International Journal of EXERCISE SCIENCE

Original Research

Prediction of Functional Threshold Power from Graded Exercise Test Data in Highly-Trained Individuals

EANNA McGRATH[†1], NICK MAHONY[‡1], NEIL FLEMING[‡1], ALESSIO BENAVOLI[‡2], and BERNARD DONNE[‡1]

[1]Human Performance Laboratory, Disciplines of Anatomy and Physiology, School of Medicine, Trinity College Dublin, IRL; [2]School of Computer Science and Statistics, Trinity College Dublin, IRL

[†]Denotes graduate student author, [‡]Denotes professional author

ABSTRACT

International Journal of Exercise Science 15(4): 747-759, 2022. The purpose of the current investigation was to derive an equation that could predict Functional Threshold Power (FTP) from Graded Exercise Test (GxT) data. The FTP test has been demonstrated to represent the highest cycling power output that can be maintained in a quasi-steady state for 60-min. Previous investigations to determine a comparable marker derived from a Graded Exercise test have had limited success to date. Consequently, the current study aimed to predict FTP from GxT data to provide an additional index of cycling performance. FTP has been reported to provide an insight not provided by a GxT and, in addition, does not require a formal exercise testing facility. The study design facilitated a deliberate and transparent sequence of statistical decisions, resolved in part from the perspective of exercise physiology. Seventy triathletes (male n=50, female n=20) completed cycling GxT and FTP tests in sequential order. Collected data (power output, blood lactate indices, VO_2peak, body mass) were analysed using stepwise regression to identify the key parameters for predicting FTP, and confirmed using a Leave One Out (LOO) cross-validation. As a consequence of wittingly including some likely transiently highly correlated parameters on the basis of a physiological argument, the model's function is limited to predicting FTP. This investigation concluded the model (FTP = -6.62 + 0.32 FBLC-4 + 0.42 BM + 0.46 Pmax) was the prediction model of choice.

KEY WORDS: Triathlon, lactate profile, modelling, stepwise regression, cross-validation

INTRODUCTION

The Graded Exercise Test (GxT) has been used both to assess clinical issues relating to health, and evaluate exercise performance (4). The measured responses to each step-increase in exercise intensity typically include; heart rate, fuel utilization, oxygen cost, and during the final stage measurement of peak oxygen consumption, and frequently in an exercise setting an athlete's blood lactate profile. A key characteristic of the GxT is that it appraises contributions from multiple individual physiological indices concurrently at each intensification. Moreover, as the derived step-data is interpolated using all of the performance markers listed above, exact cycling intensities can be identified between the predefined GxT step-increases. Commonly used indices of physical conditioning derived from a GxT such as lactate threshold (T_{Lac}), load associated with 2 mmol.L^{-1} blood lactate concentration (FBLC-2), load associated with 4 mmol.L^{-1} blood lactate concentration (FBLC-4), workload preceding a fixed rise of 1 mmol.L^{-1} in blood lactate concentration (FRBL) and the maximum distance from a curve representing ventilatory and metabolic variables (Dmax), reflect specific "micro-occurrences", which infer the capacity of a subset of actual performance. Conversely, the FTP test is a 20-min maximum effort, reflecting the sum total of whole-body energetics. The FTP test format can be viewed in Table 1. The computation of FTP is calculated by simply reducing the average power sustained over a 20-min time-trial period by 5 %. Allen and Coggan (1) suggested that the minor reduction of the 20-min power equated to an output that could be sustained for 60-min. This proposition has been demonstrated both valid and reliable (21), without these measurement qualities, a prediction equation would not be warranted. The basis of the small (5%) reduction in power between the 20- and 60-min cycling intervals is conjectured to be a consequence of the respective time periods being positioned close to the lowest point of the hyperbolic curve and therefore the 40-min differential (20-min test versus 60-min limits of tolerance) being associated with only a minor (5%) reduction in load (22).

Table 1. Test protocol for assessing FTP (1).

	Duration	Description	% FTP
Warm-up	20-min	Endurance pace	65
	3 by 1-min with 1-min recoveries	Fast pedalling 100 rev·min^{-1}	Not applicable
	5-min	Easy riding	65
	5-min	Time-trial	Maximum effort
	10-min	Easy riding	65
FTP test	20-min	Time-trial	Maximum effort
Cooldown	10 to 15-min	Easy riding	65

(FTP) – Functional Threshold Power, (min) – minute, (rev·min^{-1}) – revolutions per minute

The goal of the current investigation was to derive an empirical equation to predict FTP from the gold standard GxT. In some respects, the strengths and weaknesses of the GxT and FTP tests appear to compliment one-another. The FTP test requires a power meter but does not require the more elaborate equipment commonly associated with a GxT; namely, a metabolic cart and lactate analyzer. The FTP test does not provide any physiological data, rather power output alone. The proponents of the FTP test highlight the advantage of using power output to pace time-trial efforts as this constant analogue is unaffected by time (1). Conversely, the multiple physiological variables derived from a GxT can be influenced by a multitude of variables, for example if heart rate is used for pacing, the associated power is likely to reduce over time (9). Given that the two tests (GxT and FTP) in their current guise provide mutually exclusive information; there is a rationale for predicting FTP from GxT derived data. The purpose of the current study was to identify an effective prediction model for FTP from GxT data from a combination of both robust statistical analyses and using established physiological principles. Published articles demonstrate variations in the GxT increments and durations can affect the gleaned data (4, 17). These differences may have benefited a particular circumstance but limit inter-investigation comparisons (4). The derivation of a model from the current research would be limited to the same test protocols, ancillary calculations and highly-trained populations. The hypothesis of the current investigation was that a prediction model could be trained to predict FTP for highly-trained athletes using GxT data.

METHODS

Participants

The current study obtained ethical approval from the Faculty of Health Sciences Research Ethics Committee in Trinity College Dublin and was performed in accordance with the ethics standards of the International Journal of Exercise Science (24). An *a priori* linear multiple regression power test was conducted with a Type 1 error probability of 0.05, a power of 0.85 and a projected effect size 0.1. This analysis indicated that n = 74 would provide a statistical power of 85% for 2, 3 and 4 parameter models (G*Power v3.0.10 free software; Institute of Experimental Psychology, Heinrich Heine University, Dusseldorf, Germany). Inclusion criteria were that participants were; aged 18-35 years; healthy and injury free as assessed by medical questionnaire and medical assessment; competing in triathlon or cycling for a minimum of 2-years. Exclusion criteria included the following: outside the age range of 18-35 years old; high blood pressure or found to have high blood pressure during pre-screening assessment; a bleeding or clotting disorder; any previous history of cardiopulmonary disease; respiratory difficulties (based on spirometry data) or symptoms of colds/influenza on the day of testing; acute or chronic musculoskeletal injury limiting exercise capacity; disease that would prevent participation in an maximal exercise test; deemed unfit to participate on completion of a medical questionnaire and medical examination due to an on-going illness, or having any of the following; diabetes, hypertension, heart defects, metabolic disorders or other contraindications to maximal exercise testing.

Protocol

Participants VO_{2peak}, mass, height, BMI and age can be viewed in Table 2. Participants completed an informed consent form prior to beginning any trials. All enlisted participants attended the laboratory on two occasions, in a rested, carbohydrate loaded state to control for dietary induced elevations or reductions in BLa data (19) and consequently maintain the power versus BLa relationship (19). Athletes were requested to arrive hydrated, having abstained from alcohol and caffeine in the 24-h prior to testing. Hydration status was assessed as urine specific gravity (USG) using a mid-stream urine sample and an optical refractometry (Eclipse Professional, Bellingham & Stanley, Kent, UK). A 24-h food diary completed prior to the first trial identified that enlisted participants were consuming a training load adjusted isocaloric diet (macronutrient breakdown; ≥ 60% carbohydrate, ≤ 20% fat and ≤ 20% protein). Each participant was requested to replicate their food intake prior to both tests, or if different, to consume comparable carbohydrate quantities. When necessary, participants were assisted in planning their pre-test meals. The two trials were performed within a two-week period, at least 7-days apart. Training loads were agreed with both athletes and coaches prior to commencing the current study, with weekly training load remaining as constant as possible preceding both tests. Exercise was limited to aerobic work for 48-h prior to each test.

Table 2. Mean (± SD) VO_{2peak}, mass, height, BMI and age data for participants.

	VO_{2peak} (mL·kg⁻¹·min⁻¹)	Mass (kg)	Height (m)	BMI (kg·m⁻²)	Age (yr)
Female (*n* = 20)	58.9 ± 5.1	57.8 ± 7.6	1.69 ± 0.07	20.1 ± 1.9	29.1 ± 5.0
Male (*n* = 50)	62.7 ± 8.8	77.0 ± 10.0	1.79 ± 0.06	23.8 ± 2.5	27.4 ± 5.7

(SD) – standard deviation, (BMI) – Body mass index, (VO_{2peak}) – Peak maximum oxygen uptake, ($mL \cdot kg^{-1} \cdot min^{-1}$) - millilitres of Oxygen consumed per kilogram body mass per minute, (kg) – kilogram, ($kg \cdot m^{-2}$) – kilogram divided by metre squared, (yr) – year.

The FTP test was performed on the athlete's own bicycle using Garmin pedals (Garmin, KA, USA) to measure cycling power output, with the bicycle mounted on an indoor trainer (LeMond Revolution, WA, USA). The Lode Excalibur Sport ergometer facilitates self-paced time-trial efforts in "Linear Mode". However, this requires the practitioner to predetermine a specific power and associated cadence prior to testing. We considered the risk of bias to supersede the benefits of performing both the GxT and FTP on one ergometer, particularly given that the Garmin power pedals could be used on both ergometers and cross-referenced. Garmin pedals were calibrated prior to each trial as per the manufacturer's instructions to zero offset. Having completed the FTP test, the Garmin pedals were subsequently placed on the Lode cycle ergometer and calibrated at their ascertained FTP to mitigate against any differences in the respective devices (22). The corrected FTP (namely, that corrected to the Lode ergometer) was subsequently used for all proceeding analyses throughout the current investigation.

All participants had completed an FTP test prior to enlistment into the current study. The 20-min FTP test protocol can be viewed in Table 1. The order of the two cycle tests was not randomized as the GxT data were used to identify the appropriate warm-up intensity for the FTP test. The warm- up intensity was set at 65% of the alternate threshold index "Dmax" (derived from the GxT data) in keeping with previous research (21) and utilising the recent fitness test (< 2-weeks between tests). The line of best fit for the Dmax computation was determined using a third order curvilinear regression using VO_2 and BLa data at each workload during the GxT test. Thereafter, the maximum perpendicular distance to the straight line between the lowest and highest exercise BLa data identified load at Dmax (8). The instruction given to participants for the 20-min FTP time-trial was "a strong, steady effort for the entire 20 min. Do not start out too hard! Get up to speed (power) and then try to hold that speed (power). Your goal is to produce the highest average wattage over the entire period" (1). Subsequently, FTP was determined by reducing the mean power output across the 20-min time-trial by 5% (1).

Statistical Analysis
A list of potential FTP prediction variables from GxT data was compiled. The workloads (Watt) at the following indices were included; load at T_{Lac}, Dmax, FBLC-2, FBLC-4 and Pmax. Additionally, body mass (BM in kg) and absolute VO_2peak ($mL.min^{-1}$) were included in the initial data gathering phase, as these data were deemed relevant to both the GxT and cycling performance. The relationship between each of these potential independent variables (IV) versus FTP was first checked using scatter plots to visualize each relationship. Correlation coefficients (r) and the corresponding coefficients of determination (r^2) were calculated using Prism 9 (Graph Pad, CA, USA). Two iterations of all of these initial plots and calculations were prepared. The first used absolute data and the second used data scaled to body mass. Computed correlation coefficients of ≥ 0.84, and accompanying coefficients of determination ≥ 0.70 were determined as a minimum inclusion requirement (6). The interpretation of r^2 was also considered from the context of the specific field of application (14). In elite sport, meaningful improvements are relatively small (18), and, therefore any prospective model would need to be sensitive to small biological changes.

Each of the independent variable (IV) were then correlated with one another, the rationale here was to avoid any potential distortion of the line of best fit that could not be explained because two or more parameters were measuring the same quantity within one equation. The researchers were alert specifically to the risk of collinearity between; T_{Lac}, Dmax, FBLC-2, FBLC-4 and FRBL versus Pmax or VO_2peak. For physiological reasons explained in the discussion below, the combinations of either T_{Lac}, Dmax, FBLC-2, FBLC-4 versus Pmax or VO_2peak was permitted. This caveat was not afforded to statistical evidence of collinearity between T_{Lac}, Dmax, FBLC-2 and FBLC-4. In respect to collinearity, the following responses were considered indicators; high variation inflation factor (VIF), a sizeable drop-off between r^2 and r^2 adjusted (r^2adj) to the number of parameters included, and an increase in the p-value to > 0.05.

A stepwise regression, using an entry and exit α of $p < 0.05$ and 0.1, respectively, was applied to all of the non-redundant IV correlates versus FTP using JMP 16 (SAS Institute, NC, USA). Every permutation of the non-redundant IV correlates that passed this initial cull was further assessed as a potential parameter of a single or multi-parameter predictive equation. The following indicators were used to evaluate each equation; namely, relationship of the β-coefficient to FTP; VIF; r^2; r^2adj; the root mean square of the error with the number of parameters inserted into the equation ($s_{y.x}$); Akaike Information Criterion (AIC); and an estimation of the prediction error using a Leave One Out (LOO) cross-validation technique. The iterative process of stepwise regression facilitated the combined interpretation of statistical results with physiological tenets. The objective of this phase of the analysis was solely to identify the apparently most suitable parameters for estimating FTP. LOO cross-validation was included in the model selection criteria to validate the ability of the model to predict to unseen data (athlete). The mean-squared-error (MSE) was used as the evaluation criterion. In LOO cross-validation, for each observation (athlete) in the dataset, say the i^{th} observation, the same model is fitted keeping aside the i^{th} observation and using the remaining observations (athletes) to train the model. The MSE is then calculated from the model prediction for the i^{th} observation. Finally the average of the individual

MSE is calculated, which corresponds to the LOO cross-validation metric. For linear regression, we do not need to refit the model N-times, where N is the number of observations (16). The predictive capacity of gender on the predictive model for FTP was assessed using comparisons of the AIC and LOO cross-validation technique.

RESULTS

As only 70 participants met the strict inclusion criteria a subsequent *post-hoc* power analysis indicated that the current study achieved an overall statistical power of 83.5%. The calculated mean and standard deviation of power output data at T_{Lac}, Dmax, FBLC-2, FBLC-4, FRBL and Pmax can be viewed in Table 3. The scatter plots representing FTP versus each IV are presented in Figure 1. The correlation matrix of all IV that were used to give insight as to potential collinearity is documented in Table 4. The results of the initial regression analyses of FTP versus each IV are documented in Table 5, the four strongest correlates of FTP with the smallest $s_{y.x}$ and most favourable 95%CI were; Dmax, Pmax, FRBL and FBLC-4. These indices remained topmost when rescaled to body mass: Dmax ($s_{y.x}$ = 0.28 W.kg^{-1}, r = 0.87, 95%CI of 0.78 to 0.93); Pmax ($s_{y.x}$ = 0.26 W.kg^{-1}, r = 0.89, 95%CI of 0.81 to 0.94); FRBL ($s_{y.x}$ = 0.32 W.kg^{-1}, r = 0.74, 95%CI of 0.76 to 1.03) and FBLC-4 ($s_{y.x}$ = 0.33 W.kg^{-1}, r = 0.88, 95%CI of 0.9 to 1.2) albeit slightly inferior to their un-scaled equivalents. The results for the same analyses versus FTP for T_{Lac}, FBLC-2, VO_2peak and BM are also presented in Table 5. Four prospective model parameter options were extricated using stepwise regression. The formulae and associated $s_{y.x}$, r^2, r^2adj, VIF and AIC are presented in Table 6. The reported VIF and AIC for Model 4, see Table 6, without and with gender as a parameter were AIC 399 versus 596 and VIF 288 versus 296, respectively.

Table 3. Mean power (in W and W·kg^{-1}) associated with load at FTP, T_{Lac}, Dmax, FBLC-2, FBLC-4, FRBL and Pmax.

	FTP	T_{Lac}	Dmax	FBLC-2	FBLC-4	FRBL	Pmax
Male							
Mean power (W)	298 ± 34	297 ± 41	277 ± 32	265 ± 39	314 ± 35	274 ± 32	371 ± 40
Female							
Mean power (W)	215 ± 22	222 ± 34	207 ± 27	200 ± 33	232 ± 29	205 ± 29	267 ± 30
Male							
Mean power (W·kg^{-1})	4.0 ± 0.6	4.0 ± 0.8	3.7 ± 0.7	3.5 ± 0.7	4.2 ± 0.7	3.6 ± 0.6	4.9 ± 0.8
Female							
Mean power (W·kg^{-1})	3.8 ± 0.4	3.9 ± 0.6	3.7 ± 0.5	3.6 ± 0.7	4.1 ± 0.7	3.6 ± 0.7	4.7 ± 0.7

(W) - Watt, (W·kg^{-1}) – Watt per kilogram of body mass, (FTP) – Functional Threshold Power, (T_{Lac})- Lactate threshold, (Dmax) – Load at maximum displacement, (FBLC-2) - load associated with 2 mmol.L^{-1} blood lactate concentration, (FBLC-4) - load associated with 4 mmol.L^{-1} blood lactate concentration, (FRBL) - workload preceding a fixed rise of 1 mmol.L^{-1} in blood lactate concentration, (Pmax) - maximum workload completed on the final stage of the GxT, (GxT) – graded incremental test.

Table 4. Correlation matrix of IV.

	T_{Lac} (W)	Dmax (W)	FBLC-2 (W)	FBLC-4 (W)	FRBL (W)	VO_2peak (mL.min.$^{-1}$)	Pmax (W)	BM (kg)
T_{Lac} (W)	-	0.97	0.78	0.84	0.85	0.67	0.90	0.50
Dmax (W)		-	0.81	0.89	0.90	0.74	0.94	0.56
FBLC-2 (W)			-	0.94	0.94	0.64	0.79	0.40
FBLC-4 (W)				-	0.97	0.75	0.90	0.55
FRBL (W)					-	0.71	0.89	0.54
VO_2peak (mL.min.$^{-1}$)						-	0.76	0.75
Pmax (W)							-	0.60
BM (kg)								-

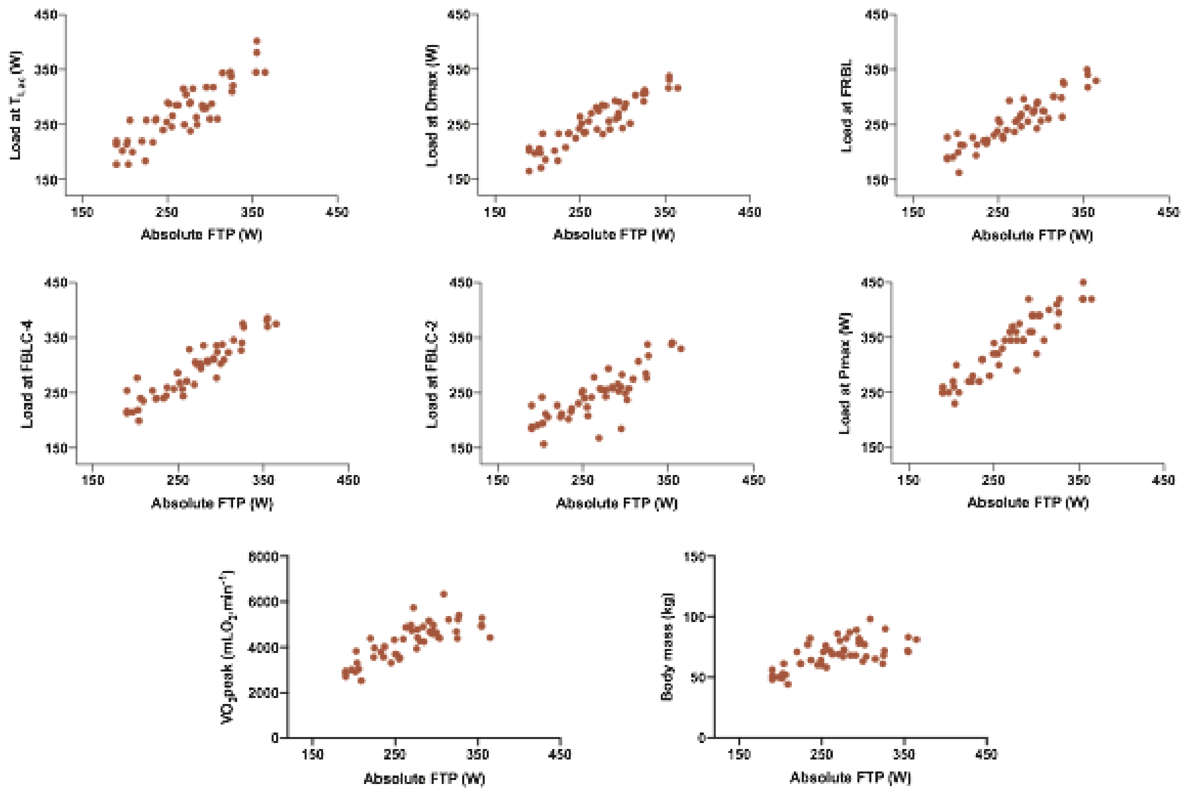

Figure 1: Scatter plots of FTP versus each individual IV.

Table 5: Regression analyses of FTP versus each individual IV.

	R	r^2	$s_{y.x}$ (W)	95%CI of r	*p*
FTP vs. T_{Lac} (W)	0.87	0.76	25	0.78 to 0.92	<0.0001
FTP vs. Dmax (W)	0.91	0.83	18	0.85 to 0.95	<0.0001
FTP vs.FBLC-2 (W)	0.83	0.68	26	0.71 to 0.90	<0.0001
FTP vs. FBLC-4 (W)	0.93	0.86	19	0.83 to 0.94	<0.0001
FTP vs. FRBL (W)	0.90	0.81	19	0.83 to 0.94	<0.0001
FTP vs. Pmax (W)	0.92	0.85	22	0.87 to 0.96	<0.0001
FTP vs. VO_2peak (mL.min.$^{-1}$)	0.78	0.61	30	0.64 to 0.87	<0.0001
FTP vs. BM (kg)	0.63	0.40	37	0.43 to 0.77	<0.0001

Table 6: Prospective parameters derived from a stepwise multiple regression analyses of non-redundant IV

	FTP prediction equation	$s_{y.x}$ (W)	r^2	r^2adj	VIF	AIC	LOO	*p*
Model-1	13.4 + 0.64 Pmax + 0.16 Dmax	18	0.87	0.87	8.6 8.6	410	345	0.0004 0.0300
Model-2	16.7 + 0.75 Pmax	18	0.86	0.86	-	409	501	0.0001
Model-3	21.6 + 0.98 Dmax	22	0.80	0.79	-	438	334	0.0001
Model-4	-6.6 + 0.32 FBLC-4 + 0.42 BM + 0.46 Pmax	15	0.89	0.89	-	399	288	0.0001

DISCUSSION

This investigation concluded that Model-4 (-6.6 + 0.32 FBLC-4 + 0.42 BM + 0.46 Pmax) was the prediction model of choice. This assertion was borne from multiple statistical decisions coupled with actualities of exercise physiology. As might be expected, the study design commenced by identifying potential correlates to be used to predict FTP, whilst remaining alert to the potential of collinearity in the instance that more than one prediction variable could be included in a final equation. As mentioned in the methods section, scope for potential collinearity was afforded to the combinations of T_{Lac}, Dmax, FBLC-2 or FBLC-4 versus Pmax or VO_2peak. This exemption was on the basis that the power output associated with these four indices changes with aerobic fitness without an obligatory concomitant change in Pmax or VO_2peak (26). Figure 2 illustrates this scenario whereby a lower power output at FBLC-4 may be observed when an athlete is in a deconditioned versus well-conditioned state, pivotally all the while Pmax conceivably remaining constant (26). In this scenario, if for example Pmax alone were used to prescribe training, the athlete would be required to train at the same intensity whether they were well-conditioned or deconditioned. Alternatively, if FBLC-4 (or T_{Lac}, Dmax, FBLC-2) were used in conjunction with Pmax (or VO_2peak), the training load would be proportionately lower for the deconditioned athlete. As the study population were highly-trained and all in competition-phase at the time of testing; T_{Lac}, Dmax, FBLC-2 and FBLC-4 were likely to equate to a similarly high fraction of Pmax or VO_2peak (21). Without variation in this fraction, these four prospective IV will exhibit a statistically linear relationship with Pmax. Importantly however, these indices still provide unique insight not afforded by Pmax alone. Similarly, it was anticipated that gender would likely enhance the predictive model given findings in the literature that female athletes have lower relative VO_2max data but higher thresholds relative to their VO_2max (27). However, the addition of gender did not significantly improve the error associated with future predictions. Notably, the number of females was limited as compared to the male group (n=20 versus 50, respectively), this may have had an impact on our results.

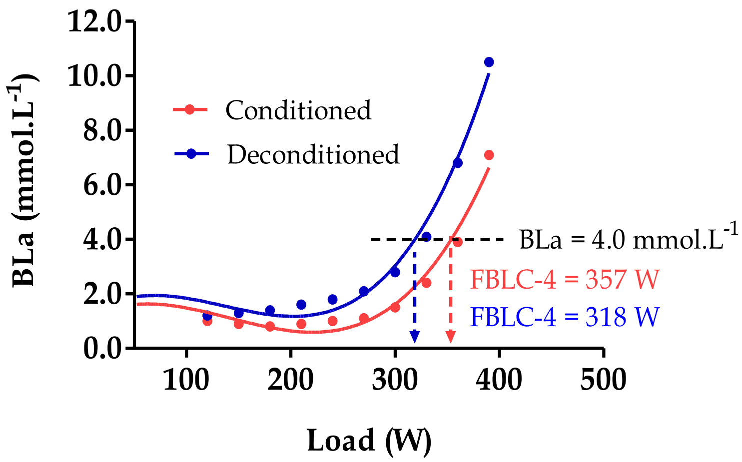

Figure 2. Illustrative data recorded in our laboratory depicting a change in power output at FBLC-4 whilst maintaining the same absolute Pmax load.

The study design used stepwise regression to identify the best predictor variables for FTP. The findings derived from each regression analysis were considered from the purview of physiology prior to being accepted or rejected as being the best prediction equation. A purely statistical comparison of Model-1 versus Model-2, would likely favor the latter equation as a consequence of the apparently similar predicative capacity and parsimony, see Table 6. However, the two parameters of Model-1 engaged two principle demarcations of physical conditioning; namely, aerobic fitness by using load at Dmax (8) and aerobic capacity *vis-à-vis* Pmax (5). Conversely, the single parameter Model-2 is limited to peak power, not necessarily an indicator of training status and not convergent with the quintessence of a GxT.

The single-parameter Model-3 was expected to reflect trained state in the guise of Dmax, a kernel marker for current training status associated with a GxT (8). However, the statistical findings herein are less favourable as the $s_{y.x}$ and AIC are higher than the other prospective models, see Table 5. Although the bivariate Model-1 contains succinct measures of training status, the results of the stepwise regression (Table 5) favor the alternative Model-4 parameters; FBLC-4, BM and Pmax. This Model-4 yielded the lowest $s_{y.x}$ and AIC (Table 5). Model-4 demonstrated the lowest LOO, indicating the best equation for predicting the performance of an athlete not included within a data set.

From a statistical perspective, caution should be taken when analysing these findings. Firstly, the β-coefficients cannot be used for explanatory purposes, as a consequence of some collinearity, although not enough to diminish the principles of the regression equation. Notably, the three explanatory parameters in Model-4 were still found to be significant, see Table 5, irrespective of the likely inflated *p*-value associated with co-variance. The IV of Model-4 can be partitioned to illustrate the redundant explanatory function of Model-4. If FBLC-4 and the Y-intercept are held constant, FTP will only increase by 0.46 W (the slope coefficient for Pmax) for every 1 W increase in Pmax. This is proportionately at odds with the relative intensity of power output at FTP versus Pmax. This can be seen in Model-2 where FTP equates to more than 75% of Pmax (the sample slope coefficient of 0.75 plus the constant of 17 W) and has been demonstrated elsewhere to occur at approximately 80% of the peak power reached during the same GxT protocol in a similarly trained athletic cohort (21). That stated, the main purpose of this regression model was to predict FTP rather than to explain the relationships of the already well-established and reliable equation parameters.

There were some decisions that were taken from the perspective of exercise science that are worth highlighting. Firstly, the GxT derived measurements of T_{Lac}, Dmax, FRBL-2 and FRBL-4 each serve similar functions physiologically; namely, to track lactate kinetics. Without any unique predictive capacity, only one of these measures was anticipated to be included in any single equation. This point did not preclude the inclusion of all four indices prior to data reduction as the purpose was to commence the regression analyses with the strongest predictors of FTP, rather than cherry-picking any particular marker. The merits of each of these four measurements (T_{Lac}, Dmax, FRPB-2 and FRBL-4) has polarised opinions (8, 15), and we acknowledged that some researchers may have a preference for one particular BLa calculation, hence the initial inclusion of each of them. Secondly, cycling data is frequently normalized to body mass as a way of making intra- and inter-individual comparisons. FTP is reported by its originators in $W.kg^{-1}$ facilitating the tabulation of categorizations of cyclist's performance capacity (1). Herein, normalized data were also assessed with the view that a relationship between variables may have existed that may not have been evident when expressed in absolute terms. However, this conjecture was not reflected in our current findings. Scaling data to body mass did not appear to have any additional predictive capacity for FTP in this cohort of well-trained triathletes. This may have been impacted by the participants having similarly low BMI data and might differ if cyclists / triathletes with a wider range of BMI data were evaluated.

Previous investigations have sought to associate GxT derived indices with FTP, apparently unsuccessfully (20), and, therefore, supporting the notion that a prediction equation is necessary. One previous investigation, by Denham *et al.* (10), generated two relevant prediction equations, one to predict FTP from GxT data and the second to predict VO_2max from FTP (scaled to BM) and age data. The models were trained using twenty-one inactive non-cyclists and nineteen self-reported recreational cyclists collectively. The age profile ranged from 19 to 55 years and included just three female participants. The model to predict FTP from GxT data was (FTP = -56.5 + 0.86 Pmax). Their computed Y-intercept will likely have a sizeable proportional effect on the data given that the reported mean FTP in their investigation was 200 ± 58.2 W (2.62 ± 0.75 $W.kg^{-1}$). This relationship might be explained in some way by the untrained group having their FTP occurring at a very low percentage of their Pmax, however, this is difficult to generalize to trained athletes. The equation stipulates that FTP has a set position in excess of 57 W lower than Pmax (the constant – 56.5 W plus the 0.86 W coefficient of Pmax using the lowest possible Pmax value of 1 W).

In respect to predicting VO_2max from FTP, Denham *et al.* (10) suggested that, although the bivariate model (using FTP and age) to predict VO_2max was trained on a combination of recreational cyclists and sedentary individuals, the model appeared to "provide robust estimates of VO_2max even for those at the upper end of the fitness spectrum". This claim was based on Denham *et al.* (10) applying their predicative equation to a single elite athlete from another researchers findings (3). Therein, Denham *et al.* (10) used the reported power at FBLC-4 as a proxy for the individual athlete's FTP (as FTP was not actually reported). The premise of this proxy calculation was that an alternate investigation by Gavin *et al.* (13) had reported that FTP and FBLC-4 were interchangeable. Again, confoundingly, the FTP reported in the Gavin *et al.* (13) study was computed using an uncontrolled 8-min field test for FTP and the FBLC-4 data were derived from altogether different GxT protocols. Specifically, Gavin *et al*. (13) commenced their test at 150 W and increased power output at a rate of 25 W every 3-min, whereas Denham *et al*. (10) commenced at 100-W and increased power output at a rate of 20 W every minute. Given that modelling already has inherent error, consistent discrepancies can only compound model inadequacies. In our investigation one particular female triathlete (swimming, cycling running) competed in the recent Olympics in Tokyo and this investigation in the same year. This provides a useful comparison with Denham *et al*. (10) proposition of using a single elite-athlete case study to test an algorithm. This female triathlete had a body mass of 45.4 kg, a measured Pmax of 270 W (5.9 $W.kg^{-1}$) and a FTP of 229 W (5 $W.kg^{-1}$), Model 4 predicted FTP. This paradigm cannot be accommodated in the proposed Denham *et al.* (10) model for FTP as the difference between Pmax and FTP is less than 56.5 W and of course the magnitude of the delta value will only increase as the coefficient of Pmax in their predictive model is 0.865. This scenario is usual where highly-trained endurance athletes have FTP data that occur at high FTP fractions of VO_2peak (21).

The approach taken herein is unusual insofar as each statistical and physiological step is described and each decision explained. A wide variety of GxT indices were included so as to create a model that was not biased to any particular GxT metric, a contentious topic ever-present in exercise science literature (8, 15). Stepwise regression afforded the combination of science and statistics. The heuristic LOO cross-validation approach permitted better usage of the limited number of high-performance athletes available, a population sample that can be more difficult to recruit for scientific research studies.

There is nothing startling in the statement that models are imperfect (14, 23) and that physiological tests of physical fitness have limitations (8). The approach of the current investigation was in the words of Anscombe (2) "weighing of evidence in the light of circumstances, available knowledge and theory". To quote Anscombe (2) a second time, "The word 'valid' should be better dropped from the statistical vocabulary. The only real validation of a statistical analysis, or of any scientific enquiry, is confirmation by independent observations." The development of this model would likely benefit from an increased number of female athletes to ensure the current analysis is accurate and gender does not enhance a model's predictive capacity. The application of the predictive model to an alternate sport such as rowing, which uses power as an analogue and does not require the athlete to carry their entire weight (25) may prove beneficial to rowers and would provide a measure of external validity of the FTP model.

ACKNOWLEDGMENTS

The authors disclose no conflicts of interest or financial arrangements related to the current research. We would like to thank all enlisted triathletes for their gracious participation in the current research.

REFERENCES

1. Allen H, Coggan A. Training and racing with a power meter. VeloPress: CO, USA; 2010.

2. Anscombe F. Topics in the investigation of linear relations fitted by method of least squares. J R Stat Soc 29(1): 1-29, 1967.

3. Bell P, Furber M, van Someren K, Swart J. The physiological profile of a multiple Tour de France winning cyclist. Med Sci Sport Exercise 49(1): 115-123, 2017.

4. Beltz N, Gibson A, Janot J, Kravitz L, Mermier C, Dalleck L. Graded exercise testing protocols for the determination of VO_2max: Historical perspectives, progress and future considerations. Sports Med 2016: 3968393, 2016.

5. Bishop D, Jenkins D, Mackinnon L. The effect of stage duration on the calculation of peak VO_2 during cycle ergometry. J Sci Med Sport 1(3): 171-178, 1998.

6. Bland J, Altman D. Comparing methods of measurements: Why plotting difference against standard difference is misleading. Lancet 346: 1085-1087, 1995.

7. Buchfuhrer M, Hansen J, Robinson T, Sue D, Wasserman K, Whipp B. Optimising the exercise protocol for cardiopulmonary assessment. J Appl Physiol: Resp Environ Exercise Physiol 55(5): 1558-1564, 1983.

8. Cheng B, Kuipers H, Snyder A, Keizer H, Jeukendrup A, Hesselink M. A new approach for the determination of ventilatory and lactate thresholds. Int J Sports Med 13(7): 518-522, 1992.

9. Coyle E, Gonzalez-Alonzo J. Cardiovascular drift during prolonged exercise: New perspectives. Exerc Sport Sci Rev 29(2): 88-92, 2001.

10. Denham J, Scott-Hamilton J, Hagstrom A, Gray A. Cycling power outputs predict Functional Threshold Power and maximum oxygen uptake. J Strength Cond Res 34: 3489-3497, 2017.

11. Earnest C, Wharton R, Church T, Lucia A (2005). Reliability of the Lode Excalibur Sport ergometer and applicability to Computrainer electromagnetically braked cycling training device. J Strength Cond Res: 19(2): 344-348, 2005.

12. Foxdal P, Sjodin B, Sjodin A, Ostman B. The validity and accuracy of blood lactate measurements for the prediction of maximal endurance running capacity. Dependency of analyzed blood media in combination with different designs of exercise test. Int J Sports Med 15(2): 89-95, 1994.

13. Gavin T, van Meter J, Brophy P, Dubis G, Potts K, Hickner R. Comparisons of a field-based test to estimate Functional Threshold Power output at lactate threshold. J Strength Cond Res 26(2): 416-421, 2012.

14. Hamburg M. Statistical analysis for decision making. 3rd ed. Harcourt Brace Jovanovich; 1983. pp. 390-391.

15. Heck H, Mader G, Hess G, Mucke S, Muller R, Hollmann W. Justification of the 4-mmol·L-1 lactate threshold. Int J Sports Med 6: 117-130, 1985.

16. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Springer, New York; 2013. pp. 178-181.

17. Jamnick N, Botella J, Pyne D, Bishop D. Manipulating graded exercise test variables affects the validity of the lactate threshold and VO_2peak. PLoS One 13(7): e0199794, 2018.

18. Lamberts R, Swart J, Woolrich R, Noakes T, Lambert M. Measurement error associated with performance testing in well-trained cyclists: Application to the precision of monitoring changes in training status. Int Sports Med J 10: 33-44, 2009.

19. Maassen N, Busse M. The relationship between lactic acid and work load: A measure for endurance capacity or an indicator of carbohydrate deficiency? Eur J Appl Physiol Occup Physiol 58: 728-737, 1989.

20. Mc Kay J, Horner K. What is known about the FTP 20 test related to cycling? A scoping review. J Sport Sci 39(23): 2735-2745, 2021.

21. McGrath, Mahony N, Fleming N, Donne B. Is the FTP test a reliable, reproducible and functional assessment tool in highly trained athletes? Int J Exerc Sci 12(4): 1334-1345, 2019.

22. McGrath, Mahony N, Fleming N, Raleigh C, Donne B. Do critical and functional threshold power equate in highly-trained athletes? Int J Exerc Sci 14(4): 45-59, 2021.

23. Morton R. The critical power and related whole-body bioenergetic models. Eur J Appl Physiol 96: 339-354, 2006.

24. Navalta J, Stone W, Lyons T. Ethical issues relating to scientific discovery in exercise science. Int J Exerc Sci 12(1): 1-8, 2019.

25. Nevill A, Ramsbottom R, Williams C. Scaling physiological measurements for individuals of different body size. Eur J Appl Physiol 65: 110-117, 1992.

26. Olbrecht J. The science of winning. Planning, periodizing and optimizing swim training. F & G Partners, Belgium; 2007. pp. 115-117.

27. Stoa E, Helgerud J, Ronnestad B, Hansen J, Ellefsen S, Oyvind S. Factors influencing running velocity at threshold in male and female runners at different levels of performance. Front Physiol 11:585267, 2020.