Journal of Electronic Gaming and Esports, (Ahead of Print)
https://doi.org/10.1123/jege.2022-0006
First Published Online: Sept. 22, 2022

Human Kinetics
RESEARCH PAPER

Competitive Racing in Virtual Cycling—Is It Possible, Realistic, and Fair?

Jonas Bjärehed[1] **and Marlene Bjärehed**[2]

[1]The Network of Sport and Exercise Psychology, Department of Psychology, Lund University, Lund, Sweden; [2]Faculty of Education, Kristianstad University, Kristianstad, Sweden

Competitive racing through virtual cycling has established itself as an entirely new discipline within cycling. This study explores what equipment racers use and examines important power metrics for racing. Data were collected from three different races from the current ranking of the most highly regulated and professionally organized race series on the virtual cycling platform Zwift. Power output data from 116 race participants, over five power durations (5 s–20 min), and two separate power measuring sources were collected and analyzed using the Bland–Altman method. The findings indicate that the physiological efforts of these races are comparable to those found in traditional competitive cycling. Furthermore, findings also support that the equipment typically used produces similar power outputs with good agreement between different power meters for most measurement points. Finally, the implications of these results for the status of virtual racing are discussed.

Keywords*:* performance data, virtual sport, virtual training, e-health, simulation, turbo trainer

Virtual cycling is a new and rapidly growing form of cycling with profound implications for the future of the sport, for both recreational and competitive riders (for reviews, see McIlroy et al., 2021; Westmattelmann et al., 2021). In virtual cycling, a stationary bike capable of interacting with computer software is used to mimic outdoor cycling (i.e., creating resistance representing the course topography, considering wind and rolling resistance, simulating realistic draft effects). The rider provides input (typically the power output generated by pedaling), and the software uses gamification elements and mixed-reality technology (Westmattelmann et al., 2021) that contribute to immersive experiences. These include realistic sound, graphics, and even tactile effects (such as the bike shaking when going over uneven surfaces or elevating the front when going uphill). The simulation is typically run on an ordinary computer, laptop, tablet, or smartphone, in the rider's own home.

There are currently several large commercial platforms available for virtual cycling. These can be differentiated from traditional digital training software and platforms also commonly used in cycling by their focus on social and interactive aspects (such as riding and racing with other users). There are no authoritative estimates of the number of people involved in virtual cycling today, but the numbers are likely in the millions worldwide. For example, the largest platform (Zwift), established as late as 2015, reported 550,000 users in 2018 (Bonnington, 2018) and 2.5 million in 2020 (Long, 2020), after a massive influx of capital and user growth. The COVID-19 pandemic has likely accelerated this development; for example, in 2020, the platform broke a new record with 35,000 simultaneous users online (Schlange, 2020). The pandemic has also forced the adaptation of many seminal cycling events (including paracycling and triathlon events) to online versions, from local to national and international championships.[1] The legendary Tour de France has been raced online (Herman, 2020), and virtual cycling has been featured as part of the Olympic Virtual Series by the International Olympic Committee (IOC, 2021). In a short time, athletes, organizers, coaches, sponsors, and fans have been catapulted into this new digital reality.

Despite the massive growth of virtual cycling and its rapid adaptation for racing purposes, almost no research has explored the numerous questions surrounding this development. Recently, however, there have been a few important attempts to describe and outline virtual cycling as a phenomenon, including its potential and implications, and suggestions for further research (McIlroy et al., 2021; Westmattelmann et al., 2021).

One crucial line for inquiry is the status of competitive racing through virtual cycling. Recently, cheating and doping in virtual racing have begun to be explored in the scientific literature, including recommendations on how the sport should manage these issues going forward (Richardson et al., 2022). However, to the best of our knowledge, only two studies have examined aspects of virtual racing, focusing on competitive cyclists who transitioned to virtual racing due to the COVID-19 restrictions (Westmattelmann et al., 2021, 2022). Both Westmattelmann et al. (2021) and Mcllroy et al. (2021) emphasize the paramount importance of this area, especially the accuracy and reliability of virtual racing, to gain a wider acceptance of and credibility for this form of cycling. Virtual racing is judged to be realistic (even by procyclists) and has been suggested to demand similar physical performance as traditional racing (Westmattelmann et al., 2021). The notable difference between virtual cycling and traditional road cycling is that the former is typically significantly shorter in duration (generally lasting about 40–60 min). Thus, virtual cycling can be compared to time-trial competitions, track cycling, and other shorter duration forms of cycling. While virtual cycling

Jonas Bjärehed https://orcid.org/0000-0002-2128-6031
Marlene Bjärehed (marlene.bjarehed@hkr.se) is corresponding author, https://orcid.org/0000-0001-5753-1310

in general leaves some room for casual users to misrepresent or manipulate their actual performance through the game (e.g., by changing their weight, length, gender, or by manipulating their equipment), the virtual racing community has developed several practices that address issues related to fairness and equality in racing. For instance, guidelines regulating the types of equipment used for racing, its reliability, and methods for validating the power generated in the game (such as using secondary power source) have been created (Zwift Cycling Esports Ruleset; Zwift, 2021). These also include procedures for ensuring the identity of the rider, the accuracy of the rider's ability, and performance characteristics. There are protocols for verifying riders' height and weight, collecting and verifying performance data from virtual as well as traditional cycling, and even performing verifications or racing while simultaneously live-streaming through video recording. However, these procedures are not familiar to most riders. Even if they are known, most riders lack any deeper insights into the reasoning behind the rules and how they can inform the riding practices for more casual cycling.

Purpose of the Present Study

The present study examines the generated power outputs and the equipment used in virtual cycling races by looking at naturalistic data from official race results. This approach is relevant for several reasons and results in two main aims for the study. First, we want to provide a general overview of the physiological and technical demands associated with this kind of cycling. For both riders, trainers, and coaches unfamiliar with virtual racing, this information can help contextualize virtual races in relation to cycling more generally, for example, if virtual racing can be incorporated into a traditional training regime. It also provides some references and insights for benchmarking ability that can guide riders in assessing their performance (e.g., choosing what racing category might be the most appropriate) and their equipment (e.g., guide in what equipment to get and how to assess whether it produces realistic power outputs). Furthermore, it can also guide training and even racing efforts, as it illustrates the critical power outputs and their durations for typical virtual racing.

Second, we want to look at the agreement between primary and secondary power meters in these data. This is relevant as one important method to assess the reliability of the equipment used in virtual cycling. However, this type of analysis can also give much more nuanced information. It can begin to construct a picture of potential systematic and variable issues with importance for the validity of the power outputs. To give some examples, there could be systematic differences between different types of power meter designs (e.g., a trainer vs. a pedal-based power meter) or even between manufacturers and models (e.g., highly accurate vs. inaccurate power meters). Furthermore, there could be variable issues that depend on the intensity and duration of the effort (e.g., power readings could be more unreliable for high vs. low outputs and long vs. short durations) or rider characteristics (such as using high cadence vs. low cadence). Any such systematic or variable differences (and the interactions of these) could be directly relevant in influencing racing outcomes.

Method

Design, Sample, and Statistical Methods

This study used data from actual racing events. Data were extracted from www.zwiftpower.com where official results and power outputs from all virtual racing events on the Zwift platform are recorded. On the Zwiftpower site, riders can submit their power output files for additional power meters. For this study, power data from three different races in the weekly men's Premier Division in the Zwift Racing League series (which is currently the highest echelon of virtual racing available) were examined. Data were collected for those riders where both primary and secondary sources for power data were available.[2] The three races took place in January of 2022, and the data comprise 76 individual riders,[3] totaling 117 observations. The races were regarded as independent observations in the analyses. In addition to power data, self-reported information on smart trainer and power meter brands and models was collected from Zwiftpower for each race. The weights of all riders ($n = 201$) in these races varied between 53.3 and 86.9 kg ($M = 70.3$, $SD = 5.6$). Finishing times in the races were typically around 40 min.

In order to examine the agreement between primary power sources (which is always the smart trainer for these races, as regulated by the Zwift e-racing ruleset; Zwift, 2021) and secondary power sources (which is a crank- or pedal-based power meter), the power output for each rider's best 5- and 15-s duration, and 1-, 5- and 20-min durations were compared from the races. This was done by calculating the mean values for each pair of measurements, the mean differences, and constructing Brand–Altman plots (Bland & Altman, 1986) for each measurement interval. The plots were then used to assess the distribution of the measurements based on 95% confidence intervals. Finally, to assess the existence of potential (proportional) bias for either of the measurement types, standard linear regressions analyses were performed using mean differences of measures as independent variables and the differences as independent variables. All analyses were conducted with the JAMOVI (version 2.2.5) statistical package (www.jamovi.org).

Results

Equipment Used and Their Power Outputs

Table 1 shows the mean values for each measurement type and duration, the combined mean for both measures, the mean difference, and the associated *SDs*. Table 2 shows an overview of the kinds of equipment used, divided by type (smart-trainer, smart bike, pedal-based power meter, or crank-based power meter).

For all durations, the secondary power source showed, on average, a slightly higher power than the primary source. Mean differences were: 3.1 W ($SD = 24.1$) for 5 s (913.0 vs. 916.1 W); 6.6 W ($SD = 21.3$) for 15 s (845.8 vs. 852.4 W); 4.99 W ($SD = 12.5$) for 1 min (570.2 vs. 575.0 W); 3.2 W ($SD = 7.8$) for 5 min (393.5 vs. 396.6); and 2.1 W ($SD = 7.2$) for 20 min (341.6 vs. 339.5 W).

Agreement Between Power Outputs

The Bland–Altman plots can be seen in Figure 1(a–e). They show the power range in watt (*x*-axis) and the difference between measures in watt (*y*-axis). Dashed lines in each plot represent the mean difference for the values (i.e., the constant bias) and the limits of agreement for the plot. The highlighted areas around these lines represent the 95% confidence intervals. Finally, the plots show regression fit (i.e., the proportional bias) of the differences of the means as a solid line, with its confidence intervals (hour glass shaped). The corresponding exact values can also be seen in Table 3, that is, the bias estimate, upper and lower limits of agreement (for 95% confidence interval) for the different durations.

In summary, the distributions in the plots indicate a reasonable agreement between primary and secondary power sources, as the

Table 1 Mean, *SD*, Minimum and Maximum Power Values (in Watt) for the Different Durations for Primary (Smart Bike or Smart Trainer), Secondary (Power Meter) Power Sources, and the Difference Between Primary and Secondary Source (*N* = 117)

		Output (W)			
Measure	Duration	*M*	*SD*	Minimum	Maximum
Primary source	5 s	912.98	171.94	569.0	1338.0
	15 s	845.84	160.86	516.0	1187.0
	1 min	570.16	75.00	428.0	743.0
	5 min	393.47	35.07	308.0	480.0
	20 min	339.54	27.31	254.0	402.0
Secondary source	5 s	916.05	174.44	580.0	1362.0
	15 s	852.39	160.42	534.0	1175.0
	1 min	575.04	76.10	428.0	745.0
	5 min	396.65	35.38	305.0	473.0
	20 min	341.63	27.78	254.0	412.0
Primary–secondary	5 s	−3.07	24.09	−68.0	44.0
source difference	15 s	−6.55	21.32	−74.0	44.0
	1 min	−4.89	12.49	−39.0	29.0
	5 min	−3.17	7.99	−27.0	16.0
	20 min	−2.09	7.17	−21.0	15.0

Table 2 Numbers and Percentages for the Different Equipment Used for Racing (*N* = 116)

	Number	% of total
Primary power sources		
Wahoo Kickr[a]	57	49.5
Wahoo Kickr Core[a]	5	4.3
Elite Direto XR[a]	15	13.0
Saris Hammer[a]	2	1.7
Saris H3[a]	12	10.5
Tacx Neo/Neo 2T[a]	11	9.6
Wahoo Kickr Bike[b]	8	7.0
Wattbike Atom[b]	2	1.7
Tacx neo smart bike[b]	2	1.7
Unknown	1	0.9
Secondary power sources		
Quarq[c]	31	26.7
P2M[c]	3	2.6
4iiii[c]	9	7.8
Stages[c]	2	1.7
FSA[c]	2	1.7
Power2Max[c]	3	2.6
Garmin[d]	7	6.0
Favero[d]	36	31.0
Power PRO[c]	1	0.9
Pioneer[c]	2	1.7
Rotor[c]	5	4.3
SRAM[c]	7	6.0
Verve[c]	8	6.9

Note.[a]Smart trainer type. [b]Smart bike type. [c]Crank-based type. [d]Pedal-based type.

vast majority of observed differences were within the estimated limits of agreement. Furthermore, together with the results presented in Table 1, they indicate what power values might be expected in different racing situations and how these can be expected to vary due to random error. However, a small number (a total of 11) of observations fall outside these limits, distributed as two values for the 5-s plot, two values for the 15-s plot, three values for the 1-min plot, two values for the 5-min plot, and two values for the 20-min plot. These should be considered values with unexpectedly large differences, possibly indicating problematic agreement between the measures.

A further 45 observations fell within the confidence intervals of the limits of agreement (10 values for the 5-s plot, nine values for the 15-s plot, six values for the 1-min plot, 12 values for the 5-min plot, and eight values for the 20-min plot). These could be considered measurements that might lay outside the limits of agreement, depending on where the true limit is found within these confidence intervals.

The plots were also examined for potential constant and proportional biases. There was no statistically significant constant bias for the 5-s plot, but a small statistically significant difference for the other durations(−6.55-W bias for 15 s, −4.89 W for 1 min, −3.17 W for 5 min, and −2.09 W for 20 min.) These effects can be seen in the plots as the confidence interval of the dashed lines does not cover zero. Hence, for 15-s to 20-min duration, there is a small but significant difference, suggesting that secondary power measures are higher than the primary sources. This difference translates to roughly between 0.5% and 1%.

Finally, a set of linear regression analyses was also conducted, one for each plot, to test for proportional biases (Table 4 shows the results for the overall models). These were nonsignificant for all five plots (*R*s ranging between .02 and .10, and all p values >.2), thus failing to confirm any proportional bias for the respective plots. A proportional bias in this context would represent a systematic over- or underreporting of one of the power sources that is dependent on the power range (e.g., if higher power values on the continuum—the right-hand side of the plot—were biased, as compared to the ones on the lower continuum—left-hand side—of the plot).

Discussion

To the best of our knowledge, this is the first attempt to study competitive racing in the new sport of virtual cycling. The lack of research in the area is puzzling, given the striking growth of popularity among users and its adoption by influential sporting organizations (such as the Union Cycliste Internationale [UCI] and the International Olympic Committee). Our best guess is that this is due to the novelty of the phenomenon and its rapid development. New technology-related aspects of this form of cycling require technical understanding, and familiarity with the rules and the racing context in order to formulate relevant research questions. Therefore, research in the area can be considered in the early stages of exploration, and formulating relevant research questions,

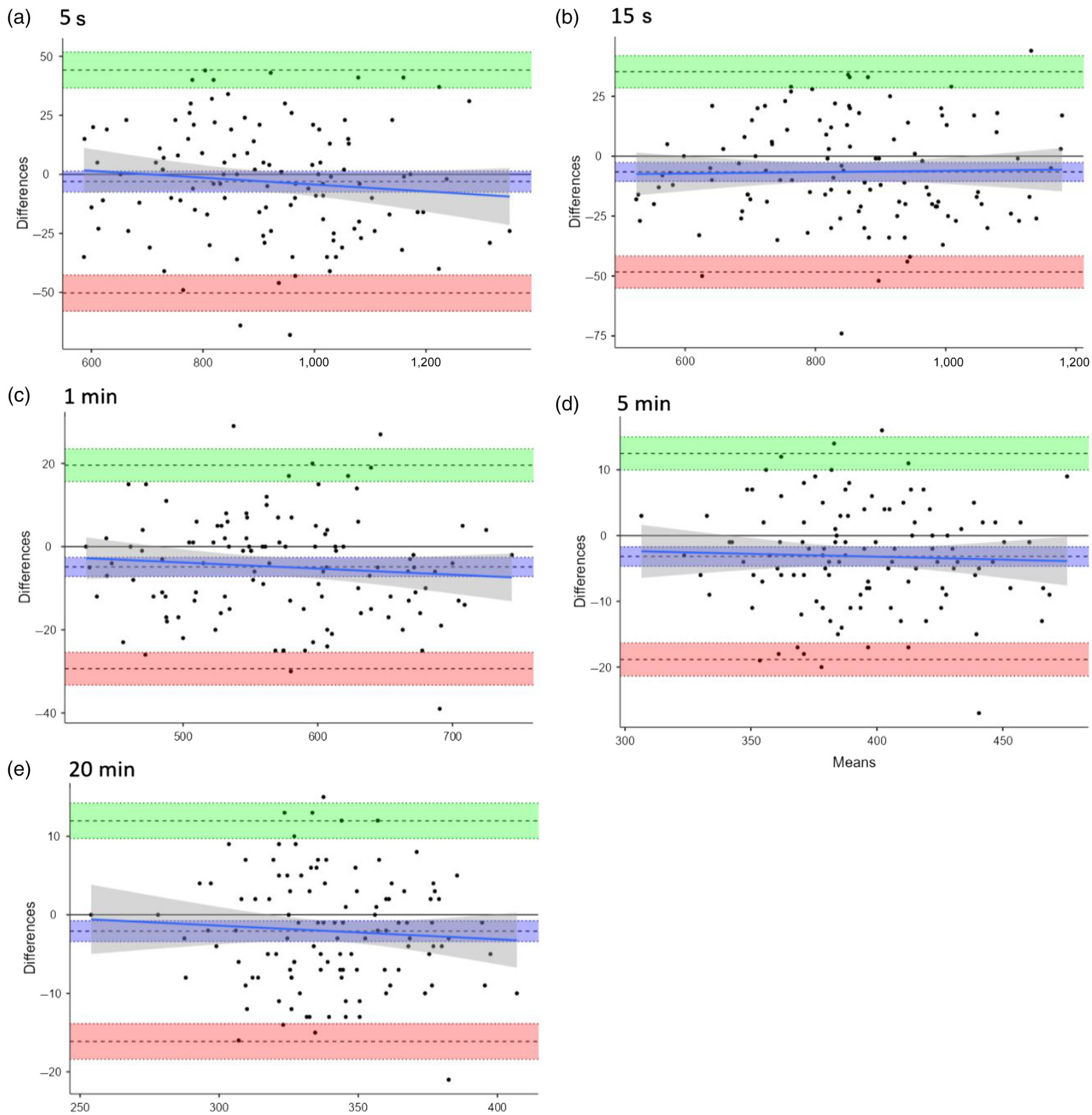

Figure 1 — Bland–Altman plots for agreement analysis ($n = 116$) of difference between primary versus secondary power meters' power output for the different durations (a–e). The bias and limits of agreement are shown as dashed lines with 95% confidence intervals highligted, and regression fit of the differences of the means is shown as a solid line with confidence interval (hour glass shaped). All values are measured in watt.

designing methodologies to answer them, and coordinating these efforts, will need further work. In that context, this study adds to the few recent attempts at broadly defining and describing the phenomena of virtual cycling (e.g., McIlroy et al., 2021; Richardson et al., 2022; Westmattelmann et al., 2021) by outlining some of the central issues related to virtual racing.

First, this study gives insights into the physical demands of this type of competition and allows an analyis of established frameworks of understanding performance compared with traditional cycling. Westmattelmann et al. (2021) have already proposed that virtual racing, from a physiological standpoint, can be compared with traditional cycling for similar durations (e.g., time

Table 3 Bland–Altman Bias Estimate, Upper and Lower Limits of Agreement (for 95% Confidence Interval) for the Compared Power Measures on Different Measurement Durations

Duration		Estimate	95% confidence interval Lower	Upper
5 s	Bias	–3.07	–7.50	1.36
	Lower limit of agreement	–50.28	–57.87	–42.68
	Upper limit of agreement	44.14	36.55	51.73
15 s	Bias	–6.55	–10.5	–2.63
	Lower limit of agreement	–48.34	–55.1	–41.62
	Upper limit of agreement	35.24	28.5	41.96
1 min	Bias	–4.89	–7.18	–2.59
	Lower limit of agreement	–29.36	–33.30	–25.43
	Upper limit of agreement	19.59	15.65	23.52
5 min	Bias	–3.17	–4.64	–1.70
	Lower limit of agreement	–18.84	–21.36	–16.32
	Upper limit of agreement	12.50	9.98	15.02
20 min	Bias	–2.09	–3.40	–0.768
	Lower limit of agreement	–16.13	–18.39	–13.872
	Upper limit of agreement	11.96	9.70	14.218

Table 4 Linear Regression Models for the 5-s, 15-s, 1-Min, 5-Min, and 20-Min Bland–Altman Plots Exploring Proportional Bias for Higher, Compared with Lower, Power Values on the Continuum

Model	*R*	*F*	*p*
5-s power plot	.10	1.25	.27
15-s power plot	.02	0.05	.82
1-min power plot	.09	0.89	.34
5-min power plot	.04	0.17	.68
20-min power plot	.07	0.50	.48

Note. Degrees of freedom = 1,114 for all *F* values.

trials) and for decisive sections in longer races (which might be a critical climb or the finishing sprint). Our findings can be compared with previously reported findings from the cycling literature to give a sense of the physical demands of this kind of racing. Virtual racing is typically decided in shorter critical sequences of the races. The top riders studied here produce power outputs comparable with the results from a study of successful road racing sprints by an international prorider and a U23 rider competing at the international level reported by Menaspa et al. (2013). In this study, average power was reported at about 1,100 W/15 s for the finishing sprint and 450 W/1 min, and 370 W/5 min for the time preceding that. In the current study, the 20-min power (i.e., 340 W for 20 min), which is likely a submaximal 20-min effort for most riders in these races, can be similarly compared. This performance is considerably higher than that reported by Karsten et al. (2021) for 20-min maximal power output in moderately trained cyclists (i.e., 262 ± 46 W) and in line with what MacInnis et al. (2019) found as maximal output in well-trained cyclists that also compete in cycling races (i.e., 335 W, *SD* = 35 W). The results can also be compared with a widely accepted, although rough, general taxonomy for the categorization of cycling performance that has been provided by Coggan (i.e., Allen et al., 2019), where performance (in watt per kilogram) is compared on a scale spanning from novice to world class. The performances found in this study would place the included riders around the power output profiles corresponding to excellent or exceptional riders (i.e., what would be expected from Category 1/domestic prolevel riders). This is expected given that the studied races are from the most competitive race series currently available in virtual cycling. A number of the included riders are recognized for competition on this level in traditional cycling. Hence, the current study supports the notion that virtual racing will put very taxing physiological demands on the rider and that the performances will be very highly correlated to riders' ability in traditional cycling and therefore should be viewed as a separate but legitimate discipline within the sport where physiological fitness is the deciding characteristic when racing. Further research could, however, explore in more detail how the demands of virtual racing specifically relate to those in traditional racing and how riders' performances typically translate between indoor and outdoor riding.

Furthermore, this study also gives an insight into the validity of the equipment and, specifically, the reliability of the power outputs it generates. The findings show what types and models of equipment are currently used in top-level racing, both the trainer/smart bikes and the secondary power meters (crank or pedal based), that are common. Today, there is a vast array of commercially available products mirrored by the many brands reported by riders here. It may be noted that the equipment used is predominantly from the higher price ranges of each brand. Although widespread among more recreational users, no low-price range models appear to be used in this setting. The manufacturers of trainers themselves attest that models in their lower price range should be expected to produce less reliable power outputs than more expensive ones. Exactly how (un)reliable different models may be in practice is unknown and would need further examination.

Pertaining to the validity of the equipment's power outputs, there does appear to be a reasonably high degree of agreement in the present study. Measuring power in cycling is recognized as complex (Bouillod et al., 2022), and differences in measures can always be expected due to random measurement errors. Previously, however, when power meters were used primarily to guide training and sometimes evaluate racing, the demands were different; beyond some basic accuracy (values attained are near the true value), repeatability and reproducibility were most crucial from a user perspective (that the measure is consistent over time, for similar conditions and over variations in conditions). The user could accept a degree of uncertainty in the true values as they were only used as points of reference for riders in relation to themselves. For training purposes, the durations and power ranges are also typically different from virtual racing, usually longer (e.g., riding a 4-min steady interval in training). The expectations and demands on equipment have changed with virtual racing; it now becomes much more crucial that measures are comparable between different riders and that the measures are accurate for short durations, respond quickly to changes, and manage the extreme ends of the power output spectrum. For the higher price range products reported in this study, manufacturers typically claim a 1%–2% accuracy. However, as Richardson et al. (2022) point out, there is no industry standard for determining these tolerances, and there is no independent testing to verify them. These will be essential aspects for manufacturers of equipment for virtual racing to consider when developing their products. Also, there might be opportunities associated with establishing collaborations with the virtual racing community and individual researchers when developing such products. Furthermore, as a recent review shows (Bouillod et al., 2017), the research literature that has evaluated different power meters shows varying results and uses methodologies that have limited generalizability to the context of virtual racing. Manufacturers of trainers and power meters should consider these new expectations of their products while also realizing that producing equipment that is not very accurate will directly hurt the credibility of the sport (and their brands).

Individual riders and coaches interested in better understanding the values produced by their equipment may use the data presented here for benchmarking purposes. For example, the higher limit for the confidence intervals for the upper limit of agreement values from the Bland–Altman plots could be proposed as a rough cutoff for acceptable differences in power values in competitive racing. Consequently, such criteria could also help racers interpret the meaning of differences in their equipment (e.g., when differences are likely insignificant). Race organizers could also use this reasoning as more objective criteria to decide whether a particular rider's equipment needs further review or to be disqualified. Such limits might, of course, be derived in several ways. However, to concretize from the reasoning above (i.e., setting a limit based on the 95% confidence intervals) and the present data, this would translate to cutoffs of about 5.7% difference for 5 s, 5.0% difference for 15 s, 4.1% difference for 1 min, 3.8% difference for 5 min, and 4.2% difference for 20-min power duration comparisons.

This study used a naturalistic approach to virtual racing in order to begin to identify, describe, and test some essential issues relevant to the new sport of virtual cycling. As such, we use a relatively small sample and a particular setting (top echelon racing), and therefore, it is difficult to generalize from the results to other categories of racers (e.g., women or racers in lower echelons) and different settings (less strictly regulated races).

Conclusions

In a very short time, competitive racing through virtual cycling has established itself as an entirely new discipline within cycling. While this development has come far in some respects (e.g., the development of rules and regulations, race organizations, and standards of practice), it has been limited to a small subset of the riders and races. For many virtual cyclists and the larger cycling populace, virtual cycling is still viewed as too unreliable to be considered a legitimate form of competition. Nevertheless, the findings presented in this study support the notion that virtual cycling can be an appropriate and relevant form of racing where actual physiological performance is the deciding factor. Equipment from the leading brands, at least from the higher price range, appears to produce power numbers that agree over two separate power measuring devices. However, further studies are warranted to explore the limits of the robustness and scope of this conclusion. To this end, the study also provides some suggestions on how one might view and assess the precision that might be expected from power output data from virtual cycling, including the general variability in measurements and the differences between two different measures of the same effort.

Notes

1. The first-ever UCI-endorsed Cycling Esports World Championships was held in December 2020. The second event was held in February 2022.
2. The Zwift Racing League is the most regulated and controlled virtual racing event, officially run by Zwift. It is team based, and participation is on an invitational basis only. The riders are registered and go through a performance verification process before the beginning of the series and they also submit power data and weight data continuously while racing. Most teams in the series have commercial sponsors, and there are significant cash prizes for the top finishing teams of each series. The races are also professionally broadcasted via YouTube for a viewing audience.
3. Some riders participated in more than one race (19 riders in two races and 11 riders in all three races).

References

Allen, H., Coggan, A., & Mcgreggor, S. (2019). *Training and racing with a power meter*. VeloPress.

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307–310. https://doi.org/10.1016/J.IJNURSTU.2009.10.001

Bonnington, C. (2018). How to have fun riding a stationary bike for your New Year's resolution. *Slate*. https://www.slate.com/technology/2018/01/how-to-have-fun-riding-a-stationary-bike-for-your-new-years-resolution.html

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. https://doi.org/10.1123/IJSPP.2016-0436

Bouillod, A., Soto-Romero, G., Grappe, F., Bertucci, W., Brunet, E., & Cassirame, J. (2022). Caveats and recommendations to assess the validity and reliability of cycling power meters: A systematic scoping review. *In Sensors, 22*(1), 386. https://doi.org/10.3390/s22010386

Herman, M. (2020). Gibbon and Tacey win opening stage of virtual Tour de France | Reuters. *Reuters*. https://www.reuters.com/article/uk-cycling-france-virtual-idUKKBN2450NT

International Olympic Committee (IOC). (2021). International Olympic Committee makes landmark move into virtual sports by announcing first-ever Olympic virtual series. *Olympic News*. https://www.olympics.com/ioc/news/international-olympic-committee-makes-landmark-move-into-virtual-sports-by-announcing-first-ever-olympic-virtual-series

Karsten, B., Petrigna, L., Klose, A., Bianco, A., Townsend, N., & Triska, C. (2021). Relationship between the critical power test and a 20-min functional threshold power test in cycling. *Frontiers in Physiology, 11,* 1877. https://doi.org/10.3389/fphys.2020.613151

Long, M. (2020). Zwift raises US$450m in Series C funding round led by KKR – SportsPro. *Sortspromedia*. https://www.sportspromedia.com/news/zwift-series-c-funding-round-kkr-specialized-investment/

MacInnis, M.J., Thomas, A.C.Q., & Phillips, S.M. (2019). The reliability of 4-minute and 20-minute time trials and their relationships to functional threshold power in trained cyclists. *International Journal of Sports Physiology and Performance, 14*(1), 38–45. https://doi.org/10.1123/IJSPP.2018-0100

McIlroy, B., Passfield, L., Holmberg, H.-C., & Sperlich, B. (2021). Virtual training of endurance cycling—A summary of strengths, weaknesses, opportunities and threats. *Frontiers in Sports and Active Living, 3,* 31. https://doi.org/10.3389/fspor.2021.631101

Menaspa, P., Quod, M., Martin, D., Victor, J., & Abbiss, C. (2013). Physiological demands of road sprinting in professional and U23 cycling. A pilot study. *Research Outputs 2013*. https://ro.ecu.edu.au/ecuworks2013/874

Richardson, A., Smith, P., & Berger, N. (2022). Zwift's anti – doping policy: Is it open to cheating? *International Journal of Esports, 1*(1). https://www.ijesports.org/article/90/html

Schlange, E. (2020). New "Peak Zwift" achieved: 34,940 | Zwift insider. https://www.zwiftinsider.com/peak-zwift-34940/

Westmattelmann, D., Grotenhermen, J.G., Sprenger, M., & Schewe, G. (2021). The show must go on – virtualisation of sport events during the COVID-19 pandemic. *European Journal of Information Systems, 30*(2), 119–136. https://doi.org/10.1080/0960085X.2020.1850186

Westmattelmann, D., Stoffers, B., Sprenger, M., Grotenhermen, J.-G., & Schewe, G. (2022). The performance-result gap in mixed-reality cycling—Evidence from the virtual tour de France 2020 on Zwift. *Frontiers in Physiology, 13*, Article 868902, 829. https://doi.org/10.3389/FPHYS.2022.868902

Zwift. (2021). Cycling esports rules and regulations v 1.0.7. https://www.zwift.com/p/zwift-cycling-esports-rules?__znl=en-eu