J Sci Cycling. Vol. 6(2), 22-26
DOI: 10.28985/171231.jsc.10

RESEARCH ARTICLE **Open Access**

Is the PowerCal device suitable for monitoring performance with competitive cyclists?

Carl D. Paton [1]✉ and Vitor P. Costa [2]

Abstract

The use of bicycle power meters is becoming an increasingly popular, although somewhat costly, means of monitoring training adaptations in competitive cyclists. The PowerCal (PCal) is a low-cost power prediction device that may offer an alternative means of monitoring cycling performance. The aim of this study was to determine the validity and reliability of the PCal in comparison to a laboratory standard ergometer, over a range of constant load powers. Nine trained male cyclists (35 ± 9 yrs, 177 ± 5 cm, 74 ± 6 kg) completed two exercise trials on a Velotron ergometer (VTron) while wearing the PowerCal (PCal) device. During each trial, participant's performed six constant-load efforts commencing at 100 W and increasing by 40 W increments up to a final exercise intensity of 300 W. Power output predicted by the PCal was significantly lower ($p < 0.05$) than that measured by the VTron in all but the 100W load condition. The groups mean bias (± 95 % CL) in the PCal prediction across all loads was - 19.3 % ± 6.4 % while individual subject bias covered a wide range from - 41.2 % to + 6.6 %. Correlations between the individuals predicted and measured power were all significant ($r \geq 0.93$). There were no significant differences ($p > 0.05$) in predicted power at any workload between repeat trials. The mean re-test typical error was similar (~ 6 W) across all measured power outputs, corresponding to a mean coefficient of variation of ~ 5 % between trials. The PCal reports substantially lower power outputs compared to the VTron ergometer. Furthermore, the variability in the PCal measures across repeated trials of ~ 5 % makes it unsuitable for monitoring performance changes that are considered worthwhile in trained competitive cyclists.

Keywords: Validity, Reliability, Cycling, Power meter, Ergometer

✉ cpaton@eit.ac.nz (C Paton)

[1] *Faculty of Health and Sport Science, Eastern Institute of Technology, Napier, New Zealand. cpaton@eit.ac.nz*
[2] *Centre for Health and Sport Science, Department of Physical Education, Santa Catarina State University, Florianopolis, Brazil.*

Received: 22 February 2017. Accepted. 2 October 2017.

Introduction

The accurate monitoring of exercise intensity is important for cyclists wishing to progress their performance through training, and also for researchers wishing to examine the efficacy of potential performance enhancing interventions. Power output is the most objective way to measure a cyclist's performance and has traditionally been measured using static ergometers in the laboratory (Abbiss et al., 2009). However, the last decade has seen a proliferation of portable power devices which are used to directly assess the cyclists' power during real cycling in either a field or laboratory setting (Bertucci et al., 2005; Duc et al., 2007). The practicality of these portable devices for monitoring cyclists power output is in little doubt; indeed most elite cyclists now use these devices to regularly monitor their power output during training and competition (Sanders et al., 2017).

The majority of the available portable power devices rely on force measures taken via strain gauges mounted in the bicycle cranks, pedals or hub of the rear wheel (Gardner et al., 2004; Bouillod et al., 2016). The accuracy and utility of several such mechanical devices have been investigated in a number of studies over the last decade. Two commonly used and highly regarded power devices are the SRM crank and the PowerTap hub both of which reportedly provide reliable power measures when compared to a gold standard calibration rig (Gardner et al. 2004); (Abbiss et al. 2009) and valid and reliable measures when compared to each other (Bertucci et al. 2005). In contrast several other commercially available portable devices are reported to provide poor levels of reliability and accuracy of power measures (Duc et al. 2007); (Bertucci et al. 2013) and as such have been deemed unsuitable for monitoring changes in power that are reportedly worthwhile to competitive cyclists (Paton and Hopkins 2006).

The complexity and cost of most portable power measuring devices currently preclude their widespread use amongst non-elite cyclists. However, a relatively recent product to the market the "PowerCal" (PCal), is designed to overcome the cost and technical limitations while providing good estimates of cycling power output. The PCal monitor is similar in construction to commonly available telemetry heart rate straps but utilises a built in propriety algorithm to convert measures of heart rate into power output. A unique feature of the PCal is its ability to respond almost instantaneously to changes in power despite no measurable change in exercise heart

rate. The procedures and algorithms that the PCal uses to convert these heart rate measures to power output is subject to trade patent and is thus unavailable from the manufacturer (personal communication).

To date, few studies exist examining the utility of the PCal device. Two recently published studies utilising the PCal have reported relatively poor reliability for power measures when performing short (< 45 s) duration high-intensity efforts (Costa et al. 2015) and also when completing longer duration but stochastic (none steady-state) time trials (Costa et al. 2017). However, to our knowledge, no studies have examined the utility of the PCal under steady-state exercise conditions, as is commonly performed during many laboratory exercise testing procedures. Therefore, in this study, we investigate the validity and reliability of the PCal over a range of constant load powers in comparison to a laboratory standard ergometer.

Materials and methods

Subjects

Nine male cyclists (Mean ± SD; age: 35 ± 9 years, mass 74 ± 6 kg, height 177 ± 5 cm, VO_2 peak 62 ± 5 $ml.kg^{-1}.min^{-1}$) volunteered to participate in this study. All cyclists gave their written informed consent to participate in the study which was approved by the Eastern institute of Technology research ethics committee. All cyclists were trained and competed regularly in cycling competitions. Further, all participants were familiar with the test procedures used in this study and had previously participated in a training study which utilised similar test procedures.

Study design

The study was a repeated measures controlled trial where cyclists completed two trials within a 7-day period. In each session, cyclists completed a series of exercise bouts at six constant load power outputs. All physiological and performance assessments were completed on a Velotron Dynafit Pro cycle ergometer (RacerMate Inc, WA, USA) calibrated in accordance with the manufacturer's instructions. The velotron ergometer has previously been shown to provide valid and reliable power measurements during steady state output in comparison to a gold standard calibration rig (Abbiss et al. 2009). Prior to testing each participant was fitted to the ergometer in a position to replicate their own racing bicycle; the fit measurements were recorded and repeated for the subsequent session. During each session, participants wore the PCal heart rate strap paired (via ANT+ protocols) to a cycle computer which recorded data at ~ 1 Hz. In the 24 hours before each testing session, participants were instructed to prepare as if it was a competition and to avoid strenuous physical activity and any potential performance altering supplements (e.g. caffeine) prior to the trials. Throughout all tests, cooling was provided via two 30 cm pedestal fans and the ambient temperature of the laboratory was controlled at ~ 20 ºC with a relative humidity of ~ 40 %.

Exercise Test Procedures

Cyclists initially completed a 15-minute warm-up; the first 10 minutes was performed at a self-selected low to moderate aerobic intensity and the final five minutes was performed at a fixed power of 150 watts (W). The VTron ergometer was then set to isokinetic mode for the test to ensure that power output remained constant regardless of any fluctuations in pedal cadence. Cyclists were requested to maintain a cadence of ~ 90 ± 5 revolutions per minute throughout the test. The test commenced at 100 W and was increased by 40 W every four minutes until the subject reached volitional exhaustion. During the test respiratory gases were continuously measured with a metabolic cart (Metalyser 3B, Cortex, Leipzig, Germany) calibrated in accordance with the manufacturer instruction using Alpha gas standards. Peak oxygen uptake (VO_2peak) was determined as the highest 30-second oxygen uptake value recorded during the test. Mean power output from the final two minutes of each exercise stage that was completed by all subjects (300 W) was downloaded from the PCal and used for subsequent data analysis and ergometer comparisons.

Statistical analysis

Simple descriptive statistics are shown as means ± standard deviation. Two made for purpose Microsoft excel spreadsheets (Hopkins, 2015) were used to determine the validity of PCal compared to the Vtron and also to assess the re-test reliability of the PCal device over repeated trials. The data were analysed as both raw values to derive absolute mean differences between devices and also as log transformed values in order to reduce any effects occurring through possible non-uniformity of error in the data and provide differences between devices as percent's (%). In addition, we used the Analysis of variance techniques (ANOVA) to determine if significant differences existed between the PCal and Vtron devices. All ANOVA analyses were performed using SPSS statistical software, version 20 for Windows (SPSS Inc, Chicago, IL) with alpha indicating significance set a priori at $p < 0.05$

Validity

The validity spreadsheet was used to determine the strength of the relationship between the group data and additionally between each individual participants VTron and PCal power output from each test stage, via the Pearson's correlation coefficient (r). The spreadsheet also provides the bias and standard error of estimate (SEE) in the PCal measures (in relation to the Vtron) as both an absolute value and as a percentage error (% ± 95 % CL) for log transformed data. In addition, a two-way analysis of variance (ANOVA) with repeated measures was also performed to determine if any significant differences existed between power output from the VTron and PCal devices for each stage of the incremental test. When significance was found, subsequent post hoc comparisons were made using Bonferroni corrections.

Reliability
Reliability between the PCal trials was determined using the appropriate spreadsheet and reported as the raw unbiased typical error in watts. In addition, we used log transformation of the data to enable error to be expressed as a percentage via the coefficient of variation (CV) along with the accompanying confidence intervals (± 95 % CL). The utilised spreadsheet also provided the intra-class correlation coefficient (ICC type 3:1) between repeat trials for the PCal. Once again, a two way repeated measures ANOVA with Bonferroni post hoc analysis was utilised to determine the presence of any significant differences between the PCal power outputs between trials.

Results

Validity
Power output was significantly lower (all $p < 0.05$) on the PCal compared to the VTron at all except at the first 100 W ($p = 0.71$) intensity. The Pearson's correlations (± 95 % CL) between the PCal and VTron power was 0.72 ± 0.14 and 0.72 ± 0.13 for trials one and two respectively. The overall mean systematic bias (± 95 % CL) across all power outputs for the PCal group data was similar (- 19.3 ± 6.4 % and 21.6 ± 6.2 %) for trials one and two. The standard error of estimate (SEE) between the Vton and Pcal power output was large at ~30 % for both trials. Figure 1 shows individual data plots (with linear regression) for participant's predicted power output versus measured power for the six exercise stages recorded (trial 1). The accompanying Pearson's correlations (r) between PCal and Vtron power for each individual subject (trial 1 only) were all significant and are reported in Table 1. Table 1 also details each individual's estimated systematic bias (from the VTron measured mean) and standard error of estimate (SEE) for the PCal predicted power. As an additional analysis, the data was also re-examined using the same procedures but with the omission of the lowest power (100 W) data. Omitting the lowest power output stage substantially reduced the estimation error in all subjects.

Reliability
There was no significant difference ($p > 0.05$) in PCal predicted power between trial one and two at any exercise intensity. Furthermore, the interclass correlations between repeated trials were identified as high ($r \geq 0.95$) across all power outputs. However overall test reliability, when reported as a coefficient of variation (table 2), was generally poor averaging ~ 5 % across all power outputs. The intraclass correlation between trials and the typical error expressed as both raw and % differences for each exercise intensity are also reported in table 2.

Discussion

The aim of this research was to determine the validity and reliability of the PCal prediction device, and in doing so assess its suitability as a monitoring tool for competitive cyclists. The main finding of this study was the PCal device does not provide valid power estimations when compared to a laboratory standard (Abiss et al. 2009) ergometer. Furthermore, the overall poor retest reliability of the PCal does not support its use for accurately monitoring the small but meaningful changes in power output that may matter to elite cyclists. It is apparent from our results that the PCal substantially under predicted power in the majority (7 out of 9) of subjects in this sample. The mean bias of ~20 % between the VTron and Pcal outputs across the subject group is large but more importantly, it is inconsistent across individual subjects. The individual bias in PCal power over the examined range was substantial, ranging from -41.2 % to + 6.6 % (table 1). The reason for the large bias range of the PCal is unknown, but it is presumably associated with each individual's unique exercise heart rate response and its interpretation by the devices built in algorithm. Unfortunately, due to the trademarked proprietary nature of the PCal it is not possible to determine any specific reasons for the large bias range. Furthermore, the large range in bias estimates between individual subjects makes it impossible to utilise any single correction algorithm in order to improve the accuracy of prediction for all athletes. However, despite the PCal limitations in overall accuracy, it did consistently report strong linear relationships with the VTron power output (Pearson's correlations $r > 0.93$) for each individual subject (see table 1). Given these strong correlations between individual predicted and measured power, it would be a relatively simple task to apply an appropriate correction equation to each individual's test data (post hoc) and in affect calibrate the PCal power output to that of the VTron or some other accurate power measuring device. Indeed, the spreadsheet utilised in this study provides such a calibration equation based on a simple linear regression plot between devices. Unfortunately, despite this ability to "calibrate" the power measure from the PCal closer to that of the Vtron

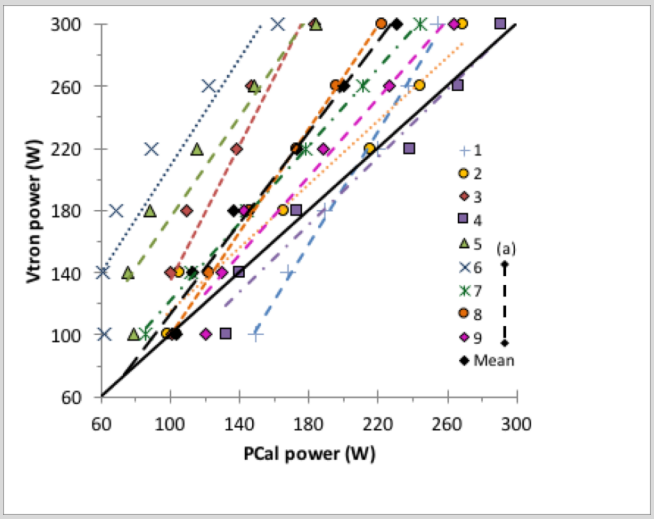

Figure 1. Data plots of VTron versus PCal power output for each individual subject: Solid line represents identity between the VTron and PCal: **(a)** indicates the approximate between subject standard deviation for the shown mean value (black dashed line).

Table 1. Pearson correlations between the VTron and PCal output from trial 1; mean bias and standard error (SEE) are reported as a CV % for each individual participant for the PCal power output.

Subject	Pearson r	Bias as % (Mean ± 95 % CL)	SEE % (6 stages) (mean x/÷ 95 % CL)	SEE % (5 stages)* (mean x/÷ 95 % CL)
1	1.0	6.6 (22.9)	5.3 (2.3)	3.1 (2.6)
2	0.98	-8.9 (9.4)	11.0 (2.3)	7.2 (2.7)
3	0.95	-27.8 (15.1)	20.5 (2.5)	8.3 (2.7)
4	0.98	5.4 (13.1)	12.3 (2.3)	5.5 (2.7)
5	0.95	-34.3 (10.4)	21.0 (2.5)	5.9 (2.7)
6	0.93	-41.2 (9.0)	22.2 (2.5)	9.0 (2.8)
7	1.0	-16.9 (2.3)	2.6 (2.2)	0.3 (2.6)
8	1.0	-15.9 (10.6)	4.6 (2.3)	1.2 (2.6)
9	0.97	-8.4 (13.9)	15.3 (2.4)	7.0 (2.7)

* Indicates SEE when 100 W stage is omitted from the analysis. x/÷ multiply and divide by this number to give the approximate 95 % confidence levels for the given value.

Table 2. Predicted PCal power from repeated trials and the associated intraclass correlation between trials.

Velotron (W)	PCal T1 mean ± SD (W)	PCal T2 mean ± SD (W)	ICC (± 95 % CL)	T1-T2 typical error (W)	T1-T2 CV % (± 95 % CL)
100	103 ± 27	95 ± 24	0.95 (0.78-0.99)	6.8 (4.6-13.0)	7.6 (5.1-15.1)
140	113 ± 33	111 ± 32	0.98 (0.90-0.99)	5.6 (3.8-10.7)	6.0 (4.0-11.8)
180	136 ± 40	133 ± 41	0.99 (0.94-1.0)	5.9 (4.0-11.3)	5.1 (3.4-10.1)
220	172 ± 50	164 ± 50	0.99 (0.94-1.0)	5.8 (3.9-11.1)	4.8 (3.2-9.3)
260	200 ± 50	194 ± 52	0.98 (0.94-1.0)	6.5 (4.4-12.3)	4.5 (3.0-8.8)
300	230 ± 45	225 ± 47	0.99 (0.97-1.0)	4.3 (2.9-8.1)	2.0 (1.4-4.0)

the random prediction error (SEE) of the corrected estimate would remain unacceptably high (> 10 %) especially when the power data from all test stages are used in the regression analysis. However, inspection of the linear regression plots indicates that the majority of error occurs at lower power outputs (<140W) and more especially with fitter individuals. To test this observation we removed the lowest PCal intensity (VTron 100 W) and repeated our analysis with the effect that our observed error was substantially reduced (to < 10 %) in all subjects and especially in those subjects with initially larger prediction errors. The large error at low power outputs is similar to that of another recent study by Hoon et al. (2016) that reports that the largest source of error of another portable power device (SRM) occurred at low power outputs. Therefore, we would recommend that any attempt at post hoc calibration with the PCal take into account each subjects personal heart rate response and commence at power outputs above 140 W especially for fitter individuals.

A concept arguably far more important than the validity of the PCal device is its retest reliability. It is the reliability of a device or measure that determines its ability to track potentially worthwhile changes in performance (Hopkins et al. 2001). A previous study by Paton and Hopkins (2006) has established that elite cyclists vary in their performance times by between 0.5 - 2.4 % across a range of cycling event. Based on this variation in elite cyclist's race times, the smallest worthwhile change in power output for an elite cyclist would be estimated to ~1-2.5 %, depending on the type of event. However, we found that the PCal produced a mean CV of ~5 % across all the powers utilised in this investigation, which is at least double the smallest worthwhile enhancement. Any device with such a large CV would therefore be unable to track any changes in power smaller than the determined error of the device itself (Hopkins et al. 2001). Interestingly the absolute error of the PCal remained consistent ~6 W across the range of measured power outputs; this has the effect of actually reducing the observed % error from 7.6 % at the lowest output to 2.0 % at the highest. The reported CV for the PCal in the current study of ~5 % is consistent with the value (4.9 %) for for the PCal device reported by Costa et al (2017) when cyclist completed repeated 20-km stochastic (variable intensity self-paced) time trials. In comparison to the PCal, several recent studies have reported that some of the more expensive mechanical cycle power monitors available produce CVs over a range of steady state power output of 0.5-2.5 % (Hoon et al. 2016; Bouillod et al. 2016; Sparks et al. 2016). A limitation of the current study is that the power range examined is somewhat restricted. The 300 W intensity was chosen as a cut-off point for this study as it was the highest that all subjects could maintain for a full four-minute test duration and while the range is limited it is representative of the typical aerobic exercise intensities seen amongst lower ability amateur cyclists for which the PCal is primarily intended. Regardless of the upper limit of power output used, the reported CV for the PCal represents a considerable inter-trial difference, and as such would limit the devices ability to track anything but large changes (<5 %) in cycling power output over time. Given the large between-trials CV of the PCal compared to other available portable power devices we would suggest that any small changes in power output of less than 5 % identified by the PCal be interpreted with caution.

Practical Applications

The PCal is a simple to use low-cost device designed to monitor power output with cyclists. While the PCal may provide some estimation of power output it lacks validity compared to other available portable power devices, and more importantly, shows poor retest reliability (CV ~ 5 %). The poor reliability of the device renders it unsuitable for monitoring the small (1-2 %) changes in power output that appear to matter to elite level cyclists. However, while the PCal might not be appropriate for monitoring small changes in power with elite cyclists it may provide useful power information to lower ability cyclists or those expected to make large improvements in their power output following training.

Acknowledgement
The authors would like to thank Professor Will Hopkins for his valuable assistance with data analysis and interpretation.

Conflict of interest
None.

References

1. Abbiss CR, Quod MJ, Levin G, Martin DT, Laursen PB (2009) Accuracy of the Velotron Ergometer and SRM Power Meter. International journal of Sports Medicine 30: 107-112
2. Bertucci W, Crequy S, Chiementin X (2013) Validity and reliability of the G-Cog BMX Powermeter. International Journal of Sports Medicine 34: 538-543
3. Bertucci W, Duc S, Villerius V, Pernin JN, Grappe F (2005) Validity and Reliability of the PowerTap Mobile Cycling Powermeter when Compared with the SRM Device. International journal of Sports Medicine 26: 868-873
4. Bouillod A, Pinot J, Soto-Romero G, Bertucci W, Grappe F (2016) Validity, Sensitivity, Reproducibility of the Powertap, stages and Garin Vector Power meters in comparison with the SRM device. International Journal of Sports Physiology and Performance. 14:1-26. [Epub ahead of print]
5. Costa VP, Guglielmo LG, Paton CD (2015) Reproducibility and validity of the PowerCal device for estimating power output during sprints in well-trained cyclists. Isokinetics and Exercise Science 23: 127-132
6. Costa VP, Guglielmo LG, Paton CD (2017) Validity and Reliability of the PowerCal Device for Estimating Power Output During Cycling Time Trials. Journal of Strength and Conditioning Research 31: 227-232
7. Duc S, Villerius V, Bertucci W, Grappe F (2007) Validity and reproducibility of the ErgomoPro powermeter when compared with the SRM and PowerTap powermeters. International Journal of Sports Physiology and Performance 2: 270-281
8. Gardner AS, Stephens S, Martin DT, Lawton W, Lee H, Jenkins D (2004) Accuracy of SRM and Power Tap power monitoring systems for bicycling. Medicine and Science in Sport and Exercise 36: 1252-1258
9. Hoon MW, Michael SW, Patton RL, Chapman PG, Areta JL (2016) A Comparison of the Accuracy and Reliability of the Wahoo KICKR and SRM Power Meter. Journal of Science and Cycling 5: 11-15
10. Hopkins WG (2015) Spreadsheets for analysis of validity and reliability. Sportscience, 19:36-42. (sportsci.org/2015/ValidRely.htm)
11. Hopkins WG, Schabort EJ, Hawley JA (2001) Reliability of power in physical performance tests. Sports Medicine 31: 211-234
12. Paton CD, Hopkins WG (2006) Variation in performance of elite cyclists from race to race. European Journal of Sport Science 6: 25-31
13. Sanders D, Myers T, Akubat I (2017). Training Intensity Distribution in Road Cyclists: Objective Versus Subjective Measures. International Journal of Sports Physiology and Performance 2:1-20. [Epub ahead of print]
14. Sparks AS, 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: 35-41