Abstract

Purpose: The purpose of this study was to assess the relationship between typical performance tests amongst elite and professional cyclists when conducted indoors and outdoors. Methods: 14 male cyclists of either UCI Continental or UCI World Tour level (mean ± SD: age 20.9 ± 2.8 y, mass 68.13 ± 7.25 kg) were recruited to participate in 4 test sessions (2 test sessions indoors, 2 test sessions outdoors) within a 14-day period, consisting of maximum mean power (MMP) testing for durations of 60s, 180s, 300s and 840s. Results: Across all MMP test durations, the trimmed mean power was higher outdoors compared to indoor testing (p < 0.05). Critical Power (CP) was higher outdoors compared to indoors (+19 W, p = 0.005) whilst no difference was observed for the work capacity above CP (W’). Self-selected cadence was 6 rev·min⁻¹ higher indoors versus outdoors for test durations of 60s (p = 0.038) and 300s (p = 0.002). Conclusions: These findings suggest that maximal power testing in indoor and outdoor settings cannot be used interchangeably. Furthermore, there was substantial individual variation in the difference between indoor and outdoor MMPs, across all time durations, further highlighting the difficulty of translating results from indoor testing to outdoor, on an individual level in elite populations.

Keywords: elite, cycling, maximum mean power, self-selected cadence
Introduction

Cycle training can be carried outdoors or indoors, typically on fixed ergometers. Due to high volume and frequency of training sessions required to achieve a high level in cycling, and the geographical and seasonal variations in weather conditions, many cyclists utilise indoor training methods to maintain optimal training consistency. Anecdotally, cyclists will make use of indoor training due to illness or injury, sub-optimal environmental conditions or logistical difficulties that may complicate or hinder outdoor sessions. However, less is known on the differences between indoor and outdoor cycling from a physiological, biomechanical and performance perspective, especially in well-trained populations. Of particular note is the limited research on the comparison of typically adopted performance assessments when performed in either indoor or outdoor settings.

When assessing physiological capacity or performance, a wide variety of both indoor and outdoor tests are available to practitioners. Furthermore, there is a large research base utilising indoor cycle ergometry to determine physiological capacities in elite cyclists. Typical measures in indoor settings (i.e. laboratory) include physiological parameters like lactate threshold, maximal oxygen consumption ($\dot{V}O_{2\text{max}}$) or exercise economy. Although cycling is predominantly an aerobic sport and there is strong evidence that elite cyclists possess a high $\dot{V}O_{2\text{max}}$ (~74 mL·kg$^{-1}$·min$^{-1}$), we know that two cyclists with the same $\dot{V}O_{2\text{max}}$ may have significantly different peak sprint powers. Further to this, the outcome of a race is often decided by the production of maximal powers for short durations. This is why the evaluation of maximum mean power outputs (MMP) over a varying range of durations (5-600s), typically referred to as ‘power profiling’ is widely adopted as an assessment of performance in cyclists. These assessments are typically performed by doing all-out trials of varying durations in the field or by extracting data from racing situations. Such approaches have proven successful in differentiating cyclists of varying levels, with various degrees of specialisation, or rider type (sprinter versus climber), but also to evaluate the demands of cycling races. Additionally, MMP data can be used in estimating Critical Power (CP), with power taken from short (120s) to longer (900s) durations being used in the calculation of CP and W’. CP is the highest sustainable rate of oxidative
metabolism above which occurs a continuous loss of metabolic (Phosphocreatine, pH), and systemic ($\bar{V}O_2$, blood lactate concentration) homeostasis. The relationship between the CP concept and cycling performance, as well as the underlying physiology behind it, has been well established. Therefore, within cycling, its application has been widely implemented as a performance measure (in both laboratory and field settings) and as a threshold to demarcate training zones. In addition, the amount of work (kJ) that can be completed above CP ($W'$) can be identified from MMP data, as long as the assumption that $W'$ is as close to being fully depleted as possible during the fixed work rate test is met. Despite the bioenergetics and underlying physiology behind $W'$ requiring further investigation, it remains an important metric in both laboratory and practical settings of training prescription, monitoring and performance assessment.

Field testing is often a favourable approach over laboratory testing to identify and determine athletes power profile and to set training intensities within high performance teams during competition periods, due to procedural ease. Additionally, compliance to complete multiple CP tests in a laboratory for athletes is low, due to the logistical challenges of completing multiple test sessions within the usual training and race schedule. Therefore, the need ‘exists’ to identify if laboratory and field assessments of power output can be used interchangeably. Previous research has attempted to investigate differences between indoor testing against outdoor cycling. For example, Quod et al. evaluated how MMP assessed over different durations in an indoor setting compares to MMP over similar durations achieved within actual competitions in elite cyclists. Whilst their findings suggested field MMP analysis directly compared with an athlete’s ability to produce peak MMP in a laboratory setting, observing maximal powers over a fixed duration in a race setting is difficult. That is, in a mass-start race setting, the terrain, format and tactics of races will directly impact the type of effort made. Athletes rarely complete a maximal effort for a fixed amount of time within a race and to infer that both a maximal MMP effort was completed and $W'$ is as close to being fully depleted is difficult. Although recent studies have determined that CP from race data and formal testing are comparable, Leo et al. found that relying on training data alone, MMP values and subsequent CP and $W'$ estimations were not sufficient markers of performance capacity in professional U23 cyclists. Therefore, performing maximal trials to establish a power profile, determine the individual’s CP and $W'$, and repeating this throughout different phases, would be an optimal approach instead of solely relying on race or training data. Additionally, the reliance on race data alone to
determine a power profile would require a larger amount of data to be collected first before that assessment could be made, losing the ability of a more ‘acute’ performance assessment. However, the comparison of performing such ‘formal testing’ (i.e. maximal trials of different durations in training) between indoor and outdoor settings is largely unknown. Findings from studies similar to Quod et al. \(^ {15}\) are often contradictory, and the majority conducted on low- to moderately-trained individuals \(^ {17,18}\). In particular, Triska and colleagues employed a dual-test protocol, with a constant-power, time-to-exhaustion and a fixed-duration test to identify indices of CP and W’, in both the laboratory and the field. Whilst they found no significant difference between laboratory- and field-derived variables, there was a high individual variation, and could not support the application of laboratory and field-derived CP tests interchangeably. Furthermore, many of the current laboratory/field comparison studies employ a constant work/speed measure of time to exhaustion, which has reduced reliability in respect to fixed-duration tests \(^ {19}\).

Therefore, the purpose of the current study is to investigate potential differences in maximum mean power outputs of varying durations, CP and W’, when performing tests either indoors or outdoors, in elite cyclists.

**Methods**

**Participants**
13 UCI Continental and 1 UCI World Tour male cyclists (mean ± SD: age 20.9 ± 2.8 y, mass 68.13 ± 7.25 kg, CP 341 ± 39 W) were recruited to participate in 4 test sessions within a 14-day period, during a European, road-cycling season (February-October). Each participant provided written, informed consent and ethics approval was granted in accordance with the Declaration of Helsinki. Throughout their participation in the study, participants managed their normal diet and daily training activities, however, diet was recorded in a morning questionnaire, and replicated in the 24 hours prior to each test session. Individual training loads were matched between test sessions over the 14-day period.

**Study Design**
Four maximum mean power (MMP) test sessions were completed over a 14-day period. MMPs were grouped in order to complete two MMP durations in one day: the test sessions
consisted of an indoor 180s and 840s MMP test, an indoor 60s and 300s MMP test, an outdoor 180s and 840s MMP test and finally, an outdoor 60s and 300s test. Each test session was preceded by a standardised warm up. A minimum of 48 hours separated the first two test sessions, followed by a minimum of five days and a maximum of eight days of the athletes usual training programme, with the subsequent two test sessions interspersed by 48 hours. Participants were randomised and placed into groups to determine the order of test completion. Indoor test sessions were completed on participants own turbo trainer (Tacx NEO, Wassenaar, The Netherlands). All tests were completed on the participants own bicycles fitted with their team-issue power meter (ROTOR INpower, Spain), all power data used in the analyses was from the ROTOR power meter. All procedures between the indoor and outdoor tests were standardised as much as possible (see details below).

**Procedures**

Prior to each test session, participants recorded their nude body weight and calibrated their power meters according to the individual manufacturers’ instructions. A modified version of the Lamberts and Lambert Submaximal Cycle Test (LSCT) with power clamped at 60, 75 and 85 % of the participants CP (as determined from their own testing in the year preceding) was used as a standardised warm-up before each trial \(^{20,21}\). Following five minutes recovery at 60% of CP, participants completed two, 10s all-out sprints from a rolling start at around 70-80 rev·min\(^{-1}\), separated by 3 minutes of recovery, to determine 1 second maximal power (\(P_{\text{max}}\)). On completion of the sprints, participants had 5 minutes of light cycling before beginning the test protocol. During each of the test sessions, feedback in the form of heart rate response (b·min\(^{-1}\)), elapsed time (min:sec), power output (W) and cadence (rev·min\(^{-1}\)) on the participants personal computers was permitted. No other external encouragement, music or other stimulus was permitted. Throughout the testing the participants could adjust their gear ratio freely. On each testing day, the two MMP efforts were separated by a standardised 40 minutes at 60% of the participants CP. Following the MMP test efforts, participants recovered for a duration of at least 20 minutes, and no longer than 90 minutes at a self-selected power before dismounting the bike and their session RPE and nude-body mass was recorded, and fluids consumed were recorded in order to estimate sweat rate for each test. Power output was exported from the participants personal computer (Garmin Edge 520 and Garmin Edge 830, Garmin International, Olathe, KS) and analysed for the average and SD for each MMP test. Total work completed was recorded from summing the work (kJ) in the 60s, 180s, 300s and 840s test for all indoor and outdoor tests.
Control of environmental factors

Due to differing geographical locations of the participants, and in order to standardise the outdoor tests, participants were instructed to use a slightly uphill section of road, that utilised a gradient of approximately 2-3%. This section of road was the same for all 4 outdoor test durations with no interruption in the form of traffic-lights, dangerous junctions, railway crossings or significant changes in gradient for at least 840s (around 11 km). Ambient temperature was recorded from the participants personal computer.

Calculation of Critical Power and W’

In addition, linear regression was used to assess critical power (CP) and work completed above CP (W’) from power collected from the fixed durations of 180s, 300s and 840s using the work-time model 13, where W’ represents the y-axis intercept and CP represents the slope of the regression line, the following equation was applied: \( P = W' + (CP \cdot t) \), where \( P = \) power output (W), \( t = \) duration of test (s).

Statistical Analysis

All data and figures are presented as trimmed means (trim = 0.2) ± trimmed standard deviation (trim = 0.2) or estimated difference in the trimmed means. Statistical significance was set at \( p \leq 0.05 \) unless corrected for multiple comparisons. The assumption of normality was tested using the Shapiro-Wilk test, Q-Q plots and Kernel density plots prior to analysis of the data. Differences in CP, W’, Total Work (kJ) and in MMP test data at 60s, 180s, 300s and 840s parameters were compared between indoor and outdoor tests using Yuen’s paired t-tests — a robust comparison of trimmed means and Winsorized variances that performs better than other robust statistics when sample sizes are small 22. For MMP test data, where multiple time periods are compared, Benjamini and Hochberg’s 23 method of controlling false discovery rates was used to adjust p-values.

Raw effect sizes are presented as estimates of the difference in trimmed means along with 95% confidence intervals. To support comparisons across measures that use different units of measurement, a robust standardised effect size is reported based on Algina, Keselman, and Penfield’s 24 robust version of Cohen’s d. To provide an approximate guide to interpreting
standardised effects, the magnitude of the effect size criteria was set at: < 0.19 trivial, 0.2 – 0.59 small, 0.6 – 1.19 moderate, 1.2 – 1.99 large, > 2 very large.

To further aid interpretation of the results the S-value (Shannon information value or surprisal (s) value) is also reported in Table 1 and Table 2. The S-value is the negative log of the p-value − log₂(p), which transfers the p-value into bits of information against the test hypothesis; the greater the S-value, the less compatible the observed data are with the null-hypothesis. By rounding the S-value to the nearest integer, it presents an intuitive metric for interpreting the evidence against the null-hypothesis, as a measure of the amount of information that would be gained from the same number of coin tosses. For example, an S-value of 4 indicates that the observed test statistic is no more surprising than getting all heads on 4 fair coin tosses.

All statistical analyses were conducted in R (R Core Team (2020), Vienna, Austria), using the WRS2 package.

Results
Differences in ambient temperature and sweat rate between the indoor and outdoor test sessions could not be established (p > 0.05), nor was there any statistically significant effect of indoor or outdoor setting on average heart rate achieved in all MMP tests (60s, 180s, 300s and 840s) (p > 0.05). The estimated difference in the trimmed means of MMP’s for durations P_max (i.e. 1s peak power) up until 840s during the indoor and outdoor tests are presented in Table 1. MMP achieved in the outdoor condition was higher (p < 0.05) compared to the indoor condition across all effort durations (P_max, 60s, 180s, 300s and 840s), with effect sizes being reported as small for all MMP time points and trivial for P_max. MMP was between 4.2 – 8.8% lower in the indoor settings compared to the outdoor setting for all effort durations (P_max, 60s, 180s, 300s and 840s). Additionally, CP determined from the outdoor trials (348 ± 39 W) was greater (p = 0.005) compared to CP determined from the indoor trials (330 ± 36 W), whilst for W’ a reliable difference could not be established between conditions (23.9 ± 4.0 kJ vs. 25.1 ±4.0 kJ; p = 0.332). The estimated difference in the trimmed means of cadence during the effort durations P_max up until 840s during the indoor and outdoor tests are presented in Table 2. Statistical significance was observed in average cadences between
indoor and outdoor MMP tests for both 60s and 300s \( (p < 0.05) \), but differences were not established for the same variables at 180s and 860s \( (p > 0.05) \).

[Table 1 about here], [Table 2 about here]

Discussion
The purpose of the current study was to evaluate if there are differences in MMPs of varying durations, CP and \( W' \), when performing tests either indoors or outdoors, in elite cyclists. Previously, the relative lack of studies investigating the aforementioned in an elite, or professional population has prevented the ability to make meaningful comparisons between indoor and outdoor tests of cycling performance.

The results of this study suggest that \( P_{\text{max}} \) and MMP produced over time durations between 60s and 840s are lower when completed indoors compared to MMPs produced over the same duration of testing outdoors, in elite male cyclists, and cannot be used interchangeably. This is in conflict with a similar study of Quod et al.\(^{15} \) which found no differences in power outputs between a laboratory peak power test and analysis of MMP data from races over the same time durations. Despite the findings from the Quod et al. study not being in agreement with the current findings, they observed power differences of between 23 and 85 W for durations ranging from 5s and 600s. This is similar to the current study with differences in the trimmed means of between -19 and -58 W when comparing indoor and outdoor settings.

The differences observed in the current study, suggest that comparing MMPs over different durations in indoor and outdoor settings should be done with a degree of caution, especially in an applied setting amongst an elite population, where the extent of changes in performance are smaller. Furthermore, these results suggest that the magnitude of the difference between power outputs produced either indoors or outdoors, varies considerably per individual as presented in Figure 1 (e.g. one subject had a difference between indoor and outdoor tests of 152, 48, 37 and 36 W for MMP time durations of 60s, 180s, 300s and 840s, respectively), suggesting there might be individual considerations that contribute to large differences observed between indoor and outdoor performance testing, and this warrants further investigation.

Potential factors contributing to this variation might be individual differences in motivation and desire to perform in tests indoors and outdoors, as well as being familiarised to training.
and performing maximal efforts indoors. There was a large variation in experience between the participants used in this study of riding indoors versus outdoors: for example, some subjects reported utilising the indoor trainer on average more than two times per week, whilst other subjects reported using the indoor trainer less than once per month.

CP estimations were different between indoor and outdoor tests, whilst the fixed energy reserve above CP \((W')\) was similar. Given that the MMP achieved over 180s, 300s and 840s were used in the estimation of CP, it is not surprising that the lower indoor powers achieved within the 180 – 840s duration subsequently also resulted in a lower CP indoor. When comparing this to previous literature, a similar study from Quod, found no effect of indoor or outdoor settings on CP and \(W'\) \((p = 0.32, 0.85,\) respectively), with trivial effect sizes for both CP and \(W'\) and absolute differences of -3 W and -0.3 kJ reported. Quod and colleagues employed a protocol which used race data to calculate CP and \(W'\). Due to the variability in race situations, it is likely that MMPs taken from racing are underestimating the true maximal capacity for each MMP duration. Furthermore, Leo and Spragg highlighted that it is unlikely \(W'\) being fully exhausted in a race scenario compared to a test setting 16, as well as Pinot and Grappe recommending the use of the term, “record power output” instead of MMP, due to the highest power output achieved in competition not being reflective of the true maximum capability of the cyclists 10. In the current study, the estimated difference of the trimmed means of CP in indoor and outdoor conditions was -19 W (-5.31%), whereas Karsten \((0.17 \pm 5.72\) W) and Quod found non-significant relative differences in CP of -0.8% between laboratory tests and MMP race data 2,15. Practitioners at an elite level should take care when interpreting these results, as studies within elite cycling have highlighted that modest variations in typical performance metrics such as peak power output can already lead to different performance outcomes (i.e. win vs. lose) or indeed differences between professional and elite-level cyclists 10.

Self-selected cadence, when producing maximal efforts was higher indoors compared to outdoors \((5 – 6 \text{ rev} \cdot \text{min}^{-1})\). Despite observing no differences for MMP efforts of 180s and 860s respectively, coaches and practitioners could interpret the differences in cadence observed in this study as being practically important, especially at an elite level, as studies have shown that elite cyclists perform better at their most efficient cadence 26 and that freely chosen cadence is variable between individuals and within individuals at different powers 27.
These results are similar to and strengthen the findings from Quod et al.\textsuperscript{15}, however, the timing of MMP efforts and gradient of road on which the peak power efforts were recorded in the Quod study were not controlled as stringently, due to the data being taken from a selection of races over varying durations and terrains. The MMP efforts and associated data (heart rate, cadence etc.) recorded in the aforementioned study were recorded during races, and therefore could have occurred at any time between 1 and 5 hours of racing. Self-selected cadences have been reported as decreasing as much as 18 rev-min\textsuperscript{-1} over 2 hours of continuous cycling, in part due to neuromuscular fatigue\textsuperscript{28}, as well as freely chosen pedalling rates varying with differing terrains\textsuperscript{20}. In the current study both indoor and outdoor MMP test efforts were completed within 2 hours of exercise, on a fixed gradient outdoors and at a fixed ergometer-resistance indoors. Therefore, it is reasonable to assume that terrain and neuromuscular fatigue were not factors influencing the drop in cadence of the participants. One potential contributing factor to the higher cadences seen indoors is different crank inertial load (CIL). In outdoor cycling, the velocity and mass of cyclist and bicycle impact kinetic energy. On a bicycle fixed to an ergometer, as used by the subjects in the current study, the rotational kinetic energy of the flywheel varies with inertia and rotational speed, dependant on the make and model of such ergometer devices, as well as gear selection. The CIL varies considerably, dependant on mass of the cyclists and gear ratio\textsuperscript{30}, with CIL being as much as 50\% lower on a Kingcycle ergometer compared to field cycling. The CIL has also been shown to affect the self-selected pedal rate of untrained and trained\textsuperscript{31} cyclists. Related to this, studies have suggested that laboratory cycling conditions elicit significantly different crank torque profiles compared to road cycling conditions, partly due to the stiffness and lack of side-to-side movement of the ergometers used in those studies\textsuperscript{5}. Although the current study utilised the cyclists own bicycles attached to the ergometer, there is still a relative lack of movement and increased stiffness and dampening of the bicycle, which could impact the crank torque profile and CIL. Hence, differences in CIL between indoor and outdoor settings may contribute to the differences in self-selected cadences observed in this study. The potential impact of on test performance (power output) between the difference conditions, especially at a shorter duration (< 60s) requires further investigation. It remains questionable as to the long-term effects of differing CIL between indoor and outdoor cycling, the subsequent impact that it would have on the physiological and biomechanical performance characteristics as a result\textsuperscript{31}, and continues to be a topic for future investigation.

\textit{Limitations}
As a result of the time constrains and locational differences where each subject conducted their outdoor testing, we were unable to control for environmental conditions. Furthermore, differences in wind speed and direction might have had an impact on power output and pacing strategies of the participants. However, to control for wind conditions below 6.6 m·s\(^{-1}\) as recommended by Karsten et al.\(^2\) would likely incur tests being completed in a greater time frame than the 14 days suggested in the methods section, and the likely drop out of participants due to their own training and racing constraints, as set out by their respective teams. Karsten et al. did find strong correlation in tests in both laboratory and field conditions for CP (\(r^2 = 0.99 \pm 0.008\)), this is also reflected in the current studies r-value for CP (\(r^2 = 0.99 \pm 0.0009\)) suggesting that environmental differences between subjects’ tests did not significantly impact CP in field conditions.

**Conclusions**

This study in elite and World Tour cyclists showed that performance tests of varying durations (60 – 840s) as well as \(P_{\text{max}}\) and CP are lower when performed indoors compared to outdoor settings. Furthermore, the large individual differences in absolute power across all time points suggests interpreting the comparison between performance trials indoors and outdoors with some degree of caution, especially in well-trained populations. MMP power profiling and the subsequent performance indicator, critical power should not be used interchangeably between indoor and outdoor testing. Further research is required to investigate the potential factors contributing to differences of indoor and outdoor cycling such as mechanistic effects of cadence, CIL and other biomechanical factors.
References


Table 1: Maximum Mean Power (MMP) for conditions Indoor and Outdoor for durations of 60-840s, peak power ($P_{\text{max}}$), total work, critical power (CP) and W’ ; n=14

Table 2: Average cadences (rev•min$^{-1}$) for indoor and outdoor conditions recorded from peak power tests from durations of 60-840s ; n=14

Figure 1: Illustration of the individual (dashed line) differences between indoor and outdoor power profiling tests and mean differences (bar) for durations of 60s (A), 180s (B), 300s (C) and 840s (D). *Significantly different ($p \leq 0.05$)