Mean stability and between-session reliability of cycling biomechanics variables in elite pursuit cyclists.

WILLIAMS, Daniel, DUNN, Marcus <http://orcid.org/0000-0003-3368-8131>, WORSFOLD, Paul, NEWTON, Deborah, FAULKNER, Steve H, D'ANDREA, Francesca <http://orcid.org/0000-0001-9678-5612> and WHEAT, Jonathan <http://orcid.org/0000-0002-1107-6452>

Mean Stability and Between-Session Reliability of Cycling Biomechanics Variables in Elite Pursuit Cyclists

Daniel Williams[1,2,3], Marcus Dunn[4], Paul Worsfold[2,5], Deborah Newton[2,3], Steve H Faulkner[3,6], Francesca d'Andrea[1], Jon Wheat[7]

1. Sports Engineering Research Group, Health Research Institute, Sheffield Hallam University, Sheffield, United Kingdom; 2. UK Sports Institute, Manchester, United Kingdom; 3. Great Britain Cycling Team, Manchester, United Kingdom; 4. School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, United Kingdom; 5. School of Allied and Public Health, University of Chester, Chester, United Kingdom; 6. Department of Engineering, School of Science & Technology, Nottingham Trent University, Nottingham, United Kingdom; 7. Department of Sport Science, School of Science & Technology, Nottingham Trent University, Nottingham, United Kingdom

Twitter/X: @danjwilliams99, @marcusdunnphd, @debssides, @shfaulkner, @wheat_jon

ORCID: Daniel Williams: 0009-0004-5226-7492, Marcus Dunn: 0000-0003-3368-8131, Paul Worsfold: 0000-0002-1712-9818, Steve H Faulkner: 0000-0003-4688-7252, Francesca d'Andrea: 0000-0001-9678-5612, Jon Wheat: 0000-0002-1107-6452

Corresponding Author: Daniel Williams; danjwilliams99@gmail.com

Key Words: biomechanics, cycling, reliability, intraclass correlation coefficient, sequential averaging

Abstract:

The purpose of this study was to determine the number of crank revolutions required to obtain stable mean values of sagittal plane biomechanics variables, and the between-session reliability of these variables, whilst cyclists used an aerodynamic position. Eighteen elite cyclists completed a 3-minute maximal bout on a cycling ergometer. Lower-limb kinematic and kinetic data were captured using 2D motion capture and force pedals. Raw data were filtered using a 4th order Butterworth low-pass filter (6 Hz) and interpolated to 100 points per revolution. The middle 60 revolutions of each trial were extracted and 37 discrete and 15 time-series variables were calculated. Mean stability was assessed in all participants and between-session reliability was analysed in a subset of 11 participants. Sequential averaging indicated more revolutions to stability than iterative intra-class correlation coefficients. Crank kinetics were more stable than joint kinematics and kinetics. For stable discrete and time-series variables, 30 and 38 revolutions are recommended, respectively. Between-day reliability for all variables was moderate to excellent, and good to excellent for crank kinetics and joint kinematics variables. Hip flexion-extension and ankle dorsiflexion kinetics were least reliable. Researchers and applied practitioners should consider these findings when planning, and interpreting results from, cycling biomechanics interventions.

Introduction

The aerodynamic cycling position (AP) requires the cyclist to lower their torso and rest their forearms on pads while they grip extended handlebars. The AP is predominantly used in competitive events such as the road time-trial, triathlon, and in the track cycling team pursuit and 1km time-trial, with cyclists benefitting from a 20 - 30% reduction in drag area compared to typical road cycling positions (Defraeye et al., 2010). Further optimisation in the AP through reduced handlebar height and greater reach can lead to an additional aerodynamic benefit of 6 - 14% (Garcia-Lopez et al., 2008; Oggiano et al., 2008). Whilst minimising resistive forces, the lower torso angles in the AP - which can be as low as 4° to 12° in elite time trialists (Cubel et al., 2022) - reduces power output and gross efficiency (Fintelman et al., 2015; Kordi et al., 2019). Aerodynamic gains often outweigh losses in power (Underwood et al., 2011) but competitive success requires a balance of reducing aerodynamic drag and maintaining or improving biomechanical power output. Limited research has reported that reduced torso angles alter pedal force application and joint kinetic contributions, associated with increased hip flexion and ankle dorsiflexion (Bini et al., 2019; Jongerius et al., 2022). However, the acute biomechanical responses to changes in handlebar position in the AP, and the optimal trade-off with aerodynamics, are not completely understood.

To maximise power output, cyclists must generate high torque through the crank at a fast but sustainable crank angular velocity (Douglas et al., 2012). Torque is the product of tangential force applied to the crank at the pedal and the crank length. Any force applied through the pedal along the crank arm (known as radial force) has no direct mechanical benefit. As such, the ratio of tangential to total force - known as the index of force effectiveness (IFE) – is considered a useful metric and has been used to compare cycling positions (Jongerius et al., 2022). Additionally, the lower-limb muscles, particularly the hip (gluteus maximus) and knee extensors (vastii), are the predominant power generators in cycling (van Ingen Schenau et al., 1990). Therefore, kinematic and kinetic analysis techniques are often implemented to understand the effects of riding position on lower extremity biomechanics (Bini et al., 2010). As this lower extremity motion occurs primarily in the sagittal plane (Umberger & Martin, 2001), two-dimensional motion analysis is commonly used to provide accurate analysis of cycling biomechanics (Garcia-Lopez & del Blanco, 2017). While previous research has described the crank kinetics, joint kinematics, and joint kinetics associated with standard cycling positions, these variables have not been fully investigated in the AP.

In cycling biomechanics studies, researchers typically arbitrarily select and capture a section of an exercise bout for further analysis, which can range from 10-seconds (Ferrer-Roca et al., 2014) to 1-minute (Jongerius et al., 2022), equating to approximately 15 to 95 pedal revolutions. However, the number of repetitions collected in a study can affect the stability of the dependent variables (Bates et al., 1992). Statistical techniques have been proposed to identify the number of repetitions of a movement required for stable mean estimates of dependent variables, including sequential averaging (SA) and iterative intra-class correlation coefficient (IICC) analyses (Bates et al., 1983; James et al., 2007). Initially these methods were applied to discrete dependent variables during overground walking and running experiments, in which it was only possible to capture one stride per trial. More recently, Taylor et al (2015) investigated the stability of time-series data, which enables further understanding of biomechanical information outside of pre-determined discrete data-points (Pataky, 2010). SA and IICC have also been applied to repetitive actions, such as continuous jumping (Racic et al., 2009) and treadmill running (Riazati et al., 2019). However, no investigation has been applied to cycling to understand the number of crank revolutions required to achieve stable dependent variables.

It is also important to understand the between-session reliability of biomechanical dependent variables during cycling, to support researchers to attribute biomechanical changes to an intervention or measurement uncertainty. Good within and between-session reliability has been reported for muscular activation (Jobson et al., 2013), pedal forces (Bini & Hume, 2013) and combined two-dimensional kinematic and kinetic variables (Burnie et al., 2020). However, these are three of only a few studies that explore the reliability of cycling biomechanics variables. To the authors' knowledge, the latter study is the only study exploring the reliability of data obtained from both motion capture and instrumented pedal techniques (and calculated joint kinetics). In addition, there are no studies that investigate the reliability of biomechanical variables captured during cycling in the AP. While Burnie et al. (2020) found good to excellent between-session reliability in their study, they investigated sprint cycling. The intensity, duration, and riding position in sprint cycling are different to that of pursuit cycling in the AP, and therefore findings may not be fully representative. An understanding of the between-session reliability of biomechanics dependent variables captured during cycling in the AP would add to existing knowledge on the reliability of cycling biomechanics data.

The aim of this study was two-fold. First, to determine the number of successive crank revolutions required to achieve stable mean estimates of discrete and continuous biomechanical dependent variables during maximal cycling in the AP. Second, to quantify the between-session reliability of such outcomes captured during a 3-minute maximal effort in the AP over two testing sessions.

Materials and Methods

Study Design

A test-retest, within-subjects design was used to understand the aims of the study. Participants were asked to complete a maximal self-paced 3-minute effort on two occasions. Mean stability of the chosen dependent variables was assessed from data collected on their first visit and between-session reliability was assessed from data collected on both visits. Ethical approval for all study procedures was granted by the Sheffield Hallam University Research Ethics Sub-Committee prior to recruitment (ER50767037).

Participants

Eighteen elite cyclists (10 females and 8 males; age: 22 ± 7 years; height: 1.74 ± 0.10 m; mass: 69.0 ± 9.5 kg) participated in this study, and a subset of 11 cyclists were included in the reliability analysis. All participants volunteered and provided informed written consent before participation. The participants were members (one former) of the national cycling team and experienced at cycling in the AP, having competed in a time-trial and/or pursuit race at senior (8), under-23 (4) or junior (6) international level. They were all free from any current injury, physical impairment or health condition that might have limited participation in the study.

Equipment

All trials were completed on a customised SRM cycling ergometer (Julich, Germany) which allowed cadence to be fixed - full details have been described previously (Burnie et al., 2020). To ensure riding position was identical, the ergometer was set-up with the same saddle and measurements from the participants' track bicycles. Instrumented force pedals (Sensix, Poitiers, France), with LOOK or Shimano pedal adapters, were used at each cyclists desired crank length to measure pedal reaction forces and an encoder on the crank (Model LM13, RLS, Komenda, Slovenia) measured crank position. Both pedal reaction forces and crank position were sampled at 200 Hz. A high-speed camera (Quintic, Birmingham, UK) recorded sagittal plane motion of retroreflective markers at 100 Hz and all data were captured within CrankCam software (Sports Engineering Research Group, Sheffield Hallam University, UK).

Experimental Protocol

Each participant was asked to attend two testing sessions, though due to training and competitive requirements only 11 were able to complete both sessions. Upon arrival, they completed a 20-minute progressive cycling warm-up on a turbo trainer – the same warm-up that they would complete before performing pursuit training or race efforts. Following this, reflective markers were placed on four anatomical landmarks of both lower limbs by the same experienced investigator on each occasion. A marker was also placed on the pedal spindle to enable calculation of an ankle angle (**Figure 1**).

** FIGURE 1 HERE ***

Before data collection, participants performed a 30-second familiarisation trial on the ergometer at the same cadence and intensity as the experimental trial. The experimental trial consisted of a maximal self-paced 3-minute bout at a fixed cadence of 100 to 105 rpm (103 ± 2 rpm). As individual pursuit distance and duration varies based on competitive age group and sex, 3-minutes was selected as a compromise to best reflect the whole participant sample. The middle of the trial was deemed the most representative section, so the middle 40-seconds (1:10 - 1:50) – containing at least 60 revolutions – were captured for analysis. This process was repeated on the participants' second visit to assess between-session reliability – approximately 6 ± 5 days later and at a similar time of day (1 ± 1 hour difference).

Data Analysis

All data processing and analysis was completed using custom Python scripts (Version 3.9, Spyder IDE). Raw sagittal plane kinetic and kinematic data were filtered with a Butterworth fourth order (zero-lag) low-pass filter. A cut-off frequency of 6 Hz was chosen following a residual analysis (Winter, 2009). The same cut-off frequency was used for kinetic and kinematic data to avoid errors in calculated joint kinetics (Bezodis et al., 2013).

Individual crank revolutions were identified using crank position data (0 - 360°) and the middle 60 revolutions were selected for further analysis. Data were interpolated to 100 data points per revolution (% crank cycle) using cubic spline interpolation. For the crank kinetics dependent variables, the peak and mean values were calculated for each revolution, whereas peak extension and peak flexion values were calculated for the joint kinematic and kinetics outcome variables. In addition, the full 100-point time-series of every revolution was recorded for each continuous dependent variable. A total of 37 discrete and 15 time-series variables were calculated for further analysis.

Crank power output was calculated as the sum of left and right crank power; both a product of the respective left and right crank torques and crank angular velocity. Due to a technical issue with the left force pedal, only right crank power output was compared between-days, whereas total crank power output was included in the stability analysis. A further issue with the force pedals data for two participants on their second visit meant that the crank kinetics data of only 9 of the 11 participants could be included in the between-session reliability protocol, while all 11 participants' joint kinematics data were analysed. All other variables included in the stability and reliability analyses were from the right limb and crank as symmetry was assumed (Garcia-Lopez et al., 2015). Effective and ineffective force were defined as force produced tangentially and radially to the direction of crank angular rotation, respectively (**Figure 1**). For each crank cycle, IFE was calculated as the ratio of the area under the raw effective force-time curve relative to the area under the resultant force-time curve, with resultant force being the vector sum of effective and ineffective force.

Kinematic variables derived from joint angles and angular velocities of the ankle, knee, and hip were calculated (**Figure 1)**. Joint moments, powers, and work were calculated using standard inverse dynamics techniques (Elftman, 1939). For joint angular velocities, change in the clockwise direction is negative and positive in the anticlockwise direction. Therefore, for angular velocities and moments, the extension peak was the greatest magnitude in the negative direction for the hip and ankle and the positive direction for the knee, and vice-versa with respect to flexion. Joint powers were calculated as a product of joint moment and joint angular velocity, and the peak magnitudes during periods of flexion or extension were extracted - based on the direction of joint angular velocity (Martin & Brown, 2009). Joint work was calculated as the integral of joint power within phases of both flexion and extension periods at each joint.

Statistical Analysis – Mean Stability

Both SA (Bates et al., 1983) and IICC (James et al., 2007) were performed to assess the mean stability of each dependent variable. For SA, an initial number of 60 reference revolutions was selected. The 60-revolution mean value for each variable was subtracted from the cumulative mean at each consecutive revolution (1 + 2, 1 + 2 + 3, etc.) to provide a mean deviation value. The mean of a variable was considered stable at one revolution greater than the revolution where the mean deviation fell, and remained, below 0.25 of the 60-revolution standard deviation (0.25SD; Bates et al., 1983). To perform SA on time-series data, all 100 points were treated as a discrete point (Taylor et al., 2015). Stability was defined as one revolution greater than the revolution where all 100 points had fell below the threshold. The effect of using different SD thresholds (0.25, 0.41, 0.49 and 0.60) and numbers of reference revolutions (20, 30, 40, 50 and 60) was also investigated (Forrester, 2015). The different SD thresholds of 0.41, 0.49 and 0.60 were identified from previous studies as they approximated IICC results (James et al., 2007; Racic et al., 2009). The relative SA score was calculated by dividing the number of revolutions indicated to be required for stability by the number of reference revolutions (Taylor et al., 2015). This was included to understand whether an increased number of repetitions to stability was related to measurement instability or a direct result of increasing reference revolutions.

The IICC method involved calculating ICCs for each variable across an iteratively increasing (by 1) number of consecutive crank revolutions up to the full 60-revolution sample (James et al., 2007). The model used was ICC (3, 1) for single measures and consistency (Trevethan et al., 2017). The maximum ICC value, the 95% confidence interval and the number of revolutions to reach the maximum ICC were calculated for each dependent variable. In addition, the number of revolutions for the ICC (and the 95% CI lower-bound as a more conservative estimate) to reach 0.80 and 0.90 were recorded, which were interpreted as thresholds for good (James et al., 2007) and excellent (Koo and Li, 2016) stability, respectively.

Statistical Analysis – Between-Session Reliability

The procedures for investigating between-session reliability followed the recommendations of Atkinson and Nevill (1998). To investigate systematic bias, paired samples *t*-tests were performed on the discrete values and on the time-series data using statistical parametric mapping (SPM; Pataky, 2010). Alpha (α) was set to 0.05 to infer statistical significance. ICC (3, 1) analysis for single measures and absolute agreement was used to determine an ICC for each discrete variable (Trevethan, 2017). Integrated pointwise indices were performed as described by Pini et al. (2022) to assess the reliability of the time-series variables. Briefly, this involved treating all fixed points along the time-series as a discrete point and performing ICC (3, 1) groupwise on each discrete point. The interpretation of the ICC was based on guidelines presented by Koo and Li (2016): less than 0.50 (poor), 0.50 to 0.75 (moderate), 0.75 to 0.90 (good) and greater than 0.90 (excellent). Standard error of measurement (SEM) was used to assess absolute reliability, after confirming on the absence of heteroscedasticity (Atkinson & Nevill, 1998). This was determined as $SD\sqrt{(1\text{-}ICC)}$, where SD is the standard deviation of the scores from all participants (Weir, 2005). The minimal detectable difference (MDD) for each measurement was calculated as SEM × 1.96 × $\sqrt{2}$ (Weir, 2005).

Results

Mean Stability:

Mean revolutions to stability reduced when the number of reference revolutions was decreased and when the SD threshold increased (**Figure 2**). The mean relative SA score generally reduced as the number of reference revolutions was increased to 30 or 40 revolutions, and then this either stabilised or began to increase again. The SA results demonstrated that time-series variables required more revolutions to reach stability than discrete variables, and this is further evidenced in **Table 1.**

The 60 reference revolution results show that all discrete variables reached stability by 47 revolutions and all time-series variables by 60 revolutions, but this differed depending on the type of variable (**Table 1**). All discrete crank kinetics and joint kinematics outcomes reached stability by 36 revolutions, while all time-series crank kinetics variables required 52 revolutions and time-series joint kinematics required 57 revolutions.

The IICC results showed that all discrete variables, excluding peak knee flexion angular velocity (32 revolutions), reached maximum stability in 11 revolutions (**Table 1**) and excellent stability (ICC ≥ 0.90) within 4 revolutions. However, peak knee flexion power did not reach excellent stability (Max. ICC = 0.837). All 37 discrete outcomes reached good stability (≥ 0.80) within 3 revolutions. Using the more conservative lower bound of the ICC 95% confidence interval excellent stability was not reached in 14 variables (8 of which were joint kinetics) – only peak knee flexion power did not reach good stability.

*****FIGURE 2 HERE *****

***** TABLE 1 HERE *****

Between-Session Reliability

No significant differences were identified between-days for the crank kinetics variables (**Table 2**). All discrete crank kinetics variables demonstrated excellent between-session reliability (ICC > 0.9), apart from IFE which displayed good reliability (ICC = 0.766).

Peak ankle plantarflexion and knee flexion angles were significantly reduced on Day 2, by 1.4° and 1.0° respectively (**Table 2**). However, no other discrete joint variables were found to be significantly different between days. The ICC results showed that the joint kinematic variables had good to excellent between-session reliability. Peak joint angles had MDDs from 0.9 to 3.2°, and peak joint angular velocities had MDDs from 4.6 to 19.6°/s.

Joint moments showed moderate to excellent reliability, with ankle dorsiflexion, hip flexion and hip extension moments having relatively large MDDs of 2.5, 13.1 and 17.3 Nm, respectively. All peak joint powers displayed good to excellent reliability, apart from peak hip extension power with only moderate reliability (ICC = 0.581 and MDD = 59.1 W). Joint work outcomes had moderate to good between-session reliability. Knee flexion work had a large MDD of 10.2 J, compared to a magnitude of only 4.9 and 2.4 J on Day 1 and 2, respectively.

***** TABLE 2 HERE *****

Across the whole crank cycle, no significant differences were found between-sessions for right crank power or effective and ineffective force time-series **(Figure 3)**. Crank power and effective force showed good time-series reliability, with mean ICC values of 0.794 and 0.807, respectively. The pointwise ICC (3, 1) was excellent during the downstroke phase (13 - 58%) and moderate to good during the rest of the crank cycle. Ineffective force also showed good reliability (mean ICC = 0.862), with good to excellent reliability for the whole crank cycle, apart from 75 - 93%.

***** FIGURE 3 HERE *****

For the joint time-series data, only the ankle angle time-series showed a significant difference between-days, with a significant reduction on Day 2 between 90% to 1% of the crank cycle, indicated by the shaded region (**Figure 4**).

**** FIGURE 4 HERE *****

Ankle joint angle and angular velocity showed excellent mean ICCs (**Figure 5**). However, knee and hip kinematics demonstrated good mean reliability (ICC > 0.75). Mean reliability for all joint moments was good, but during periods of the crank cycle this dropped to moderate and to poor for the knee in one period (61 - 68%). Ankle and knee joint powers had good mean ICC, with periods of the cycle in which reliability was moderate or poor. Hip joint power had only moderate mean ICC with moderate reliability during most of the crank cycle.

**** FIGURE 5 HERE *****

Discussion and Implications

The first aim of this study was to define the number of crank revolutions required to achieve stable biomechanical dependent variables during cycling in the AP. We demonstrated the first application of SA and IICC in cycling biomechanics research. There were differences between the SA and IICC techniques and between categories of dependent variables (crank kinetics, joint kinematics and joint kinetics) in the number of revolutions identified as being required for stability. The second aim was to quantify the between-session reliability of cycling biomechanics variables in the AP. All variables showed moderate to excellent between-session reliability, with crank kinetics and joint kinematics demonstrating good to excellent reliability.

SA was most conservative in assessing stability, with all discrete and time-series variables reaching stability by 47 and 60 revolutions, respectively, when using 60 reference revolutions (**Table 1**). As expected, the number of revolutions SA indicated were required for stability reduced with fewer reference revolutions (Forrester, 2015; Taylor et al., 2015) and greater SD thresholds (**Figure 2**; James et al., 2007; Racic et al., 2009). Despite this, increasing the SD threshold to 0.60SD was still more conservative than IICC when using 60 reference revolutions. As in previous research, the number of revolutions IICC indicated were required for stability was lower than SA (James et al., 2007). All discrete variables – except peak knee flexion velocity – reached maximum stability within 11 revolutions. IICC therefore indicated that, counterintuitively, when the number of reference revolutions increased beyond 11, outcome stability did not improve. Previous research has also shown IICC does not demonstrate a pattern of increased stability when the number of repetitions of a movement is increased (Chua et al., 2017). IICC may have assigned stability too soon - when the mean deviation was still converging (Taylor et al., 2015). Previous research has also found IICC underestimates the number of trials to stability when high between-participant variability is present (James et al., 2007; Coleman et al., 2023). Due to these limitations, further discussion and recommendations are based on the SA results.

The relative SA score (Taylor et al., 2015) was highest with 20 reference revolutions in all variables - except the joint kinematics time-series variables (**Figure 2**). The relative SA score then reduced with 30 and 40 reference revolutions before stabilising from 50 to 60 reference revolutions in all variables - except for the crank and joint kinetics time-series variables. When the relative SA score remains unchanged, any increase in the number of trials to stability is a direct result of increasing the number of reference trials (Taylor et al., 2015). For the crank and joint kinetics time-series variables, the relative score was lowest at 30 and 40 but instead increased at 50 and 60 reference revolutions, respectively. This suggests these variables became less stable with greater than 40 revolutions. During a maximal self-paced cycling effort, it is reasonable to assume that increasing the number of reference revolutions will, at some point, reduce the stability of dependent variables because of fatigue (Galindo-Martinez et al., 2021). Taylor et al., (2015) recommended that there should be an upper limit on the number of reference revolutions when performing SA. They suggested 30 trials as the maximum for completing SA during an overarm throwing task because the relative SA score did not reduce from 20 to 30 reference repetitions. In our study, 60-revolutions appears to have been an excessive initial selection because relative SA did not reduce for reference revolution sizes greater than 40, and in some cases increased. Using a reference revolution size of 40, all discrete and all time-series variables reached stability at 30 and 38 revolutions, respectively.

In the between-session reliability analysis, maximum ankle plantarflexion and knee flexion demonstrated a systematic difference between testing sessions, but the mean differences were small. In addition, the SPM analysis identified that ankle joint angle was lower on Day 2 than Day 1 between 90% to 1% of the crank cycle (**Figure 4**). The reasons for these systematic differences are unclear. They could be related to a learning or fatigue effect - as the activity of participants between sessions could not be controlled - but it is not clear why this would systematically affect these variables. A learning effect would appear unlikely due to the high experience level of the participants. On the other hand, the differences were small (1.0° and 1.4° for maximum ankle plantarflexion and knee flexion, respectively, and approximately 1.5° for ankle joint angle time series; **Table 2 & Figure 4**) and there is a potential for a Type 1 error given the large number of variables analysed and relatively small sample of participants (Brysbaert, 2019).

No variables showed poor between-session reliability (ICC < 0.50). This is in accordance with previous research that reported moderate to excellent between-session reliability for 2D sagittal plane cycling biomechanics variables (Burnie et al., 2020). In the present study, crank kinetics variables were most reliable, again in agreement with previous research that indicated pedal forces are highly reliable between-days (Bini et al., 2013). Joint kinematics reliability was good to excellent, with good reliability for knee and hip kinematics and excellent reliability for ankle kinematics variables - except for peak ankle dorsiflexion angle (ICC = 0.898; **Table 2**). The reduced reliability at the knee and hip might be caused by increased marker misplacement error and soft tissue movement artefact at the hip and pelvis (Stagni, 2000; Li et al., 2017). Li et al. (2017) found that soft tissue artefact was greater for the greater trochanter than the lateral femoral epicondyle and malleolus during cycling. This increased soft tissue also contributes to greater marker misplacement (Stagni, 2000). Contrary to our findings, Burnie et al. (2020) reported that, in sprint cycling, ankle kinematics were less repeatable than knee and hip kinematics. However, the poorer reliability at the ankle was attributed to the prominent role of the ankle in power transfer, as well as generation, in maximal sprint cycling (Burnie et al., 2020).

Several joint kinetics variables demonstrated only moderate reliability with relatively large minimal detectable differences. All hip joint kinetics variables, except maximum hip flexion power (ICC = 0.772), showed only moderate between-session reliability (**Table 2**). This might partly be explained by the increased soft tissue artefact and marker misplacement error at the hip. During gait, 30 mm anterior-posterior greater trochanter location error leads to differences in hip flexion-extension moments of ~22 - 25% (Stagni, 2000). In addition, the hip is at the proximal end of the lower extremity kinetic model. An implication of the iterative Newton-Euler method of inverse dynamics calculations used in our data analysis (and standard across biomechanics research) is that errors from calculations at the ankle and knee propagate up the kinetic chain to the hip (Kuo, 1998). Other joint kinetics variables - ankle dorsiflexion moment and work, and knee flexion work - also had only moderate reliability. Since ankle kinematics were highly reliable, it is more likely that poorer reliability for ankle dorsiflexion kinetics variables is due to the small magnitude and trivial contribution to overall limb work (-1 J; **Table 2**). The small magnitude of ankle dorsiflexion kinetics was accompanied by low between-participant variability, which is associated with smaller ICCs (Portney & Watkins, 2000; Russek, 2004). This would also explain the period of very poor reliability in the ankle power time-series (87 - 97%) that occurred when there was low magnitude (~0 W) and low between-participant variability. Similarly, between-participant variability in knee power was low at 42 - 44% where pointwise reliability was poor, and at 62 - 67% values were of low magnitude and both positive or negative (**Figure 4 & 5**). This might also explain the moderate reliability identified for knee flexion work. The lower reliability and large MDDs associated with these joint kinetics suggests researchers should be cautious when interpreting changes in these variables.

Stability and between-session reliability were not consistent across different categories of dependent variable. Crank kinetics were more stable and reliable than joint kinematics or kinetics. They reached stability in 21 and 30 revolutions for discrete and time-series outcomes, compared to 26 and 38 for joint kinematics and 30 and 33 revolutions for joint kinetics, respectively. This is likely to be because joint kinematics and kinetics are subject to larger measurement errors. Joint kinematics are affected by error from optical measurement systems, such as soft tissue artefact and marker misplacement. What is more, joint kinetics are influenced by measurement errors from both kinematics and kinetic sources (Camomilla et al., 2017). Interestingly, previous research has shown joint movements to exhibit greater within participant variability than performance outcome measures, suggesting compensatory joint movements may be necessary to achieve consistent outcomes (Bradshaw et al., 2007). Our results are congruent with this and hint that elite cyclists vary their lower-limb movement patterns to achieve lower variability in crank kinetics.

The sample of participants included in this study, and the chosen protocol, could have affected the findings, and the applicability of the mean stability and repeatability results in future experiments. For example, a maximal self-paced intensity was used in the experimental trials rather than a lower, fixed intensity. A lower, fixed intensity might have reduced within-participant and cycle-to-cycle variability which would lead to mean stability being achieved in fewer revolutions and a greater between-session reliability. In addition, while the study did not investigate sex differences, the inclusion of male and female cyclists could have increased the between-participant variability. This could have inflated the calculated ICC statistics in both the stability and reliability analyses (Weir, 2005), leading to mean stability being reached sooner and a greater estimation of between-session reliability. Similarly, while all cyclists were considered to be elite, there was a range of competitive age groups and performance capabilities within the participant cohort. Again, this might have increased between-participant variability and influenced calculated ICC values, and a more homogenous participant group might have indicated poorer mean stability and between-session reliability. Researchers should consider these factors when designing future experiments based on the recommendations in this study.

While 18 participants represent a substantial sample size in the context of intrasession stability research and research involving elite cyclists, this study is limited by the relatively small subset of participants (n = 11) in the assessment of between-session reliability. Previous cycling biomechanics reliability and repeatability studies have included between 10 and 24 participants (Bini & Hume, 2013;

Dorel et al., 2008; Burnie et al., 2022; Jobson et al., 2013), so our sample of 11 and 9 cyclists for the kinematic and kinetic analysis, respectively, is relatively low. As the participants in our study were from a small elite cycling population, training and competitive demands made completion of a second, repeated data collection difficult. This smaller sample size might have reduced the statistical power of the tests of between session differences (Brysbaert, 2019). Relatedly, it was also not feasible to control participants' activity between testing sessions. All tests were completed at a similar time of day (1 ± 1 hour difference) and with the same warm-up protocol, but some participants may have been more fatigued on a given testing session, potentially altering their pedalling biomechanics. The difficulties in recruitment and control of extraneous variables when conducting research with elite athletes are well understood, yet the novelty and potential benefits and applications of research on those with elite skill and fitness levels should not be ignored (Sands et al., 2005).

Conclusions

Based on a reference revolution size of 40 pedal revolutions, we recommend that a minimum of 30 and 38 consecutive crank revolutions are required to achieve stable discrete and time-series sagittal plane dependent variables, respectively, during cycling in the AP. Fewer revolutions are required if solely interested in crank kinetics - 21 and 30 revolutions for discrete and time-series variables, respectively. Future investigations should consider these recommendations; however, they might not apply to investigations of other cycling disciplines and further research could explore whether mean stability is influenced by fatigue, intensity, skill-level or riding position. All dependent variables showed moderate to excellent between-session reliability, but we suggest caution is required when interpreting differences in hip joint flexion-extension kinetics, ankle dorsiflexion kinetics and knee flexion work. Researchers and practitioners should consider these reliability findings when interpreting biomechanical results in future interventions.

Funding Details

This study is part of a PhD programme of research jointly funded by British Cycling and Sheffield Hallam University.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26, 217-238.

Bates, B.T., Dufek, J.S. & Davis, H.P. (1992). The effect of trial size on statistical power. *Medicine & Science in Sports & Exercise*, 24(9), 1059-1065.

Bates, B.T., Osternig, L.R., Sawhill, J.A. & James, S.L. (1983). An assessment of subject variability, subject-shoe interaction, and the evaluation of running shoes using ground reaction force data. *Journal of Biomechanics*, 16(3), 181-191.

Bezodis, N. E., Salo, A. I., & Trewartha, G. (2013). Excessive fluctuations in knee joint moments during early stance in sprinting are caused by digital filtering procedures. *Gait & Posture*, 38(4), 653-657.

Bini, R.R., Tamborindeguy, A.C. & Mota, C.B. (2010). Effects of saddle height, pedaling cadence, and workload on joint kinetics and kinematics during cycling. *Journal of Sport Rehabilitation*, 19(3), pp.301-314.

Bini, R.R. & Hume, P.A. (2013). Between-day reliability of pedal forces for cyclists during an incremental cycling test to exhaustion. *Isokinetics and Exercise Science*, 21(3), 203-209.

Bini, R. R., Daly, L., & Kingsley, M. (2019). Muscle force adaptation to changes in upper body position during seated sprint cycling. *Journal of Sports Sciences*, 37(19), 2270-2278.

Bradshaw, E. J., Maulder, P. S., & Keogh, J. W. (2007). Biological movement variability during the sprint start: Performance enhancement or hindrance? *Sports Biomechanics*, 6(3), 246-260.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1).

Burnie, L., Barratt, P., Davids, K., Worsfold, P., & Wheat, J. (2023). Biomechanical measures of short-term maximal cycling on an ergometer: a test-retest study. *Sports Biomechanics*, 22(8), 997-1015.

Camomilla, V., Cereatti, A., Cutti, A. G., Fantozzi, S., Stagni, R., & Vannozzi, G. (2017). Methodological factors affecting joint moments estimation in clinical gait analysis: a systematic review. *Biomedical Engineering Online*, 16, 1-27.

Chua, Y. K., Quek, R. K., & Kong, P. W. (2017). Basketball lay-up–foot loading characteristics and the number of trials necessary to obtain stable plantar pressure variables. *Sports Biomechanics*, 16(1), 13-22.

Coleman, J. L., Huynh, M., & Middleton, K. J. (2023). Number of trials necessary to achieve a representative performance of accuracy and timing during combat shooting. *Scandinavian Journal of Medicine & Science in Sports*, 33(10), 2039-2045.

Cubel, C., Piil, J. F., & Nybo, L. (2022). Time Trial positioning in elite cyclists-exploring the physiological effects of adapting to a lower torso position. *Journal of Science and Cycling*, 11(3), 67-75.

Defraeye, T., Blocken, B., Koninckx, E., Hespel, P., & Carmeliet, J. (2010). Aerodynamic study of different cyclist positions: CFD analysis and full-scale wind-tunnel tests. *Journal of Biomechanics*, 43(7), 1262-1268.

Dorel, S., Couturier, A., & Hug, F. (2008). Intra-session repeatability of lower limb muscles activation pattern during pedaling. *Journal of Electromyography and Kinesiology*, 18(5), 857-865.

Douglas, J., Ross, A., & Martin, J. C. (2021). Maximal muscular power: lessons from sprint cycling. *Sports Medicine-Open*, 7, 1-15.

Elftman, H. (1939). Forces and energy changes in the leg during walking. *American Journal of Physiology-Legacy Content*, 125(2), 339-356.

Ferrer-Roca, V., Bescós, R., Roig, A., Galilea, P., Valero, O. & García-López, J. (2014). Acute effects of small changes in bicycle saddle height on gross efficiency and lower limb kinematics. *The Journal of Strength & Conditioning Research,* 28(3), 784-791.

Fintelman, D. M., Sterling, M., Hemida, H., & Li, F. X. (2015). The effect of time trial cycling position on physiological and aerodynamic variables. *Journal of Sports Sciences*, 33(16), 1730-1737.

Forrester, S. E. (2015). Selecting the number of trials in experimental biomechanics studies. *International Biomechanics*, 2(1), 62-72.

Galindo-Martínez, A., López-Valenciano, A., Albaladejo-García, C., Vallés-González, J. M., & Elvira, J. L. (2021). Changes in the trunk and lower extremity kinematics due to fatigue can predispose to chronic injuries in cycling. *International Journal of Environmental Research and Public Health*, 18(7), 3719.

García-López, J., Rodríguez-Marroyo, J. A., Juneau, C. E., Peleteiro, J., Martínez, A. C., & Villa, J. G. (2008). Reference values and improvement of aerodynamic drag in professional cyclists. *Journal of Sports Sciences*, 26(3), 277-286.

García-López, J., Díez-Leal, S., Larrazabal, J., & Ogueta-Alday, A. (2015). No bilateral asymmetry during pedalling in healthy cyclists of different performance levels. In *ISBS-Conference Proceedings Archive*.

García-López, J., & del Blanco, P. A. (2017). Kinematic analysis of bicycle pedalling using 2d and 3d motion capture systems. *ISBS Proceedings Archive*, 35(1), 125.

James, C. R., Herman, J. A., Dufek, J. S., & Bates, B. T. (2007). Number of trials necessary to achieve performance stability of selected ground reaction force variables during landing. *Journal of Sports Science & Medicine*, *6*(1), 126.

Jobson, S.A., Hopker, J., Arkesteijn, M. & Passfield, L. (2013). Inter-and intra-session reliability of muscle activity patterns during cycling. *Journal of Electromyography and Kinesiology*, 23(1), 230-237.

Jongerius, N., Wainwright, B., Walker, J. & Bissas, A. (2022). The biomechanics of maintaining effective force application across cycling positions. *Journal of Biomechanics*, 138, 111103.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.

Kordi, M., Fullerton, C., Passfield, L., & Parker Simpson, L. (2019). Influence of upright versus time trial cycling position on determination of critical power and W′ in trained cyclists. *European Journal of Sport Science*, 19(2), 192-198.

Kuo, A. D. (1998). A least-squares estimation approach to improving the precision of inverse dynamics computations. *Journal of Biomechanical Engineering*, 120(1), 148-159.

Li, J. D., Lu, T. W., Lin, C. C., Kuo, M. Y., Hsu, H. C., & Shen, W. C. (2017). Soft tissue artefacts of skin markers on the lower limb during cycling: effects of joint angles and pedal resistance. *Journal of Biomechanics*, 62, 27-38.

Martin, J. C., & Brown, N. A. (2009). Joint-specific power production and fatigue during maximal cycling. *Journal of Biomechanics*, 42(4), 474-479.

Oggiano, L., Leirdal, S., Sætran, L., & Ettema, G. (2008). Aerodynamic optimization and energy saving of cycling postures for international elite level cyclists. *Eng. Sport*, *7*(1), 597-604.

Pataky, T. C. (2010). Generalized n-dimensional biomechanical field analysis using statistical parametric mapping. *Journal of Biomechanics*, 43(10), 1976-1982.

Pini, A., Markström, J. L., & Schelin, L. (2022). Test–retest reliability measures for curve data: An overview with recommendations and supplementary code. *Sports Biomechanics*, 21(2), 179-200.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: applications to practice.* (2nd edition). Upper Saddle River, NJ: Prentice Hall.

Racic, V., Pavic, A., & Brownjohn, J. M. (2009). Number of successive cycles necessary to achieve stability of selected ground reaction force variables during continuous jumping. *Journal of Sports Science & Medicine*, 8(4), 639.

Riazati, S., Caplan, N. & Hayes, P.R. (2019). The number of strides required for treadmill running gait analysis is unaffected by either speed or run duration. *Journal of Biomechanics*, 97, 109366.

Russek, L. (2004). Factors affecting interpretation of reliability coefficients. *Journal of Orthopaedic & Sports Physical Therapy*, 34(6), 341-349.

Sands, W. A., McNeal, J. R., & Stone, M. H. (2005). Plaudits and pitfalls in studying elite athletes. *Perceptual and Motor skills*, 100(1), 22-24.

Stagni, R., Leardini, A., Cappozzo, A., Benedetti, M. G., & Cappello, A. (2000). Effects of hip joint centre mislocation on gait analysis results. *Journal of Biomechanics*, 33(11), 1479-1487.

Taylor, P.G., Lee, K.Y., Landeo, R., O'Meara, D.M. & Millett, E. (2015). Determining optimal trial size using sequential analysis. *Journal of Sports Sciences*, 33(3), 300-308.

Trevethan, R. (2017). Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Health Services and Outcomes Research Methodology*, 17(2), 127-143.

Umberger, B. R., & Martin, P. E. (2001). Testing the planar assumption during ergometer cycling. *Journal of Applied Biomechanics*, 17(1), 55-62.

Underwood, L., Schumacher, J., Burette-Pommay, J. & Jermy, M. (2011). Aerodynamic drag and biomechanical power of a track cyclist as a function of shoulder and torso angles. *Sports Engineering*, 14, 147-154.

van Ingen Schenau, G. J., Van Woensel, W. W. L. M., Boots, P. J. M., Snackers, R. W., & De Groot, G. (1990). Determination and interpretation of mechanical power in human movement: application to ergometer cycling. *European Journal of Applied Physiology and Occupational Physiology*, 61, 11-19.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, 19(1), 231-240.

Winter, D. A. (2009). *Biomechanics and motor control of human movement*. John Wiley & Sons.