

Kent Academic Repository

Full text document (pdf)

Citation for published version

Mattioni Maturana, Felipe and Fontana, Federico Y. and Pogliaghi, Silvia and Passfield, Louis and Murias, Juan M. (2017) Critical power: How different protocols and models affect its determination. *Journal of Science and Medicine in Sport* . ISSN 1440-2440.

DOI

<https://doi.org/10.1016/j.jsams.2017.11.015>

Link to record in KAR

<http://kar.kent.ac.uk/64941/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

Critical power: How different protocols and models affect its determination

Felipe Mattioni Maturana^a, Federico Y. Fontana^b, Silvia Pogliaghi^b, Louis Passfield^c, Juan M. Murias^a

^a Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada; ^b Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy; ^c School of Sport & Exercise Sciences, University of Kent, Chatham Maritime, Kent, United Kingdom

Corresponding author: Juan M. Murias

Faculty of Kinesiology, University of Calgary

KNB 434, 2500 University Dr NW

Calgary, AB, Canada, T2N 1N4

e-mail: jmmurias@ucalgary.ca

tel. +1 403 220 7955, fax. +1 403 220 0448

Word count: 2898

Abstract count: 246

Tables: 2

Figures: 1

1 **Critical power: How different protocols and models affect its determination**

2

3 **Abstract**

4 In cycling, critical power (CP) and **work above CP** (W') can be estimated through linear and nonlinear
5 models. Despite the concept of CP representing the upper boundary of sustainable exercise,
6 overestimations may be made as the models possess inherent limitations and the protocol design is not
7 always appropriate. **Objectives:** to measure and compare CP and W' through the exponential (CP_{exp}), 3-
8 parameter hyperbolic (CP_{3-hyp}), 2-parameter hyperbolic (CP_{2-hyp}), linear (CP_{linear}), and linear 1/time
9 ($CP_{1/time}$) models, using different combinations of TTE trials of different durations (approximately 1 to 20
10 min). **Design:** repeated measures. **Methods:** Thirteen healthy young cyclists (26 ± 3 yrs; 69.0 ± 9.2 kg;
11 174 ± 10 cm; 60.4 ± 5.9 mL·kg⁻¹·min⁻¹) performed five TTE trials on separate days. CP and W' were
12 modeled using two, three, four, and/or five trials. All models were compared against a **riterion method**
13 (CP_{3-hyp} with five trials; confirmed using the leaving-one-out cross-validation analysis) using smallest
14 worthwhile change (SWC) and concordance correlation coefficient (CCC) analyses. **Results:** CP was
15 considerably overestimated when only trials lasting less than 10 min were included, independent of the
16 mathematical model used. Following CCC analysis, a number of alternative methods were able to predict
17 our **riterion method** with almost a perfect agreement. However, the application of other common
18 approaches resulted in an overestimation of CP and underestimation of W' , typically these methods only
19 included TTE trials lasting less than 12 min. **Conclusions:** Estimations from CP_{3-hyp} were found to be the
20 most accurate, independently of TTE range. Models that include two trials between 12 and 20 min
21 provide good agreement with the **riterion method** (for both CP and W').

22 **Key-Words:** power-time relationship; time-to-exhaustion; linear model; nonlinear model; exercise
23 intensity domains.

24 **Introduction**

25 Since first introduced by Monod and Scherrer¹ as the maximal capacity of a muscle, or muscle group to
26 perform work for a prolonged period of time, the concept of critical power (CP) has been widely used as
27 it presents a useful approximation of the endurance capacity of an individual²⁻⁴. Typically CP is
28 determined from a series of 5 time-to-exhaustion trials (TTE) conducted at severe exercise intensities⁵⁻⁷.
29 However, several studies suggest that estimates of CP can vary and are influenced by the test protocol
30 design. Factors such as the particular model used, and the duration of the TTE trials can change the CP
31 calculated from the model⁸⁻¹¹. Researchers use varying models to estimate CP, which are derived from a
32 range of two to seven TTE trials that are not standardized in terms of their duration. Although we note
33 that the most commonly used method is probably one employing four to five trials and fitted with the
34 two-parameter hyperbolic model (CP_{2-hyp})⁷.

35 Different studies have focused on either the influence of changing the mathematical model, or the number
36 of repetitions on the derived value for CP. For example, Gaesser et al.⁸, Bull et al.¹⁰, and Bergstrom et al.
37 ¹¹ investigated the influence of different mathematical models such as exponential (CP_{exp}), three
38 parameter hyperbolic (CP_{3-hyp}), CP_{2-hyp} , linear (CP_{linear}), and linear 1/time ($CP_{1/time}$) **on the determination of**
39 **CP and the work above CP (W')**. These three studies found CP_{3-hyp} and CP_{exp} result in different
40 estimations of CP. Bishop et al.⁹ asked their participants to perform five TTE trials ranging from 1 to 10
41 minutes in duration in order to evaluate the influence of the length of TTE trials on CP parameter
42 predictions. Using data from only three of the five trials CP was modelled with CP_{linear} and CP_{2-hyp} . Bishop
43 et al. found that a significant difference in modelled CP when the three shortest trials (i.e., $CP_{1,2,3}$), the
44 three longest trials (i.e., $CP_{3,4,5}$), or the first, the third, and the fifth trials (i.e., $CP_{1,3,5}$) were selected.
45 Consequently, the authors suggested that TTE trials of widely varying duration should be used to
46 minimize the influence of shorter trials when modelling CP. However, this investigation did not fit the
47 data from all five TTE trials, and was also limited modelling CP using TTE rides of less than 10 min,
48 about half the longest duration recommended by Morton⁷. Moreover, the aforementioned studies lacked

49 comparisons of the effects of using different mathematical methods and range of TTE trials on W'
50 outcomes.

51 Given the variety of approaches used in the literature and the effects of different models and combinations
52 of TTE trials, and the lack of a complete comparison of estimations of CP and W', the present
53 investigation aimed to examine the effect of number and range of TTE trials, and equation model on CP
54 and W'. Specifically we modelled CP and W' using combinations of two to five TTE trials with a variety
55 of different mathematical approaches (CP_{exp}, CP_{3-hyp}, CP_{2-hyp}, CP_{linear}, and CP_{1/time}).

56

57 **Methods**

58 Thirteen healthy young participants (9 men and 4 women; mean \pm SD values: age, 26 ± 3 yr; body mass,
59 69.0 ± 9.2 kg; height, 174 ± 10 cm) volunteered and gave written informed consent to participate in the
60 study. All participants had previous recreational or competitive cycling experience at the provincial level.
61 Participants were nonsmokers, with no musculoskeletal or cardiorespiratory conditions. The full testing
62 protocol was completed in 3 ± 1 weeks and consisted of: i) a preliminary maximal ramp incremental test
63 for determination of maximal $\dot{V}O_2$ ($\dot{V}O_{2max}$), and peak power output (PO_{peak}); and ii) five TTE trials for
64 estimation of CP. All procedures were conducted in an environmentally controlled laboratory (i.e.
65 temperature $\sim 21^\circ\text{C}$, relative humidity $\sim 36\%$), at a similar time of the day for each participant, with each
66 test performed on separate days, with a minimum interval of 24 h and a maximum interval of 72 h (most
67 typically 48 h) between tests to ensure appropriate recovery between trials. Participants were instructed to
68 keep their water and carbohydrate intake consistent throughout the protocol, and they were requested not
69 to engage in vigorous physical activity for 24 h prior to each test. Participants were asked not to consume
70 caffeine less than 12 h prior to the test. This study was approved by the Conjoint Health Research Ethics
71 Board of the University of Calgary. The results from CP_{2-hyp} using five TTE trials have been published as
72 part of a separate study comparing CP with the maximal lactate steady-state ⁵.

73 All exercise tests were performed on an electromagnetically braked cycle ergometer (Velotron Dynafit
74 Pro, Racer Mate, Seattle, WA, USA). Breath-by-breath pulmonary gas exchange, ventilation and heart
75 rate (HR) were continuously measured using a metabolic cart (Quark CPET, COSMED, Rome, Italy), as
76 previously described¹². Calibration was performed before each test as recommended by the manufacturer.
77 Breath-by-breath $\dot{V}O_2$ data were edited as follows: data points that were 3 SD from the local mean were
78 considered outliers and then removed¹³; trials were time-aligned to the onset of exercise (i.e. time zero
79 representing the onset of the ramp incremental exercise), and averaged into 30-s time bins. $\dot{V}O_{2max}$ was
80 considered as the highest 30-s $\dot{V}O_2$ average throughout the ramp incremental test. PO_{peak} was established
81 as the highest power output achieved at the end of the ramp incremental test.

82 For the ramp incremental test, the baseline consisted of participants cycling at 50 W for 4 min, as
83 suggested by Boone and Bourgois¹⁴, followed by either 1 W every 2 s ($30 \text{ W}\cdot\text{min}^{-1}$) (men) or 1 W every
84 2.4 s ($25 \text{ W}\cdot\text{min}^{-1}$) (women) increase in PO.

85 For the estimation of CP, each participant performed five constant-power output trials to exhaustion
86 which ranged from approximately 1 – 20 min, as recommended by Morton⁷. The first three TTE trials
87 were performed at 80, 95 and 110% of PO_{peak} (as determined from the preliminary ramp incremental test).
88 The order of the tests was randomly assigned. Subsequently, the other two power outputs were
89 determined to generate an even distribution of TTE between the five trials. Each test was preceded by a 4-
90 min baseline at 20 W, followed by a square-wave transition to the predetermined PO.

91 For all TTE trials, participants cycled at their preferred pedal cadence (range, 70-105 rpm), which was
92 determined during the preliminary ramp incremental test. The moment of exhaustion was deemed to
93 occur when participants failed to maintain the cadence within 5 rpm of their preferred rate for more than 5
94 s despite strong verbal encouragement. Participants were blinded to the elapsed time, but they received
95 visual feedback on their pedal cadence.

96 CP was modelled as follows:

- 97 i. $CP_{\text{exp}} \rightarrow PO = CP + (P_{\text{max}} - CP) * \exp(-t / \tau)$ Hopkins et al. ¹⁵
- 98 ii. $CP_{3\text{-hyp}} \rightarrow t = (W' / PO - CP) + (W' / CP - P_{\text{max}})$ Morton ¹⁶
- 99 iii. $CP_{2\text{-hyp}} \rightarrow t = W' / (PO - CP)$ Hill ¹⁷
- 100 iv. $CP_{\text{linear}} \rightarrow W_{\text{lim}} = W' + CP * t$ Moritani et al. ¹⁸
- 101 v. $CP_{1/\text{time}} \rightarrow PO = W' * (1 / t) + CP$ Whipp et al. ¹⁹

102 where P_{max} is the maximal instantaneous power (in watts), τ an undefined time constant, and W_{lim} is the
 103 work done (i.e., $PO * t$) in each predictive trial (in Joules).

104 When the model was fitted using four trials, two combinations were used: trials 1 to 4, and trials 2 to 5.
 105 Using three trials, four combinations were performed: trials 1, 2, 3; trials 1, 3, 5; trials 2, 3, 4; and trials 3,
 106 4, 5. Finally, when using two trials in the linear models, four combinations were tested: trials 1 and 2;
 107 trials 1 and 5; trials 3 and 4; and trials 4 and 5. Importantly, not every possible combination was reported
 108 to avoid superfluous comparisons that would not add predicting value to the model. Instead, we selected
 109 the combination of methods that would result in a wide combination of TTE, as well as those often used
 110 in the literature. See Table 1 for details on the exercise intensities and durations of the aforementioned
 111 TTE trials.

112 All data editing, processing, and modeling were performed using OriginLab version 9.2 (OriginLab,
 113 Northampton, MA).

114 Data are presented as means \pm SD. 90 % confidence intervals were calculated and used as a measure of
 115 uncertainty (the likely limit of the true value in the population ²⁰) around each CP and W' values derived
 116 from the different methods proposed. Differences between methods were quantified by calculating
 117 chances that the true value of a difference was substantial or greater than the smallest worthwhile change
 118 (see below). To perform these calculations, we assumed that a substantial difference (in either direction,
 119 positive or negative) was larger than 8 W (3.2 %) and 1500 J (6.5%) (these are calculated as a constant
 120 factor (0.2) multiplied by the between-subjects standard deviation ²⁰ around the **critereion-method** average

121 CP and W' values of the 3-parameter hyperbolic method using the all trials (i.e., $CP_{3\text{-hyp}(1,2,3,4,5)}$) as
122 described below. The above calculated thresholds were defined as the smallest worthwhile changes
123 perceived to be practically meaningful for both CP and W' . Thresholds for assigning qualitative terms to
124 chances of substantial effects were as follows: <1 %, almost certainly not; 1-5 %, very unlikely; 5-25 %,
125 unlikely or probably not; 25-50 %, possibly not; 50-75 %, possibly; 75-95 %, likely or probably; 95-99 %
126 very likely; >99 %, almost certain^{20,21}. Here the criterion value chosen to declare an effect as
127 likely/possibly vs unclear is based on a probabilistic approach. In Figure 1 and Figure 2, the exact
128 probability of the difference is reported. Effect sizes of each difference (Cohen's d, ranked as trivial (0-
129 0.19), small (0.20-0.49), medium (0.50-0.79) and large (0.80 and greater)²² are also reported as objective
130 and standardized measures of magnitude of effects and as alternative meaningfulness metrics²³. In effect
131 size calculation, the SD of $CP_{3\text{-hyp}(1,2,3,4,5)}$ was used to standardize the mean difference for each contrast²⁴.
132 The appropriateness of the $CP_{3\text{-hyp}(1,2,3,4,5)}$ model as "criterion method" for our data was determined by
133 testing how well this model fitted the observed data. CP parameter estimates of each method as well as
134 the ability of each model to generalize to new data were tested using the leave-one-out cross-validation
135 (LOOCV) approach. The model that fits the data most closely for both CP and W' was confirmed as $CP_{3\text{-}$
136 $\text{hyp}(1,2,3,4,5)$.

137 The measurement agreement between the criterion method and each other models or number of trials was
138 assessed by evaluating Lin's concordance correlation coefficient (CCC)²⁵. Concisely, this metric indicates
139 the degree to which the relationship between two variables approximates the perfect agreement (i.e. line-
140 of-identity)²⁶. The CCC was interpreted using the following criterion ranges: almost perfect agreement
141 ($CCC > 0.99$), substantial agreement ($0.95 > CCC < 0.99$), moderate agreement ($0.90 < CCC > 0.95$), and
142 poor agreement ($CCC < 0.90$)²⁷. Additionally, the RMSE and the slope/intercept resulting from the above
143 regression analyses were used to i) indicate the typical error that may be expected when using any
144 "inadequate" model to directly estimate the criterion model (i.e. $CP_{3\text{-hyp}(1,2,3,4,5)}$) and ii) to understand

145 whether this error was better or worse at high or low values of CP (see Supplementary Material 1 for the
146 results of the above mentioned analysis).

147 The statistical analysis was performed using STATA (Version 14, Texas, USA) and α was set in advance
148 at the 0.05 level; statistical significance was accepted when $p < \alpha$.

149

150 **Results**

151 Group mean absolute and relative $\dot{V}O_{2\max}$ were $4.17 \pm 0.68 \text{ L}\cdot\text{min}^{-1}$ (range: 2.85 – 5.08 $\text{L}\cdot\text{min}^{-1}$) and 60.4
152 $\pm 5.9 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ (range: 50.7 – 68.1 $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$), respectively. Group mean PO_{peak} was $376 \pm 54 \text{ W}$
153 (range: 274 – 448 W).

154 Group mean duration, corresponding exercise intensities of TTE trials, and mechanical work (i.e., time *
155 PO) for CP estimations are summarized in Table 1. The mean duration of trials ranged from 1.7 to 19.4
156 min. Group mean parameter estimates (i.e., CP and W') from the combinations performed between
157 number of trials vs mathematical models are displayed in Table 2.

158 Based on the LOOCV analysis, the model that predicts data most accurately was confirmed as the CP_{3-}
159 $_{\text{hyp}(1,2,3,4,5)}$ ($R^2 = 0.99$, 95% CI [248 255], RMSE = 26.5 W). Figures 1 and 2 show the mean difference
160 between each model and the criterion model for CP and W' , respectively. The difference for the majority
161 of the alternative methods was declared unclear for CP. However, for the methods CP_{2-} $_{\text{hyp}(1,2,3)}$, $CP_{\text{linear}(1,2)}$,
162 $CP_{\text{linear}(1,2,3)}$, $CP_{1/\text{time}(1,2,3)}$, $CP_{1/\text{time}(1,2,3,4)}$, and $CP_{1/\text{time}(1,2,3,4,5)}$ the difference in relation to the criterion method
163 was considered likely positive (i.e., overestimation). When using CP_{exp} , the CP estimates were
164 consistently higher than those observed in the criterion method, with the chance of an overestimation
165 declared very likely positive (Figure 1).

166 The difference in W' among almost all the alternative methods (20 out of 31) was considered likely
167 negative (i.e., underestimation) (Figure 2). With respect to the **critierion method** (CP_{3-} $_{\text{hyp}(1,2,3,4,5)}$) a number

168 of alternative methods **resulted** in a very small chance of underestimating W' : $CP_{3\text{-hyp}(1,2,3,4)}$, $CP_{3\text{-hyp}(2,3,4,5)}$,
169 $CP_{2\text{-hyp}(3,4,5)}$, $CP_{2\text{-hyp}(2,3,4,5)}$, $CP_{2\text{-hyp}(1,2,3,4,5)}$, as well as CP_{linear} and $CP_{1/\text{time}}$ using the trials (3,4), (4,5), and
170 (3,4,5). In Figure 2 it is notable that the inclusion of trials lasting less than 10 min (i.e., trials 1 – 3)
171 caused a substantial underestimation of W' , whereas the inclusion of trial 5 (approximately 20 min, on
172 average) led to the best approximation to the criterion method.

173 The results of the CCC analysis are presented in Supplementary Material 1.

174 **Discussion**

175 The main findings of this study were that: i) using TTE trials lasting less than 10 min (i.e., trials 1-3) to
176 model CP resulted in consistently higher values than those using the criterion method, and a considerable
177 underestimation of W' ; ii) when longer TTE trials were included in the model (between approximately 12
178 to 20 min), the estimations of CP were similar to those observed for the criterion **method**; and iii) $CP_{1/\text{time}}$
179 may provide an accurate CP and W' estimation, as long as TTE trials lasting less than 7 min are not
180 included in the mathematical model.

181 Given the popularity of CP as a measure of sustainable exercise intensity, accurate determination
182 of CP is important. Since its introduction more than 60 years ago¹, different protocol designs have been
183 used to determine CP. For example, whereas Poole²⁸ stated that predictive trials should range between 1
184 and 10 min, Morton⁷ suggested that longer trials, ranging from approximately 1 to 20 min, should be
185 included in order to model a power output that more realistically predicts the upper boundary of
186 sustainable endurance exercise.

187 Housh et al.²⁹ studied the effect of different combinations of TTE trials when modelling CP. The authors
188 compared CP modelled using two and three TTE trials against a pre-determined **criterion method** (four
189 trials), in an attempt to find the optimal protocol. These researchers found that when using the shortest
190 (~1 min) and the longest (~10 min) trials, the estimation was the most accurate and presented the lowest
191 standard error of the estimate. Therefore, they suggested that CP could be accurately estimated using two

192 trials, lasting 1 and 10 min. Bishop et al.⁹ conducted five TTE trials and estimated CP using different
193 combinations of three trials. The authors found significant differences in the CP values when using any
194 combination of three TTE trials, across a range of 1 to 10 min. The present study shows that using this
195 range of predictive trials (i.e., 1 to 10 min) the modelled CP is consistently higher than that obtained using
196 the criterion method. Furthermore, this effect is independent of the mathematical model used. In fact,
197 such differences in relation to the criterion method were always substantially higher than our minimum
198 detectable difference of 8 W. This finding implies that a time range within 1 to 10 min will likely result in
199 an overestimation of CP irrespective of the model selected (Figure 1). Additionally, CCC analysis
200 (Supplementary Material 1) also revealed that models including only shorter trials (i.e. less than 10 min)
201 resulted in: i) poor agreement with the criterion method ($CCC < 0.90$) and ii) a disproportionately higher or
202 lower estimation of CP (as indicated by the reported slope and intercept values). This may also have
203 implications for the interpretation of previous studies where the CP has been modelled only with TTE
204 trials lasting less than 10 min.

205 Based on the results from this and previous investigations, the mathematical model can have a significant
206 impact on the predicted CP^{8,30}. In other words, for a given set of data points (e.g., three TTE trials),
207 different CP predictions may be generated by each of the models. Our results show that several models
208 allow the predicted CP value to be overestimated (with the probability ranging from 75 to 99%, see
209 Figure 1). In contrast, the 3-parameter hyperbolic model (i.e., CP_{3-hyp}) appears to provide the most
210 accurate approach, regardless of the TTE trials modelled. This supports the contention that CP_{3-hyp}
211 overcomes the limitation of other linear and nonlinear models that assume an infinite power as time
212 approaches zero¹⁶. This limitation is addressed by adding a third-parameter to the CP model, the so-called
213 maximal instantaneous power (P_{max}). Interestingly, CP_{2-hyp} and CP_{linear} only overestimated CP, when the
214 range of predictive trials all had a duration of less than 10 min. This suggests that accurate predictions of
215 CP can be achieved provided longer TTE trials are included. Lastly, and in contrast with previous
216 investigations^{8,11}, $CP_{1/time}$ may provide an accurate measure of CP if trials longer than 10 min are included

217 in the model (e.g., $CP_{1/time(4,5)}$ and $CP_{1/time(1,5)}$). Importantly, it should be noted that CP may not always
218 reflect the highest boundary of physiological steady-state with prolonged exercise, as shown in previous
219 investigations^{5,31}. However, by using the most appropriate testing method (i.e., model and range of TTE
220 trials), it is likely that the CP value will more closely approximate the highest PO associated with a
221 metabolic steady-state, and will provide better estimations of TTE for any given intensity above CP for
222 performance prediction.

223 The accurate prediction of CP is possible with a range of different CP models provided that longer TTE
224 trials are included. This means that an accurate prediction of CP is possible using simpler mathematical
225 models and fewer tests (see also Supplementary Material 1). In this context, based on the present results,
226 CP may be estimated using either the CP_{linear} or $CP_{1/time}$ models with as few as two predictive trials if they
227 range from approximately 7 to 20 min (e.g., $CP_{linear(3,4)}$, $CP_{1/time(4,5)}$). These data are most relevant in "field
228 conditions" where time-efficiency (i.e. reducing the number of repetitions to minimize the time
229 commitment of athletes to testing) is a priority and where testing results can be combined with perception
230 of effort³² towards the fine-tuning of training intensity. However, when maximal accuracy and
231 repeatability are required, such as in a longitudinal research design, researchers should use several TTE
232 trials and a model that possesses high accuracy (i.e., hyperbolic) for CP estimation.

233 Alongside CP, the accuracy of W' is important for performance, as it delineates exercise capacity in the
234 severe-intensity domain³³. As CP models are often used for predicting the optimal time for a given
235 distance, a reliable measure of W' becomes crucial for the success of coaches and sport scientists in the
236 final outcome of a race. As observed in the present results, inclusion of TTE trials lasting less than 10 min
237 results in an underestimation of W' , whereas the inclusion of two TTE trials ranging between 12 and 20
238 min in the model yielded the most accurate W' estimations when compared to our **criterion method**.

239 In conclusion, estimations from CP_{3-hyp} provided the most accurate and generalizable approach for CP and
240 W' calculation (i.e., the model that was the least affected by protocol design). Accurate estimations of CP
241 can be made with models that use fewer exercise tests and simpler analyses, such as CP_{2-hyp} , CP_{linear} , and

242 $CP_{1/time}$. However, for these methods to express their accuracy, longer TTE trials ranging from
243 approximately 7 to 20 min should be included in the model. Modeling only TTE trials lasting less than 10
244 min may lead to a considerable overestimation of CP especially when using CP_{linear} and $CP_{1/time}$, as well as
245 underestimation of W' .

246 **Practical Implications**

- 247 • The use of only TTE trials lasting less than 12 min not only overestimates CP, but also
248 underestimates W' . Scientists and coaches should ensure that two TTE trials ranging between 12
249 and 20 min are performed to provide more accurate CP and W' estimations;
- 250 • CP_{linear} and $CP_{1/time}$ models using two TTE trials lasting between 7 and 20 min may provide
251 accurate estimations of CP and, combined with self-perception of effort, can offer a valid option
252 when time-efficient models are preferable;
- 253 • CP_{3-hyp} is not affected by protocol design and CP_{2-hyp} only overestimates CP when only trials
254 lasting less than 10 min are used. This suggests that these models might be more appropriate for
255 research purposes where reducing measurement variability is critical.

256

257 **References**

- 258 1. Monod H, Scherrer J. The Work Capacity of a Synergic Muscular Group. *Ergonomics*. 1965; 8(1-
259 4):329-338.
- 260 2. Keir DA, Fontana FY, Robertson TC, et al. Exercise Intensity Thresholds: Identifying the
261 Boundaries of Sustainable Performance. *Med Sci Sports Exerc*. 2015; 47(9):1932-1940.
- 262 3. Poole DC, Burnley M, Vanhatalo A, et al. Critical Power: An Important Fatigue Threshold in
263 Exercise Physiology. *Med Sci Sports Exerc*. 2016.
- 264 4. Poole DC, Ward SA, Gardner GW, et al. Metabolic and respiratory profile of the upper limit for
265 prolonged exercise in man. *Ergonomics*. 1988; 31(9):1265-1279.
- 266 5. Mattioni Maturana F, Keir DA, McLay KM, et al. Can measures of critical power precisely
267 estimate the maximal metabolic steady-state? *Appl Physiol Nutr Metab*. 2016; 41(11):1197-1203.
- 268 6. Vanhatalo A, Doust JH, Burnley M. Determination of critical power using a 3-min all-out cycling
269 test. *Med Sci Sports Exerc*. 2007; 39(3):548-555.
- 270 7. Morton RH. The critical power and related whole-body bioenergetic models. *Eur J Appl Physiol*.
271 2006; 96(4):339-354.
- 272 8. Gaesser GA, Carnevale TJ, Garfinkel A, et al. Estimation of critical power with nonlinear and
273 linear models. *Med Sci Sports Exerc*. 1995; 27(10):1430-1438.
- 274 9. Bishop D, Jenkins DG, Howard A. The critical power function is dependent on the duration of the
275 predictive exercise tests chosen. *Int J Sports Med*. 1998; 19(2):125-129.
- 276 10. Bull AJ, Housh TJ, Johnson GO, et al. Effect of mathematical modeling on the estimation of
277 critical power. *Med Sci Sports Exerc*. 2000; 32(2):526-530.
- 278 11. Bergstrom HC, Housh TJ, Zuniga JM, et al. Differences among estimates of critical power and
279 anaerobic work capacity derived from five mathematical models and the three-minute all-out test.
280 *J Strength Cond Res*. 2014; 28(3):592-600.

281 12. De Roia G, Pogliaghi S, Adami A, et al. Effects of priming exercise on the speed of adjustment of
282 muscle oxidative metabolism at the onset of moderate-intensity step transitions in older adults.
283 *Am J Physiol Regul Integr Comp Physiol.* 2012; 302(10):R1158-1166.

284 13. Lamarra N, Whipp BJ, Ward SA, et al. Effect of interbreath fluctuations on characterizing
285 exercise gas exchange kinetics. *J Appl Physiol.* 1987; 62(5):2003-2012.

286 14. Boone J, Bourgois J. The oxygen uptake response to incremental ramp exercise: methodological
287 and physiological issues. *Sports Med.* 2012; 42(6):511-526.

288 15. Hopkins W, Edmond I, Hamilton B, et al. Relation between power and endurance for treadmill
289 running of short duration. *Ergonomics.* 1989; 32(12):1565-1571.

290 16. Morton RH. A 3-parameter critical power model. *Ergonomics.* 1996; 39(4):611-619.

291 17. Hill DW. The critical power concept. A review. *Sports Med.* 1993; 16(4):237-254.

292 18. Moritani T, Nagata A, deVries HA, et al. Critical power as a measure of physical work capacity
293 and anaerobic threshold. *Ergonomics.* 1981; 24(5):339-350.

294 19. Whipp B, Huntsman D, Storer T, et al. A constant which determines the duration of tolerance to
295 high-intensity work. *Federation Proceedings. Vol 41: FEDERATION AMER SOC EXP BIOL*
296 *9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998; 1982:1591-1591.*

297 20. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Sportscience.*
298 *2005; 9:6-14.*

299 21. Hopkins WG. A scale of magnitudes for effect statistics. *A new view of statistics.* 2002; 502.

300 22. Cumming G. The new statistics: Why and how. *Psychol. Sci.* 2014; 25(1):7-29.

301 23. Winter EM, Abt GA, Nevill AM. Metrics of meaningfulness as opposed to sleights of
302 significance: Taylor & Francis; 2014.

303 24. Field A. *Discovering statistics using IBM SPSS statistics,* Sage; 2013.

304 25. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.*
305 *1989:255-268.*

- 306 26. Leo JA, Sabapathy S, Simmonds MJ, et al. The Respiratory Compensation Point is Not a Valid
307 Surrogate for Critical Power. *Med Sci Sports Exerc.* 2017.
- 308 27. McBride G. A proposal for strength-of-agreement criteria for Lin's concordance correlation
309 coefficient. NIWA Client Report: HAM2005-062. 2005.
- 310 28. Poole D. Measurements of the anaerobic work capacity in a group of highly trained runners. *Med*
311 *Sci Sports Exerc.* 1986; 18(6):703-705.
- 312 29. Housh DJ, Housh TJ, Bauge SM. A methodological consideration for the determination of critical
313 power and anaerobic work capacity. *Res Q Exerc Sport.* 1990; 61(4):406-409.
- 314 30. Jones AM, Vanhatalo A, Burnley M, et al. Critical power: implications for determination of
315 VO₂max and exercise tolerance. *Med Sci Sports Exerc.* 2010; 42(10):1876-1890.
- 316 31. McLellan TM, Cheung KS. A comparative evaluation of the individual anaerobic threshold and
317 the critical power. *Med Sci Sports Exerc.* 1992; 24(5):543-550.
- 318 32. Mattioni Maturana F, Keir DA, McLay KM, et al. Critical power testing or self-selected cycling:
319 Which one is the best predictor of maximal metabolic steady-state? *J Sci Med Sport.* 2017;
320 20(8):795-799.
- 321 33. Jones AM, Vanhatalo A. The 'Critical Power' Concept: Applications to Sports Performance with
322 a Focus on Intermittent High-Intensity Exercise. *Sports Med.* 2017:1-14.

323

324

325 **Figure 1.**

326 Absolute difference (watt) between each proposed method (see method section) and the selected criterion
327 ($CP_{3\text{-hyp}(1,2,3,4,5)}$) are presented along with 90% confidence interval around each difference along with the
328 effect size calculation. The two vertical lines, placed along the x-scale, highlight the Smallest Worthwhile
329 Change (± 8 W) and represent the boundaries for a substantially negative or positive difference (displayed
330 as grey areas). Numbers shown are the chances (%) that each difference is negative, trivial and positive.
331 Inferences are also reported in qualitative terms as almost certainly not, very unlikely, unlikely or
332 probably not, possibly not, possibly, likely or probably, very likely and almost certain.

333

334 **Figure 2.**

335 Absolute difference (Joules) between each proposed method (see method section) and the selected
336 criterion ($CP_{3\text{-hyp}(1,2,3,4,5)}$) are presented along with 90% confidence interval around each difference along
337 with the effect size calculation. The two vertical lines, placed along the x-scale, highlight the Smallest
338 Worthwhile Change (± 1500 J) and represent the boundaries for a substantially negative or positive
339 difference (displayed as grey areas). Numbers shown are the chances (%) that each difference is negative,
340 trivial and positive. Inferences are also reported in qualitative terms as almost certainly not, very unlikely,
341 unlikely or probably not, possibly not, possibly, likely or probably, very likely and almost certain.

Table 1. Group mean percent PO_{peak} , duration, absolute PO, and mechanical work during the five time-to-exhaustion trials for estimation of critical power.

Trial	%PO_{peak}	Duration (min)	Absolute PO (W)	Mechanical Work (kJ)
1	110 ± 0	1.7 ± 0.4	413 ± 60	42.7 ± 13.4
2	95 ± 0	3.2 ± 0.8	354 ± 59	66.5 ± 16.6
3	80 ± 0	7.1 ± 1.8	303 ± 50	127.9 ± 31.2
4	75 ± 4	12.5 ± 1.9	281 ± 47	209.8 ± 44.0
5	72 ± 4	19.4 ± 3.4	271 ± 46	312.1 ± 57.5

Table 2. Critical Power (CP) parameter estimates for the models CP_{exp} , CP_{3-hyp} , CP_{2-hyp} , CP_{linear} , and $CP_{1/time}$ using different combinations of time-to-exhaustion

		5	4		3			2				
			1,2,3,4	2,3,4,5	1,2,3	1,3,5	2,3,4	3,4,5	1,2	1,5	3,4	4,5
CP_{exp}	CP	275 ± 47	281 ± 47	270 ± 46								
	SEE	5.7 ± 2.1	8.0 ± 2.8	2.9 ± 1.9								
CP_{3-hyp}	CP	252 ± 44	250 ± 41	250 ± 43								
	SEE	3.4 ± 1.8	5.9 ± 5.8	4.5 ± 4.6								
	W'	23.1 ± 7.6	25.8 ± 12.6	24.9 ± 10.0								
	SEE	3.7 ± 2.0	5.9 ± 6.3	6.3 ± 7.3								
CP_{2-hyp}	CP	253 ± 44	256 ± 42	253 ± 44	263 ± 45	254 ± 43	256 ± 42	252 ± 44				
	SEE	1.6 ± 1.1	2.8 ± 2.2	2.3 ± 1.7	5.0 ± 6.9	1.6 ± 1.5	3.4 ± 3.2	2.6 ± 2.6				
	W'	20.3 ± 5.9	18.7 ± 6.7	20.1 ± 6.0	16.4 ± 5.7	19.8 ± 7.0	19.2 ± 7.1	21.2 ± 6.5				
	SEE	1.7 ± 1.0	1.9 ± 1.4	2.2 ± 1.5	1.6 ± 2.0	1.8 ± 1.5	2.3 ± 2.1	2.6 ± 2.3				
CP_{linear}	CP	256 ± 45	259 ± 44	255 ± 45	265 ± 47	256 ± 45	256 ± 45	253 ± 44	272 ± 50	257 ± 45	252 ± 46	251 ± 44
	SEE	2.3 ± 1.1	3.9 ± 2.7	2.4 ± 1.4	6.1 ± 6.5	3.1 ± 2.5	4.2 ± 3.5	2.4 ± 2.1				
	W'	17.9 ± 5.7	17.1 ± 5.7	19.2 ± 5.8	15.9 ± 5.7	17.7 ± 5.9	18.7 ± 5.9	21.0 ± 7.1	14.8 ± 6.5	16.3 ± 5.9	21.7 ± 8.8	22.3 ± 7.6
	SEE	1.5 ± 0.8	1.7 ± 1.1	1.8 ± 1.0	1.6 ± 1.5	2.4 ± 2.0	2.1 ± 1.8	2.0 ± 1.7				
$CP_{1/time}$	CP	261 ± 45	263 ± 45	256 ± 45	268 ± 47	260 ± 45	257 ± 45	253 ± 45	272 ± 50	257 ± 45	252 ± 46	251 ± 44
	SEE	4.4 ± 2.4	5.8 ± 3.4	3.1 ± 1.9	7.9 ± 5.6	4.6 ± 3.5	4.4 ± 3.6	2.4 ± 1.8				
	W'	16.1 ± 6.0	15.8 ± 5.9	18.4 ± 5.6	15.2 ± 6.0	16.0 ± 5.9	18.2 ± 5.7	21.1 ± 7.6	14.8 ± 6.5	16.3 ± 5.9	21.7 ± 8.8	22.4 ± 7.5
	SEE	0.8 ± 0.4	1.0 ± 0.6	1.0 ± 0.6	1.1 ± 0.8	0.8 ± 0.6	1.3 ± 1.0	1.5 ± 1.2				

(TTE) trials.

CP_{exp} , exponential model; CP_{3-hyp} , 3-parameter hyperbolic model; CP_{2-hyp} , 2-parameter hyperbolic model; CP_{linear} , linear model; $CP_{1/time}$, 1/time linear model; CP, critical power; SEE, standard error of the estimation; W', anaerobic work capacity. Numbers on the top first and second row identify the number of trials and their corresponding durations (Table 1), respectively.

Figure 1
[Click here to download high resolution image](#)

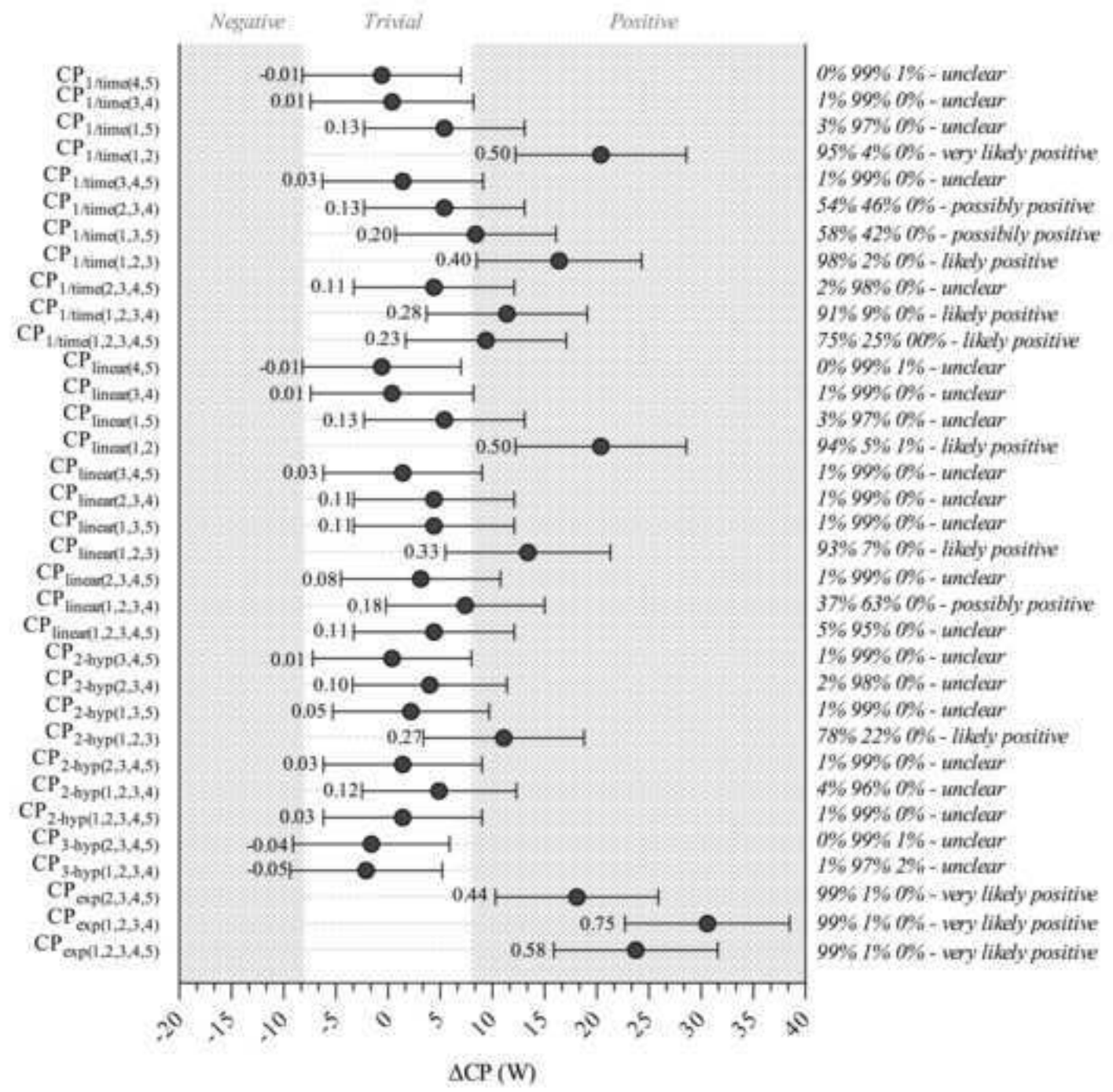
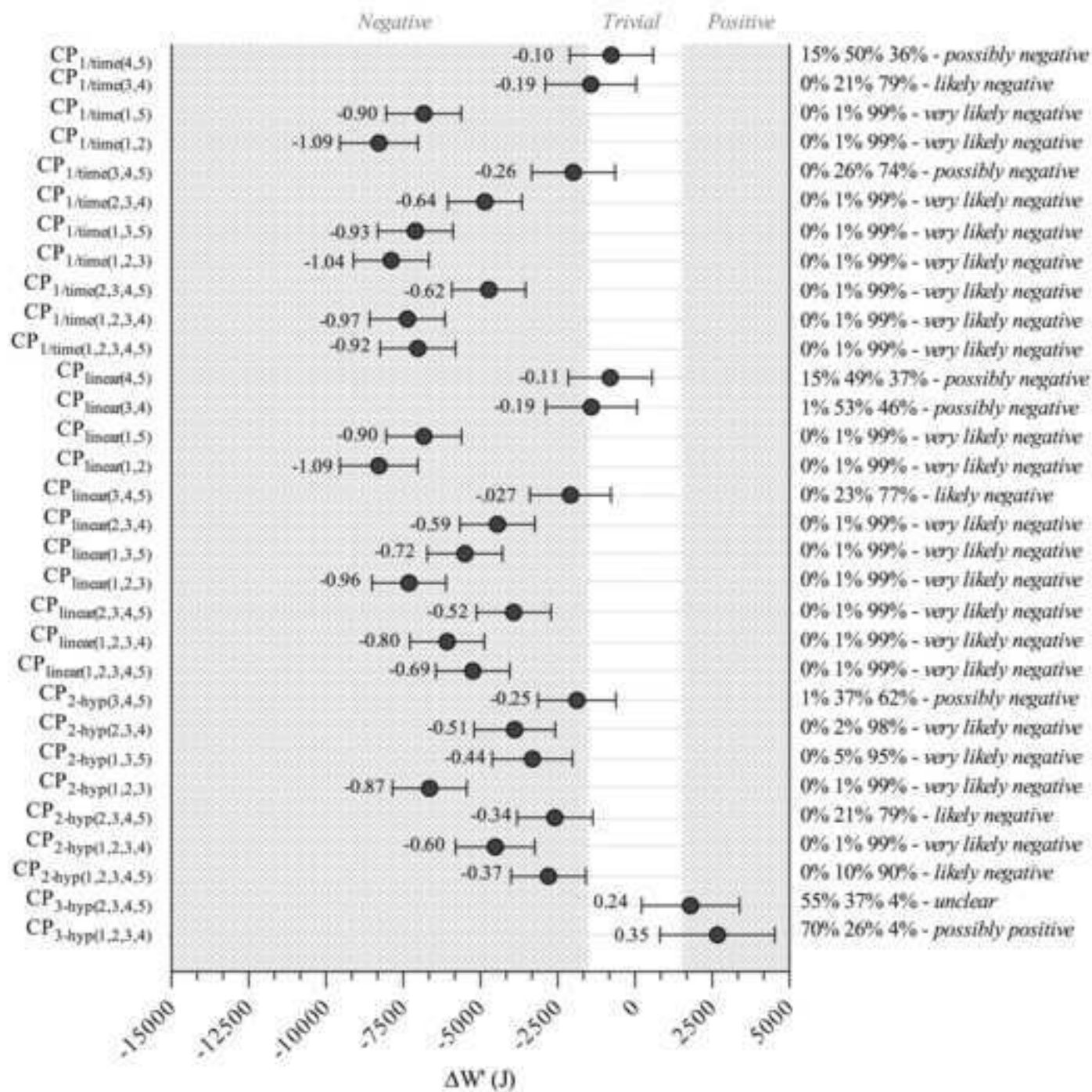


Figure 2

[Click here to download high resolution image](#)



Supplementary Material 1. Concordance correlation coefficient (CCC) analysis.

	#	R ²	RMSE (W)	CCC	Slope (b ₁)	Intercept (b ₀)
				95% CI [LL UL]	95% CI [LL UL]	95% CI [LL UL]
CP _{exp} (1,2,3,4,5)	5	0.98	6.80	0.86 ^d [0.75 0.97]	0.93 [0.84 1.02]	-5.47 [-31.10 20.10]
CP _{exp} (1,2,3,4)	4	0.96	9.70	0.79 ^d [0.64 0.92]	0.92 [0.79 1.05]	-7.13 [-44.40 30.10]
CP _{exp} (2,3,4,5)	4	0.99	4.60	0.91 ^c [0.84 0.98]	0.96 [0.89 1.02]	-6.73 [-23.90 10.50]
CP _{3-hyp} (1,2,3,4)	4	0.96	9.60	0.97 ^b [0.94 1.00]	1.06 [0.91 1.22]	-14.39 [-52.50 23.70]
CP _{3-hyp} (2,3,4,5)	4	0.98	6.29	0.99 ^a [0.98 1.00]	1.02 [0.93 1.11]	-3.51 [-27.10 20.20]
CP _{2-hyp} (1,2,3,4,5)	5	0.99	2.86	0.99 ^a [0.99 1.00]	0.99 [0.95 1.03]	-0.10 [-10.60 10.40]
CP _{2-hyp} (1,2,3,4)	4	0.98	5.47	0.98 ^b [0.97 1.00]	1.04 [0.96 1.12]	-15.91 [-37.30 5.48]
CP _{2-hyp} (2,3,4,5)	4	0.99	3.03	0.99 ^a [0.99 1.00]	1.01 [0.96 1.04]	-2.03 [-13.24 9.16]
CP _{2-hyp} (1,2,3)	3	0.91	14.10	0.92 ^c [0.84 1.00]	0.94 [0.74 1.14]	3.59 [-49.50 56.70]
CP _{2-hyp} (1,3,5)	3	0.99	3.01	0.99 ^a [0.99 1.00]	1.01 [0.97 1.06]	-6.67 [-18.20 4.70]
CP _{2-hyp} (2,3,4)	3	0.98	5.47	0.98 ^b [0.97 1.00]	1.05 [0.96 1.13]	-17.20 [-38.60 4.32]
CP _{2-hyp} (3,4,5)	3	0.99	3.51	0.99 ^a [0.99 1.00]	0.99 [0.94 1.04]	0.72 [-12.10 13.60]
CP _{linear} (1,2,3,4,5)	5	0.99	3.60	0.99 ^a [0.98 1.00]	0.98 [0.93 1.03]	-0.43 [-13.70 12.80]
CP _{linear} (1,2,3,4)	4	0.98	5.66	0.98 ^b [0.95 1.00]	0.99 [0.91 1.10]	-4.45 [-25.60 16.75]
CP _{linear} (2,3,4,5)	4	0.99	2.44	0.99 ^a [0.99 1.00]	0.98 [0.95 1.01]	0.34 [-8.60 9.30]
CP _{linear} (1,2,3)	3	0.93	11.98	0.92 ^c [0.84 1.00]	0.90 [0.74 1.06]	12.75 [-30.20 55.70]
CP _{linear} (1,3,5)	3	0.99	3.57	0.99 ^a [0.98 1.00]	0.98 [0.93 1.03]	-0.90 [-14.00 12.00]
CP _{linear} (2,3,4)	3	0.98	5.72	0.99 ^a [0.97 1.00]	0.98 [0.90 1.06]	-0.18 [-21.30 20.90]
CP _{linear} (3,4,5)	3	0.99	2.44	0.99 ^a [0.99 1.00]	0.99 [0.96 1.03]	0.53 [-8.40 9.40]
CP _{linear} (1,2)	2	0.75	22.88	0.78 ^d [0.58 0.98]	0.77 [0.47 1.06]	42.55 [-37.60 122.60]
CP _{linear} (1,5)	2	0.99	3.97	0.99 ^a [0.98 1.00]	0.98 [0.93 1.04]	-1.91 [-16.50 12.80]
CP _{linear} (3,4)	2	0.98	6.56	0.99 ^a [0.98 1.00]	0.95 [0.86 1.04]	11.00 [-12.20 34.20]
CP _{linear} (4,5)	2	0.97	7.80	0.98 ^b [0.97 1.00]	0.99 [0.88 1.10]	2.10 [-26.60 30.70]
CP _{1/time} (1,2,3,4,5)	5	0.97	7.50	0.96 ^b [0.93 1.00]	0.97 [0.87 1.10]	-3.20 [-31.20 12.70]
CP _{1/time} (1,2,3,4)	4	0.96	9.13	0.94 ^c [0.88 1.00]	0.96 [0.83 1.07]	-1.23 [-35.50 32.90]
CP _{1/time} (2,3,4,5)	4	0.98	5.53	0.98 ^b [0.97 1.00]	0.97 [0.89 1.05]	2.28 [-17.50 22.50]
CP _{1/time} (1,2,3)	3	0.92	13.26	0.89 ^d [0.79 0.99]	0.91 [0.72 1.08]	8.80 [-40.10 57.60]
CP _{1/time} (1,3,5)	3	0.98	7.10	0.97 ^b [0.94 1.00]	0.97 [0.87 1.07]	-0.10 [-26.50 26.40]
CP _{1/time} (2,3,4)	3	0.97	8.43	0.98 ^b [0.95 1.00]	0.96 [0.85 1.06]	3.80 [-27.10 34.70]
CP _{1/time} (3,4,5)	3	0.99	2.70	0.99 ^a [0.99 1.00]	0.99 [0.96 1.03]	0.06 [-9.80 9.90]
CP _{1/time} (1,2)	2	0.75	22.90	0.78 ^d [0.58 0.98]	0.76 [0.48 1.05]	43.20 [-36.70 123.23]
CP _{1/time} (1,5)	2	0.99	3.97	0.99 ^a [0.98 1.00]	0.99 [0.93 1.04]	-1.90 [-16.60 12.80]
CP _{1/time} (3,4)	2	0.98	6.56	0.99 ^a [0.98 1.00]	0.95 [0.86 1.04]	11.10 [-12.20 34.20]
CP _{1/time} (4,5)	2	0.97	7.80	0.98 ^b [0.97 1.00]	0.99 [0.88 1.10]	2.10 [-26.60 30.70]

Relationships between the gold standard method (CP_{3-hyp} (1,2,3,4,5)) and all the other methods are expressed as R², root mean square error (RMSE), concordance correlation coefficient (CCC), slope and intercept.

95% CI around statistics are expressed as [LL UL]. ^a, ^b, ^c, ^d represent the CCC descriptors namely almost perfect agreement (CCC > 0.99), substantial agreement (0.95 > CCC < 0.99), moderate agreement (0.90 < CCC < 0.95), and poor agreement (CCC < 0.90) respectively.

Acknowledgements

No sources of external funding were received for this work.

The authors declare no conflicts of interest.