J Sci Cycling. Vol. 7(1), 3-8
DOI: 10.28985/180130.jsc.02

RESEARCH ARTICLE

Open Access

Agreement between the Stages Cycling and PowerTap Powermeter

Patrick Schneeweiss.[1]✉, Samuel Haerlen[1], Marc-Daniel Ahrend[1], Andreas Michael Niess[1], Inga Krauss[1]

Abstract

Several powermeters for almost every type of bicycle are available. The PowerTap (CycleOps) quantifies cycling power and cadence in the rear wheel hub and has already been validated in previous studies. The Stages Cycling Powermeter (Stages Cycling) is lower-priced and more flexible for usage as it measures in the left crank arm. The aim of this study was to determine the agreement between these two devices.

38 participants performed laboratory tests on a stationary roadbike. Power output and cadence were recorded with PowerTap and Stages simultaneously. Differences in power output and cadence were determined. The agreement between methods was quantified by use of mean differences and limits of agreement.

Stages Powermeter underestimates power output by -1.9 ± 4.0% in comparison to the PowerTap (limits of agreement: 5.9% to -9.7%). Considering cadence, Stages calculates 0.94± 0.16) revolutions per minute more than the PowerTap (limits of agreement: -0.4 to 2.3 rpm). Mean coefficients of variation for power output (50.1%) and cadence (14.2%) estimate good reliability of Stages compared to PowerTap (50.3% and 14.3%).

Despite a systematic bias, Stages can be considered a suitable alternative to measure power output. However, limitations regarding power output measurement have to be respected, especially when cycling with high intensities.

Keywords: bicycle; powermeter; validity; reliability

✉ **Contact email:** patrick.schneeweiss@med.uni-tuebingen.de (P Schneeweiss)

[1] *Department of Sports Medicine, Medical Clinic, University of Tuebingen, Germany*

Received: 22 January 2018. Accepted. 28 April 2018.

Introduction

Performance measures are important instruments to control training intensities and to monitor physical demands within competition. In this regard the heart rate has been the gold standard instrument for a long time. However, high-intensity intermittent cycling can only be controlled inadequately by heart rate because it is considerably affected by internal and external factors and it's respond to exercise intensity modifications is delayed (Theobald 2015). In contrast, mobile powermeters instantaneously assess cyclist's power output and cadence and thus allow a continuous monitoring. Measuring the PO or the work accomplished during a training session or race is therefore the most precise way to control the intensity in cycling (Allen and Coggan 2010). As the most economical cadence increases with increasing workload in elite cyclists (Foss and Hallen 2004), mobile powermeters can help tremendously to optimize cyclists' training. This is why mobile powermeters became indispensable for professional athletes and get increasingly attractive for amateur athletes as well.

Powermeters can be integrated into the rear hub, crank, bottom bracket or pedal of the bike and are applicable for road and off-road bikes as well as for BMX bikes. However, some of the offered systems seem to be less valid though partly reliable (Bertucci 2012; Bertucci et al. 2005; Bouillod et al. 2017; Duc et al. 2007; Gardner et al. 2004; Hurst and Atkins 2006; Hurst et al. 2015; Kirkland et al. 2008; Miller et al. 2016; Millet et al. 2003; Sparks et al. 2015). Up to now, the SRM crank set (SRM, Welldorf, Germany) is considered the most accurate mobile powermeter (Balmer et al. 2004; Jones and Passfield 1998; Lawton et al. 1999) and therefore used as "gold standard" to measure cycling performance. The PowerTap (PT; CycleOps, Madison, USA) is also well established and quantifies PO via strain gauges in the rear wheel hub. It has been compared with the SRM system in previous studies (Bertucci et al. 2005; Duc et al. 2007; Gardner et al. 2004). The Stages Cycling Powermeter (SCP; Stages Power, Boulder, CO, USA) is characterized by a low additional weight of 20g, comparatively low purchase cost and a more flexible usage as PO is quantified via strain gauges in the left crank arm. According to the manufacturer, it provides a measurement accuracy of ± 2% and is one of the most frequently used powermeters. However and to the best of our knowledge, only three independent studies investigated the measurement accuracy of SCP (Bouillod et al. 2017; Hurst et al. 2015; Miller et al. 2016).

Hurst (Hurst et al. 2015) and Bouillod (Bouillod et al. 2017) compared SCP and other powermeters with the SRM system during different laboratory and off-road cycling tests that were performed by a single experienced male cyclist. Miller (Miller et al. 2016) stated strong agreement during steady cycling under laboratory conditions but significant differences of SCP compared to PT and Quarq during a varied off-road test performed by eight nationally competitive XCO-MTB athletes. Summarizing the results of these studies, SCP underestimates PO when compared with other powermeters though it seems to be reliable. However, these findings are only based on very few subjects and reference data are related to non-standardized test settings.

Thus, the aim of this case study was to determine the agreement of SCP and PT at different cycling intensities under standardized laboratory conditions with an adequate sample size.

Methods

Subjects

38 students (33 male, age: 23.8 ± 3.5 years, height: 179.5 ± 6.7 cm, body weight 74.5 ± 8.2 kg) were recruited from the University and voluntarily participated in this study. Inclusion criteria for study participation were physical health and a self-reported capacity to conduct an incremental cycling test. Subjects were excluded from study participation in case of any pulmonary, cardiac, acute or chronic disease. The study was approved by the local University Ethics Committee and conducted in accordance with the ethical standards as required by the journal (Harriss and Atkinson 2013). All participants gave informed written consent.

Methodology

Tests were performed on an ordinary road bike (Scott CR 1; medium frame size). The bike was mounted on a stationary cycle trainer (Tacx Flow T2200, Tacx B.V., Wassenaar, Netherlands) and fitted to participants' preferences. The handlebar computer of the stationary cycle trainer controlled the resistance of its electromagnetic brake and displayed PO and cadence. Tires were inflated to 700-750 kPa. The used SCP (Shimano FC-6800; 175 mm) replaced the original left crank arm. The manufacturer does not give information on the number of integrated strain gauges. To determine the pedalling frequency (cadence), SCP uses an integrated accelerometer that identifies each crank revolution.

The used PT was a G3 road model with eight strain gauges placed in the rear wheel hub. The PT uses the repetitive PO peaks, caused by the maximum pressure phase of the respective leg to identify the position of the crank and thus the cadence. Each, SCP and PT were paired via ANT+ (radio frequency signal) to Garmin Edge computers (Edge 705 & Edge 800; Garmin, Lenexa, Kansas, USA) which recorded data once per second. Prior to each test, both powermeters and the cycle trainer were calibrated as described by the manufacturers.

Before exercise testing, participants answered a short medical questionnaire and a physician carried out a medical examination. If no pulmonary, cardiac, acute or chronic diseases were found, participants were included into the study. PO and cadence of SCP (POS, CS) and PT (POPT, CPT) were recorded simultaneously once per second by the Garmin devices.

Test protocol

Prior testing, participants had to answer some questions related to their health status and were subsequently examined by a medical doctor to state physical eligibility. The incremental test consisted of seven levels with increasing loads from 100 to 400 watt (W), each lasting 120 seconds (s). Levels were separated from each other by recovery periods at 70 W (90 s). The submaximal sprint test followed immediately after the incremental test and started at 70 W. After 180 s, resistance increased to 600 W and participants attempted to maintain this load for 30 s. The test protocol was cut short if a participant could no longer perform the required performance at any level of the test. Then, a 180 s lasting cool down period completed the test.

The nominal PO was defined according to the cycle trainer. This led to an overall duration of 19 to 33 minutes, depending on participants' individual cycling performance. During the whole exercise testing, participants had to stay seated and maintain a cadence between 70 and 110 revolutions per minute (rpm). The reference value for the cyclist to hit the predefined resistance was defined according to the PO of the cycle trainer and displayed in the computer at the handlebar. PO and cadence of SCP (POS, CS) and PT (POPT, CPT) were recorded simultaneously once per second by the Garmin devices throughout the test.

Statistical Analysis

Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) was used to import the data from the Garmin devices, whereas the statistical analysis was done with IBM SPSS Statistics (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.).

Levels up to 400 W were included into data analysis if athletes could perform at least 30 s of the given level. For sprint testing (600 W), the minimum period for data analysis was set to 15 s. The first and last ten seconds of each level were not analysed as adaptation efforts to changed resistance or breaking off the test ahead of schedule could influence cycling PO substantially.

Absolute differences in PO ($PO_{Diff} = PO_S - PO_{PT}$) and Cadence ($C_{Diff} = C_S - C_{PT}$) between methods were determined. Data were further normalized to the corresponding PT values (PO_{nDiff} and C_{nDiff}). Agreement between methods was quantified by use of mean differences, 95% confidence intervals (CI) and 95% limits of agreement (LoA) (Bland and Altman 1986, 1999). Heteroscedasticity was checked by a linear regression model (standardized predicted values against standardized residuals). The Shapiro-Wilk test was used to test for normality for relative and absolute differences. Paired-sample t-tests ($P< 0.05$) were used for parametric data to test for a statistically significant systematic error between methods. Coefficients of variation (CV) as the ratio of standard deviation to the mean were calculated for PO_S, C_S and PO_{PT}, C_{PT}, respectively.

Results

No participant had to be excluded from the study for any reason. All participants completed the first four levels for at least 30 s. Some participants could not perform subsequent levels, resulting in a decreasing number of analysed records from 38 records (100 W) to 15 records (400 W) due to participants' subjective exhaustion. Nine participants completed the sprint test; another 15 riders could perform for at least 15 seconds.

Absolute differences of PO between PT and SCP were not normally distributed, while normalized differences were normally distributed. Relationship between cadence (C_{mean}) and PO_{Diff} was weak (r = 0.12; P=.059) and therefore not analysed in detail. Mean PO measured via SCP and PT both were higher than the nominal resistance predetermined by the cycle trainer: 6.7 W and 12.7 W respectively. SCP calculated systematically lower PO values than PT.

Table 1 shows PO data per level for PT and SCP. The t-test showed a significant difference between the systems for normalized values (PO_{nDiff}), indicating an overall systematic bias of -1.9% (P<.001) (see also figure 2).

Table 1. Cycling power output: comparison of Stages and PowerTap powermeter.

Level [W]	PO_{PT} [W]	PO_S [W]	PO_{nDiff} [%]	LoA [%]	CV_{PT} [%]	CV_S [%]
100	109 ± 10	108 ± 12	-0.7 ± 5.8	-12.1 - 10.7	9.4	10.9
150	164 ± 11	160 ± 11	-2.0** ± 3.7	-9.1 - 5.2	6.8	7.0
200	212 ± 15	210 ± 14	-0.9 ± 3.5	-7.8 - 5.9	7.0	6.5
250	264 ± 12	258 ± 11	-2.1** ± 3.4	-8.8 - 4.5	4.6	4.3
300	317 ± 14	309 ± 14	-2.7** ± 3.1	-8.8 - 3.4	4.5	4.4
350	365 ± 14	355 ± 14	-2.6** ± 3.7	-9.9 - 4.7	3.8	3.9
400	420 ± 13	407 ± 16	-3.1* ± 4.1	-11.2 - 5.0	3.2	4.1
600	604 ± 52	592 ± 46	-1.8* ± 3.4	-8.5 - 5.0	8.6	7.8

PO_{PT}: power output PowerTap; PO_S: power output Stages; PO_{nDiff}: normalized PO difference ((PO_S – PO_{PT})/PO_{PT}); *P< 0.05; **P< 0.01; LoA: 95% limits of agreement (PO_{nDiff}); CV: coefficient of variation

Table 2. Cycling cadence: comparison of Stages and PowerTap powermeter

Level [W]	C_{PT} [rpm]	C_S [rpm]	C_{nDiff} [%]	LoA [%]	CV_{PT} [%]	CV_S [%]
100	97.8 ± 8.3	98.9 ± 8.1	1.1** ± 0.8	-0.4 ± 2.6	8.5	8.2
150	91.0 ± 8.6	91.8 ± 8.5	0.9** ± 0.8	-0.6 ± 2.5	9.4	9.2
200	87.6 ± 8.0	88.3 ± 8.0	0.9** ± 0.6	-0.2 ± 2.0	9.1	9.1
250	85.8 ± 8.5	86.8 ± 8.5	1.2** ± 0.7	-0.2 ± 2.7	9.9	9.8
300	84.6 ± 8.6	85.6 ± 8.6	1.3** ± 1.1	-0.8 ± 3.4	10.1	10.1
350	81.1 ± 7.7	82.1 ± 7.7	1.2** ± 0.9	-0.5 ± 2.9	9.5	9.4
400	80.4 ± 12.1	81.3 ± 11.7	1.1** ± 0.4	0.3 ± 2.0	15.1	14.4
600	86.4 ± 9.5	87.4 ± 10.1	1.1** ± 0.9	-0.6 ± 2.8	11.0	11.6

C_{PT}: cadence PowerTap; C_S: cadence Stages; rpm: revolutions per minute; C_{nDiff}: cadence' normalized difference ((C_S – C_{PT})/C_{PT}); *P< 0.05; **P< 0.01; LoA: 95% limits of agreement (CnD_{iff}); CV: coefficient of variation

The difference plots show pairwise scattering PO data, the mean bias and 95% limits of agreement as proposed by Bland and Altman (Bland and Altman 1986, 1999). Heteroscedasticity of absolute PO data is shown as PO difference increased with increasing PO values (r = 0.328**; figure 1). However, the normalized differences PO_{nDiff} were independent of cycling PO (r = 0.108ns; figure 2). As a consequence, interpretation of bias and levels of agreement are subsequently related to normalized PO values only.

Participants' self-determined cadence increased with increasing PO (r = 0.68, P<.0001), while SCP calculated slightly but statistical significant higher values on each level than the PT (table 2). Differences between the systems were almost consistent across the levels, while 95% LoA ranged in total from -0.5% to 2.7% or -0.4 rpm to 2.3 rpm (figure 3). Differences in cadence showed a homoscedastic distribution, (r = 0.006_{ns}; figure 3). However, to proceed in the same way as for the power output data, we normalized data to the corresponding PT values (r = 0.085ns) for further calculations.

The CV was calculated for POS, CS and POPT, CPT respectively (table 1 and 2). This should allow an assessment regarding the reliability of SCP through the comparison to the CV of PT. Mean CVs of 6.0% and 6.1% for SCP and PT express mean variation of PO, whereas determination of cadence caused CVs of 10.3% and 10.2% respectively.

Discussion

Previous studies indicated acceptable validity and reliability of PT compared to SRM during sub-maximal intensities up to 450 W (Bertucci et al. 2005; Duc et al. 2007; Miller et al. 2016). Thus, in our view the use of PT as a reference instrument is reasonable. Our results show that SCP significantly underestimates PO compared to PT. Particularly with regard to very high loads in short intensive intervals, absolute differences between powermeters are more pronounced. The mean difference of 1.9 % between SCP and PT seems reasonable. However, 95% of the individual PO differences account for up to 10%. Reducing this to practice, differences in PO may account for up to 50 W between PT and SCP when cycling at higher intensities. These differences between the methods should therefore be considered in performance testing and interpretation of training loads.

Hurst (Hurst et al. 2015) investigated an experienced male cyclist who performed 15 timed short off-road climbs. The authors stated that SCP underestimates PO (-8%) but represents an affordable solution to determine PO. Furthermore, they assume that cyclists' bilateral imbalances may have potential influence on PO calculations as SCP measures PO single-sided. Miller (Miller et al. 2016) recently examined the agreement between SCP, PT and Quarq (Quarq, Spearfish, SD, USA). They stated strong agreement during steady cycling under laboratory conditions with similar CVs for each device (about ± 2.1%) but significant differences of

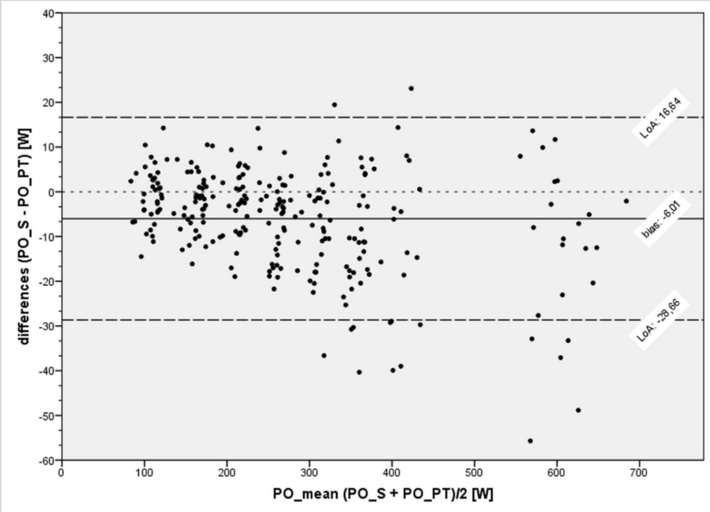

Figure 1. Differences of power output between Stages (PO_S) and PowerTap (PO_PT) with mean bias (-6.01 W) and 95% limits of agreement (16.64 W to -28.66 W).

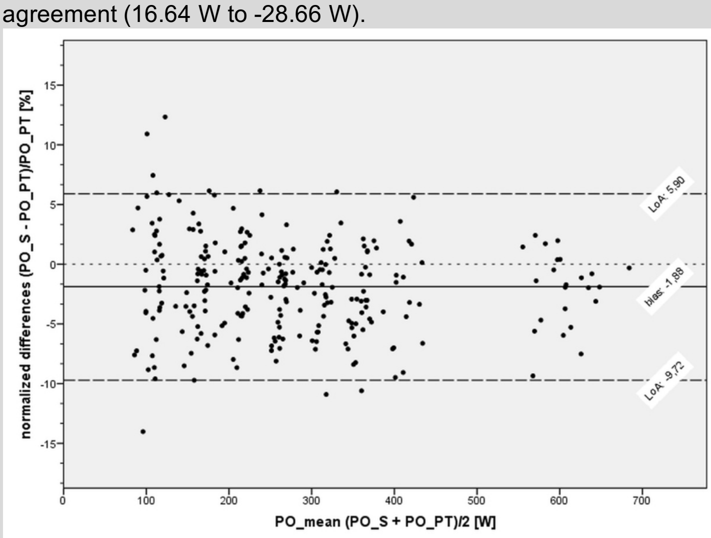

Figure 2: Normalized differences of power output between Stages (PO_S) and PowerTap (PO_PT) with mean bias (-1.88%) and 95% limits of agreement (5.90% to -9.66%).

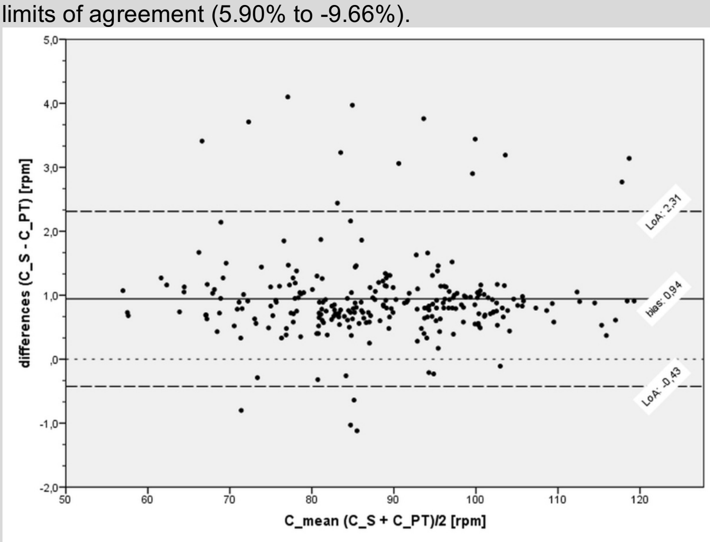

Figure 3. Differences of cadence between Stages (C_S) and PowerTap (C_PT) with mean bias (0.94 rpm) and 95% limits of agreement (-0.43 rpm to 2.31 rpm). rpm = revolutions per minute

SCP compared to PT and Quarq during a varied off-road test (P<.05).

With 38 participants, this present investigation was able to show that individual differences between SCP and PT regarding PO are remarkable as limits of agreement show a broad range of PO calculations. This may be in part due to the algorithms used to calculate PO and the potential influence that bilateral imbalances may have on these calculations. SCP calculates PO by simply doubling the values of the left leg while PT sums up pedalling forces of both legs. A divergent use of the legs could significantly affect the accuracy of SCP. Nevertheless, it is unknown how much of the variation in PO is attributable to bilateral asymmetries. In their review, Carpes et al. (Carpes et al. 2010) treat the influence of bilateral asymmetries on different variables during running and cycling. They conclude that there are differences in the expression of skills and abilities between the legs but the impact of these differences on PO and how they are highly individual or generalizable is vague. As long as the influence of muscular differences on cycling performance is not sufficiently clarified, a single-sided measurement of pedalling forces remains questionable and needs further investigation. Assuming that in right-handed people the right leg is more powerful than the left leg (Carpes et al. 2010) would explain lower PO_S compared to PO_{PT}.

Furthermore, the location and number of used strain gauges could cause differences between tested systems. As the manufacturer does not give a statement on it, the number of strain gauges used in SCP is unknown. As Stages Cycling assume by their own account that the number of used strain gauges does not affect measurement accuracy of mobile powermeters, one might suspect rather few integrated strain gauges in the SCP. However, we do not know how this affects measurement accuracy. Precise training at high intensities can therefore only be warranted if precedent performance tests are carried out with the same powermeter. The SCP measures the power, unlike the PT in the crank. This could theoretically result in a lower power output for the PT due to the mechanical loss in the drive train. In contrast to the results of Bertucci et al. (Bertucci et al. 2005), who also compared a crank based power meter with a rear wheel hub system, this did not lead to a lower PO_{PT} compared to the PO_S. It is conceivable that the manufacturers of the mobile power meters will take mechanical losses into account when calculating the cycling PO, or that these losses are so small and present in any system, that they do not systematically influence the measurement of cycling PO.

Taking into account that there is only a weak correlation between cadence and differences in power output, this reliance was not analysed in detail. However, former investigations confirm this result (Bertucci et al. 2005; Duc et al. 2007). Despite different measurement procedures concerning the calculation of cadence, results show no practical relevant differences. SCP calculates about one revolution per minute more than the PT, what might not reduce its practical use.

We did not perform a real sprint test, as the inclusion criteria presumably led to a rather heterogeneous sample and the technical design was not optimally suited for this (essential gear changes, strong vibrations on the bicycle trainer). Instead, we chose the 600 W level to achieve comparable high performance in all areas. Only nine subjects were able to maintain the load of 600 W for the entire 30 s, while 15 participants still reached 15 s. This means that this test is comparable with a maximum sprint test for most subjects.

It can be seen as a limitation of the study that the comparison of the two systems only took place under laboratory conditions. Under the given conditions, comparable results were achieved, but whether this would also be the case with measurements in the field cannot be answered. However, some other studies compared mobile powermeters under laboratory and field conditions e.g. (Bertucci et al. 2005; Bouillod et al. 2017; Duc et al. 2007; Miller et al. 2016) or solely under field-based conditions (Hurst et al. 2015). In summary, it can be said that under field-based conditions there might be a higher variability between and within the tested systems but the results are not entirely consistent. The variability of the mobile power meters rather tends to increase under high sprint loads than being excessively influenced by the environmental conditions, at least during road cycling.

The most important finding of this study is that there are relevant differences in power output data between Stages and PowerTap powermeters, especially when cycling at high intensities. Despite a lower PO and slightly higher cadence compared to the PT, SCP is an affordable alternative to well established powermeters such as SRM or PT.

Practical applications

Precise training of high intensive intervals cannot be controlled by heart rate but by power output. In order to attain the highest possible accuracy of the data, performance tests should be carried out with the same powermeter that is used for training or competition. Although the variability is relatively high, the mean PO difference of less than 2% underpins the practical use of SCP.

Acknowledgements

We thank all the participants who volunteered to take part in the study.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The manuscript has been read and approved by all the listed co-authors and meets the requirements of co-authorship as specified in the Authorship Guidelines. We have no conflicts of interest to disclose.

References

Allen H, Coggan A (2010) Training and racing with a power meter. VeloPress

2. Balmer J, Bird S, Davison RC, Doherty M, Smith P (2004) Mechanically braked Wingate powers: agreement between SRM, corrected and conventional methods of measurement. J Sports Sci 22: 661-667

3. Bertucci W (2012) Analysis of the agreement between the Fortius cycling ergometer and the PowerTap powermeter PO during time trials of 6 and 30 min. Comput Methods Biomech Biomed Engin 15: 212-214

4. 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. Int J Sports Med 26: 868-873

5. Bland JM, Altman DG (1986) Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. Lancet 1: 307-310

6. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8: 135-160

7. Bouillod A, Pinot J, Soto-Romero G, Bertucci W, Grappe F (2017) Validity, Sensitivity, Reproducibility, and Robustness of the PowerTap, Stages, and Garmin Vector Power Meters in Comparison With the SRM Device. Int J Sports Physiol Perform 12: 1023-1030

8. Carpes FP, Mota CB, Faria IE (2010) On the bilateral asymmetry during running and cycling - a review considering leg preference. Phys Ther Sport 11: 136-142

9. Duc S, Villerius V, Bertucci W, Grappe F (2007) Validity and reproducibility of the ErgomoPro power meter compared with the SRM and Powertap power meters. Int J Sports Physiol Perform 2: 270-281

10. Foss O, Hallen J (2004) The most economical cadence increases with increasing workload. Eur J Appl Physiol 92: 443-451

11. Gardner AS, Stephens S, Martin DT, Lawton E, Lee H, Jenkins D (2004) Accuracy of SRM and power tap power monitoring systems for bicycling. Med Sci Sports Exerc 36: 1252-1258

12. Harriss DJ, Atkinson G (2013) Ethical standards in sport and exercise science research: 2014 update. Int J Sports Med 34: 1025-1028

13. Hurst HT, Atkins S (2006) Agreement between polar and SRM mobile ergometer systems during laboratory-based high-intensity, intermittent cycling activity. J Sports Sci 24: 863-868

14. Hurst HT, Atkins S, Sinclair J, Metcalfe J (2015) Agreement Between the Stages Cycling and SRM Powermeter Systems during Field-Based Off-Road Climbing. Journal of Science and Cycling 4: 21-27
15. Jones S, Passfield L (1998) The dynamic calibration of bicycle power measuring cranks. The Engineering of Sport 0: 265-274
16. Kirkland A, Coleman D, Wiles JD, Hopker J (2008) Validity and reliability of the Ergomopro powermeter. Int J Sports Med 29: 913-916
17. Lawton E, Martin D, Lee H (1999) Validation of SRM power cranks using dynamic calibration. 5th IOC World Congress
18. Miller MC, Macdermid PW, Fink PW, Stannard SR (2016) Agreement between Powertap, Quarq and Stages power meters for cross-country mountain biking. Sports Technology 8: 44-50
19. Millet GP, Tronche C, Fuster N, Bentley DJ, Candau R (2003) Validity and reliability of the Polar S710 mobile cycling powermeter. Int J Sports Med 24: 156-161
20. Sparks SA, Dove B, Bridge CA, Midgely AW, McNaughton LR (2015) Validity and reliability of the look Keo power pedal system for measuring power output during incremental and repeated sprint cycling. Int J Sports Physiol Perform 10: 39-45
21. Theobald U (2015) Leistungsanforderungen und Trainingsmittel in der Radsportdisziplin Mountainbike Cross-Country. Leistungssport 1: 20-24