Kent Academic Repository

ORIGINAL ARTICLE

A comparison of different methods to analyse data collected during time-to-exhaustion tests

Andrea Nicolò[1] · Massimo Sacchetti[1] · Michele Girardi[1] · Alister McCormick[2] · Luca Angius[3] · Ilenia Bazzucchi[1] · Samuele M. Marcora[3,4]

Abstract

Purpose Despite their widespread use in exercise physiology, time-to-exhaustion (TTE) tests present an often-overlooked challenge to researchers, which is how to computationally deal with between- and within-subject differences in exercise duration. We aimed to verify the best analysis method to overcome this problem.

Methods Eleven cyclists performed an incremental test and three TTE tests differing in workload as preliminary tests. The TTEs were used to derive the individual power–duration relationship needed to set the workload (corresponding to an estimated TTE of 1200 s) for four identical experimental TTE tests. Within individuals, the four tests were subsequently rank ordered by performance. Physiological and psychological variables expected to change with performance were analysed using different methods, with the main aim being to compare the traditional "group isotime" method and a less-used "individual isotime" method.

Results The four tests, ranked from the best to the worst, had a TTE of 1526 ± 332, 1425 ± 313, 1295 ± 325, and 1026 ± 265 s. Ratings of perceived exertion, minute ventilation, respiratory frequency, and affective valence were sensitive to changes in performance when their responses were analysed with the "individual isotime" method ($P < 0.022$, $\eta_p^2 > 0.144$) but not when using the "group isotime" method, because the latter resulted in partial data loss.

Conclusions The use of the "individual isotime" method is strongly encouraged to avoid the misinterpretation of the phenomenon under study. Important implications are not limited to constant-workload exercise, but extend to incremental exercise, which is another commonly used test of exercise tolerance.

Keywords Exercise tolerance · Method of analysis · Variability in time-to-exhaustion · Cycling · Endurance performance

✉ Samuele M. Marcora
s.m.marcora@kent.ac.uk

1 Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Piazza Lauro De Bosis 6, 00135 Rome, Italy

2 School of Sport, Health and Wellbeing, Plymouth Marjon University, Derriford Road, Plymouth PL6 8BH, UK

3 Endurance Research Group, School of Sport and Exercise Sciences, University of Kent at Medway, Chatham Maritime, Kent ME4 4AG, UK

4 Department of Biomedical and Neuromotor Sciences, University of Bologna, Via Ugo Foscolo 7, Bologna 40123, Italy

Introduction

Time-to-exhaustion (TTE) tests are extensively used in exercise physiology to evaluate exercise tolerance. They require a person to sustain a fixed workload for the longest time possible, which makes the execution of a TTE test feasible even for some clinical patients [1]. Nevertheless, TTE tests have been widely criticized for a number of (somewhat debatable) reasons. The main argument against the use of TTE tests is the supposedly poor within-subject reliability when TTE tests are compared to time trials [2]. However, TTE tests and time trials have similar within-subject reliability when the curvilinear relationship between exercise intensity and duration is taken into account [3]. Moreover, they have a similar sensitivity to changes in endurance performance [4]. Notwithstanding conflicting views on this issue, there are experimental conditions where the use of TTE tests is

invaluable, such as when the aim is to exclude the potentially confounding factor of varying workload (as it occurs during time trials) in the evaluation of the effects of a given experimental intervention on physiological and psychological responses.

Beyond the reliability issue, the well-documented between- and within-subject variability in TTE [5–7] offers another overlooked challenge to researchers, which is how to computationally deal with differences in exercise duration. Indeed, exercise tests with different durations prove difficult to analyse, especially when the time course of physiological or psychological responses must be compared across tests. Therefore, it is imperative to find appropriate ways to analyse data collected during TTE tests. The analysis most used so far (here termed "group isotime") includes only the portion of the TTE test that is available for all the participants in the group being analysed [8–16], which is limited by the participant with the shortest TTE. Disappointingly, this analysis results partial data loss for all the other participants, and the extent of data loss increases with the difference between the TTE of a given participant and that of the participant with the shortest TTE. This may negatively impact the interpretation of the phenomenon under study. To avoid any data loss, it is common to normalize data to the TTE of each test for each individual [17]. However, this analysis compares different TTE tests at the same relative duration (this analysis is here termed "relative isotime") instead of absolute duration, and consequently, it cannot be used to assess the effects of an experimental intervention. For instance, muscle glycogen depletion [18] or different environmental conditions [19] have no effect on rating of perceived exertion (RPE) if data are expressed against relative time instead of absolute time.

There is a third, less-appreciated method of analysis (here termed "individual isotime") which allows for the assessment of the effect of an experimental intervention while limiting data loss. Indeed, within individuals, it normalizes data to the shorter TTE and maintains absolute isotime comparisons between different TTE tests [20–22]. Nevertheless, the "individual isotime" has not been widely used by researchers. This may be because its potential advantages over traditional methods have not been demonstrated empirically, and possibly because the method has not been explained in sufficient detail to be easily reproduced.

The present study compared the three aforementioned methods of analysis when processing the same set of data collected during TTE tests, with the main aim of comparing the "individual isotime" and the "group isotime" methods. Participants performed four TTE tests at the same workload within individuals, and these four performances were rank ordered from best to worst to have four different levels of performance within individuals (within-subject performance ranking). Physiological and psychological variables expected to change with performance were measured, with a particular interest in the RPE and respiratory frequency (f_R) responses. RPE and f_R reflect effort during exercise [23–25], their rate of increase is strongly associated with TTE [17, 26], and they are sensitive to interventions that affect TTE test performance [10, 11, 27–29]. We hypothesized that measured variables would be more sensitive to performance ranking when using the "individual isotime" method compared to the "group isotime" method, because the latter results in partial data loss, thereby preventing full exploration of the data available. We used "within-subject performance ranking" as the independent variable in the present study to demonstrate that variability in TTE can only be reduced to a limited extent, which means that researchers need to select methods of TTE analysis that deal with between- and within-subject differences in exercise duration.

Methods

Subjects

Eleven male participants (mean ± SD: age 22 ± 3 years, stature 1.74 ± 0.09 m, body mass 66 ± 8 kg) volunteered to participate in this study. They were well-trained competitive cyclists with a minimum of 3-years cycling experience and 200 km training per week. This study was approved by the Ethics Committee of the University of Rome Sapienza in compliance with the *Declaration of Helsinki*. Written informed consent was obtained from all of the participants.

Study protocol

Participants reported to the laboratory on six separate occasions over a 3-week period, with visits separated by at least 48 h. On the first visit, participants performed a ramp incremental exercise test, followed by a TTE test, with the two tests separated by 30 min of recovery. On the second visit, participants performed two TTE tests separated by 30 min of recovery. The 30-min recovery time used in both visits is in line with previous studies performing multiple performance tests in the same visit with the aim to obtain the power–duration relationship [30, 31]. The three TTE tests differed in the exercise intensity and, therefore, exercise duration. This allowed us to obtain the power–duration relationship for each individual, which was used to set the exercise intensity for the experimental TTE tests. Four identical experimental TTE tests were then performed on separate days (visits 3–6). Based on the power–duration relationship, the exercise intensity corresponding to a TTE of 1200 s was selected for these four tests. All the protocols were performed on an electromagnetically-braked cycle ergometer (Lode Excalibur

Sport, Groningen, The Netherlands). The positions of the ergometer seat and handlebar during the first visit were recorded for each participant and reproduced in the following visits.

Ramp incremental test

The ramp incremental test was preceded by a 5 min warm-up at 100 W, 3 min of rest, and 2 min pedalling at 20 W. The test consisted of a continuous ramped increase in work rate of 30 W min^{-1}, starting from 20 W. Preferred pedalling cadence was selected by each participant and was kept constant throughout the test, which terminated when cadence fell by more than 10 rpm, despite strong verbal encouragement. The peak power output (PPO) was defined as the highest power output achieved at exhaustion, registered to the nearest 1 W, and the $\dot{V}O_{2\,peak}$ as the highest value of a 30-s average.

Before the ramp incremental test, participants were given standard instructions for providing RPE using the Borg 6–20 scale [32]. During the ramp incremental test, participants were asked to rate their perceived exertion on the RPE scale every minute during exercise and, retrospectively, at exhaustion. This procedure served as a familiarization with the scale.

Preliminary TTE tests

Three preliminary TTE tests were performed during visits 1 and 2 to obtain the power–duration relationship for each individual. These three TTE tests were performed on average at $87 \pm 0.3\%$, $76 \pm 1.3\%$, and $70 \pm 2.7\%$ of the ramp PPO, to result in TTEs of approximately 4, 10, and 18 min, although between-subject variability was expected. This choice was made to have exercise durations suitable for a good prediction of the power output corresponding to a TTE of 20 min. The 76% PPO test was performed during the first visit after the ramp incremental test, while the 70% PPO test was performed before the 87% PPO test during the second visit. The three performance data points were then used to derive the power–duration relationship for each individual using a power law mathematical function with the following equation: $Y = cX^b$, where Y (power output) and X (time) are the two variable quantities, c is the theoretical maximal power output at time zero, and b is the scaly exponent describing the decrease in power output over time. For more detailed information on the use of the power law function in endurance sports, see García-Manso et al. [33]. In the present study X was fixed to 1200 s, and the corresponding power output was obtained for each individual.

Preferred pedalling cadence was selected by each participant before the first preliminary TTE test. The participant was asked to maintain the cadence within a range of preferred cadence ± 7 rpm during the test. This was done to reduce potential changes in physiological and psychological variables due to changes in pedalling cadence. The participant had been informed that a 10-s countdown would have started whenever pedalling cadence fell outside the predefined range. If the cadence returned to a value within that range before the countdown was completed, the test continued; otherwise, participants were judged to have reached exhaustion, which corresponded to the end of the 10-s countdown. This objective exhaustion criteria allowed us to register TTE to the nearest second. Preferred cadence was kept constant for a given participant throughout both preliminary and experimental TTE tests. The participant did not receive any feedback or encouragement during any of the TTE tests performed in the present study.

Every 2 min during the three preliminary TTE tests, RPE and affective valence (i.e., pleasure/displeasure experienced during exercise, measured using the Feeling Scale) were collected to allow the participants to thoroughly familiarise with the two scales. The Feeling Scale was first presented to participants in the first visit before the 76% PPO test, and standard instructions were provided [34]. Feeling Scale scores can range from $+5$ (the exercise feels "very good") to -5 (the exercise feels "very bad").

Experimental TTE tests

On visits 3–6, participants performed an identical TTE test at a power output corresponding to a predicted TTE of 1200 s, as detailed in the previous section, with the aforementioned exhaustion criteria. Power output was prescribed based on the individual power–duration relationship in an attempt to reduce the relatively high between-subject variability in TTE that is commonly observed when other methods of exercise prescription are used. The four experimental TTE tests were preceded by a standardized warm-up. This consisted of 3 min at 100 W, 6 min at 50% of PPO, 1 min at 60% of PPO, and 1 min at 100 W. Tests were then preceded by 3 min of rest and 2 min pedalling at 20 W. During all the tests, f_R, minute ventilation ($\dot{V}_E$) and heart rate (HR) were measured breath-by-breath using a metabolic cart (Quark b2, Cosmed, Rome, Italy). Appropriate calibration procedures were performed following the manufacturer's instructions. RPE and affective valence were reported every 2 min.

Control of factors potentially confounding performance

In an attempt to limit the within-subject variability in TTE, a number of potential confounding factors were controlled, to limit their influence on performance. All testing was completed in the laboratory with a room temperature of 20–22 °C and at the same time of day (± 1 h) within

participants. Participants were asked to refrain from caffeine and alcohol at least for the 3 h and 24 h, respectively, preceding each test. They were asked to record food intake on the day of the first experimental TTE test as well as the day before, to replicate it before the subsequent experimental TTE tests. Participants were also asked to standardise their training routine and to avoid strenuous exercise the day before the test. At each visit to the laboratory, participants were asked to complete a pre-test checklist to verify that they had complied with the instructions given to them. They were also asked to confirm that they were not in a state of mental fatigue, physical fatigue, sleep deprivation, and that they were free of injury and under no medical treatment. A single test was rescheduled, because the participant failed to meet some of the requirements. To reproduce a competitive setting and favour the achievement of a maximal effort in all the tests, a performance-based prize (£ 200 voucher) was offered for the participant with the longest average TTE considering all the four tests. No feedback on performance was provided to participants until all the tests were completed.

Data analysis

Data were analysed with MATLAB (R2016a, The Mathworks, Natick, MA, USA). Before performing the three different TTE analyses, breath-by-breath ventilatory data were filtered for errant breaths (i.e., values resulting from sighs, swallows, coughs, etc.) by deleting values greater than 3 standard deviations from the local mean [35]. Subsequently, breath-by-breath ventilatory data were interpolated with a linear function and extrapolated every second. Data were then smoothed by a moving average of 60 s. RPE and affective valence data collected every 2 min were interpolated with a linear function and subsequently extrapolated to have continuous values every second. For each individual, the four tests were rank ordered from the best to the worst based on TTE. This was done to have different levels of performance within individuals, reflective of within-subject variability in TTE.

Group isotime method

When data were processed with the "group isotime" method, the worst test of the participant with the shorter TTE (i.e., participant 2; Fig. 1) was selected to identify the timepoints in which to segment all the tests of all the participants. This test lasted 530 s. To obtain 10 equally spaced timepoints characterizing each test, the following timepoints were considered: 53, 106, 159, etc. up to 530 s. This method is termed "group isotime", because all the tests of all the participants are analysed considering the same absolute timepoints. The extent of data loss (EDL) that occurs with this method was calculated as: EDL = (avgTTE − isotime
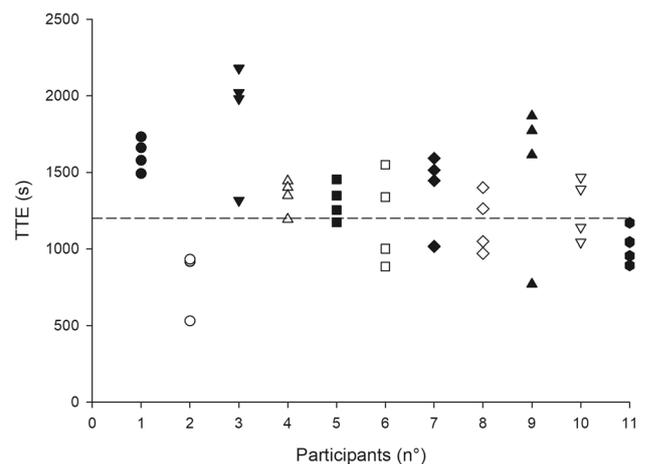
Fig. 1 TTE of the four tests for each participant. The dashed line indicates the 1200 s value. A test for participant 2 is hidden by other two tests

duration)/avgTTE × 100, where avgTTE is the average TTE of the group in seconds (e.g., 1026 s in the present study; worst test), while isotime duration corresponds to the last timepoint (in seconds), where all the participants were represented (e.g., 530 s in the present study; worst test). When data were available, the same formula was used to calculate the EDL from the previous studies that used the "group isotime" method, for a comparison with the present study.

Individual isotime method

When data were processed with the "individual isotime" method, each participant was considered in isolation when segmenting the tests in timepoints, hence the name "individual isotime". For each participant, the worst test was taken into account for identifying ten timepoints in which the four tests were segmented. Considering again participant 2, exactly the same timepoints used for the "group isotime" analysis (53, 106, 159, etc. up to 530 s) were selected. However, the timepoints identified for the other participants differed from each other on the basis of the TTE of their worst test. For instance, the worst test of participant 5 had a TTE of 1173 s. Hence, the timepoints considered for the four tests of that participant were 117, 235, 352, etc. up to 1173 s. Importantly, with this analysis, the worst test of all participants did not result in any data loss. This means that a greater portion of data was included in the between-test comparison, relative to the "group isotime" analysis. For further clarification on this analysis, please note that each data point corresponds to different absolute time values between participants. This can be depicted graphically by adding horizontal error bars. However, horizontal error bars are identical across conditions when using the "individual isotime" method, apart from the test end value. Therefore, we

opted for a graphical representation of the horizontal error bars, which preserves the quality of the graph and avoids redundancy (Figs. 2, 3, 4). To promote a full understanding of the "individual isotime" analysis as well as the "relative isotime" analysis described below, we have made available the codes used to run the two analyses as Supplementary material (Online Resource 1).

Relative isotime method

When data were processed with the "relative isotime" method, each test of each participant was segmented into ten timepoints on the basis of the TTE of the test analysed. Again, for participant 2, the worst test was segmented in the following timepoints: 53, 106, 159, etc. up to 530 s, as for the other two analyses. However, the best test of participant 2 was segmented in the following timepoints: 93, 187, 280, etc. up to 933 s because of the longer TTE. The same procedure was applied for the other tests of participant 2 as well as for all the tests of the other participants. With this method, there is no data loss, but different tests are not compared at the same absolute timepoints within participants but at the same percentages of TTE. Therefore, this method is here termed "relative isotime".

Statistical analysis

An a priori power analysis was performed using G*Power (version 3.1.9.2; Kiel University, Kiel, Germany). Expecting a large effect size for the sensitivity of f_R and RPE to within-subject performance ranking, a sample size of 7 was required based on $1 - \beta = 0.80$ and $\alpha = 0.05$. Eleven participants were recruited to account for potential dropping out.

Statistical analyses were conducted using IBM SPSS Statistics 20 (SPSS Inc, Chicago, IL, USA) unless otherwise stated. Data were checked for normality prior to analysis. The reliability in TTE was quantified by means of the log-transformed coefficient of variation (CV) with 90% confidence limits using a published open-source spreadsheet in Microsoft Excel (Microsoft Corp.) [36]. For TTE data, the within- and between-subject variance components were calculated as percentages of the total variance by means of a linear mixed model based on the restricted maximum likelihood estimates approach, where participants served as a random between-subjects factor and test as a fixed within-subject factor [7]. More specifically, the within- and between-subject variance components were summed to obtain the total variance, and their percentage contributions to the total variance were calculated. A one-way repeated-measures ANOVA was used to compare the end value of f_R, $\dot{V}_E$, HR, RPE, and affective valence across the four TTE tests. A two-way repeated-measures ANOVA (rank × time) was used to analyse the effect of rank on f_R, $\dot{V}_E$, HR, RPE,
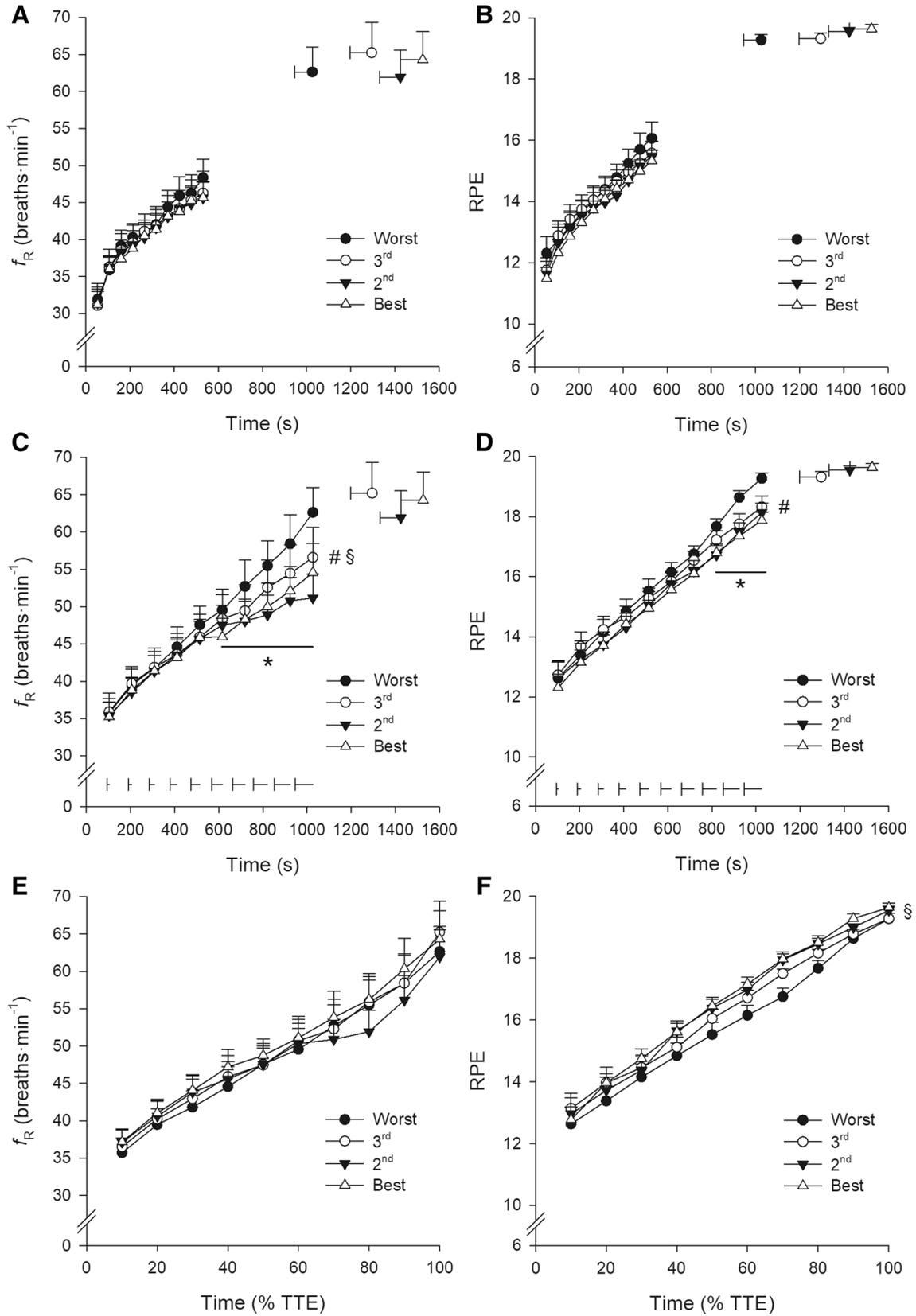
and affective valence responses. The same statistical analysis was used for the three methods of data processing under study, i.e., the "group isotime", "individual isotime", and "relative isotime". When the sphericity assumption was violated, the Greenhouse–Geisser adjustment was performed. For the main effect of rank, the main effect of time, and the interaction, partial eta squared (η_p^2) effect sizes were calculated; an effect of $\eta_p^2 \geq 0.01$ indicates a small effect, $\eta_p^2 \geq 0.059$ a medium effect, and $\eta_p^2 \geq 0.138$ a large effect [37]. When a significant main effect of rank was found, the Bonferroni test was used as follow-up analysis. When a significant interaction was found, a one-way repeated-measures ANOVA was used to test the simple main effect of rank at different timepoints.

Within-subject correlation coefficients (r) were computed for the correlations between RPE and f_R, using the method described by Bland and Altman [38]. This method adjusts for repeated observations within participants, using multiple regression with "participant" treated as a categorical factor using dummy variables. A correlation coefficient and a P value were obtained considering the four tests together, as well as for each test considered separately. These correlations were computed using data analysed with the "relative isotime" method, because it is the only analysis method which results in no data loss for any test. A P value < 0.05 was considered statistically significant in all analyses. The results are expressed as mean $\pm$ SD in text and as mean $\pm$ SE in figures.

Results

$\dot{V}O_{2\,peak}$ and the PPO measured during the ramp incremental test were 4341 ± 623 mL min^{-1} (66 ± 6 mL kg^{-1} min^{-1}) and 422 ± 42 W, respectively. The power outputs of the three preliminary tests were 368 ± 37 W, 323 ± 33 W, and 294 ± 34 W, and the corresponding TTEs were 286 ± 56 s, 547 ± 56 s and 1049 ± 213 s. The power output for the experimental TTE tests was 288 ± 40 W.

Figure 1 depicts the TTE for the four tests of each participant. On average, the TTE (1319 ± 356 s) was higher than that predicted by the power–duration relationship. The TTE CV (with 90% confidence limits) was 25.3% (20.2, 35.2). The within- and between-subject variance components contributed to 52.9% and 47.1%, respectively, of the total variance. No significant differences were observed between the four tests when these were ordered from the first to the last experimental visit (1300 ± 284 s, 1288 ± 381 s, 1360 ± 374 s, and 1328 ± 418 s). These findings indicate that there was no order effect. When the four tests were rank ordered by performance (from the best to the worst), a significant effect of rank was found

($P < 0.001$; $\eta_p^2 = 0.712$). The TTEs were: 1526 ± 332 s; 1425 ± 313 s, 1295 ± 325 s and 1026 ± 265 s, and pairwise comparisons were all significantly different ($P < 0.044$).

Table 1 reports the P value and effect size (η_p^2) for the main effect of rank, the main effect of time and the rank × time interaction for f_R, $\dot{V}_E$, HR, RPE, and affective valence, analysed with the three different methods. While no main effect of rank and no interaction was found for any variable when data were analysed with the "group isotime" method, all the variables except for HR showed a main effect of rank and/or an interaction when data were analysed with the "individual isotime" method. Considering the two variables for which a main effect of rank was found when using the "individual isotime" method, the follow-up analyses revealed a significant difference ($P < 0.022$) between the worst test and the best and second best tests for f_R, while only a statistical trend was found for $\dot{V}_E$ (worst vs. best, $P = 0.095$; worst vs. second best, $P = 0.073$).

The differences observed between the "group isotime" and "individual isotime" methods are due to the data loss that occurs with the "group isotime" method, with the extent of data loss being 48%. This is evident from the time courses of f_R, $\dot{V}_E$, HR, RPE and affective valence, as depicted in Figs. 2, 3, and 4. When a significant interaction was found, these figures show where a simple main effect of rank was found.

No significant between-test differences were found when comparing the end value of f_R (from the best to the worst test: 64 ± 13, 62 ± 12, 65 ± 14, and 63 ± 11 breaths min^{-1}), $\dot{V}_E$ (147 ± 30, 148 ± 29, 145 ± 28, and 147 ± 29 L min^{-1}), HR (185 ± 10, 186 ± 11, 186 ± 12, and 185 ± 10 bpm), and affective valence (0.2 ± 3.3, 0.6 ± 3.1, 0.9 ± 2.8, and 1.0 ± 3.0), while a statistical trend ($P = 0.052$) was found for RPE (19.6 ± 0.5, 19.5 ± 0.5, 19.3 ± 0.6, and 19.3 ± 0.6).

A strong correlation was found between RPE and f_R when the four tests were considered together ($P < 0.001$, $r = 0.80$), as well as when the tests were considered separately ($P < 0.001$; from the best to the worst test: $r = 0.84$, $r = 0.81$, $r = 0.85$ and $r = 0.84$). These correlations are reported in Fig. 5.

Discussion

This is the first study to compare different methods of analysis to empirically determine which is the most appropriate way to analyse data collected during TTE tests. The findings show that the choice of method dramatically influences the magnitude and the statistical significance of the effect of the independent variable (within-subject performance ranking) on the dependent variables (physiological and psychological responses). Specifically, f_R, $\dot{V}_E$, RPE, and affective valence were sensitive to performance ranking when data were analysed with the "individual isotime" method, but not when the traditional "group isotime" method was used. This emphasises how the method of analysis influences the interpretation of the phenomenon under study.

To reduce the between-subject variability in TTE, we prescribed exercise based on the power–duration relationship, which is, at least in principle, an ideal method to reduce between-subject variability in TTE, because the power output can be selected on the basis of the desired TTE (20 min in the present study). Using the power–duration relationship, we found that the % contribution of the between-subject variability (47.1%) to the overall variability in TTE was lower than the 59.4% reported by Faude et al. [7] during constant-load cycling at the maximal lactate steady state. This suggests that the prescription modality used in the present study limited the extent of between-subject variability in TTE, despite average TTE being higher than the desired TTE. We also attempted to reduce the within-subject variability in TTE by controlling potential confounding factors. However, we found similar reliability values of TTE (CV % = 25.3) to those reported by Faude et al. [7] (CV % = 24.6), because within-subject variability is an inherent characteristic of performance. Our findings suggest that the overall variability in TTE can only be reduced to a limited extent. Therefore, it is imperative to use appropriate methods of analysis that deal with between- and within-subject differences in performance duration.

With the aim to identify the most appropriate method to analyse data collected during TTE tests, we purposely selected physiological and psychological variables expected to change according to variations in performance, with a special interest in the responses of RPE and f_R. The rates of increase in RPE and f_R are correlated with TTE at least during high-intensity exercise [17, 26], and RPE and f_R are sensitive to experimental interventions that affect TTE performance including muscle fatigue [10] and damage [28], and increase in body temperature [39] and hypoxia [27]. Moreover, the linear increase in RPE and f_R over time makes the distinction between a control and an
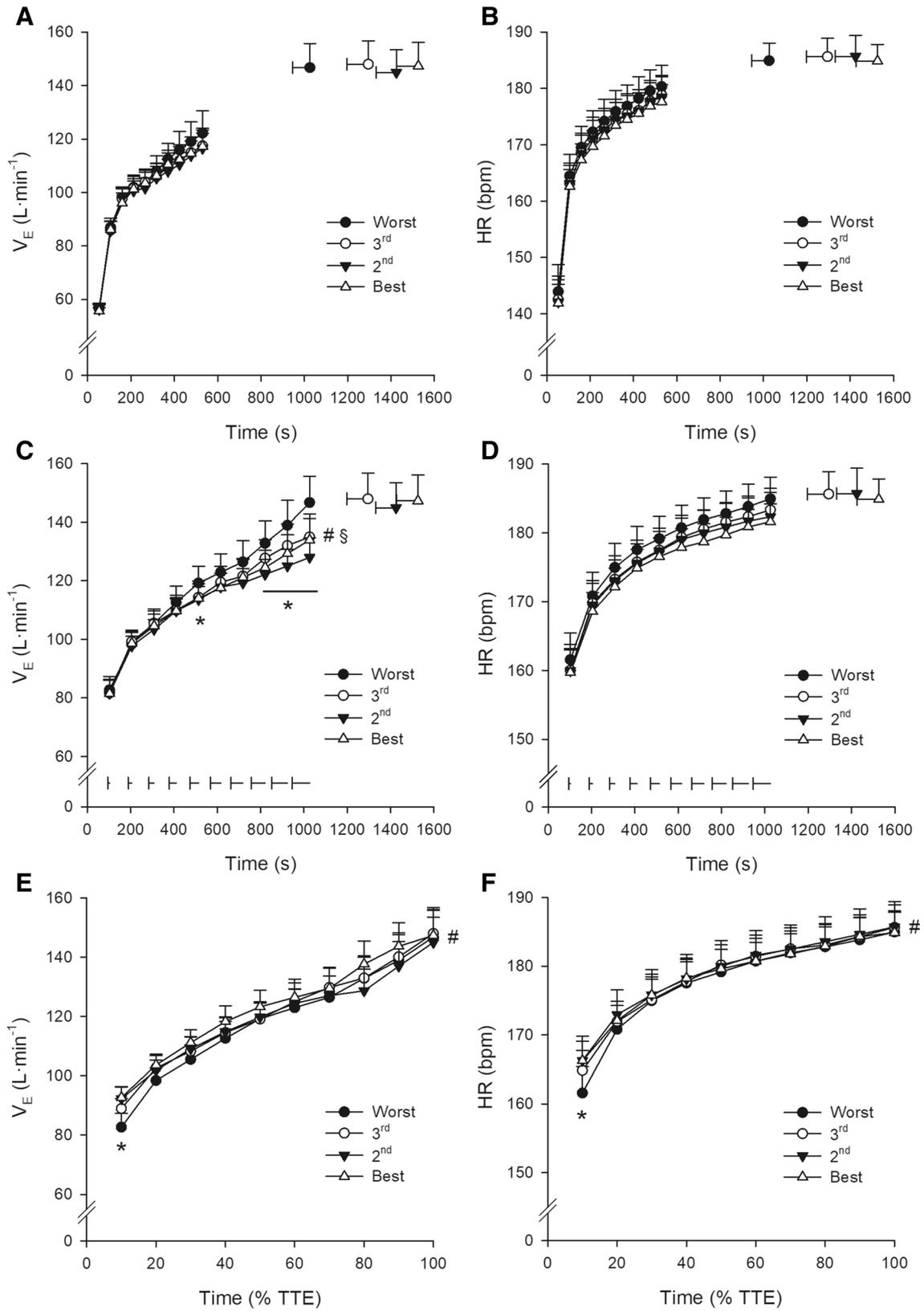
◀**Fig. 3** $\dot{V}_E$ and HR responses for the worst (filled circles), third (open circles), second (filled triangles), and best (open triangles) test analysed with the "group isotime" (**a**, **b**), "individual isotime" (**c**, **d**), and "relative isotime" (**e**, **f**) methods. To preserve the clarity of panels **c** and **d**, the horizontal error bar is depicted in the lower part of the two panels, being the time error identical across the four tests when using the "individual isotime" analysis. #Significant interaction ($P < 0.05$). §Significant main effect of rank ($P < 0.05$). *Significant simple main effect of rank ($P < 0.05$)

experimental condition more evident in the second half of a TTE test [10]. Therefore, the partial data loss that occurs when using the "group isotime" method may considerably affect the sensitivity of RPE and f_R to an experimental intervention or another independent variable. This hypothesis was supported in the present study.

When analysed with the "group isotime" method, no significant differences in f_R, $\dot{V}_E$, RPE, and affective valence were found across tests. Conversely, when analysed with the "individual isotime" method, f_R, $\dot{V}_E$ and RPE values were higher, while affective valence values were lower, in the worst test compared to the other tests, and the extent of the differences increased over time. This emphasises how the data loss that occurs with the "group isotime" method profoundly affects the results of the study, and this is evident in Figs. 2, 3, and 4. Therefore, as expected, f_R, $\dot{V}_E$, RPE, and affective valence were sensitive to performance ranking, but only when the "individual isotime" method was used. This shows how the interpretation of the phenomenon under study may change dramatically depending on the analysis conducted. Our findings strongly suggest that researchers should use the "individual isotime" instead of the traditional "group isotime" method.

The importance of our findings can be further appreciated if they are read against the previous studies that used the traditional "group isotime" method. The extent of data loss found in the present study when using the "group isotime" method (48%) is similar to that found in the previous research [8–12, 14–16], with these studies having values ranging from 42% [11] to 53% [10]. This suggests that the use of the "group isotime" method may have affected the results and interpretation of a number of the previous studies. Therefore, quantifying the extent of data loss is useful when interpreting data from the previous studies, because the higher the extent of data loss, the higher the probability that the use of the "group isotime" method may have influenced the results. For instance, Marcora et al. [10] found no effect of muscle fatigue on RPE in the first of two similar studies, where 53% of data loss occurred as a consequence of the use of the "group isotime" method. Conversely, in the second study, the authors [10] found a higher RPE in the muscle fatigue condition when the control condition was performed, for each individual, at exactly the same duration of the muscle fatigue condition

to avoid any data loss. Therefore, our findings should be considered when interpreting the previous results obtained with the "group isotime" method.

Our findings clearly show that the "relative isotime" analysis is not an alternative method to the individual "isotime analysis" and that it cannot be used to assess the effects of an independent variable. Indeed, very different effects were found for physiological and psychological variables when comparing the "relative isotime" with the "individual isotime" method (Table 1). For instance, the large differences observed across tests for f_R when using the "individual isotime" method were not revealed with the use of the "relative isotime" method. Furthermore, when analysed with the "individual isotime" method, RPE was higher over the last minutes of the worst TTE test compared to the other TTE tests, while it was, conversely, lower when analysed with the "relative isotime" method. On the other hand, it is not surprising that variables representing effort show similar responses across different TTE tests when values are compared at the same relative distances from the point where a maximal effort is exerted (i.e., using the "relative isotime" analysis), as also found in other studies [18, 19]. Therefore, the "relative isotime" method is not informative of the between-test differences that may occur at the same absolute timepoints. However, isotime comparisons are needed in most of the studies using TTE tests. For instance, the effects of different environmental conditions [19] or muscle glycogen depletion [18] on RPE are not revealed if data are expressed against relative time instead of absolute time. Nevertheless, the "relative isotime" method can be used for correlating the responses of different variables (as done in the present study for RPE and f_R), as it is the only method that results in no data loss.

In the light of our findings, it is surprising how the "individual isotime" method has received limited attention so far. As it had not previously been compared with other methods of analysis, there may be limited awareness of its importance. In addition, there are a number of discrepancies in the way that the "individual isotime" method has been reported in the previous studies [20–22, 40, 41], sometimes leaving uncertainty over which method of analysis was used. Therefore, we provide some guidelines for TTE analysis reporting that researchers are encouraged to follow. First, we suggest that researchers use the terminology adopted in the present manuscript, where the rationale for the terms "group isotime", "individual isotime", and "relative isotime" have been explained. Second, we discourage expressing data processed with the "individual isotime" method as a % of TTE (e.g., [40]), because this may lead researchers to confuse the "individual isotime" with the "relative isotime" method. Rather, we suggest expressing data against absolute time and adding horizontal error bars (see Figs. 2, 3, 4 for an example). Third, we suggest to analyse data collected during TTE tests

Fig. 4 Affective valence response for the worst (filled circles), third ▶ (open circles), second (filled triangles), and best (open triangles) tests analysed with the "group isotime" (**a**), "individual isotime" (**b**), and "relative isotime" (**c**) methods. To preserve the clarity of the panel **b**, the horizontal error bar is depicted in the lower part of the panel, being the time error identical across the four tests when using the "individual isotime" analysis. [#]Significant interaction ($P < 0.05$). *Significant simple main effect of rank ($P < 0.05$)

using the MATLAB code provided here as Supplementary material (Online Resource 1). This would clarify the analysis used, and would avoid any error due to manual data processing.

While we used a classic constant-workload exercise test, the present findings also apply to a variety of exercise protocols characterized by variable workloads that are performed to exhaustion. Among these, the incremental exercise test is commonly used to evaluate exercise tolerance and measure key physiological parameters. Intermittent TTE tests are also widely used, particularly in the field of neuromuscular physiology [42]. Furthermore, our findings can be extended to animal studies, where TTE tests are the preferential exercise protocols used to evaluate exercise tolerance [43]. Collectively, TTE tests, in their various forms, have contributed substantially to our understanding of the mechanisms underlying exercise tolerance and fatigue in humans and animals. However, for a deeper understanding of the physiological and psychological mechanisms of exercise tolerance, further research should pay careful attention to the method used to analyse data collected during TTE tests.

Conclusion

The present study shows that the method used to process data collected during TTE tests dramatically affects the magnitude and the statistical significance of the effect of the independent variable on the dependent variables. Investigating the effect of within-subject performance ranking on physiological and psychological variables, we found that f_R, $\dot{V}_E$, RPE, and affective valence are sensitive to performance ranking, but this was the case only when the "individual isotime" method was used. This method greatly reduces the partial data loss that occurs when the traditional "group isotime" method is used. We also provided detailed information on how to use the "individual isotime" method to encourage its use in future studies. Based on our findings, researchers are strongly encouraged to use the "individual isotime" method instead of the "group isotime" method, to correctly
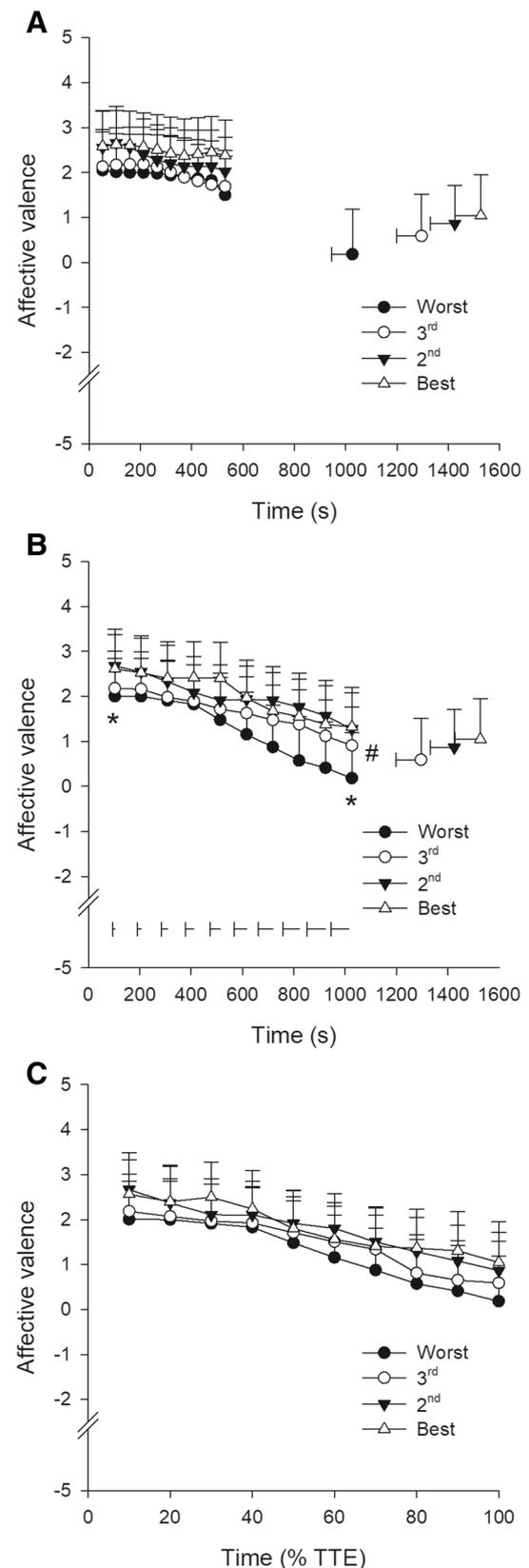
Table 1 Physiological and psychological variables analysed with the three methods

Analysis	Variable	Rank		Time		Interaction	
		P	η_p^2	P	η_p^2	P	η_p^2
Group isotime	f_R (breaths min^{-1})	0.547	0.067	<0.001[†]	0.853	0.677	0.079
	$\dot{V}_E$ (L min^{-1})	0.362	0.100	<0.001[†]	0.894	0.123	0.119
	HR (bpm)	0.310	0.111	<0.001[†]	0.950	0.995	0.040
	RPE	0.437	0.085	<0.001[†]	0.858	0.114	0.120
	Affective valence	0.218	0.135	<0.001[†]	0.439	0.411	0.094
Individual isotime	f_R (breaths min^{-1})	0.001[†]	0.417	<0.001[†]	0.862	<0.001[†]	0.294
	$\dot{V}_E$ (L min^{-1})	0.004[†]	0.349	<0.001[†]	0.827	<0.001[†]	0.281
	HR (bpm)	0.125	0.171	<0.001[†]	0.849	0.801	0.071
	RPE	0.083	0.196	<0.001[†]	0.919	0.015[†]	0.148
	Affective valence	0.094	0.189	<0.001[†]	0.467	0.021[†]	0.144
Relative isotime	f_R (breaths min^{-1})	0.199	0.142	<0.001[†]	0.885	0.057	0.131
	$\dot{V}_E$ (L min^{-1})	0.278	0.119	<0.001[†]	0.812	0.005[†]	0.162
	HR (bpm)	0.722	0.043	<0.001[†]	0.887	<0.001[†]	0.266
	RPE	0.031[†]	0.253	<0.001[†]	0.932	0.054	0.131
	Affective valence	0.265	0.122	<0.001[†]	0.498	0.721	0.076

Rank main effect of rank, *Time* main effect of time, *Interaction* rank×time interaction, f_R respiratory frequency, $\dot{V}_E$ minute ventilation, *HR* heart rate, *RPE* rating of perceived exertion

[†]Significant at $P < 0.05$
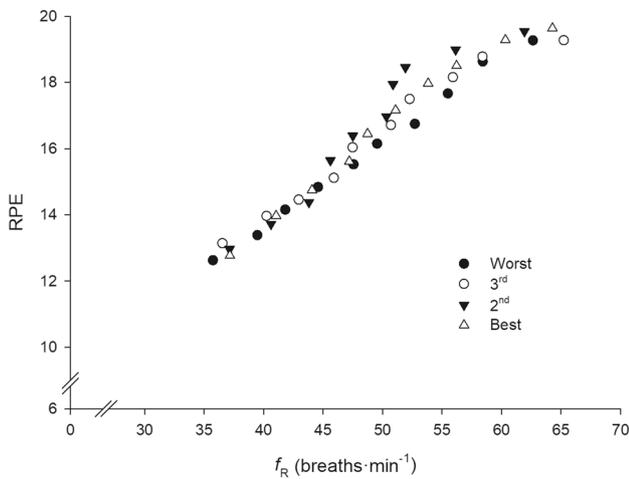
Fig. 5 Correlations between RPE and f_R for the worst (filled circles), third (open circles), second (filled triangles), and best (open triangles) tests. Each symbol represents the mean value of all participants at each percentage of the TTE

interpret the phenomenon under study. The arising implications extend to incremental exercise, which is another commonly used test of exercise tolerance.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee (Ethics Committee of the University of Rome Sapienza) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Whipp BJ, Ward SA (2009) Quantifying intervention-related improvements in exercise tolerance. Eur Respir J 33:1254–1260. https://doi.org/10.1183/09031936.00110108

2. Jeukendrup A, Saris WH, Brouns F, Kester AD (1996) A new validated endurance performance test. Med Sci Sports Exerc 28:266–270

3. Hinckson EA, Hopkins WG (2005) Reliability of time to exhaustion analyzed with critical-power and log-log modeling. Med Sci Sports Exerc 37:696–701. https://doi.org/10.1249/01.MSS.0000159023.06934.53

4. Amann M, Hopkins WG, Marcora SM (2008) Similar sensitivity of time to exhaustion and time-trial time to changes in endurance. Med Sci Sports Exerc 40:574–578. https://doi.org/10.1249/MSS.0b013e31815e728f

5. Blondel N, Berthoin S, Billat V, Lensel G (2001) Relationship between run times to exhaustion at 90, 100, 120, and 140% of vVO_{2max} and velocity expressed relatively to critical velocity and maximal velocity. Int J Sports Med 22:27–33. https://doi.org/10.1055/s-2001-11357

6. Mann T, Lamberts RP, Lambert MI (2013) Methods of prescribing relative exercise intensity: physiological and practical considerations. Sports Med 43:613–625. https://doi.org/10.1007/s40279-013-0045-x

7. Faude O, Hecksteden A, Hammes D et al (2017) Reliability of time-to-exhaustion and selected psycho-physiological variables during constant-load cycling at the maximal lactate steady-state. Appl Physiol Nutr Metab 42:142–147. https://doi.org/10.1139/apnm-2016-0375

8. Fulco CS, Lewis SF, Frykman PN et al (1996) Muscle fatigue and exhaustion during dynamic leg exercise in normoxia and hypobaric hypoxia. J Appl Physiol 81:1891–1900

9. Marcora SM, Staiano W, Manning V (2009) Mental fatigue impairs physical performance in humans. J Appl Physiol 106:857–864. https://doi.org/10.1152/japplphysiol.91324.2008

10. Marcora SM, Bosio A, de Morree HM (2008) Locomotor muscle fatigue increases cardiorespiratory responses and reduces performance during intense cycling exercise independently from metabolic stress. Am J Physiol Regul Integr Comp Physiol 294:R874–R883. https://doi.org/10.1152/ajpregu.00678.2007

11. Taylor BJ, Romer LM (2008) Effect of expiratory muscle fatigue on exercise tolerance and locomotor muscle fatigue in healthy humans. J Appl Physiol 104:1442–1451. https://doi.org/10.1152/japplphysiol.00428.2007

12. Yoon T, Schlinder-Delap B, Keller ML, Hunter SK (2012) Supraspinal fatigue impedes recovery from a low-intensity sustained contraction in old adults. J Appl Physiol 112:849–858. https://doi.org/10.1152/japplphysiol.00799.2011

13. Girard O, Racinais S (2014) Combining heat stress and moderate hypoxia reduces cycling time to exhaustion without modifying neuromuscular fatigue characteristics. Eur J Appl Physiol 114:1521–1532. https://doi.org/10.1007/s00421-014-2883-0

14. Bastos-Silva VJ, de Melo A, Lima-Silva AE et al (2016) Carbohydrate mouth rinse maintains muscle electromyographic activity and increases time to exhaustion during moderate but not high-intensity cycling exercise. Nutrients 8:49. https://doi.org/10.3390/nu8030049

15. Astokorki AHY, Mauger AR (2017) Transcutaneous electrical nerve stimulation reduces exercise-induced perceived pain and improves endurance exercise performance. Eur J Appl Physiol 117:483–492. https://doi.org/10.1007/s00421-016-3532-6

16. Bowtell JL, Mohr M, Fulford J et al (2018) Improved exercise tolerance with caffeine is associated with modulation of both peripheral and central neural processes in human participants. Front Nutr 5:6. https://doi.org/10.3389/fnut.2018.00006

17. Pires FO, Noakes TD, Lima-Silva AE et al (2011) Cardiopulmonary, blood metabolite and rating of perceived exertion responses to constant exercises performed at different intensities until exhaustion. Br J Sports Med 45:1119–1125. https://doi.org/10.1136/bjsm.2010.079087

18. Noakes T (2004) Linear relationship between the perception of effort and the duration of constant load exercise that remains. J Appl Physiol 96:1571–1572. https://doi.org/10.1152/japplphysiol.01124.2003 **(author reply 1572–1573)**

19. Crewe H, Tucker R, Noakes TD (2008) The rate of increase in rating of perceived exertion predicts the duration of exercise to fatigue at a fixed power output in different environmental conditions. Eur J Appl Physiol 103:569–577. https://doi.org/10.1007/s00421-008-0741-7

20. Barbosa TC, Machado AC, Braz ID et al (2015) Remote ischemic preconditioning delays fatigue development during handgrip exercise. Scand J Med Sci Sports 25:356–364. https://doi.org/10.1111/sms.12229

21. Gagnon P, Bussières JS, Ribeiro F et al (2012) Influences of spinal anesthesia on exercise tolerance in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 186:606–615. https://doi.org/10.1164/rccm.201203-0404OC

22. Blanchfield AW, Hardy J, De Morree HM et al (2014) Talking yourself out of exhaustion: the effects of self-talk on endurance performance. Med Sci Sports Exerc 46:998–1007. https://doi.org/10.1249/MSS.0000000000000184

23. Nicolò A, Bazzucchi I, Haxhi J et al (2014) Comparing continuous and intermittent exercise: an "isoeffort" and "isotime" approach. PLoS One 9:e94990. https://doi.org/10.1371/journal.pone.0094990

24. Nicolò A, Marcora SM, Bazzucchi I, Sacchetti M (2017) Differential control of respiratory frequency and tidal volume during high-intensity interval training. Exp Physiol 102:934–949. https://doi.org/10.1113/EP086352

25. Nicolò A, Marcora SM, Sacchetti M (2016) Respiratory frequency is strongly associated with perceived exertion during time trials of different duration. J Sports Sci 34:1199–1206. https://doi.org/10.1080/02640414.2015.1102315

26. Pires FO, Lima-Silva AE, Bertuzzi R et al (2011) The influence of peripheral afferent signals on the rating of perceived exertion and time to exhaustion during exercise at different intensities. Psychophysiology 48:1284–1290. https://doi.org/10.1111/j.1469-8986.2011.01187.x

27. Koglin L, Kayser B (2013) Control and sensation of breathing during cycling exercise in hypoxia under naloxone: a randomised controlled crossover trial. Extrem Physiol Med 2:1. https://doi.org/10.1186/2046-7648-2-1

28. Davies RC, Rowlands AV, Eston RG (2009) Effect of exercise-induced muscle damage on ventilatory and perceived exertion responses to moderate and severe intensity cycle exercise. Eur J Appl Physiol 107:11–19. https://doi.org/10.1007/s00421-009-1094-6

29. Nicolò A, Massaroni C, Passfield L (2017) Respiratory frequency during exercise: the neglected physiological measure. Front Physiol 8:922. https://doi.org/10.3389/fphys.2017.00922

30. Karsten B, Jobson SA, Hopker J et al (2015) Validity and reliability of critical power field testing. Eur J Appl Physiol 115:197–204. https://doi.org/10.1007/s00421-014-3001-z

31. Galbraith A, Hopker J, Lelliott S et al (2014) A single-visit field test of critical speed. Int J Sports Physiol Perform 9:931–935. https://doi.org/10.1123/ijspp.2013-0507

32. Borg G (1998) Borg's perceived exertion and pain scales. Human Kinetics, Champaign

33. García-Manso JM, Martín-González JM, Vaamonde D, Da Silva-Grigoletto ME (2012) The limitations of scaling laws in the prediction of performance in endurance events. J Theor Biol 300:324–329. https://doi.org/10.1016/j.jtbi.2012.01.028

34. Hardy CJ, Rejeski WJ (1989) Not what, but how one feels: the measurement of affect during exercise. J Sport Exerc Psychol 11:304–317. https://doi.org/10.1123/jsep.11.3.304

35. Lamarra N, Whipp BJ, Ward SA, Wasserman K (1987) Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. J Appl Physiol 62:2003–2012. https://doi.org/10.1152/jappl.1987.62.5.2003

36. Hopkins WG (2015) Spreadsheets for analysis of validity and reliability. Sportscience 19:36–42. http://sportsci.org/2015/ValidRely.htm

37. Cohen J (1988) Statistical power analysis for the behavioural sciences, 2nd edn. Lawrence Earlbaum Associates, Hillsdale

38. Bland JM, Altman DG (1995) Calculating correlation coefficients with repeated observations: part 1—correlation within subjects. BMJ 310:446

39. Hayashi K, Honda Y, Ogawa T et al (2006) Relationship between ventilatory response and body temperature during prolonged submaximal exercise. J Appl Physiol 100:414–420. https://doi.org/10.1152/japplphysiol.00541.2005

40. Mauger AR, Taylor L, Harding C et al (2014) Acute acetaminophen (paracetamol) ingestion improves time to exhaustion during exercise in the heat. Exp Physiol 99:164–171. https://doi.org/10.1113/expphysiol.2013.075275

41. Blanchfield A, Hardy J, Marcora S (2014) Non-conscious visual cues related to affect and action alter perception of effort and endurance performance. Front Hum Neurosci 8:967. https://doi.org/10.3389/fnhum.2014.00967

42. Bigland-Ritchie B, Furbush F, Woods JJ (1986) Fatigue of intermittent submaximal voluntary contractions: central and peripheral factors. J Appl Physiol 61:421–429

43. Matsumoto K, Ishihara K, Tanaka K et al (1996) An adjustable-current swimming pool for the evaluation of endurance capacity of mice. J Appl Physiol 81:1843–1849. https://doi.org/10.1152/jappl.1996.81.4.1843