# The reproducibility of 20-min time-trial performance on a virtual cycling platform

**Garcia Matta, G., Edwards, A., Roelands, B., Hettinga, F. and Hurst, P.**



Matta G, Edwards A, Roelands B, Hettinga F, Hurst P. The reproducibility of 20-min time-trial performance on a virtual cycling platform. Int J Sports Med. 2022 May 10. doi: 10.1055/a-1848-8478. Epub ahead of print. PMID: 35537457.

# The reproducibility of 20-min time-trial performance on a virtual cycling platform

**Article** *in* International Journal of Sports Medicine · May 2022

DOI: 10.1055/a-1848-8478



Guilherme G. Matta
Canterbury Christ Church University


Andrew M Edwards
Canterbury Christ Church University


Bart Roelands
Vrije Universiteit Brussel


Florentina Johanna Hettinga
Northumbria University

International Journal of Sports Medicine
Thieme

# The reproducibility of 20-min time-trial performance on a virtual cycling platform.

| | |
|---|---|
| Journal: | *International Journal of Sports Medicine* |
| Manuscript ID | IJSM-02-2022-9444-tt.R2 |
| Manuscript Type: | Training & Testing |
| Key word: | Reliability, Virtual training, Remote exercise, Exercise testing, Test-retest |
| Abstract: | This study aimed to analyse the reproducibility of mean power output during 20-min cycling time-trials, in a remote home-based setting, using the virtual-reality cycling software, Zwift. Forty-four cyclists (11 women, 33 men; 37 ± 8 years old, 180 ± 8 cm, 80.1 ± 13.2 kg) performed 3 x 20-min time-trials on Zwift, using their own setup. Intra-class correlation coefficient (ICC), coefficient of variation (CV) and typical error (TE) were calculated for the overall sample, split into 4 performance groups based on mean relative power output (25% quartiles) and sex. Mean ICC, TE and CV of mean power output between time-trials were 0.97 [0.95—0.98], 9.36 W [8.02—11.28 W], and 3.7% [3.2—4.5], respectively. Women and men had similar outcomes (ICC: 0.96 [0.89—0.99] vs 0.96 [0.92—0.98]; TE: 8.30 W [6.25—13.10] vs. 9.72 W [8.20—12.23]; CV: 3.8% [2.9—6.1] vs. 3.7% [3.1—4.7], respectively), although cyclists from the first quartile showed a lower CV in comparison to the overall sample (Q1: 2.6% [1.9—4.1] vs. overall: 3.7% [3.2—4.5]). Our results indicate that power output during 20-minute cycling time-trials on Zwift are reproducible and provide sports scientists, coaches and athletes, benchmark values for future interventions in a virtual-reality environment. |
| | |

**INTERNATIONAL JOURNAL OF SPORTS MEDICINE**

# The reproducibility of 20-min time-trial performance on a virtual cycling platform.

Guilherme Matta[1], Andrew Edwards[1], Bart Roelands[2], Florentina Hettinga[3] & Philip Hurst[1]

*[1] School of Psychology and Life Sciences, Faculty of Science, Engineering and Social Sciences, Canterbury Christ Church University, Canterbury, United Kingdom; [2] Human Physiology and Sports Physiotherapy Research Group, Faculty of Physical Education and Physiotherapy, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium; [3] Department of Sport, Exercise and Rehabilitation, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom.*

**Running head:** Reproducibility of virtual cycling performance

**Abstract word count: 202**

**Text-only word count: 2856**

**Correspondence:**

Guilherme Matta

School of Psychology and Life Sciences

Canterbury Christ Church University

North Holmes Road, Canterbury

CT11QU, Kent

United Kingdom

+44 (0)7958690687

g.matta392@canterbury.ac.uk

# The reproducibility of 20-min time-trial performance on a virtual cycling platform


## Abstract

This study aimed to analyse the reproducibility of mean power output during 20-min cycling time-trials, in a remote home-based setting, using the virtual-reality cycling software, Zwift. Forty-four cyclists (11 women, 33 men; 37 ± 8 years old, 180 ± 8 cm, 80.1 ± 13.2 kg) performed 3 x 20-min time-trials on Zwift, using their own setup. Intra-class correlation coefficient (ICC), coefficient of variation (CV) and typical error (TE) were calculated for the overall sample, split into 4 performance groups based on mean relative power output (25% quartiles) and sex. Mean ICC, TE and CV of mean power output between time-trials were 0.97 [0.95—0.98], 9.36 W [8.02—11.28 W], and 3.7% [3.2—4.5], respectively. Women and men had similar outcomes (ICC: 0.96 [0.89—0.99] vs 0.96 [0.92—0.98]; TE: 8.30 W [6.25—13.10] vs. 9.72 W [8.20—12.23]; CV: 3.8% [2.9—6.1] vs. 3.7% [3.1—4.7], respectively), although cyclists from the first quartile showed a lower CV in comparison to the overall sample (Q1: 2.6% [1.9—4.1] vs. overall: 3.7% [3.2—4.5]). Our results indicate that power output during 20-minute cycling time-trials on Zwift are reproducible and provide sports scientists, coaches and athletes, benchmark values for future interventions in a virtual-reality environment.

## Introduction

In early- to mid-2020, to prevent the spread of COVID-19, sport and exercise science laboratories worldwide ceased all activity, and social distancing measures were put into force to prevent transmission of the virus [1]. While the pandemic begins to recede, such measures still exist and cycling research is presented with an ethical and practical challenge of examining outcome measures in laboratories, while at the same time ensuring the health and safety of both researchers and participants. A need, therefore, exists in identifying innovative means to gain meaningful outcome measures that can be conducted in an environment that do not increase the risk of COVID-19 infection. One potential alternative is developing remote-design studies using online cycling platforms that allow for social distancing and might provide insightful information about cyclists' performance. However, for such studies to be designed and to provide meaningful inferences, outcomes must be reproducible.

Among several online cycling platforms [2], Zwift is one of the most popular with over 3-million users registered [3] in more than 190 countries [2]. It consists of a virtual-reality game/software that allow cyclists to ride their bikes on a stationary trainer, replicating training/competitive environments, while presenting an opportunity for remote social interaction, competition, training and intervention studies. To our knowledge, no research has examined the reproducibility of cycling performance on such virtual platforms Given that cyclists and researchers have been heavily impacted by the restrictions caused by the COVID-19 pandemic, this research is timely, which will provide important information for cyclists, sports scientists and coaches aiming to examine performance outcomes in a remote-based environment.

Reproducibility is a measure that informs the consistency of performance tests in repeated trials for the same athlete [4]. Nimmerichter, Williams, Bachl, et al. [5] and MacInnis, Thomas and Phillips [6] found high reproducibility of mean power output during 20-min field- and laboratory-based time-trials, reporting intraclass coefficient correlations (ICC) of 0.98 (95%CL of 0.95—0.99) and 0.99 (95%CL of 0.95—1.0), respectively. In a review of exercise performance measures, Currell and Jeukendrup [7] reported that coefficients of variation (CV) are usually lower than 5% for cycling time-trials in the field and the laboratory. However, Hopkins, Schabort and Hawley [4] suggested that reproducibility is affected by athletes' performance level and sex. To our knowledge, only two studies have analysed how performance level affects the reproducibility of mean power output [8, 9]. Both studies reported lower typical errors (TE) and CVs for top-ranked cyclists during 40- [9] and 20-km [8] laboratory-based time-trials, which was explained by higher cycling experience. The differences between women and men, on the other hand, have received little attention. In an early study, Bishop [10] analysed the reproducibility of 60-min cycling time-trials in women and reported a mean ICC of 0.97, but they did not compare this against men. Although the reproducibility of laboratory- and field-based cycling time-trials is well established, it is yet to be determined how it is affected by performance groups and sex in a virtual-reality environment.

The aims of our study were twofold. First to examine the reproducibility (i.e., intra-subject reproducibility where there is consistency between time-trials for the same cyclist) of mean power output during 20-min time-trials on a virtual cycling platform. Second to examine whether reproducibility is similar between different performance levels and sex.

## Methods

*Participants (n = 44)*

After advertisements on social media (e.g., Facebook), 44 trained cyclists (11 women, 33 men; 37 ± 8 years old, 180 ± 8 cm, 80.1 ± 13.2 kg) volunteered to participate. Eligibility criteria stipulated participants were between 18 and 55 years old, free of injury, had used Zwift for more than 4 months and had not experienced COVID-19 symptoms (i.e., high temperature, a new, continuous cough and a loss or change to a sense of smell or taste) in the 2 months preceding participation. The lead author's institutional human research ethics committee approved the study in compliance with the Declaration of Helsinki (ref.: ETH2021-0133) and all participants provided digital informed consent prior to participation.

*Study design*

We used a within-participant, repeated measures, remote-research design whereby participants performed 3 x 20-min time-trials on a virtual cycling platform (i.e., Zwift) interspersed by 5-7 days each at the same time of the day (± 2 h). The 20-min time-trial was chosen as it is a standard performance measure among cyclists [6] and most performance tests on virtual platforms involve this time-trial duration.

*20-min cycling time-trials and procedures*

All time-trials were performed on participants' own setup, of which they navigated their on-screen avatar through the virtual road that simulated outdoor conditions. Each time-trial was performed at the "Tempus Fugit" course, which is available to all Zwift users and was designed as an out and back flat course, containing 17.3 km and 16 m of elevation gain. The time-trial protocol (see below) was developed by the research team, which was exported as a workout file (.zwo) and sent to participants' e-mail, who then imported the file to their accounts. Participants were provided with detailed instructions, containing a step-by-step guide about how to import and export files.

Before each 20-min time-trial, participants performed a 10-min warm-up at their habitual self-selected intensity (*i.e.*, defined during the first time-trial and replicated throughout), followed by 5-min rest. They were instructed to standardise their diet, fluid intake, equipment (*i.e.,* bike and/or trainer) and environment (*i.e.,* the position of a fan, place and starting time) during each time-trial, whereas also avoiding high-intensity and long-duration exercises 48-h beforehand. Participants performed all time-trials individually and used their time-trial virtual bike—which removes the drafting effect feature, caused by overtaking other riders. The day before the start of each time-trial, participants were e-mailed instructions described previously and requested to calibrate their equipment according to the manufacturer's instructions.

After completion, participants exported the time-trial file in a Flexible and Interoperable Data Transfer (FIT) format and sent it to the main investigator's e-mail. Given that there might be differences in the performance data generated by distinct power meters devices attached to participants' bikes and the virtual platform, they were requested to export the FIT file generated from the folder in their device (e.g., laptop or tablet) instead of the file from other potential sources. The participants also indicated which type of trainer they used. The detailed description of the trainers used by the participants can be found in Table 1, along with corresponding studies that investigated the reproducibility of those available [11-14].

[Table 1]

*Statistical analysis*

Descriptive data are reported as mean ± standard deviation, unless otherwise stated. The mean power output, cadence, and heart rate achieved in each time-trial were extracted from the FIT file generated by the virtual platform using a training-analysis software (TrainingPeaks WKO+ v3.0, PeaksWare, Lafayette, Colorado, USA). Within-participant differences in mean power output, cadence and heart rate between time-trials were analysed using two-way repeated-measures ANOVAs with Bonferroni pairwise comparisons.

The overall reproducibility of mean power output across the time-trials was reported by calculating ICC, CV and TE between each time-trial and as percentages derived from log-transformed data

[15]. To examine whether reproducibility was similar between athletes from different performance levels, participants were ranked into 4 performance groups (*i.e.,* 25% quartiles; Q1, Q2, Q3, Q4; each group n = 11) based on the mean relative power output (W/kg) produced during their best time-trial. They were also split between women and men to analyse whether reproducibility was similar between sex.

Data analyses were performed using SPSS (26.0, IBM, Armonk, USA) and an online published spreadsheet [15] (Microsoft Office 365, Excel, Microsoft, Redmond, USA). Statistical significance was set at $P \leq .05$ and effect sizes were calculated as partial eta-squared ($\eta_p^2$), of which $\eta_p^2 = 0.01$, 0.06 and 0.14 indicates a small, medium and large effect, respectively [16].

## Results

### *Overall results*

Individual values for power output, heart rate and cadence in each time-trial are shown on Figure 1. There were no differences in mean power output (256 ± 52, 254 ± 51 and 255 ± 52 W; $F = .95$, $P = .391$, $\eta_p^2 = .02$), and heart rate (161 ± 13, 160 ± 13 and 161 ± 13 bpm; $F = 1.57$, $P = .215$, $\eta_p^2 = .04$) between time-trials 1 to 3 respectively. However, we found an interaction effect for cadence (87 ± 9, 86 ± 9 and 86 ± 8 rpm for time-trials 1 to 3, respectively; $F = 5.81$, $P = .007$, $\eta_p^2 = .81$), and pairwise comparisons showed a difference between time-trials 1-3 ($P = .006$), but not between trials 2-3 ($P = .230$). During their best time-trial, women and men achieved 2.92 ± 0.47 vs 3.47 ± 0.74 W/kg, respectively; performance groups Q1 to Q4 achieved 4.17 ± 0.45, 3.60 ± 0.18, 3.11 ± 0.17, 2.44 ± 0.40 W/kg, respectively.

[Figure 1]

### *Reproducibility analysis*

The ICC, TE and CV of mean power output along with 95%CL between trials 2-1 and 3-2 for the overall sample and split by performance groups and sex are presented in Table 1. Women and men had similar outcomes, although Q1 showed a lower CV (2.6% [1.9—4.1%]) in comparison to the overall sample (3.7% [3.2—4.5%]). When we analysed the reproducibility for the participants who have been using the virtual platform for more than 24 months, we found higher reproducibility for the more experienced riders with a mean ICC, TE and CV of 0.99 [0.98—1.00], 6.7 W [5.29—9.82 W] and 2.6% [2.0—3.8%] against 0.96 [0.93—0.97], 10.17 W [8.76—12.29 W] and 4.0% [3.4—4.9%] for those using for less than 24 months, respectively.

[Table 2]

## Discussion

This is the first study to show that cycling performance during 20-min time-trials performed on a virtual platform is reproducible. We showed that the CV for mean power output between time-trials was lowest for top-ranked participants (i.e., top 25%). However, our results do not support the notion that sex affects reproducibility. Our findings are likely to assist sports scientists, coaches and athletes aiming to measure cycling performance during online virtual software.

We found that mean power output and heart rate were not different between time-trials, although cadence was lower in the third time-trial compared to the first (87 ± 9 vs. 86 ± 8 rpm, respectively) but not to the second (86 ± 9). However, a difference of 1 rpm is unlikely to represent a real effect and might have not influenced the participants' performance. In fact, Stone et al. [17], analysed the reproducibility of cadence during 4-km time-trials and found a larger variability in comparison to mean power output, which may explain the differences we found between the third and first time-trial.

The ICC values found in our study (0.97 [CL95% 0.95—0.98]), are similar with the results of Nimmerichter et al. [5], who reported high reproducibility of mean power output during field-based 20- and 4-min time-trials (0.98 [CL95% 0.95—0.99] and 0.98 [CL95% 0.92—0.99] respectively). It also agrees with MacInnis [6], who found ICC values of 0.99 [CL95% 0.95—1.00] and 0.98 [CL95% 0.91—1.00] during laboratory-based 20- and 4-min time-trials, respectively. While MacInnis, Thomas and Phillips [6] reported a mean CV of 1.4% during the 20-min time-trials, which was lower than the CV of 3.7% found in our study. However, this is most likely explainable due to the homogenous population of elite athletes used in their study [6]. The frequent exposure to high-intensity exercise they are exposed to can reduce variability in performance [4], which is also supported by our findings showing that the top-ranked participants had the lowest CV. The ICC values we found suggest that cycling performance during 20-min time-trials on a virtual platform is reproducible and similar to laboratory- and field-based cycling time-trials. We suggest that the use of exercise in a home-based setting via virtual platforms can be useful for engaging with others in a community while remote, enhancing motivation and providing a stable environment for recording outcomes that are not unduly affected by day-to-day variation. These do not replace laboratory reproducibility studies on standardised equipment but do provide a means for gaining meaningful data for athletes, coaches and researchers where the reproducibility of an individual's performance on their own setup is of value.

We found that top-ranked participants had a lower CV (2.6%) than the overall sample (3.7%) for mean power output between time-trials. This finding is consistent with the results of Zavorsky, Murias, Gow, et al. [8] who analysed the reproducibility of 20-km cycling time-trials and their top-ranked participants demonstrated a mean CV of 2.5%, against 3.7% reported for the overall sample. As suggested by Hopkins et al. [18], trained athletes might have more competitive and training experience, which might explain why the top-ranked cyclists in our study displayed lower variation in performance. Indeed, Laursen, Shing and Jenkins [9], found higher reproducibility of performance during 40-km time-trials for their top-ranked participants and found that they had significantly more cycling experience than the slower ones. It is noteworthy that the TE between Q1, Q3 and Q4 was similar, although the CV was lower for Q1. This might be explained considering that higher values of power output achieved by Q1 might have yielded higher TEs [19], although performance varied to a lesser extent. Surprisingly, Q2 showed a higher variation of performance evidenced by the CV and TE. Although we do not have enough data to provide reasonable explanations for this, we might assume that cycling experience played a role [9].

The reproducibility analysis between women and men in our study yielded similar results. Contrary to our findings, Hopkins and Hewson [18] analysed the results of official running races, including cross-country, road, half-marathon and marathon races and found that female runners display lower variability in performance in comparison to males. In another study [4], the authors reviewed the literature and identified the factors that might affect reproducibility. They suggested that variability in performance might be higher in non-athletic females than in non-athletic males, and deduced that the non-athletic females might be less active and that the menstrual cycle might also play a role. However, our results do not support those assumptions and suggest that the reproducibility of performance during 20-min time-trials between women and men is similar. Our results agree with Bishop [10] who reported a mean ICC of 0.97 for women during 60-minute cycling time-trials, which is similar to our study and the ICC found in previous studies with male cyclists [5, 7, 20]. However, there is a clear sex bias in the sports sciences research, of which women are underrepresented [21]. Although we aimed to recruit both women and men, the differences in the sample size must be considered when interpreting our results.

**Practical implications**

Our results are particularly important in times when face-to-face activities might be impacted due to restrictions caused by COVID-19 and sports scientists, coaches and athletes might necessarily incorporate virtual training into their routine. This has important implications for experimental designs where participants may reside in remote, rural communities and be unable to attend training or laboratory sessions. Therefore, having a reproducible and remote system [22] is beneficial for those aiming to understand performance measures without having to increase the risk of transmitting COVID-19 to participants and researchers.

We showed that technology could be useful for a variety of experimental studies examining cycling performance using remote designs. Studies that are performed in the athletes' own environment is important for researchers and athlete support personnel (e.g., coaches) aiming to monitor and evaluate sport performance outcomes. The originality of our work identifies the potential application of remote exercise and doing so in a reproducible way that is of ecological importance. Given the impact of the COVID-19 pandemic on athletes' training behaviours [23], our results are likely to be used in assisting coaches and athletes in their virtual training monitoring and the development of new remote-study designs by sports scientists.



## Limitations

Our study has reported novel findings, but these should be interpreted considering some limitations. First, it is important to note that on most virtual platforms, cyclists usually share the virtual road with other users which may have influenced the performance of our participants [24]. While we did instruct participants to not compete against and avoid others in the virtual platform, performance may have been affected by the presence of others. Second, although the reproducibility of mean power output was high, we could not examine the accuracy and the validity of power outputs generated by the participants' trainer, rather than how consistently they were reproduced by the individual riders. Given the potential differences in types of trainers used, discrepancies across models/devices might be expected [25, 26]. However, as suggested by Atkinson and Nevill [27], the reproducibility of any new measurement tool should be tested before its validity, as it is unlikely that it will be valid if not adequately consistent. Future research should therefore examine the validity of home-based training setups.

## Conclusions

In summary, the results of our study suggest that mean power output during 20-min cycling time-trials performed on a virtual platform is reproducible and similar for both women and men. Top-ranked and experienced cyclists might display higher reproducibility of performance between time-trials. The results of this study provide sports scientists, coaches and athletes, benchmark values for future interventions in a virtual-reality environment.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Data availability

The authors are happy to make the raw data of this study and the Zwift workout file used available on reasonable request.

## References

1. de Boer DR, Hoekstra F, Huetink KI et al. Physical activity, sedentary behavior and well-being of adults with physical disabilities and/or chronic diseases during the first wave of the covid-19 pandemic: a rapid review. Int J Environ Res Public Health 2021; 18: 6342. doi: 10.3390/ijerph18126342

2. McIlroy B, Passfield L, Holmberg H-C et al. Virtual training of endurance cycling–a summary of strengths, weaknesses, opportunities and threats. Front Sports and Act Living 2021; 3: 631101. doi: 10.3389/fspor.2021.631101

3. Forbes / Reed R. Do you even Zwift? The indoor cycling platform is having a moment (Feb 17, 2021). In Internet: https://www.forbes.com/sites/robreed/2021/02/17/do-you-even-zwift-the-indoor-cycling-platform-is-having-a-moment/?sh=644014073f86; (Accessed Jun 1, 2021).

4. Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med 2001; 31: 211-234. doi: 10.2165/00007256-200131030-00005

5. Nimmerichter A, Williams C, Bachl N et al. Evaluation of a field test to assess performance in elite cyclists. Int J Sports Med 2010; 31: 160-166. doi: 10.1055/s-0029-1243222

6. MacInnis MJ, Thomas AC, Phillips SM. The reliability of 4-min and 20-min time trials and their relationships to functional threshold power in trained cyclists. Int J Sports Physiol Perform 2018; 29: 1-27. doi: 10.1123/ijspp.2018-0100

7. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. Sports Med 2008; 38: 297-316. doi: 10.2165/00007256-200838040-00003

8. Zavorsky G, Murias J, Gow J et al. Laboratory 20-km cycle time trial reproducibility. Int J Sports Med 2007; 28: 743-748. doi: 10.1055/s-2007-964969

9. Laursen P, Shing C, Jenkins D. Reproducibility of a laboratory-based 40-km cycle time-trial on a stationary wind-trainer in highly trained cyclists. Int J Sports Med 2003; 24: 481-485. doi: 10.1055/s-2003-42012

10. Bishop D. Reliability of a 1-h endurance performance test in trained female cyclists. Med Sci Sports Exerc 1997; 29: 554-559. doi: 10.1097/00005768-199704000-00019

11. Wainwright B, Cooke CB, O'Hara JP. The validity and reliability of a sample of 10 Wattbike cycle ergometers. J Sports Sci 2017; 35: 1451-1458. doi: 10.1080/02640414.2016.1215495

12. Hopker J, Myers S, Jobson S et al. Validity and reliability of the Wattbike cycle ergometer. Int J Sports Med 2010; 31: 731-736. doi: 10.1055/s-0030-1261968

13. Zadow EK, Kitic CM, Wu SS et al. Validity of power settings of the Wahoo KICKR Power Trainer. Int J Sports Physiol Perform 2016; 11: 1115-1117. doi: 10.1123/ijspp.2015-0733

14. Zadow EK, Kitic CM, Wu SS et al. Reliability of power settings of the Wahoo KICKR power trainer after 60 hours of use. Int J Sports Physiol Perform 2018; 13: 119-121. doi: 10.1123/ijspp.2016-0732

15. Hopkins WG. Spreadsheets for analysis of validity and reliability. Sportscience 2017; 21. doi: https://www.sportsci.org/2015/ValidRely.htm

16. Cohen J. Statistical power analysis for the behavioral. Sciences Hillsdale (NJ): Lawrence Erlbaum Associates 1988. 18-74. doi: https://doi.org/10.4324/9780203771587

17. Stone MR, Thomas K, Wilkinson M et al. Consistency of perceptual and metabolic responses to a laboratory-based simulated 4,000-m cycling time trial. Eur J Appl Physiol 2011; 111: 1807-1813. doi: 10.1007/s00421-010-1818-7

18. Hopkins WG, Hewson DJ. Variability of competitive performance of distance runners. Med Sci Sports Exerc 2001; 33: 1588-1592. doi: 10.1097/00005768-200109000-00023

19. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000; 30: 1-15. doi: 10.2165/00007256-200030010-00001

20. Jeukendrup A, Saris WH, Brouns F et al. A new validated endurance performance test. Med Sci Sports Exerc 1996; 28: 266-270. doi: 10.1097/00005768-199602000-00017

21. Cowley ES, Olenick AA, McNulty KL et al. "Invisible sportswomen": the sex data gap in sport and exercise science research. Women Sport Phys Act J 2021; 29: 146-151. doi: https://doi.org/10.1123/wspaj.2021-0028

22. Bird JM, Karageorghis CI, Baker SJ et al. Ready Exerciser One: Effects of music and virtual reality on cycle ergometer exercise. Br J Health Psychol 2021; 26: 15-32. doi: 10.1111/bjhp.12445

23. Washif JA, Farooq A, Krug I et al. Training during the COVID-19 lockdown: knowledge, beliefs, and practices of 12,526 athletes from 142 countries and six continents. Sports Med 2021. 1-16. doi: 10.1007/s40279-021-01573-z

24. Hettinga FJ, Konings MJ, Pepping G-J. The science of racing against opponents: affordance competition and the regulation of exercise intensity in head-to-head competition. Front Physiol 2017; 8: 118. doi: 10.3389/fphys.2017.00118

25. Bouillod A, Soto-Romero G, Grappe F et al. Caveats and recommendations to assess the validity and reliability of cycling power meters: a systematic scoping review. Sensors 2022; 22: 386. doi: 10.3390/s22010386

26. Passfield L, Hopker JG, Jobson S et al. Knowledge is power: Issues of measuring training and performance in cycling. J Sports Sci 2017; 35: 1426-1434. doi: 10.1080/02640414.2016.1215504

27. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998; 26: 217-238. doi:10.2165/00007256-199826040-00002

## LIST OF TABLES

**Table 1.** Description of trainers used by the participants (n = 44) in this study.

**Table 2.** Mean power output (W) within-subject intraclass correlation coefficients, absolute typical errors and typical errors as coefficients of variation between time-trials for the overall sample, and split by performance groups and sex. Data are presented as mean [CL95%].

## LIST OF FIGURES

**Figure 1.** Individual values for mean power output, heart rate and cadence for each athlete during the time-trials. Each bar represents the mean values for each time-trial. * Denotes difference from time-trial 1 (*P* = .006).

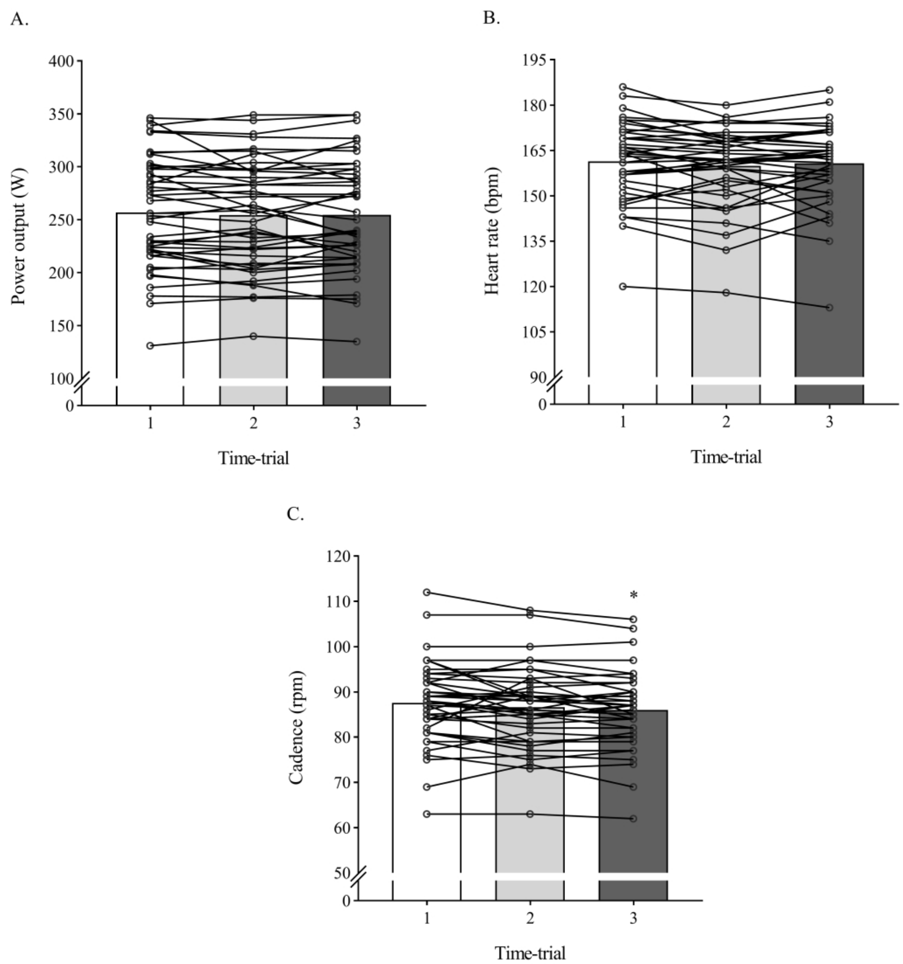


Fig 1: Individual values for mean power output, heart rate and cadence for each athlete during the time-trials. Each bar represents the mean values for each time-trial. * Denotes difference from time-trial 1 ($P$ = .006).

197x204mm (600 x 600 DPI)

**Table 1.** Description of trainers used by the participants (n = 44) in this study.

| n | Manufacturer | Country | Models (n) | Type |
|---|---|---|---|---|
| 7 | Elite | Italy | Direto (6); Suito (1) | Direct-drive |
| 16 | Tacx | Netherlands | Neo 2T (5); Neo (4); Flux S Smart (3); Flux 2 (2); Satori (1); Vortex (1) | Direct-drive, wheel-on |
| 5 | Wattbike[a,b] | United Kingdom | Atom (3); Pro (2) | Indoor bike |
| 14 | Wahoo[c,d] | United States | Kickr Core (8); Kickr (5), Snap (1) | Direct-drive, wheel-on |
| 1 | Saris | United States | H3 (1) | Direct-drive |
| 1 | Bkool | Spain | Pro 2 (1) | Wheel-on |

[a] Previous studies reported high reproducibility of Wattbike Ergometers [11, 12]; [b] Previous studies reported high reproducibility of Wahoo Ergometers [13, 14].

**Table 2.** Mean power output (W) within-subject intraclass correlation coefficients, absolute typical errors (W) and typical errors as coefficients of variation (%) between time-trials for the overall sample, and split by performance groups and sex. Data are presented as mean [CL95%].

| | | **Performance groups** | | | | **Sex** | |
|---|---|---|---|---|---|---|---|
| | **Overall (n = 44)** | **Q1 (n = 11)** | **Q2 (n = 11)** | **Q3 (n = 11)** | **Q4 (n = 11)** | **Women (n = 11)** | **Men (n = 33)** |
| **ICC$^{(TT2-TT1)}$** | 0.97 [0.95—0.99] | 0.96 [0.86—0.99] | 0.85 [0.54—0.96] | 0.99 [0.95—1.00] | 0.97 [0.89—0.99] | 0.97 [0.91—0.99] | 0.96 [0.92—0.98] |
| **ICC$^{(TT3-TT2)}$** | 0.97 [0.94—0.98] | 0.96 [0.85—0.99] | 0.86 [0.57—0.96] | 0.97 [0.91—0.99] | 0.95 [0.84—0.99] | 0.95 [0.82—0.99] | 0.96 [0.91—0.98] |
| **ICC$^{(mean)}$** | **0.97 [0.95—0.98]** | **0.96 [0.88—0.99]** | **0.87 [0.66—0.96]** | **0.98 [0.94—0.99]** | **0.96 [0.89—0.99]** | **0.96 [0.89—0.99]** | **0.96 [0.92—0.98]** |
| **TE$^{(TT2-TT1)}$** | 9.11 [7.53—11.55] | 7.66 [5.35—13.44] | 13.35 [9.33—23.44] | 7.22 [5.04—12.67] | 6.16 [4.30—10.81] | 6.56 [4.59—11.52] | 9.77 [7.85—12.92] |
| **TE$^{(TT3-TT2)}$** | 9.61 [7.94—12.17] | 7.29 [5.10—12.80] | 12.38 [8.65—21.72] | 9.52 [6.65—16.71] | 7.91 [5.52—13.87] | 9.73 [6.80—17.08] | 9.68 [7.78—12.80] |
| **TE$^{(mean)}$** | **9.36 [8.02—11.28]** | **7.48 [5.63—11.80]** | **12.88 [9.70—20.32]** | **8.45 [6.37—13.34]** | **7.09 [5.34—11.18]** | **8.30 [6.25—13.10]** | **9.72 [8.20—12.23]** |
| **CV$^{(TT2-TT1)}$** | 3.5 [2.9—4.5] | 2.6 [1.8—4.5] | 4.7 [3.2—8.3] | 3.1 [2.2—5.6] | 3.2 [2.2—5.7] | 3.3 [2.3—5.8] | 3.5 [2.8—4.7] |
| **CV$^{(TT3-TT2)}$** | 4.0 [3.3—5.0] | 2.5 [1.8—4.5] | 4.7 [3.3—8.4] | 4.1 [2.9—7.3] | 3.9 [2.7—6.9] | 4.4 [3.0—7.8] | 3.9 [3.1—5.1] |
| **CV$^{(mean)}$** | **3.7 [3.2—4.5]** | **2.6 [1.9—4.1]** | **4.7 [3.5—7.5]** | **3.7 [2.7—5.8]** | **3.6 [2.7—5.7]** | **3.8 [2.9—6.1]** | **3.7 [3.1—4.7]** |

ICC - intraclass correlation coefficient; TE - typical error; CV - coefficient of variation; TT - time-trial; Q1-4 - performance groups split by quartiles.