

Validity and reproducibility of commercial cycling power meters in hot and cold environmental temperatures

Robert Shute, Roksana Zak, Dustin Slivka ✉

Abstract

Power meters are a training tool used to help cyclists improve performance by objectively monitoring intensity. Some power meters are well established and validated, whereas others are relatively new. Most power meters have been tested for validity and reliability in laboratory and field settings of similar conditions; however, the reproducibility of these power meters across different temperatures has not been established. To examine the potential differences of the CompuTrainer, PowerTap, Stages, and Vector power meters in hot and cold compared to a room temperature environment. Recreationally trained male (n=7) and female (n=3) participants each completed three incremental cycling trials in hot (33°C), cold (7°C), or room temperature (RT, 20°C) conditions. The power meters were placed on a standard road bicycle and power output was logged and recorded. The CompuTrainer was higher in the room temperature trial compared to the cold and the hot, but not between the hot and cold trial. The PowerTap was not different in RT and cold, but was lower in hot compared to RT and compared to cold. The Stages was not different between RT and cold, but was lower in the hot compared to RT and compared to cold. The Vector was not different between RT and cold but was lower in the hot compared to RT and compared to cold. These data indicate that environmental temperature may affect the reproducibility of power meters. It is important to recognize the potential differences between temperatures when choosing a power meter.

Keywords: Vector, PowerTap, CompuTrainer, Stages, exercise

✉ **Contact email:** rshute@unomaha.edu (R. Shute)

¹ University of Nebraska at Omaha, EE.UU

Received: 01 March 2019. Accepted: 09 October 2019.

Introduction

Power meters have become an increasingly popular tool for recreational and competitive cyclists to monitor intensity due to advancement in technology and increased affordability. Several models of power meters exist for bicycles. This provides an opportunity for cyclists to monitor power output outside of the laboratory, offering an objective measure of workout intensity. Measuring cycling power output may have advantages over simply monitoring heart rate (Achten & Jeukendrup, 2003). For instance, power meters provide an independent and direct method of monitoring exercise intensity unaffected by the day to day variation that an individual may have when using heart rate alone (Jeukendrup & Diemen, 1998). Using heart rate may not be ideal if conditions such as environmental temperature are not controlled (Jeukendrup & Diemen, 1998; Swart, Lamberts, Derman, & Lambert, 2009).

If cyclists are going to make an investment in a power meter, it is important that it is valid and reliable to ensure proper exercise prescription and progression. Different locations of the power meter on the bicycle may lead to

variation, thus affecting the agreement or reliability (Atkinson & Nevill, 1998). In addition, these devices should be reproducible, which refers to the variation in repeated measurements made by a subject in changing conditions (Bartlett & Frost, 2008). Hourly, daily, and seasonal variations in temperature occur over the course of training. In order to maximize the adaptive response of a training session, it is important to obtain accurate information from training devices. Determination of the reproducibility of power meters in different environmental temperatures is of importance for those that exercise in climates that may have a wide range within a relatively short time frame or that constantly cycle in either a hot or a cold environment.

The PowerTap G3, Garmin Vector, Stages crankset, and the CompuTrainer are some of the more popular power meters available today. Each of these power meters are placed on a unique location of the bicycle. Specifically, the PowerTap is located in the hub of the rear wheel, the Stages in the crank set, the Vector in the pedals, and the CompuTrainer in the trainer that the bicycle is placed. The PowerTap and CompuTrainer have been manufactured and tested for well over a decade (Bertucci, Duc, Villerius, Pernin, & Grappe, 2005; Clark, Gartner, Williams, & Pettitt, 2015; Davison, Corbett, & Ansley, 2009; Earnest, Wharton, Church, & Lucia, 2005; Gardner et al., 2004; Sparks et al., 2016) but the Stages crank set and the Garmin Vector pedals are relatively new in the cycling market (Bouillod, Pinot,



Soto-Romero, Bertucci, & Grappe, 2017; Hurst, Atkins, Sinclair, & Metcalfe, 2015; Hutchison et al., 2017; Miller, Macdermid, Fink, & Stannard, 2015; Novak & Dascombe, 2016).

Much of the research involving power meters includes the reliability and validity of the power meters (Bertucci et al., 2005; Bouillod et al., 2017; Clark et al., 2015; Gardner et al., 2004; Hurst et al., 2015; Hutchison et al., 2017; Miller et al., 2015). However, limited research exists with temperature variations with these power meters. It is important to assess the reproducibility of these meters in different environmental temperatures in a controlled laboratory setting (Davison et al., 2009; Gardner et al., 2004). Therefore, the purpose of this study is to determine the reproducibility of four commercially available power meters using a standard road bicycle in hot and cold environmental temperature compared to a room temperature environment. In addition, the agreement of the four power meters can be evaluated with comparisons to each other within a given temperature. This study will help establish the practicality of using power meters throughout differing conditions.

Methods

Participants

The University of Nebraska Medical Center Institutional Review Board approved all testing and protocols. Recreationally active male ($n = 7$) and female ($n = 3$) participants were recruited for this study. Before testing began, participants read and signed the Institutional Review Board approved Informed Consent. Participants were asked to report to the laboratory on three separate occasions to complete trials in a hot (33°C), cold (7°C) or room temperature environment (20°C). Trials were completed in a randomized, counter-balanced order in a temperature and humidity controlled environmental chamber (Darwin, St. Louis MO).

Participant height, weight, and body composition were assessed on the participants' first visit. Height was measured with a 213 Stadiometer (Seca; United Kingdom) and a PS-660 ST digital scale (Befour; Saukville, WI) measured weight. Body density was measured by hydrostatic weighing using an electronic load cell-based system (Exertech; Dresbach, MN) with correction for estimated residual lung volume and gastrointestinal air volume (Quanjer et al., 1993). Body density from hydrostatic weighing was converted to percent body fat using the Siri equation (Siri, 1961).

Procedures

Before the participants began, all devices were calibrated according to manufacturer's instructions. To assess calibration of devices in different environmental temperatures and effect of changing temperature on calibration, calibrations were performed in the same temperature that the trials were performed. Cycling shoes with LOOK Kéo-compatible cleats (Look Cycle International, France) were provided for each participant in order to use the Garmin Vector pedals.

In order to assess a variety of workloads, participants performed an incremental $\text{VO}_{2\text{peak}}$ test in each of the three temperatures. The protocol began at 95 W and increased by 35 W after each 3 min stage. Participants were encouraged to keep pedal cadence over 60 rpm and the protocol was completed to volitional exhaustion defined by a rating of perceived exertion (RPE) greater than 18 or until pedal cadence was below 60 rpm. Maximum power associated with $\text{VO}_{2\text{peak}}$ was calculated by multiplying the portion of time spent in the final stage by 35 W and adding this portion to the W in the last completed stage. In order to assess the workloads within a given subject, the lowest final stage completed between the three temperature trials was used as a reference and any stage completed past this in a different environment was eliminated. Heart rate was monitored continuously with a chest strap (Polar Electronic, Lake Success, NY) and VO_2 was monitored with a gas and flow-calibrated metabolic cart (ParvoMedics TrueOne 2400 Metabolic System, Sandy, UT) continuously over the course of the trial.

The iterative calculation process generated tables of results that revealed the association between the three environmental variables and the subsequent changes in cycling velocity. The results were analysed to determine correction factors that could be used to conveniently determine the net effect of changes in environmental conditions on velocity, for a fixed average power output in a cycling event. The magnitude of these correction factors are themselves dependent upon velocity, which is why the entire analysis was completed for three velocities (45, 50 and 55 km/h). Equations were then derived to accurately calculate these correction factors, based on velocity.

Across the range of values for each of the environmental factors and for all three velocities modelled, the relationships between each of the environmental factors and velocity were effectively linear, achieving Pearson correlation coefficients of $r=0.9938$ to 0.9999 . Figures 1A, B & C. present the relationships between the environmental factors and cycling velocity, modelled for three different velocities. The slope of each of the three lines in each figure represents the rate at which velocity changes, in response to changes in each of the environmental factors. In other words, the slopes of the lines represent the magnitude of the factors that can be used to determine the effects of environmental conditions on cycling velocity.

Instruments and Calibration

All power meters were placed on a standard road bicycle (Lemond Composites, Oak Ridge, TN). The power meters were placed at different locations, which allowed for simultaneous measurement of all four power meters. The power meters were located in the bicycle trainer (CompuTrainer), in the rear wheel hub (PowerTap), in the crank arms (Stages), and in the pedals (Vector).

Prior to testing, the CompuTrainer (Racermate Inc., Seattle, WA) was calibrated according to manufacturer's instructions. Briefly, a 10 min warm-up at 150 W was performed on the bicycle by an investigator to prepare

the bicycle tire for the CompuTrainer Load Generator. After the warm-up, bicycle tires were inflated to 827.4 kPa (120 psi) and a rolling resistance calibration was performed by pedaling to a speed of 11.2 m · s⁻¹ then immediately discontinuing pedaling so the bicycle tire was allowed to coast to a stop. Rolling Resistance was repeated 3 times and calibration was set between .98 and 1.18 kg (1.11± .02 kg). All subjects' rolling resistance was replicated to within 0.02 kg for subsequent trials. A static zero calibration was performed on the Vector (Garmin Ltd, Olathe, KS) before each trial according to manufacturer's instructions (Garmin Ltd). Briefly, the bicycle was placed in the upright position with no pedal load applied (zero torque). Successful calibration was assumed when a display value of zero was indicated. For the Stages Shimano Ultegra 6800 crankset (Stages; Portland, OR), a zero offset calibration was performed by placing the left crank perpendicular with the ground with no pedal load applied to the bicycle. Successful calibration occurred with a display value of zero. The PowerTap G3 rear hub (PowerTap; Madison, WI) zero offset calibration was performed on the power meter by pairing with no pedal load applied to the bicycle.

Cycling computers

The Edge 500 cycling computer (Garmin Ltd) recorded and logged the Stages, PowerTap, and Vector power recording during the trial. Each power meter was paired to a separate Edge 500 computer and all computers were started simultaneously and sampled at a frequency of once per second (1 Hz). The Coaching Software (Racermate Inc.) logged the CompuTrainer power recordings. The watt load was set and clamped by the Racermate software as the CompuTrainer resistance can be adjusted by the Coaching Software program.

Statistical analysis

A three-way split-plot ANOVA was performed to indicate differences between power meters outputs (Temperature x Power meter x Stage). If significance was found, a Fisher's protected least significant difference post-hoc test was used to evaluate where significance occurred. A probability of type I error of less than 5% was considered significant (p < 0.05). Partial eta squared was used to calculate effect size within an ANOVA and Cohen's d was used to calculate effect size in pairwise comparisons. Bland-Altman plots and 95% limits of agreement were calculated to determine agreement of power meters within a given temperature (Bland & Altman, 1995). Since no true criterion measure was taken and more than two methods were used for comparison, the collective average was taken between the four power meters and this was used as the reference method. Data for Bland-Altman plots were tested for heteroscedasticity by plotting the absolute difference between a given power meter and mean of the power meters against the mean of the power meters

(Atkinson & Nevill, 1998). Coefficient of Variation (CV) was calculated to assess variation across temperature within a given power meter. All statistical data were analyzed using the Statistical Package for Social Sciences software (SPSS 24.0, Chicago, IL).

Results

Participant characteristics of the 10 recreationally trained subjects are shown in Table 1. Cadence was higher in the room temperature trial (92 ± 2 rpm)

Table 1. Participant Descriptive Data

Age (y)	24 ± 1
Height (cm)	176.0 ± 6.1
Weight (kg)	75.4 ± 10.0
W_{max} (W)	256 ± 41
Body Fat (%)	15.5 ± 5.6
VO_{2max} (ml · kg⁻¹ · min⁻¹)	56.6 ± 8.3

Data are mean ± standard deviation.

compared to the hot trial (88 ± 1 rpm, p = 0.009, Cohen's d = 0.351). Cadence was not different between room temperature and cold trials (89 ± 2 rpm, p = 0.119, Cohen's d = 0.233) or between the hot and cold trials (p = 0.237, Cohen's d = 0.112).

Power meters

There was not a 3-way interaction between temperature, power meter, and stage (p > 0.999). Additionally, there was not an interaction between power meter and stage (p = 0.348) or temperature and stage (p = 0.857, Table 2).

Table 2. Comparison of power meters during each trial.

Stage (Watt load)	CompuTrainer (W)	PowerTap (W)	Stages (W)	Vector (W)
Room Temperature (RT)				
95 (n = 10)	94 ± 3	89 ± 5	88 ± 9	93 ± 6
130 (n = 10)	128 ± 3	122 ± 5	123 ± 7	126 ± 7
165 (n = 10)	163 ± 6	154 ± 12	157 ± 9	163 ± 12
200 (n = 10)	199 ± 2	195 ± 9	194 ± 8	201 ± 8
235 (n = 8)	233 ± 8	231 ± 14	229 ± 11	236 ± 7
270 (n = 5)	255 ± 19	253 ± 19	259 ± 17	259 ± 25
305 (n = 2)	302 ± 2	307 ± 3	305 ± 3	311 ± 4
Cold				
95 (n = 10)	94 ± 5	93 ± 10	91 ± 11	102 ± 14
130 (n = 10)	128 ± 5	123 ± 12	125 ± 10	133 ± 12
165 (n = 10)	162 ± 8	159 ± 14	162 ± 10	168 ± 16
200 (n = 10)	195 ± 10	195 ± 14	194 ± 15	205 ± 20
235 (n = 8)	223 ± 12	228 ± 10	229 ± 14	235 ± 21
270 (n = 5)	249 ± 24	250 ± 22	260 ± 28	258 ± 44
305 (n = 2)	301 ± 4	307 ± 2	313 ± 6	326 ± 4
Hot				
95 (n = 10)	90 ± 12	75 ± 15	77 ± 12	83 ± 16
130 (n = 10)	126 ± 10	108 ± 13	113 ± 9	119 ± 12
165 (n = 10)	162 ± 7	145 ± 12	152 ± 6	156 ± 9
200 (n = 10)	199 ± 3	183 ± 13	183 ± 10	194 ± 9
235 (n = 8)	232 ± 2	214 ± 18	219 ± 12	229 ± 9
270 (n = 5)	249 ± 28	242 ± 22	247 ± 27	248 ± 25
305 (n = 2)	297 ± 7	290 ± 11	290 ± 11	297 ± 14

Data are mean ± standard deviation.

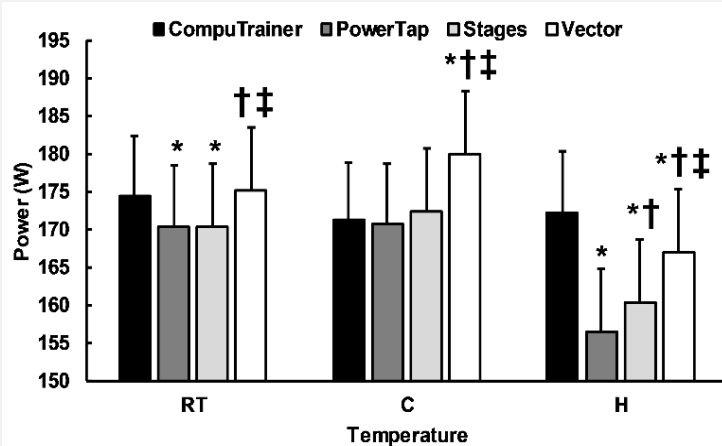


Figure 1. Comparison of average power output of each power meter within a given temperature. Data are mean \pm SE. * $p < 0.05$ from CompuTrainer; † $p < 0.05$ from PowerTap; ‡ $p < 0.05$ from Stages. RT = Room Temperature, C = Cold, H = Hot.

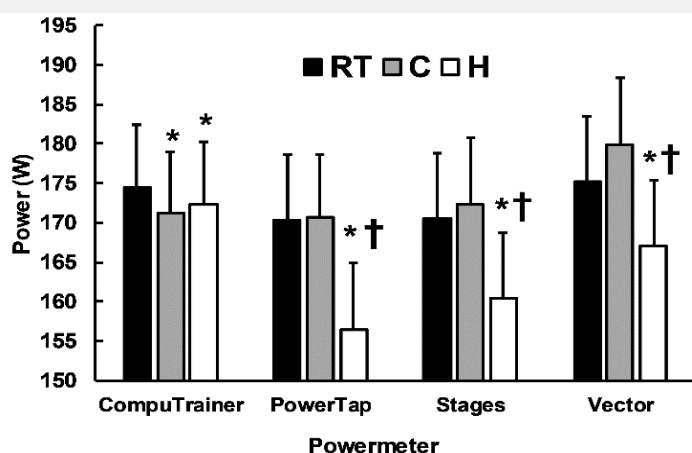


Figure 2. Comparison of the average power output during a trial with every power meters at each temperature. Data are mean \pm SE. * $p < 0.05$ from RT, † $p < 0.05$ from C. RT = 20°C Room Temperature, C = 7°C Cold, H = 33°C Hot. Mean CV of the CompuTrainer, PowerTap, Stages, and Vector between all three trials were 2.31%, 8.15%, 6.28%, and 7.73%, respectively.

An interaction occurred between temperatures and power meters ($p < 0.001$, $\eta^2 = 0.205$). The average power output for the PowerTap and Stages power meters were lower in the room temperature trial compared to the CompuTrainer ($p < 0.001$, Cohen's $d = 0.071$; $p = 0.001$, Cohen's $d = 0.069$; respectively) and compared to the Vector ($p < 0.001$, Cohen's $d = 0.083$; $p = 0.002$, Cohen's $d = 0.081$; respectively). The CompuTrainer and Vector were not different ($p = 0.425$, Cohen's $d = 0.014$) and the PowerTap and Stages were not different ($p = 0.957$, Cohen's $d = 0.001$) in the room temperature trial. In the cold temperature trial, the Vector power output was higher than the CompuTrainer ($p < 0.001$, Cohen's $d = 0.151$), PowerTap ($p < 0.001$, Cohen's $d = 0.158$), and Stages ($p = 0.003$, Cohen's $d = 0.126$). The CompuTrainer was not different from the PowerTap and Stages ($p = 0.640$, Cohen's $d = 0.011$; $p = 0.510$, Cohen's $d = 0.020$) and the PowerTap and Stages were not different ($p = 0.281$, Cohen's $d = 0.030$) in the cold trial. The CompuTrainer power output was higher than the PowerTap ($p < 0.001$, Cohen's $d = 0.269$), Stages ($p < 0.001$, Cohen's $d = 0.204$), and Vector ($p < 0.001$, Cohen's $d = 0.090$) in the hot temperature trial. The

Stages and Vector power meters were higher ($p = 0.015$, Cohen's $d = 0.065$; $p < 0.001$, Cohen's $d = 0.176$ respectively) than the PowerTap in the heat. The Vector power output was also higher than the Stages ($p < 0.001$, Cohen's $d = 0.112$) in the heat (Figure 1).

The CompuTrainer had the highest power output in the room temperature trial compared to the cold ($p = 0.006$, Cohen's $d = 0.057$) and hot ($p = 0.047$, Cohen's $d = 0.038$) trials. The cold and hot trials were not different on the CompuTrainer ($p = 0.374$, Cohen's $d = 0.019$). The PowerTap power output was lower in the hot trial compared to the room temperature ($p < 0.001$, Cohen's $d = 0.234$) and the cold trial ($p < 0.001$, Cohen's $d = 0.240$) but not between the room temperature and cold trial ($p = 0.875$, Cohen's $d = 0.005$). The Stages power output was lower in the hot trial compared to the room temperature ($p < 0.001$, Cohen's $d = 0.169$) and cold trial ($p < 0.001$, Cohen's $d = 0.202$) but not between the room temperature and cold trial ($p = 0.234$, Cohen's $d = 0.032$). The Vector was also lowest in hot trial compared to the room temperature ($p < 0.001$, Cohen's $d = 0.134$) and cold trials ($p < 0.001$, Cohen's $d = 0.210$) but not between the room temperature and cold trial ($p = 0.067$, Cohen's $d = 0.120$; Figure 2).

Agreement

Bland-Altman analysis was performed for equivalence testing between power meters (Figure 3a-c). Heteroscedasticity was not found among the power meters so the data were not log transformed. In the room temperature trial, the CompuTrainer power output was 1.8 ± 4.0 W higher than the other power meter outputs and the limits of agreement ranged from -6.1 to 9.8 W. The PowerTap power output was -2.3 ± 5.2 W lower than the other power meters and the limits of agreement ranged from -12.5 to 8.0 W. The Stages power meter output was -2.2 ± 6.3 W lower than the other power meters and limits of agreement -14.5 to 10.1 W. The Vector bias was 2.6 ± 5.9 W higher than the average of the other power meters and limits of agreement ranged from -9.0, 14.2 W (Figure 3a). In the cold temperature trial, the bias of the CompuTrainer was -2.3 ± 6.6 W and limits of agreement of -15.2 to 10.5 W. The PowerTap bias was -2.9 ± 6.5 W and limits of agreement were from -15.6 to 9.8 W; Stages bias was -1.2 ± 9.2 W and limits of agreement were from -19.2 to 16.9 W; and Vector bias was 6.4 ± 10.7 W and limits of agreement were from -14.6, 27.3 W. (Figure 3b). In the hot temperature trial, the bias of the CompuTrainer was 8.2 ± 5.7 W and limits of agreement

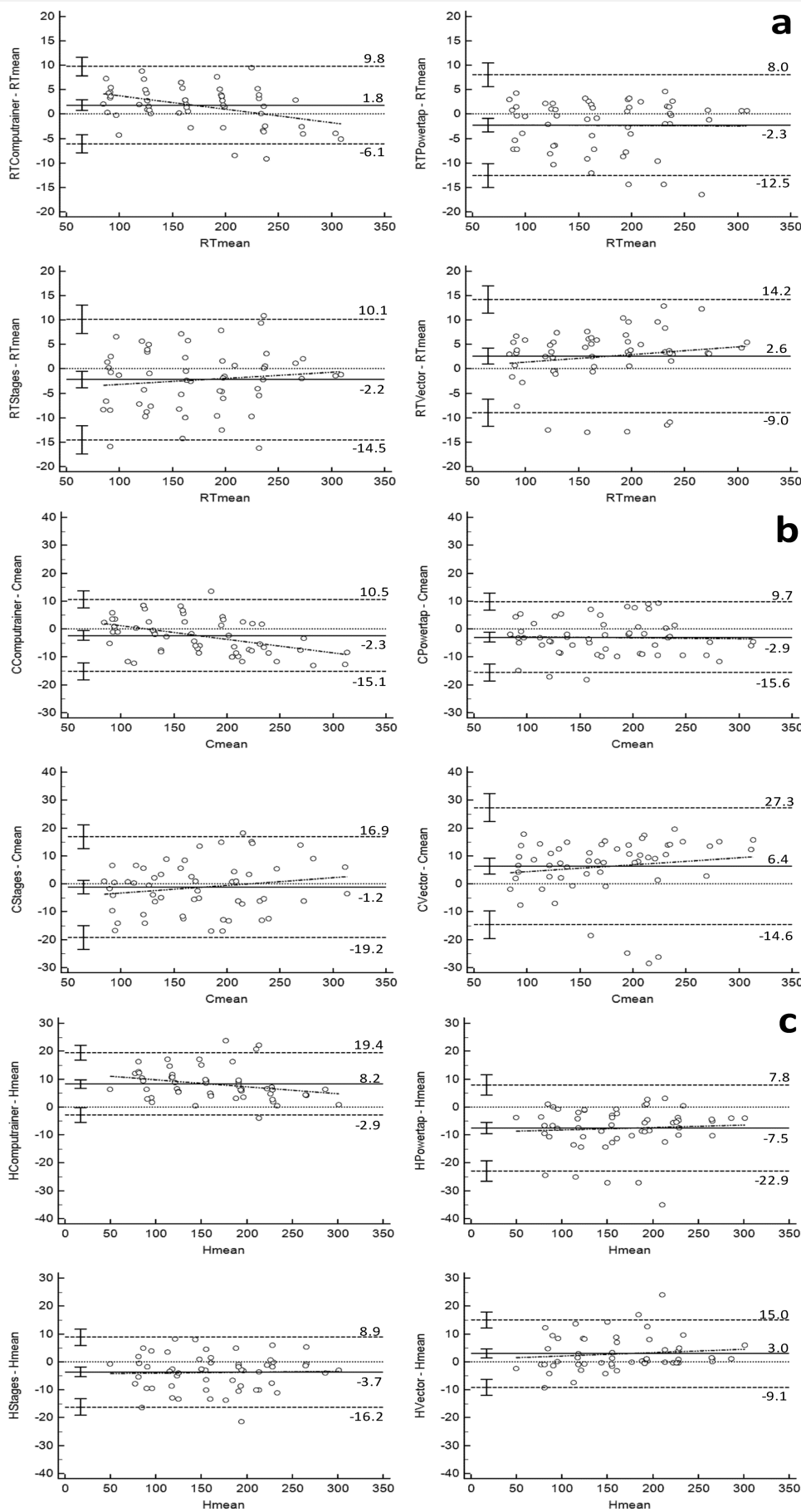


Figure 3a-c. Bland-Altman plots of difference between power outputs of an individual power meter versus the collective mean in a - room temperature trial, b - cold temperature trial, c - hot temperature trial.

were from -2.9 to 19.4 W. The PowerTap bias was -7.5 ± 7.8 W and limits of agreement were from -22.9 to 7.8 W; Stages bias was -3.7 ± 6.4 W and limits of agreement were from -16.2 to 8.9 W. The Vector bias was 3.0 ± 6.2 W and limits of agreement were from -9.1 to 15.0 W. (Figure 3c).

Discussion

It is important to consider the differing conditions that one will cycle in throughout a training period. The main finding of this study was that the power output of the power meters was affected by temperature. Specifically, power output was lower when cycling in the hot trial compared to the cold and room temperature trials. Another finding was that agreement among the power meters differ depending on the environmental temperature.

CompuTrainer

The CompuTrainer power meter allows for training indoors and the accompanying software is capable of configuring differing courses that may include differing terrains. It is important that the CompuTrainer display accurate power output to translate into cycling outside of the laboratory. In the room temperature and cold trial, the

CompuTrainer was in agreement with all other power meters. Conversely, in the hot trial, the CompuTrainer power output was higher and did not agree with the collective average power output of the other power meters. Additionally, the CompuTrainer was significantly higher in the hot trial compared to all other power meters. Since the workload was set by CompuTrainer software, this may indicate that the CompuTrainer workload was not reading accurately and that the actual workload was lower on the CompuTrainer leading all other power meters to read lower.

The CompuTrainer has been used for cycling performance tests in a laboratory setting (Clark, Gartner, Williams, & Pettitt, 2016; Peveler, 2013; Sparks et al., 2016). However, some have questioned the test-retest reliability with given cardiorespiratory variables in the CompuTrainer compared to the Lode cycle ergometer (Earnest et al., 2005). One study validated the calibration of the CompuTrainer in a range of temperatures from 15°C to 38°C and determined that the warm-up period is critical for proper calibration (Davison et al., 2009). They recommend an alternative warm-up consisting of repeated 2 min bouts of 200 W be used for calibration to enhance reliability and validity rather than that of the manufacturer's recommendation. Additionally, they conclude that at higher temperatures, the CompuTrainer was not different from the criterion measure but at lower temperatures of 15°C and 20°C, the CompuTrainer underestimated power compared to a SRM power meter. This was only regarding power reading during the calibration period and not for prolonged usage for training or performance tests. In the current study, careful preparation of calibration procedures and tire pressure was taken before each trial. Discrepancy in the CompuTrainer may be due to an increase in heat on the electromagnetic brake, which may affect the performance (Davison et al., 2009). Since the CompuTrainer is an indoor-based trainer, it may not have practical use for varying degrees of temperature. Outdoor use may be limited to warm-up and cool-down before and after a competition, respectively.

PowerTap

The PowerTap power meter is a hub-based power meter located in the rear wheel of the bicycle. The estimated accuracy of the PowerTap is around 2-3% and it has been reported to be a valid and reliable measure of power output in laboratory settings (Bertucci et al., 2005; Bouillod et al., 2017; Duc, Villerius, Bertucci, & Grappe, 2007; Gardner et al., 2004; Miller et al., 2015). The PowerTap was in agreement with the collective average of the other power meters in the room temperature and cold conditions but not in the hot trial. Perturbations exist with the PowerTap in cold temperatures of 8°C producing dissimilar average error compared to room temperature (Gardner et al., 2004). At 5°C, the range of power output is slightly underestimated, although not significantly (Bertucci et al., 2005). Similarly, the current study demonstrates that the PowerTap on average had a lower power output when compared to the all others collectively as

evidenced by the bias in the Bland-Altman plots and ANOVA. It should be noted that this was not significantly different from the Stages power meter. In the current study, the PowerTap was not different in the room temperature and cold condition but was lower in the hot condition. Additionally, the PowerTap power output was similar to the Stages in the room temperature trial and the Stages and CompuTrainer in the cold but different from all other power meters in the hot condition. More research is needed to examine the reason that the PowerTap power output reads lower in the hot condition.

Stages

The Stages power meter is a crank-based power meter with the strain gauges located in the left crank arm. The accuracy of the Stages is reported by the manufacturer to be around 2%. The Stages power meter was in agreement with the collective average of power meters in the room temperature, cold, and hot trials. The Stages power meter was lowest in the hot condition. The zero offset during calibration should account for temperature variations (Hurst et al., 2015), but this was not the case in the current study. During smooth steady state riding the Stages power recordings are similar to that of a Quarq and PowerTap power meter (Miller et al., 2015). However, the reliability of the Stages power meter has been questioned with off-road cycling (Hurst et al., 2015; Miller et al., 2015) and labeled as invalid during incremental cycling (Bouillod et al., 2017). Additionally, the Stages may underestimate both average and peak power when compared to the SRM (Bouillod et al., 2017; Hurst et al., 2015). In the current study, the Stages power output was lower than the Vector throughout each trial and higher than the PowerTap, although only significantly in the hot trial.

Vector

The Vector power meter is a relatively new pedal-based power meter with a reported accuracy of around 2%. A unique feature of the Vector power meter is that force is measured independently in each pedal with eight strain gauges within each pedal. In the current study, the Vector was not in agreement with the other power meters in the room temperature or cold trial, but was in agreement with the collective average in the hot trial. The Vector power output tended to read higher in each trial than the other power meters with the exception of the CompuTrainer in the room temperature trial. Like the location of the Stages and PowerTap, this may be due to a more direct transfer of force from the foot of the individual to the pedal with minimal power transfer lost (Novak & Dascombe, 2016). Although, underestimation of the Vector power meter compared to the SRM has been observed (Bouillod et al., 2017; Hutchison et al., 2017), overestimation of power is expected to occur and has been reported (Novak & Dascombe, 2016). More discrepancy exists with the validity and reliability of the Vector power meter compared to the other power meters tested (Bouillod et al., 2017; Hutchison et al., 2017; Novak & Dascombe, 2016).

Conclusion

The power meters used in the current study display a varying degree of agreement and temperature appears to affect the power output with the power meters. Specifically, the power meters displayed a lower power output in the hot trials compared to the room temperature trials and three of the four were lower in the hot trials compared to the cold trials. This may suggest that the power meters do not display reproducibility across different temperatures or that the CompuTrainer set workload is affected in the heat. This may be of concern when prescribing training intensities based off power for cyclists when training outdoors. Like previous studies, the zero offset of the devices was performed after a warm-up period lasting approximately 10 min in all trials (Gardner et al., 2004). The strain gauges are the mechanisms responsible for detecting force and may be displaying different properties in the heat. In the cold, the 10-minute warm-up period may be increasing the temperature of the power meter through mechanical heat buildup before zero offset to allow the internal temperature of the devices to be similar to that of room temperature. Future studies should further examine the accuracy and validity of these power meters in a variety of temperatures. Determination of the source for the differences in power output in the heat is also warranted. Limitations of this study include the use of 3 minute stages, which may have caused earlier fatigue in the participants inhibiting their ability to reach higher workloads. Shorter time periods in each stage or higher increases in power output may be warranted in future studies to represent more fit and elite level cyclists. Cadence was not controlled in the current study. Participants were instructed to maintain a cadence over 60 rpm. As a result, cadence in the hot trial was slower than cadence in the room temperature trial. Future studies should tightly regulate the cadence in order to eliminate a source of measurement error. Further work is also required to ensure the inaccuracy of the CompuTrainer in warmer environmental conditions. Higher statistical power may also be required to ensure consistent results in a similar methods comparison study.

Practical application

Consistent training and progression within an exercise program can be facilitated by using reliable and accurate exercise equipment. Monitoring power output during cycling provides an objective method for training intensity. However, little is known about how power meters perform across a variety of temperatures. The present study aimed to assess the use of four commercially available power meters across a range of power outputs in hot (33 °C) and cold (7 °C) environments when compared to a room temperature environment (20 °C). This was accomplished through simultaneous use of the four devices (Garmin Vector, Powertap G3, Stages 7800 crankset, and Racermate CompuTrainer).

The main finding from this study is that the power meters display less agreement when hot

environmental temperatures despite zero-offset calibration occurring in this temperature. This may have been due to an error in power output from the CompuTrainer or from error in the other power meters. Specifically, the CompuTrainer may not be suitable for outdoor use when the weather is hot due to inaccuracy of the set workload in that temperature. Conversely, cyclist may need to interpret power output of the Vector, PowerTap, and Stages with caution when cycling outdoors in the hot. Examination of the reason for these differences may be of value and selection of power meters that are capable of consistent readings across temperatures may be of use to cyclists in changing climates.

Conflicts of Interest

The authors declare no conflicts of interest with this study.

Funding

The authors report no funding for this project.

Acknowledgments

We would like to thank Rebecca Cuthbert, Caleb Ross, Cory Poehling, Jayson Henrickson, and Morgan Busboom for their contributions in data collection and processing.

References

1. Achten, J., & Jeukendrup, A. E. (2003). Heart rate monitoring. *Sports Medicine*, 33(7), 517-538.
2. Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26(4), 217-238.
3. Bartlett, J., & Frost, C. (2008). Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 31(4), 466-475.
4. Bertucci, W., Duc, S., Villerius, V., Pernin, J., & Grappe, F. (2005). Validity and reliability of the PowerTap mobile cycling powermeter when compared with the SRM device. *International Journal of Sports Medicine*, 26(10), 868-873.
5. Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: Why plotting difference against standard method is misleading. *The Lancet*, 346(8982), 1085-1087.
6. 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. *International Journal of Sports Physiology and Performance*, 12(8), 1023-1030.
7. Clark, I. E., Gartner, H. E., Williams, J. L., & Pettitt, R. W. (2015). Validity of the 3-min all-out exercise test on the CompuTrainer. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, doi:10.1519/JSC.0000000000001169 [doi]
8. Clark, I. E., Gartner, H. E., Williams, J. L., & Pettitt, R. W. (2016). Validity of the 3-minute all-out exercise test on the CompuTrainer. *Journal of Strength and Conditioning Research*, 30(3), 825-829. doi:10.1519/JSC.0000000000001169 [doi]

9. Davison, R. R., Corbett, J., & Ansley, L. (2009). Influence of temperature and protocol on the calibration of the computrainer electromagnetically-braked cycling ergometer. *International SportMed Journal*, 10(2), 66-76.
10. Duc, S., Villerius, V., Bertucci, W., & Grappe, F. (2007). Validity and reproducibility of the ergomo® pro power meter compared with the SRM and powertap power meters. *International Journal of Sports Physiology and Performance*, 2(3), 270.
11. Earnest, C. P., Wharton, R. P., Church, T. S., & Lucia, A. (2005). Reliability of the lode excalibur sport ergometer and applicability to computrainer electromagnetically braked cycling training device. *Journal of Strength and Conditioning Research*, 19(2), 344-348. doi:R-15714 [pii]
12. Gardner, A. S., Stephens, S., Martin, D. T., Lawton, E., Lee, H., & Jenkins, D. (2004). Accuracy of SRM and power tap power monitoring systems for bicycling. *Medicine and Science in Sports and Exercise*, 36(7), 1252-1258.
13. Hurst, H. T., 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(1), 21.
14. Hutchison, R., Klaphor, G., Edwards, K., Bruneau, K., Mocko, G., & Vahidi, A. (2017). Validity and reproducibility of the garmin vector power meter when compared to the SRM device. *Journal of Sports Science*, 5, 235-241.
15. Jeukendrup, A., & Diemen, A. V. (1998). Heart rate monitoring during training and competition in cyclists. *Journal of Sports Sciences*, 16(sup1), 91-99.
16. Miller, M. C., Macdermid, P. W., Fink, P. W., & Stannard, S. R. (2015). Agreement between powertap, quarq and stages power meters for cross-country mountain biking. *Sports Technology*, 8(1-2), 44-50.
17. Novak, A. R., & Dascombe, B. J. (2016). Agreement of power measures between garmin vector and SRM cycle power meters. *Measurement in Physical Education and Exercise Science*, 20(3), 167-172.
18. Peveler, W. W. (2013). The accuracy of simulated indoor time trials utilizing a CompuTrainer and GPS data. *The Journal of Strength & Conditioning Research*, 27(10), 2823-2827.
19. Quanjer, P. H., Tammeling, G. J., Cotes, J. E., Pedersen, O. F., Peslin, R., & Yernault, J. C. (1993). Lung volumes and forced ventilatory flows. *The European Respiratory Journal*, 6 Suppl 16, 5-40. doi:10.1183/09041950.005s1693 [doi]
20. Siri, W. E. (1961). Body composition from fluid spaces and density: Analysis of methods. *Techniques for Measuring Body Composition*, 61, 223-244.
21. Sparks, A., Williams, E., Massey, H., Bridge, C., Marchant, D., & Mc Naughton, L. (2016). Test-retest reliability of a 16.1 km time trial in trained cyclists using the CompuTrainer ergometer. *Journal of Science and Cycling*, 5(3), 35-41.
22. Swart, J., Lamberts, R. P., Derman, W., & Lambert, M. I. (2009). Effects of high-intensity training by heart rate or power in well-trained cyclists. *The Journal of Strength & Conditioning Research*, 23(2), 619-625.