

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

3 Daniel Williams^{1,2,3}, Marcus Dunn⁴, Paul Worsfold^{2,5}, Deborah Newton^{2,3}, Steve H Faulkner^{3,6}, Francesca
4 d'Andrea¹, Jon Wheat⁷

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

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

13 ORCID: Daniel Williams: 0009-0004-5226-7492, Marcus Dunn: 0000-0003-3368-8131, Paul Worsfold:
14 0000-0002-1712-9818, Steve H Faulkner: 0000-0003-4688-7252, Francesca d'Andrea: 0000-0001-
15 9678-5612, Jon Wheat: 0000-0002-1107-6452

16 Corresponding Author: Daniel Williams; danjwilliams99@gmail.com

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

18 **Abstract:**

19 The purpose of this study was to determine the number of crank revolutions required to obtain stable
20 mean values of sagittal plane biomechanics variables, and the between-session reliability of these
21 variables, whilst cyclists used an aerodynamic position. Eighteen elite cyclists completed a 3-minute
22 maximal bout on a cycling ergometer. Lower-limb kinematic and kinetic data were captured using 2D
23 motion capture and force pedals. Raw data were filtered using a 4th order Butterworth low-pass filter
24 (6 Hz) and interpolated to 100 points per revolution. The middle 60 revolutions of each trial were
25 extracted and 37 discrete and 15 time-series variables were calculated. Mean stability was assessed in
26 all participants and between-session reliability was analysed in a subset of 11 participants. Sequential
27 averaging indicated more revolutions to stability than iterative intra-class correlation coefficients.
28 Crank kinetics were more stable than joint kinematics and kinetics. For stable discrete and time-series
29 variables, 30 and 38 revolutions are recommended, respectively. Between-day reliability for all
30 variables was moderate to excellent, and good to excellent for crank kinetics and joint kinematics
31 variables. Hip flexion-extension and ankle dorsiflexion kinetics were least reliable. Researchers and
32 applied practitioners should consider these findings when planning, and interpreting results from,
33 cycling biomechanics interventions.

34 Introduction

35 The aerodynamic cycling position (AP) requires the cyclist to lower their torso and rest their forearms
36 on pads while they grip extended handlebars. The AP is predominantly used in competitive events
37 such as the road time-trial, triathlon, and in the track cycling team pursuit and 1km time-trial, with
38 cyclists benefitting from a 20 - 30% reduction in drag area compared to typical road cycling positions
39 (Defraeye et al., 2010). Further optimisation in the AP through reduced handlebar height and greater
40 reach can lead to an additional aerodynamic benefit of 6 - 14% (Garcia-Lopez et al., 2008; Oggiano et
41 al., 2008). Whilst minimising resistive forces, the lower torso angles in the AP - which can be as low as
42 4° to 12° in elite time trialists (Cubel et al., 2022) - reduces power output and gross efficiency
43 (Fintelman et al., 2015; Kordi et al., 2019). Aerodynamic gains often outweigh losses in power
44 (Underwood et al., 2011) but competitive success requires a balance of reducing aerodynamic drag
45 and maintaining or improving biomechanical power output. Limited research has reported that
46 reduced torso angles alter pedal force application and joint kinetic contributions, associated with
47 increased hip flexion and ankle dorsiflexion (Bini et al., 2019; Jongerius et al., 2022). However, the
48 acute biomechanical responses to changes in handlebar position in the AP, and the optimal trade-off
49 with aerodynamics, are not completely understood.

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

64 In cycling biomechanics studies, researchers typically arbitrarily select and capture a section of an
65 exercise bout for further analysis, which can range from 10-seconds (Ferrer-Roca et al., 2014) to 1-
66 minute (Jongerius et al., 2022), equating to approximately 15 to 95 pedal revolutions. However, the

67 number of repetitions collected in a study can affect the stability of the dependent variables (Bates et
68 al., 1992). Statistical techniques have been proposed to identify the number of repetitions of a
69 movement required for stable mean estimates of dependent variables, including sequential averaging
70 (SA) and iterative intra-class correlation coefficient (IICC) analyses (Bates et al., 1983; James et al.,
71 2007). Initially these methods were applied to discrete dependent variables during overground walking
72 and running experiments, in which it was only possible to capture one stride per trial. More recently,
73 Taylor et al (2015) investigated the stability of time-series data, which enables further understanding
74 of biomechanical information outside of pre-determined discrete data-points (Pataky, 2010). SA and
75 IICC have also been applied to repetitive actions, such as continuous jumping (Racic et al., 2009) and
76 treadmill running (Riazati et al., 2019). However, no investigation has been applied to cycling to
77 understand the number of crank revolutions required to achieve stable dependent variables.

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

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

96

97

98

99 **Materials and Methods**

100 **Study Design**

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

107 **Participants**

108 Eighteen elite cyclists (10 females and 8 males; age: 22 ± 7 years; height: 1.74 ± 0.10 m; mass: $69.0 \pm$
109 9.5 kg) participated in this study, and a subset of 11 cyclists were included in the reliability analysis. All
110 participants volunteered and provided informed written consent before participation. The participants
111 were members (one former) of the national cycling team and experienced at cycling in the AP, having
112 competed in a time-trial and/or pursuit race at senior (8), under-23 (4) or junior (6) international level.
113 They were all free from any current injury, physical impairment or health condition that might have
114 limited participation in the study.

115 **Equipment**

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

125 **Experimental Protocol**

126 Each participant was asked to attend two testing sessions, though due to training and competitive
127 requirements only 11 were able to complete both sessions. Upon arrival, they completed a 20-minute
128 progressive cycling warm-up on a turbo trainer – the same warm-up that they would complete before
129 performing pursuit training or race efforts. Following this, reflective markers were placed on four

130 anatomical landmarks of both lower limbs by the same experienced investigator on each occasion. A
131 marker was also placed on the pedal spindle to enable calculation of an ankle angle (**Figure 1**).

132 **** FIGURE 1 HERE *****

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

142 **Data Analysis**

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

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

155 Crank power output was calculated as the sum of left and right crank power; both a product of the
156 respective left and right crank torques and crank angular velocity. Due to a technical issue with the left
157 force pedal, only right crank power output was compared between-days, whereas total crank power
158 output was included in the stability analysis. A further issue with the force pedals data for two
159 participants on their second visit meant that the crank kinetics data of only 9 of the 11 participants
160 could be included in the between-session reliability protocol, while all 11 participants' joint kinematics
161 data were analysed. All other variables included in the stability and reliability analyses were from the

162 right limb and crank as symmetry was assumed (Garcia-Lopez et al., 2015). Effective and ineffective
163 force were defined as force produced tangentially and radially to the direction of crank angular
164 rotation, respectively (**Figure 1**). For each crank cycle, IFE was calculated as the ratio of the area under
165 the raw effective force-time curve relative to the area under the resultant force-time curve, with
166 resultant force being the vector sum of effective and ineffective force.

167 Kinematic variables derived from joint angles and angular velocities of the ankle, knee, and hip were
168 calculated (**Figure 1**). Joint moments, powers, and work were calculated using standard inverse
169 dynamics techniques (Elftman, 1939). For joint angular velocities, change in the clockwise direction is
170 negative and positive in the anticlockwise direction. Therefore, for angular velocities and moments,
171 the extension peak was the greatest magnitude in the negative direction for the hip and ankle and the
172 positive direction for the knee, and vice-versa with respect to flexion. Joint powers were calculated as
173 a product of joint moment and joint angular velocity, and the peak magnitudes during periods of
174 flexion or extension were extracted - based on the direction of joint angular velocity (Martin & Brown,
175 2009). Joint work was calculated as the integral of joint power within phases of both flexion and
176 extension periods at each joint.

177 **Statistical Analysis – Mean Stability**

178 Both SA (Bates et al., 1983) and IICC (James et al., 2007) were performed to assess the mean stability
179 of each dependent variable. For SA, an initial number of 60 reference revolutions was selected. The
180 60-revolution mean value for each variable was subtracted from the cumulative mean at each
181 consecutive revolution (1 + 2, 1 + 2 + 3, etc.) to provide a mean deviation value. The mean of a variable
182 was considered stable at one revolution greater than the revolution where the mean deviation fell,
183 and remained, below 0.25 of the 60-revolution standard deviation (0.25SD; Bates et al., 1983). To
184 perform SA on time-series data, all 100 points were treated as a discrete point (Taylor et al., 2015).
185 Stability was defined as one revolution greater than the revolution where all 100 points had fell below
186 the threshold. The effect of using different SD thresholds (0.25, 0.41, 0.49 and 0.60) and numbers of
187 reference revolutions (20, 30, 40, 50 and 60) was also investigated (Forrester, 2015). The different SD
188 thresholds of 0.41, 0.49 and 0.60 were identified from previous studies as they approximated IICC
189 results (James et al., 2007; Racic et al., 2009). The relative SA score was calculated by dividing the
190 number of revolutions indicated to be required for stability by the number of reference revolutions
191 (Taylor et al., 2015). This was included to understand whether an increased number of repetitions to
192 stability was related to measurement instability or a direct result of increasing reference revolutions.

193 The IICC method involved calculating ICCs for each variable across an iteratively increasing (by 1)
194 number of consecutive crank revolutions up to the full 60-revolution sample (James et al., 2007). The

195 model used was ICC (3, 1) for single measures and consistency (Trevethan et al., 2017). The maximum
196 ICC value, the 95% confidence interval and the number of revolutions to reach the maximum ICC were
197 calculated for each dependent variable. In addition, the number of revolutions for the ICC (and the
198 95% CI lower-bound as a more conservative estimate) to reach 0.80 and 0.90 were recorded, which
199 were interpreted as thresholds for good (James et al., 2007) and excellent (Koo and Li, 2016) stability,
200 respectively.

201 **Statistical Analysis – Between-Session Reliability**

202 The procedures for investigating between-session reliability followed the recommendations of
203 Atkinson and Nevill (1998). To investigate systematic bias, paired samples *t*-tests were performed on
204 the discrete values and on the time-series data using statistical parametric mapping (SPM; Pataky,
205 2010). Alpha (α) was set to 0.05 to infer statistical significance. ICC (3, 1) analysis for single measures
206 and absolute agreement was used to determine an ICC for each discrete variable (Trevethan, 2017).
207 Integrated pointwise indices were performed as described by Pini et al. (2022) to assess the reliability
208 of the time-series variables. Briefly, this involved treating all fixed points along the time-series as a
209 discrete point and performing ICC (3, 1) groupwise on each discrete point. The interpretation of the
210 ICC was based on guidelines presented by Koo and Li (2016): less than 0.50 (poor), 0.50 to 0.75
211 (moderate), 0.75 to 0.90 (good) and greater than 0.90 (excellent). Standard error of measurement
212 (SEM) was used to assess absolute reliability, after confirming on the absence of heteroscedasticity
213 (Atkinson & Nevill, 1998). This was determined as $SD\sqrt{1-ICC}$, where SD is the standard deviation of
214 the scores from all participants (Weir, 2005). The minimal detectable difference (MDD) for each
215 measurement was calculated as $SEM \times 1.96 \times \sqrt{2}$ (Weir, 2005).

216 **Results**

217 **Mean Stability:**

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

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

226 while all time-series crank kinetics variables required 52 revolutions and time-series joint kinematics
227 required 57 revolutions.

228 The IICC results showed that all discrete variables, excluding peak knee flexion angular velocity (32
229 revolutions), reached maximum stability in 11 revolutions (**Table 1**) and excellent stability (ICC \geq 0.90)
230 within 4 revolutions. However, peak knee flexion power did not reach excellent stability (Max. ICC =
231 0.837). All 37 discrete outcomes reached good stability (\geq 0.80) within 3 revolutions. Using the more
232 conservative lower bound of the ICC 95% confidence interval excellent stability was not reached in 14
233 variables (8 of which were joint kinetics) – only peak knee flexion power did not reach good stability.

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

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

236 **Between-Session Reliability**

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

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

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

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

252 Across the whole crank cycle, no significant differences were found between-sessions for right crank
253 power or effective and ineffective force time-series (**Figure 3**). Crank power and effective force showed
254 good time-series reliability, with mean ICC values of 0.794 and 0.807, respectively. The pointwise ICC
255 (3, 1) was excellent during the downstroke phase (13 - 58%) and moderate to good during the rest of

256 the crank cycle. Ineffective force also showed good reliability (mean ICC = 0.862), with good to
257 excellent reliability for the whole crank cycle, apart from 75 - 93%.

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

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

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

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

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

270 **Discussion and Implications**

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

279 SA was most conservative in assessing stability, with all discrete and time-series variables reaching
280 stability by 47 and 60 revolutions, respectively, when using 60 reference revolutions (**Table 1**). As
281 expected, the number of revolutions SA indicated were required for stability reduced with fewer
282 reference revolutions (Forrester, 2015; Taylor et al., 2015) and greater SD thresholds (**Figure 2**; James
283 et al., 2007; Racic et al., 2009). Despite this, increasing the SD threshold to 0.60SD was still more
284 conservative than IICC when using 60 reference revolutions. As in previous research, the number of
285 revolutions IICC indicated were required for stability was lower than SA (James et al., 2007). All discrete
286 variables – except peak knee flexion velocity – reached maximum stability within 11 revolutions. IICC

287 therefore indicated that, counterintuitively, when the number of reference revolutions increased
288 beyond 11, outcome stability did not improve. Previous research has also shown IICC does not
289 demonstrate a pattern of increased stability when the number of repetitions of a movement is
290 increased (Chua et al., 2017). IICC may have assigned stability too soon - when the mean deviation was
291 still converging (Taylor et al., 2015). Previous research has also found IICC underestimates the number
292 of trials to stability when high between-participant variability is present (James et al., 2007; Coleman
293 et al., 2023). Due to these limitations, further discussion and recommendations are based on the SA
294 results.

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

312 In the between-session reliability analysis, maximum ankle plantarflexion and knee flexion
313 demonstrated a systematic difference between testing sessions, but the mean differences were small.
314 In addition, the SPM analysis identified that ankle joint angle was lower on Day 2 than Day 1 between
315 90% to 1% of the crank cycle (**Figure 4**). The reasons for these systematic differences are unclear. They
316 could be related to a learning or fatigue effect - as the activity of participants between sessions could
317 not be controlled - but it is not clear why this would systematically affect these variables. A learning
318 effect would appear unlikely due to the high experience level of the participants. On the other hand,
319 the differences were small (1.0° and 1.4° for maximum ankle plantarflexion and knee flexion,
320 respectively, and approximately 1.5° for ankle joint angle time series; **Table 2 & Figure 4**) and there is

321 a potential for a Type 1 error given the large number of variables analysed and relatively small sample
322 of participants (Brysbaert, 2019).

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

337 Several joint kinetics variables demonstrated only moderate reliability with relatively large minimal
338 detectable differences. All hip joint kinetics variables, except maximum hip flexion power ($ICC = 0.772$),
339 showed only moderate between-session reliability (**Table 2**). This might partly be explained by the
340 increased soft tissue artefact and marker misplacement error at the hip. During gait, 30 mm anterior-
341 posterior greater trochanter location error leads to differences in hip flexion-extension moments of
342 $\sim 22 - 25\%$ (Stagni, 2000). In addition, the hip is at the proximal end of the lower extremity kinetic
343 model. An implication of the iterative Newton-Euler method of inverse dynamics calculations used in
344 our data analysis (and standard across biomechanics research) is that errors from calculations at the
345 ankle and knee propagate up the kinetic chain to the hip (Kuo, 1998). Other joint kinetics variables -
346 ankle dorsiflexion moment and work, and knee flexion work - also had only moderate reliability. Since
347 ankle kinematics were highly reliable, it is more likely that poorer reliability for ankle dorsiflexion
348 kinetics variables is due to the small magnitude and trivial contribution to overall limb work (-1 J ; **Table**
349 **2**). The small magnitude of ankle dorsiflexion kinetics was accompanied by low between-participant
350 variability, which is associated with smaller ICCs (Portney & Watkins, 2000; Russek, 2004). This would
351 also explain the period of very poor reliability in the ankle power time-series (87 - 97%) that occurred
352 when there was low magnitude ($\sim 0 \text{ W}$) and low between-participant variability. Similarly, between-
353 participant variability in knee power was low at 42 - 44% where pointwise reliability was poor, and at
354 62 - 67% values were of low magnitude and both positive or negative (**Figure 4 & 5**). This might also

355 explain the moderate reliability identified for knee flexion work. The lower reliability and large MDDs
356 associated with these joint kinetics suggests researchers should be cautious when interpreting changes
357 in these variables.

358 Stability and between-session reliability were not consistent across different categories of dependent
359 variable. Crank kinetics were more stable and reliable than joint kinematics or kinetics. They reached
360 stability in 21 and 30 revolutions for discrete and time-series outcomes, compared to 26 and 38 for
361 joint kinematics and 30 and 33 revolutions for joint kinetics, respectively. This is likely to be because
362 joint kinematics and kinetics are subject to larger measurement errors. Joint kinematics are affected
363 by error from optical measurement systems, such as soft tissue artefact and marker misplacement.
364 What is more, joint kinetics are influenced by measurement errors from both kinematics and kinetic
365 sources (Camomilla et al., 2017). Interestingly, previous research has shown joint movements to
366 exhibit greater within participant variability than performance outcome measures, suggesting
367 compensatory joint movements may be necessary to achieve consistent outcomes (Bradshaw et al.,
368 2007). Our results are congruent with this and hint that elite cyclists vary their lower-limb movement
369 patterns to achieve lower variability in crank kinetics.

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

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

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

399 **Conclusions**

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

411 **Funding Details**

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

414 **Disclosure Statement**

415 The authors report there are no competing interests to declare.

416 **References**

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

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

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

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

426 Bini, R.R., Tamborindéguy, A.C. & Mota, C.B. (2010). Effects of saddle height, pedaling cadence, and
427 workload on joint kinetics and kinematics during cycling. *Journal of Sport Rehabilitation*, 19(3), pp.301-
428 314.

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

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

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

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

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

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

442 Chua, Y. K., Quek, R. K., & Kong, P. W. (2017). Basketball lay-up—foot loading characteristics and the
443 number of trials necessary to obtain stable plantar pressure variables. *Sports Biomechanics*, 16(1), 13-
444 22.

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

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

450 Defraeye, T., Blocken, B., Koninckx, E., Hespel, P., & Carmeliet, J. (2010). Aerodynamic study of different
451 cyclist positions: CFD analysis and full-scale wind-tunnel tests. *Journal of Biomechanics*, 43(7), 1262-
452 1268.

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

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

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

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

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

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

466 Galindo-Martínez, A., López-Valenciano, A., Albaladejo-García, C., Vallés-González, J. M., & Elvira, J. L.
467 (2021). Changes in the trunk and lower extremity kinematics due to fatigue can predispose to chronic
468 injuries in cycling. *International Journal of Environmental Research and Public Health*, 18(7), 3719.

469 García-López, J., Rodríguez-Marroyo, J. A., Juneau, C. E., Peleteiro, J., Martínez, A. C., & Villa, J. G.
470 (2008). Reference values and improvement of aerodynamic drag in professional cyclists. *Journal of*
471 *Sports Sciences*, 26(3), 277-286.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

520 Underwood, L., Schumacher, J., Burette-Pommay, J. & Jermy, M. (2011). Aerodynamic drag and
521 biomechanical power of a track cyclist as a function of shoulder and torso angles. *Sports Engineering*,
522 14, 147-154.

523 van Ingen Schenau, G. J., Van Woensel, W. W. L. M., Boots, P. J. M., Snackers, R. W., & De Groot, G.
524 (1990). Determination and interpretation of mechanical power in human movement: application to
525 ergometer cycling. *European Journal of Applied Physiology and Occupational Physiology*, 61, 11-19.

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

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