	159 \pm 11	166 \pm 11	166 \pm 9	168 \pm 10	169 \pm 9	174 \pm 9	175 \pm 7
Lactate (mmol·L ⁻¹)	1.4 \pm 0.4	1.2 \pm 0.4	10.2 \pm 3.1	10.7 \pm 4.1	12.0 \pm 3.9	11.8 \pm 3.8	13.2 \pm 4.2	12.8 \pm 4.1	14.6 \pm 3.2	14.5 \pm 2.9
RPE-O (AU)	-		15 \pm 2	15 \pm 2	16 \pm 2	16 \pm 2	17 \pm 1	18 \pm 1	19 \pm 1	19 \pm 1
RPE-L (AU)	-	-	15 \pm 2	15 \pm 2	17 \pm 1	16 \pm 1	18 \pm 1	18 \pm 1	19 \pm 1	19 \pm 1

Table 2. Mean (\pm SD) data for the time to volitional tolerance (T_{LIM}) test.

Variable	Time Point (min)								
	Rest	10	20	30	40	50	60	70	End
Participants* (n)	22	22	21	16	13	8	5	2	22
VO ₂ (L·min ⁻¹)	0.54 \pm 0.13	3.62 \pm 0.47	3.60 \pm 0.42	3.56 \pm 0.35	3.50 \pm 0.35	3.64 \pm 0.47	3.75 \pm 0.20	3.49 \pm 0.22	3.68 \pm 0.40
VCO ₂ (L·min ⁻¹)	0.57 \pm 0.19	3.58 \pm 0.46	3.48 \pm 0.30	3.36 \pm 0.25	3.39 \pm 0.21	3.52 \pm 0.26	3.47 \pm 0.19	3.22 \pm 0.22	3.48 \pm 0.40
Heart rate (b·min ⁻¹)	66 \pm 7	158 \pm 10	164 \pm 11	163 \pm 12	165 \pm 9	165 \pm 8	167 \pm 12	165 \pm 11	175 \pm 8
Lactate (mmol·L ⁻¹)	1.2 \pm 0.4	7.0 \pm 3.3	8.4 \pm 4.1	8.0 \pm 3.8	8.2 \pm 3.5	8.4 \pm 2.7	10.8 \pm 4.3	11.0 \pm 4.0	13.7 \pm 2.6
RPE-O (AU)	-	14 \pm 2	15 \pm 2	16 \pm 2	16 \pm 2	17 \pm 2	17 \pm 2	16 \pm 1	19 \pm 1
RPE-L (AU)	-	16 \pm 3	16 \pm 2	17 \pm 1	18 \pm 2	18 \pm 2	18 \pm 2	18 \pm 0	19 \pm 1

4 Discussion

This study aimed to assess the reliability and construct validity of the FTP20 to predict 60-min power in a group of recreational cyclists. This study reports that whilst FTP20 displays excellent reliability, the construct validity of this test to predict 60-min power output is poor. Based on the excellent reliability of the FTP20 test, this could be used as a method to detect adaptations from training schedules, particularly given the practical nature of the test. Many participants were far below the 60-min predicted T_{LIM} and based on no participants achieving 20-mins above this

threshold, it suggests the calculated FTP (95% of MPO from a FTP20) is too intense for recreational cyclists to sustain for 60-mins. This is likely to be a result of gross mechanical efficiency differences, which have previously been reported between different levels of trained cyclists (Sallet et al., 2006). Consequently, coaches and cyclists need to be cautious when employing this type of test to assist with creating thresholds for efforts around 60-mins. Future research should look to investigate the percentage of FTP20 which most appropriately represents a 60-min effort and assess if this further improves the construct validity of the test.

The reliability of FTP20 was excellent in the current study, which suggests it could be a useful tool for performance analysis (MacInnis et al., 2019; McGrath et al., 2019; Lillo-Bevia et al., 2022) and determining the effectiveness of training and intervention strategies. The difference between the two trials was less than 2% and only one participant reported an intra-individual CV of >5% (5.6%) in the current study. These findings corroborate with multiple studies such as MacInnis et al. (2019) who reported similar excellent reliability, albeit in a small sample size, and McGrath et al. (2019) who reported an ICC of 0.98, and a mean bias of 2 W in limits of agreement (LOA) analysis. The reliability was slightly stronger in MacInnis et al. (2019) and McGrath et al. (2019) compared to the present study, however, this is likely due to the higher training status of participants who generally can produce more repeatable results (Currell and Juekendrup, 2008). Nonetheless, the current study adds novelty in that it shows excellent reliability in a much larger sample size than MacInnis et al. (2019) ($n = 22$ vs. 8) and in recreational cyclists, as opposed to elite cyclists. The current study is therefore more generalisable to a larger group of cyclists, many of whom use FTP to influence their training through commercially available software that is more accessible than laboratory testing. It is suggested based on the study's findings that FTP20 could be used to detect changes that have occurred from training, rather than to predict any performance capability beyond 20-mins. It is also worth noting that as this test is primarily aerobic it offers little for anaerobic testing and therefore more testing protocols will be required.

The findings of the present study, however, question the construct validity of FTP20 testing in that this test can predict 60-min maximal power output in recreational cyclists (Allen and Coggan, 2010). Most of the participants in this study could not reach the 60-min threshold, suggesting that the intensity of

FTP20 predicts is too high. This agrees with findings from Borszcz et al. (2018) who reported similar T_{LIM} results to the current study (42 ± 17 min vs. 51 ± 16 min). In contrast, McGrath et al. (2019) reported that 17 out of 19 participants were able to reach the 60-min timepoint at FTP20. Reasons for this might have been due to the training status of participant or the methodological approach to the exhaustion test in each respective study. Specifically, the current study and Borszcz et al. (2018) had participants of a similar training status, whereas McGrath et al. (2019) had participants with a higher training status. This might explain why participants in the McGrath et al. (2019) study could reach the 60 min threshold as they would be more accustomed to these types of efforts. Moreover, McGrath et al. (2019) opted to stop participants once 60-mins had been reached, therefore it is unclear whether 17 of the 19 participants who completed 60-mins could have sustained FTP for a longer duration. It is clear from the current study nonetheless, that in lesser trained individuals the 95% of MPO is too intense, and therefore future research may wish to explore the use of 90% or 85% of MPO to increase the validity of this construct for recreational cyclists.

Given our participants were also “trained”, there is a chance that in more recreational cyclists the construct validity of FTP may be even weaker. The participants in the current study had a $\text{VO}_{2\text{max}}$ of $\sim 60 \text{ ml.kg}^{-1}.\text{min}^{-1}$ BM, which is far above the typical recreational level in cyclists (see McKay et al., 2022). This is likely given that previous research (Sitko et al., 2022) showed that the TTE at FTP20 was ~ 7 min less in recreational versus trained cyclists. This may be partly explained by factors such as athlete durability (Maunder et al., 2021), but it does raise concern of the suitability of FTP testing for certain groups of cyclists.

A limitation of this study is that the recommended warm-up by Allen and Coggan (2010) was not used, as it was preferred that a self-selected warm-up was conducted. This could have impacted the validity, as the warm-up of Allen and Coggan is a pre-fatigue protocol and therefore lowered the FTP20. This is a topic of discussion anecdotally amongst cyclists and coaches that following the warm-up suggested by Allen and Coggan (2010) might improve the validity of the ability to sustain FTP20 for 60-mins. However, in pilot work participants ($n = 25$) highlighted that they would still opt for their own warm-up when attempting the FTP20. Moreover, in a study using the recommended warm up, a 60-min TTE was still highly variable (Borzc et al., 2018). It was therefore appropriate to assess the validity of this test in the current study using real world cyclist's warm-up strategies. A further limitation was the lack of second T_{LIM} test, as this would have assessed whether practice/familiarisation to a 60-min TT would have improved the validity. This could have been suitable for those that were within 10-min of the 60-min threshold, as this could be, at least theoretically, variation that is eliminated through more attempts at the protocol. Conversely, this would be the case for only 6/22 participants in the current study, therefore for most this would likely not be possible.

5 Practical Applications

- Test-retest reliability of the FTP20 test is high across all participants.
- The FTP20 test could be used to track progress of cyclists training.
- The construct validity of the FTP20 test to predict 60-min power is poor with many cyclists unable to reach this time frame, therefore, the intensity is likely too high.
- Future work could explore the role of mental state and motivation factors to explain the poor construct validity.

- Coaches and cyclists should be cautious of using the FTP20 test to predict 60-min power and consider other approaches.

6 Conclusions

The current study reports that the construct validity of the FTP20 test to predict 60-min power is poor, and coaches and male cyclists should exercise caution when using such forms of testing to develop exercise prescription for certain cycling populations. A positive finding of the study is that a 20-min TT test is reliable so this test could be used to assess training adaptations. It could also be used for training prescription purposes and monitoring of cyclists (Borzc et al., 2018). Future research should attempt to assess whether a lower threshold (i.e., FTP20 – 10% or 15% MPO) leads to an improved ability for the FTP20 to predict 60-min TT performance more suitably in recreational cyclists.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the participants for their efforts and volunteering to complete the study. In addition, the technical staff for their assistance in supporting data collection over the course of the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Allen, H., and Coggan, A. (2010) *Training and Racing with a Power Meter*, 2nd ed.; Velo Press: Boulder, CO, USA.
2. Bird, S. R., & Davison, R. (1997). BASES physiological testing guidelines.
3. 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*, 39(10), 737-742.

4. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates.
5. Coyle, E. F., Feltner, M. E., & Kautz, S. A. (1992). Physiological and Biomechanical Factors Associated with Elite Endurance Cycling Performance. *Medicine and Science in Sports and Exercise*, 23(1):93-107.
6. Currell, K., and Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine*, 38(4), 297-316.
7. Gough, L. A., Deb, S. K., Sparks, A. S., & McNaughton, L. R. (2017). The reproducibility of blood acid base responses in male collegiate athletes following individualised doses of sodium bicarbonate: a randomised controlled crossover study. *Sports Medicine*, 47(10), 2117-2127.
8. Hopker, J., Myers, S., Jobson, S. A., Bruce, W., & Passfield, L. (2010). Validity and reliability of the Wattbike cycle ergometer. *International Journal of Sports Medicine*, 31(10), 731-736.
9. Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
10. Leo, P., Simon, D., Hovorka, M., Lawley, J., Mujika, I. (2022). Elite versus non-elite cyclist – Stepping up to the international/elite ranks from U23 cycling. *Journal of Sports Sciences*, 40(16), 1874-1884.
11. Lillo-Beviá, J. R., Courel-Ibáñez, J., Cerezuela-Espejo, V., Morán-Navarro, R., Martínez-Cava, A., & Pallarés, J. G. (2022). Is the functional threshold power a valid metric to estimate the maximal lactate steady state in cyclists? *Journal of Strength and Conditioning Research*, 36(1), 167-173.
12. Mackey, J and Horner, K. (2021). What is known about the FTP20 test related to cycling? A scoping review. *Journal of Sports Sciences*, 39(23), 2735-2745.
13. Matta, G., Edwards, A., Roelands, B., Hettinga, F., Hurst, P. (2022). Reproducibility of 20-min time trial performance on a virtual cycling platform. *International Journal of Sports Medicine*, E-pub ahead of print, 2022 Aug 11, doi: 10.1055/a-1848-8478.
14. Maunder, E., Seiler, S., Mildenhall, M. J., Kilding, A. E., & Plews, D. J. (2021). The importance of ‘durability’ in the physiological profiling of endurance athletes. *Sports medicine*, 51, 1619-1628.
15. McGrath, E., Mahony, N., Fleming, N., & Donne, B. (2019). Is the FTP test a reliable, reproducible and functional assessment tool in highly-trained athletes? *International Journal of Exercise Science*, 12(4), 1334.
16. MacInnis, M. J., Thomas, A. C., & Phillips, S. M. (2019). The reliability of 4-min and 20-min time trials and their relationships to functional threshold power in trained cyclists. *International journal of sports physiology and performance*, 14(1), 38-45.
17. Passfield, L., Hopker, J. G., Jobson, S., Friel, D., & Zabala, M. (2017). Knowledge is power: Issues of measuring training and performance in cycling. *Journal of sports sciences*, 35(14), 1426-1434.
18. Reer, R. (2021). Do you even Zwift? The indoor cycling platform is having a moment. *Forbes* (Feb 17, 2021). Available at <https://www.forbes.com/sites/robreed/2021/02/17/do-you-even-zwift-the-indoor-cycling-platform-is-having-a-moment/?sh=644014073f86>. Accessed 02/012/2022.

19. Sallet P, Mathieu R, Fenech G, Baverel G. (2006). Physiological differences of elite and professional road cyclists related to competition level and rider specialization. *Journal of Sports Medicine and Physical Fitness*, 46(3):361-5.
20. Sitko, S., Cirer-Sastre, R., & López-Laval, I. (2022). Time to exhaustion at estimated functional threshold power in road cyclists of different performance levels. *Journal of Science and Medicine in Sport*, 25(9), 783-786.