Wearable Assessment of Level and Uphill Running at Critical Intensity

Nikolai Kirner[1, 2], Wouter Hoogkamer[1]

[1] Integrative Locomotion Laboratory, Department of Kinesiology, University of Massachusetts, Amherst, MA, USA

[2] Department of Health and Sport Sciences, TUM School of Medicine and Health, Technical University of Munich, Munich, Germany

Corresponding Author:

Nikolai Kirner (E-Mail: nikolai.kirner@tum.de)

ORCID numbers of the authors:

Nikolai Kirner: 0009-0005-7748-3500

Wouter Hoogkamer: 0000-0003-0524-0196

Author Contributions

Nikolai Kirner (N.K.) and Wouter Hoogkamer (W.H.) conceived and designed research; N.K. performed experiments; N.K. and W.H. analyzed data; N.K. and W.H. interpreted results of experiments; N.K. prepared figures; N.K. and W.H. drafted the manuscript; N.K. and W.H. edited and revised the manuscript; N.K. and W.H. approved the final version of the manuscript.

Acknowledgments

We thank all runners for participating in the study and Caroline Kolmodin, Keegan Butler, Nadia Surin, and Nathan Hammerschmitt Le Gal for their help with data collection. We also thank Dr. Fabian Stöcker for providing the Stryd foot pod.

Abstract

Purpose:

We used commercially available wearable devices that measure running power (foot- vs. wrist-based) and muscle oxygenation to investigate time-to-exhaustion during level and uphill running at critical intensity.

Methods:

14 healthy recreational runners participated in this study. They first ran a 3-minute all-out test to derive their critical running intensity. They then performed two time-to-exhaustion trials (TTEs) at this intensity at separate slopes to assess time-to-exhaustion, muscle oxygenation, and rating of perceived exertion.

Results:

We observed a shorter time-to-exhaustion (539 ± 160 vs. 936 ± 336 s), a faster decline in muscle oxygenation over time (-2.48 ± 1.32 vs. -1.62 ± 2.11 $\%\cdot min^{-1}$), a higher rating of perceived exertion at matched time points, and a higher estimated metabolic rate (19.51 ± 1.99 vs. 17.82 ± 1.62 $W\cdot kg^{-1}$) for the level time-to-exhaustion test (TTE_{Level}) compared to the uphill time-to-exhaustion test ($TTE_{Incline}$) at a 2.86° (5%) slope.

Conclusion:

Our data suggests that athletes should be careful when using running power at critical intensity to guide their racing and training in non-level conditions.

Keywords: *Critical Power, Critical Speed, Time-to-exhaustion, Running Power, 3-minute all-out test, Trail running*

Abbreviations

3MT	3-minute all-out test
$Borg_{3min}$	Rating of Perceived Exertion using the Borg scale at 3-min elapsed time
$Borg_{6min}$	Rating of Perceived Exertion using the Borg scale at 6-min elapsed time
$Borg_{9min}$	Rating of Perceived Exertion using the Borg scale at 9-min elapsed time
$Borg_{End}$	Rating of Perceived Exertion using the Borg scale at the end of the test
CP / CP_{Coros}	Critical Power / Critical Power derived from the Coros Watch
CS / CS_{Coros}	Critical Speed / Critical Speed derived from the Coros Watch
D’	D-Prime (a proxy for anaerobic working capacity above Critical Speed)
HTK	Hoogkamer et al.’s (2014) equation for estimating metabolic rate
NaN	Not a number
$Power_{TTE}$	Running power of the time-to-exhaustion test
RE3	Looney et al.’s (2023) equation for estimating metabolic rate
RPE	Rating of Perceived Exertion
$Time_{TTE}$	Time of the time-to-exhaustion test
TTE	Time-to-exhaustion test
$TTE_{Incline}$	Uphill time-to-exhaustion test
TTE_{Level}	Level time-to-exhaustion test
W’	W-Prime (a proxy for anaerobic working capacity above Critical Power)

Introduction

In 1986, Ulrich Schoberer developed the first spider-based power meter for cycling to measure an athlete's work rate during training and racing. The system revolutionized the sport by enabling cyclists to quantify their effort in a single metric independent of weather conditions, terrain, and physiological considerations using their own bikes. The spider-based power meter uses strain gauges to measure the force applied on the chainring during pedaling, considering the crank's angular velocity (cadence) (Montalvo-Pérez et al. 2021). Further systems, such as pedal- or crank-based power meters, have expanded the technology over the last decades (Passfield et al. 2017) since early adopters such as the East German Cycling National Team and Tour de France Champions Greg LeMond started to use the system in the 1990s, with LeMond attributing his victories to the use of a power meter (SRM GmbH 2024).

Based on this success, engineers, inventors, and scientists have tried to develop a similar device for running that provides an effort metric going beyond heart rate or speed measurements, which are substantially affected by factors such as temperature (Rojas-Valverde et al. 2020), slope (Margaria et al. 1963; Brill and Kram 2021), physiological response times or caffeine intake (Turley 2016; Stadheim et al. 2021) to highlight a few. With the introduction of running power devices in the mid-2010s, estimating power is now accessible using commercially available sensors worn at the wrist (e.g., Garmin and Coros watches) or foot (e.g., Stryd and RunScribe). However, running power is more controversial than cycling power (van Rassel et al. 2023).

In cycling, riders produce mechanical power by applying torque on the crank, achieving rotational velocity (Passfield et al. 2017), which is highly correlated with metabolic rate (Gaesser and Brooks 1975). Any fluctuations in external power demands due to drivetrain efficiency, tire pressure, road surface, air resistance, or slope at a constant power output will affect cycling speed, but not metabolic rate. Hence, a rider's power output reflects their metabolic effort to reach desired speeds. Contrarily, in running, mechanical power output is poorly correlated to metabolic rate (Heglund et al. 1982). Running energetics are characterized by negative work during the first half of the stance phase, followed by positive work during the second half of the stance phase (Cavagna et al. 1964), but most of that work is performed through elastic strain energy storage and return in passive tissues rather than by active muscle fiber work (Fenn 1930; Cavagna and Kaneko, 1977; Lai et al. 2014; Swinnen et al. 2019; Monte et al. 2020). To keep these elastic tissues under tension, muscles must produce force isometrically without performing work (Fenn 1930). Therefore, the metabolic cost of running is dominated by the metabolic cost of generating force (Kram and Taylor, 1990). Furthermore, while the positive work in cycling is done by active muscle fiber work at a fairly constant efficiency, the running efficiency of any actively performed

mechanical work will differ between positive and negative work (Margaria 1968). These differences in elastic energy storage and return and varying efficiencies for positive vs. negative work complicate the use of mechanical power to assess metabolic effort in running.

To circumvent these challenges, commercially available running sensors use proprietary algorithms to estimate the positive mechanical work and the associated metabolic energy expenditure. One of the most popular devices is Stryd's foot pod (Stryd Inc. Boulder, CO, USA), which uses a 6-axis inertial motion sensor (3-axis gyroscope and 3-axis accelerometer). Stryd's internal testing suggests that their mechanical power output estimation ($W \cdot kg^{-1}$) is highly correlated to oxygen uptake ($ml \cdot min^{-1} \cdot kg^{-1}$) ($R^2 = 0.96$), allegedly confirmed by third-party research (Stryd Team 2017). Peer-reviewed research has reported more mixed and controversial results (see Aubry et al. (2018) and Snyder et al. (2018)). Taboga et al. (2022) reported that the sensor's power estimates correlate with force plate-based power and metabolic rate across speeds for level treadmill running. Similarly, Imbach et al. (2020) observed correlations between Stryd running power and oxygen uptake for six individuals over a range of speeds during overground running. For running power to be more beneficial than running speed for measuring running intensity, validity must be assessed beyond speed dependence. Unfortunately, Stryd's running power appears to be not adequately sensitive to differences in metabolic effort between advanced footwear technology and traditional running shoes (Hudgins et al. 2024). Furthermore, the validity of running power during uphill and downhill running is widely overlooked in the current literature (Apte et al. 2023). Altogether, it remains to be determined whether the Stryd Food Pod can accurately reflect changes in metabolic rate due to differences in running conditions other than speed (Austin et al. 2018).

Interestingly, the Stryd foot pod has been extensively tested to estimate critical power (CP) and critical speed (CS) using time-trial efforts (Dearing & Paton 2023; Hunter et al. 2023; van Rassel et al. 2024) or a 3-minute all-out test (3MT) (Hunter et al. 2023). CP estimations using Stryd's foot pod have also been compared and validated to anaerobic threshold concepts (Ruiz-Alias et al. 2022; van Rassel et al. 2023). Hence, from a practical perspective, the validity of running power concerning metabolic rate in varying conditions appears secondary to its validity in providing consistent critical power estimates to inform training intensities and race strategies.

We investigated the Stryd foot pod running power estimation for level and uphill running using the CS and CP paradigms. We assessed time-to-exhaustion during level and uphill running at critical power, used near-infrared spectroscopy sensors to evaluate the physiological responses, measured rating of perceived exertion (RPE) to objectify the psycho-physiological strain on the runners, and estimated metabolic rate.

Methods

Fourteen healthy runners participated in the study (2 females; identified via self-reported sex assigned at birth; 70.7 ± 15.3 kg; 1.79 ± 0.13 m; 25.1 ± 5.7 years old). Participants gave their written informed consent before completing the first of three study visits as approved by the University of Massachusetts, Amherst Institutional Review Board (Protocol 5639). All study sessions, comprised of one session on the track and two lab sessions on the treadmill, were completed within 2 weeks, with a maximum of 1 week between each visit. Participants were recreational or trained runners who ran at least 10 miles weekly (up to 70 miles) with no lower limb injuries three months prior and during study participation.

Session 1

Participants completed a physical activity readiness questionnaire (PAR-Q, Chisholm et al. 1975) before we recorded their body mass (portable scale) and height (self-reported). We took exact height measurements later at the start of the first lab visit (session 2). We placed a latest-generation Stryd foot pod centrally on the participant's left shoelaces, according to the manufacturer's guidelines. Additionally, participants wore a GPS running watch (Apex 2, Coros Wearables Inc. Shenzhen, China) on their left arm to record speed and watch-based running power with the display on the opposite side of their wrist. We set the participant's mass and height before each trial using the Stryd App (Version 8.17.64) in Outdoor Run Mode and the Coros Watch in Track Mode. After a short device initialization, participants ran a self-paced 8-minute warmup, including two sub-maximal sprints of 20 meters. Following a passive rest of at least 5 minutes, participants performed a self-paced 3MT in lane 1. An examiner gave strong verbal encouragement to the participants riding alongside on a bike and ended their test after ~3:10 minutes; a minimum test duration of 3:05 minutes ensured sufficient data sampling (Pettitt 2016). We calculated CS and CP as the average running speed and running power from the Stryd foot pod over the last 30 seconds of the 3MT (Hunter et al. 2023), starting 150 seconds after the first recorded data point for power, which matched the first recorded positive speed value. We also calculated and compared CS and CP for Coros' Speed and Power data by calculating the average running speed and power of the last 30 seconds of each data set, starting with the first positive index of each speed and power vector (CS_{Coros} and CP_{Coros}, respectively). CS marked the reference point for sessions 2 and 3. Before each 3MT, an examiner narrated an oral script with an analogy on how to approach the 3MT, enhancing compliance and standardizing the trial across participants (see separate "Supplementary Material" file). Herein, participants acknowledged reaching maximal speed and sustaining the highest speed possible throughout the 3MT without looking at the watch.

Sessions 2 & 3

For sessions 2 and 3, at least 48 hours after session 1, and separated by at least 48 hours, participants performed a time-to-exhaustion test (TTE) on a level (TTE_{Level}) and an inclined treadmill ($TTE_{Incline}$). Both sessions entailed the exact same procedures except for the treadmill's incline, and we randomly assigned the order of these sessions in counterbalance. We used Stryd's Indoor Treadmill Mode for both TTEs.

Following height and mass measurements and foot pod installation, we placed a near-infrared spectroscopy muscle oxygenation sensor (Moxy Monitor, Fortiori Design LLC, Hutchinson, Minnesota, USA) on the participant's left and right rectus femoris muscles, halfway between the anterior superior iliac spine and the top part of the patella (Feldmann et al. 2019). We used the VO2 Master Manager App (Version 0.41.0, VO2 Master Health Sensors Inc. Vernon, BC, Canada), which allows the sampling of up to three Moxy Monitors for data recording. After a self-paced 8-minute warmup, participants performed two initialization trials of maximal 1 minute each, first on a level treadmill (0°) and after a 5-minute passive rest on the same treadmill inclined to 2.86° (5% grade) (Treadmetrix, Park City, UT, USA). The initialization trials determined the treadmill speeds to match the power and prepared the participants for the TTEs. The first initialization trial was set to the participant's CS. The second initialization trial corresponded to the speed at which the participants averaged the same running power as within the first initialization trial. Following another passive rest of 5 minutes, participants ran at either CS for the TTE_{Level} on a 0° level treadmill or up a 2.86° incline on the same treadmill at the speed derived from the second initialization trial, matching the same power output of the first initialization trial for the $TTE_{Incline}$ until exhaustion. We used CS as the reference marker instead of CP because Stryd's Indoor Mode does not account for their reported Air Power, which estimates the air resistance at a given speed considering the wind conditions affecting Stryd's running power estimate outdoors. According to Stryd's white paper, this can lead to a power difference of up to 10 W or more between Outdoor and Indoor Mode (Stryd Team 2020). Moreover, during our pilot testing phase, participants ran at 105% CP for both TTEs, for which the speed had to be increased to around 110% CS from the 3MT.

The following criteria standardized the end time of each TTE: a) Participants asked or gestured the researchers to stop the treadmill; b) Participants grabbed the safety railing of the treadmill unable to keep up the speed; c) Participants drifted backward beyond a zone we had marked on the treadmill with tape for more than 3 seconds; d) Researches stopped the treadmill as the participants showed severe signs of fatigue (e.g. losing balance, slipping) for safety, or e) Researches ended the test after 30 minutes of elapsed time. We measured RPE in 3-minute intervals (Scherr et al. 2013) ($Borg_{3min}$, $Borg_{6min}$, and so on) utilizing the 6-20 original Borg-Scale (Borg and

Dahlström 1962). Additionally, examiners asked for RPE ratings at random intervals to keep the participants unaware of the elapsed time. These additional ratings were not considered for the analysis.

Data Analysis

We used MATLAB (Version R2023b, The MathWorks Inc., Natick, MA, USA) for data and statistical analysis. The primary outcome variable was the time of both TTEs ($Time_{TTE}$). Because of initial treadmill acceleration and power fluctuations were pronounced towards the end of each TTE, we defined $Time_{TTE}$ as the total time [s] spent above a power threshold [W] two standard deviations lower than the participant's mean power [W] from 30 seconds into the TTE until 10 seconds before the first NaN (Not a number) power reading. This enabled standardization without the subjectivity of hand-clocked stopwatch times between participants and examiners. We also calculated the average trial power ($Power_{TTE}$) for each TTE using the same interval as for $Time_{TTE}$.

We collected and exported $\%SmO_2$ from both Moxy sensors at 1 Hz (Matthews et al. 2023) and averaged sensor data of the two monitors. We removed a $\%SmO_2$ data set from the analysis if it contained more than 5% of NaN data or entailed more than two jumps of 5% between two consecutive data points over the entire trial length. We then averaged the subsequent $\%SmO_2$ raw data set over 10 seconds (Kirby et al. 2021; Matthews et al. 2023). We calculated the average trial $\%SmO_2$ as the mean $\%SmO_2$ over the entire exercise trial and the end trial $\%SmO_2$ as the final 5% of the time of each TTE (Kirby et al. 2021). We calculated the "delayed" slope ($\Delta\%SmO_2 \cdot \Delta min^{-1}$) from data obtained 60 seconds after the start (beyond the onset phase) to the end of each trial (Kirby et al. 2021).

We estimated the metabolic rate during both TTEs using the RE3 equation for the TTE_{Level} for level running (eq. 1; Looney et al. 2023) and the HTK equation for uphill running for the $TTE_{Incline}$ (eq. 2; Hoogkamer et al. 2014; HTK based on the last name initials of the authors):

$$4.43 + 1.51 \cdot S + 0.37 \cdot S^2 \qquad \textit{eq. 1}$$

$$2.70 + 0.674 \cdot e^{-18.24 \cdot \sin(\theta)} + \frac{g}{0.294} \cdot \sin(\theta)$$

eq. 2

For calculating the gross metabolic rate of the $TTE_{Incline}$, we multiplied by speed and added the standing metabolic rate of 1.53 $W \cdot kg^{-1}$ as measured in Hoogkamer et al. (2014).

Statistical Analysis

We performed paired t-tests for $Time_{TTE}$, $Power_{TTE}$, Average Trial $\%SmO_2$, End Trial $\%SmO_2$, delayed slope, RPE ($Borg_{3min}$, $Borg_{6min}$, $Borg_{9min}$, and $Borg_{End}$), and estimated metabolic rate between the TTE_{Level} and the $TTE_{Incline}$. Additionally, we compared CS and CP from the Stryd foot pod with CS_{Coros} and CP_{Coros} using paired t-tests. We included mean RPE time points of the TTE_{Level} and the $TTE_{Incline}$ if more than ten paired observations were recorded. We excluded $\%SmO_2$ data for two participants from the statistical analysis according to the criteria listed above. An alpha level of 5% was set a priori.

Results

3MT CS and CP

CS and CP during the 3MT were 4.65 ± 0.41 $m \cdot s^{-1}$ and 353 ± 83 W, respectively, for the Stryd foot pod. CS_{Coros} and CP_{Coros} were 4.99 ± 0.52 $m \cdot s^{-1}$ and 371 ± 93 W for Coros, both significantly higher than the results from the Stryd foot pod ($p < 0.001$ and $p < 0.01$, respectively). Fig. 1 details the Stryd's and Coros' speed and power curves of the 3MT for all participants.

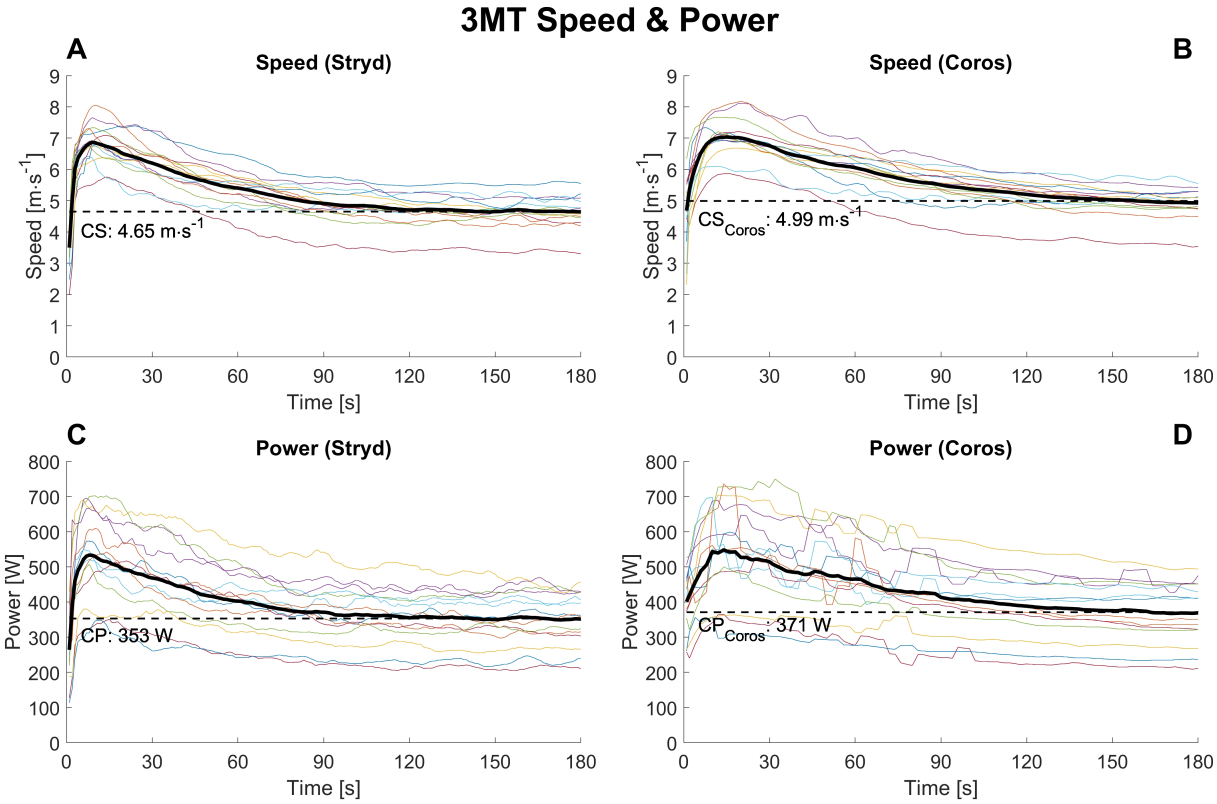

Fig. 1 Speed [m·s^{-1}] and power [W] recording of the Stryd foot pod (Fig. 1a and 1c) and the Coros Apex 2 watch (Fig. 1b and 1d). The black line represents the averaged speed and power data across participants. The dashed line represents the participants' averaged CS, CP, CS_{Coros}, and CP_{Coros}. Each color represents one participant's data

$Time_{TTE}$, $Power_{TTE}$, and Speed

All 14 participants performed both TTEs. For 26 out of the 28 total trials, participants asked the researchers to stop the treadmill (criterion a). For one trial, we stopped the treadmill because the participant was drifting out of the marked zone for more than 3 seconds (criterion c); for another trial, we stopped the treadmill for safety as the participant showed severe signs of fatigue (criterion d).

The participants ran 397 s (-42%) shorter in TTE_{Level} (539 ± 160 s) than in $TTE_{Incline}$ (936 ± 336 s) ($p < 0.001$; Fig. 2). $Power_{TTE}$ was similar between the trials with 331 ± 73 and 329 ± 68 W, respectively ($p = 0.26$). The average speed was 4.65 ± 0.41 m·s^{-1} for the TTE_{Level} and 3.51 ± 0.35 m·s^{-1} for the $TTE_{Incline}$. Participants ran at 94.1 ± 3.3 and 93.8 ± 3.6 % of their CP during the TTE_{Level} and the $TTE_{Incline}$ ($p = 0.56$).

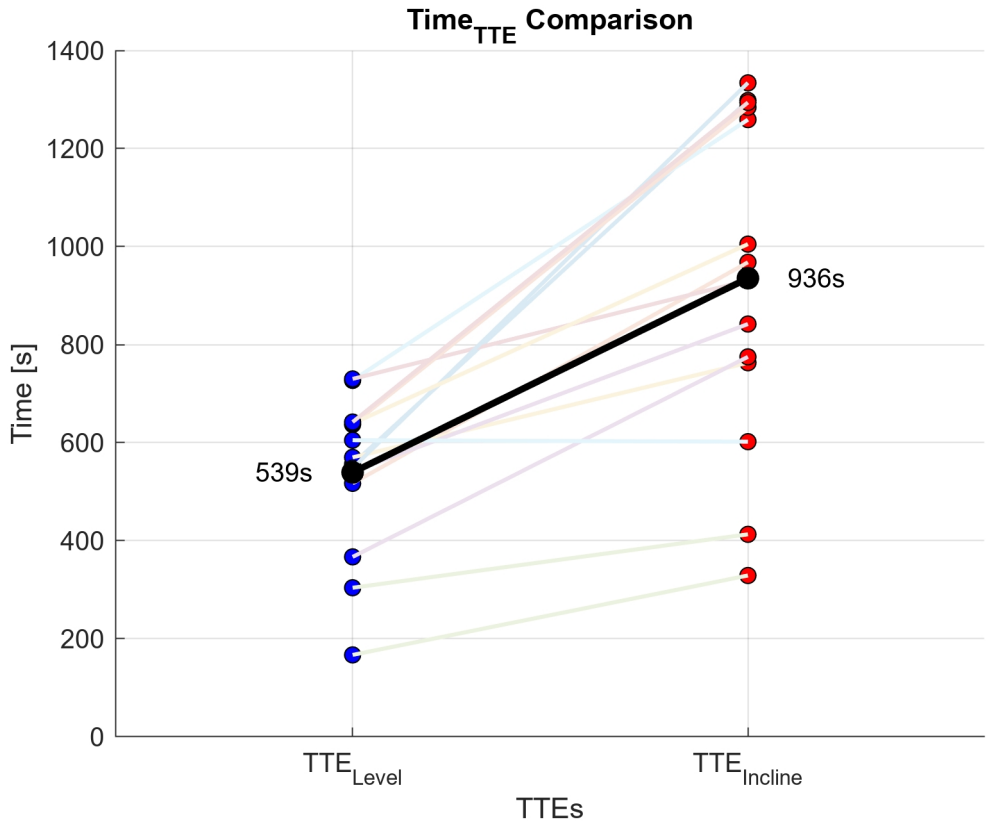

Fig. 2 The participants ran substantially shorter in the level time-to-exhaustion test (TTE_{Level}) than in the uphill time-to-exhaustion test ($TTE_{Incline}$). The black line represents the averaged group data

Muscle O_2 saturation

Average trial $\%SmO_2$ for the TTE_{Level} was 33.3 ± 10.9 % compared to 37.9 ± 11.5 % for the $TTE_{Incline}$ ($p < 0.01$), with an end trial $\%SmO_2$ of 21.3 ± 10.7 % vs. 29.3 ± 11.4 % ($p < 0.001$), respectively. The delayed slope for the TTE_{Level} was -2.48 ± 1.32 $\%\cdot min^{-1}$ compared to -1.62 ± 2.11 $\%\cdot min^{-1}$ for the $TTE_{Incline}$ ($p < 0.05$). Fig. 3 displays the SmO_2 of the $TTE_{Incline}$ and the TTE_{Level} normalized to 100 data points.

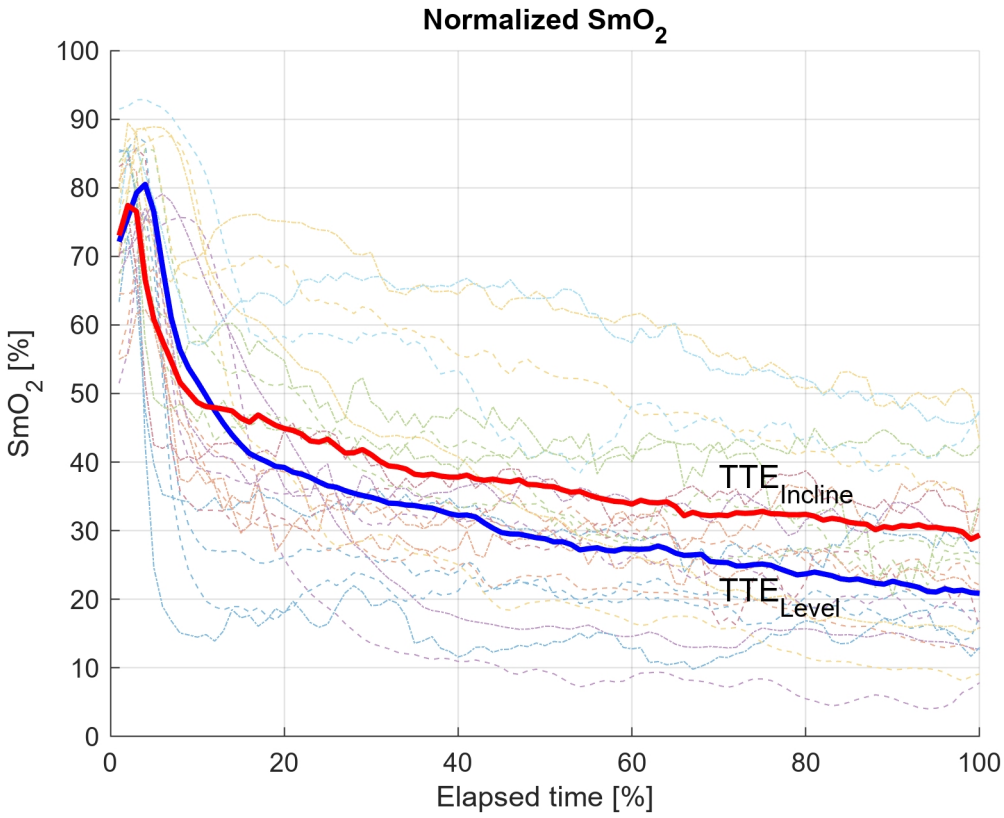

Fig. 3 Muscle oxygenation values (SmO_2 [%]) normalized to trial time decreased relatively faster and more during the level time-to-exhaustion test (TTE_{Level}) than during the uphill time-to-exhaustion test ($TTE_{Incline}$). The bold blue line represents the group's averaged %SmO2 data for the TTE_{Level} and the bold red line the group's averaged %SmO2 data for the $TTE_{Incline}$. Individual data for the TTE_{Level} and $TTE_{Incline}$ are matched in color (TTE_{Level}: Dashed line; $TTE_{Incline}$: dash-dotted line)

RPE (Borg)

Tab. 1 contains a summary of the RPE values at the analyzed time points. $Borg_{End}$ for the TTE_{Level} was 19.5 ± 0.7, similar to 19.4 ± 0.8 for the $TTE_{Incline}$ ($p = 0.34$). For both trials, all participants reported a $Borg_{End}$ of 18 or more. Specifically, eight participants (57%) stated a 20/20 $Borg_{End}$ for both TTEs, four participants (29%) reported a $Borg_{End}$ of 19 for both TTEs, one participant (7%) reported a $Borg_{End}$ of 18 for both TTEs, and one participant (7%) reported a $Borg_{End}$ of 19 for the TTE_{Level}, but 18 for the $TTE_{Incline}$. At each of the 3-minute intervals, the Borg values were significantly higher for TTE_{Level} than for the $TTE_{Incline}$, except for $Borg_{End}$.

Tab. 1 Rating of perceived exertion of the time-to-exhaustion tests (TTEs) using the 6-20 Borg Scale assessed at 3, 6, and 9-minute elapsed time was higher during the level TTE than during the uphill TTE at matched time points, but similar at the end of each TTE ($Borg_{End}$)

RPE (TTE)	TTE_{Level}	$TTE_{Incline}$	p-value
$Borg_{3min}$ (n = 14)	14.6 ± 1.9	13.3 ± 1.2	<0.01
$Borg_{6min}$ (n = 12)	16.6 ± 1.8	14.5 ± 1.3	<0.001
$Borg_{9min}$ (n = 10)	18.6 ± 1.0	16.0 ± 1.5	<0.001
$Borg_{End}$ (n = 14)	19.5 ± 0.7	19.4 ± 0.8	0.34

Estimated Metabolic Rate

The estimated metabolic rate based on running speed and slope was significantly higher during TTE_{Level} (19.51 ± 1.99 $W \cdot kg^{-1}$) than during $TTE_{Incline}$ (17.82 ± 1.62 $W \cdot kg^{-1}$) ($p < 0.001$; Fig. 4).

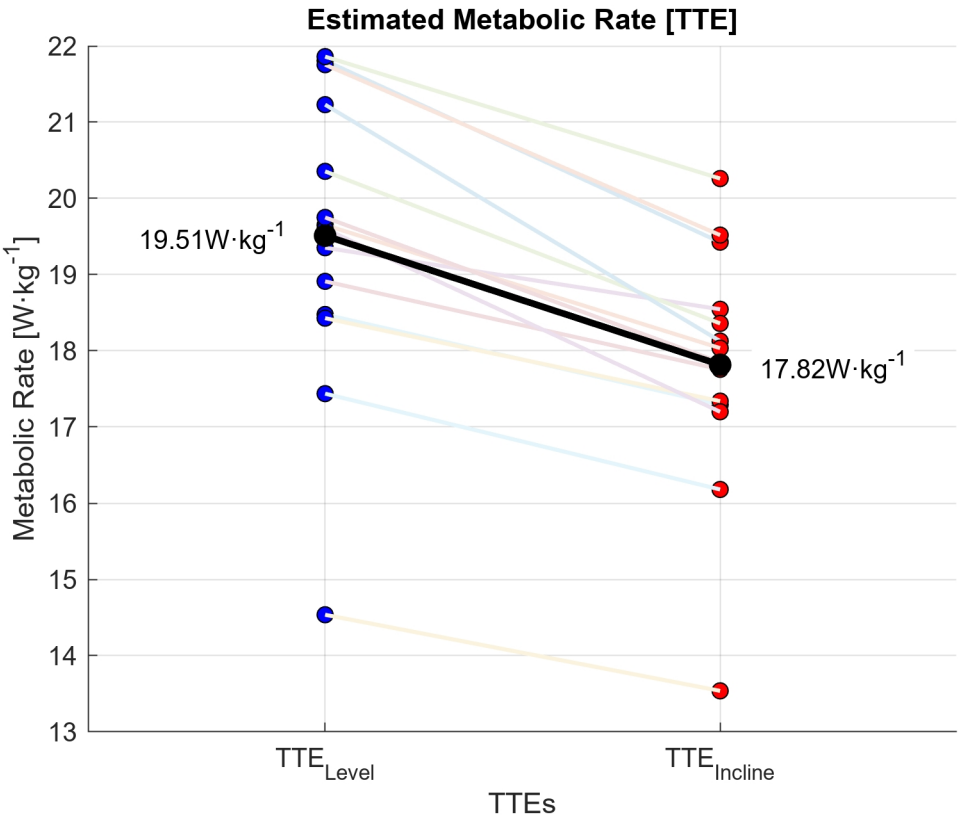

Fig. 4 Using the average speed of both time-to-exhaustion trials, the estimated metabolic rate [W·kg^{-1}] was higher for the level time-to-exhaustion test (TTE_{Level}) than for the uphill time-to-exhaustion test ($TTE_{Incline}$). Individual metabolic rates are connected by a colored solid line. The black line represents averaged group data

Discussion

Running to exhaustion at matching power outputs using Stryd's foot pod resulted in significantly shorter $Time_{TTE}$ during level running than uphill running. This was accompanied by higher muscle oxygen demands measured by the near-infrared spectroscopy sensors, perceived effort (RPE), and estimated metabolic rate during the TTE_{Level} than during the $TTE_{Incline}$.

Our findings imply that when exercising at a comparable effort between the level and uphill trials as assessed by Stryd running power, one actually performs at a lower intensity during uphill running. From a practical perspective, estimating running power mainly adds value beyond running speed when used for altered conditions, such as uphill running or running into a headwind. Stryd running power has been validated against metabolic rate across a wide range of speeds in level running (Imbach et al. 2020; Taboga et al. 2022). We did not directly measure metabolic rate, but using existing empirical equations for metabolic rate during level and uphill running, we estimated significantly higher metabolic rates at critical speed for the level running trial. We focused on CS and

CP as the boundaries of the heavy to the severe exercise intensity domain (Poole et al. 2016) to ensure comparability between both trials at an intensity representative for race performances and critical intensity-based workouts (Clark et al. 2013; Figueiredo et al. 2023).

To adjust the gradient for the $TTE_{Incline}$ in the Stryd app, we used the app's Indoor Mode for both TTEs. For outdoor running, Stryd assesses elevation from barometric pressure changes to correct their power estimation (Stryd Outdoor Mode), but as barometric pressure changes do not occur during uphill treadmill running, a runner needs to manually set the treadmill grade in the app in Indoor Mode. We conclude that the incline factor of Stryd's Indoor Mode does not sufficiently capture the alteration in mechanical power of uphill running.

We estimated the metabolic rate of the TTE_{Level} and the $TTE_{Incline}$ and found significantly higher values for the TTE_{Level} (19.51 ± 1.99 $W \cdot kg^{-1}$) compared to the $TTE_{Incline}$ (17.82 ± 1.62 $W \cdot kg^{-1}$) at matching running power outputs. We used the HTK equation (Hoogkamer et al. 2014) to estimate the metabolic rate for the $TTE_{Incline}$ and the RE3 equation (Looney et al. 2023) for the TTE_{Level}. It is important to note that the HTK equation was developed with speeds of up to 3 $m \cdot s^{-1}$, yet our study's average $TTE_{Incline}$ speed was 3.51 $m \cdot s^{-1,}$ with only one participant running slower than 3 $m \cdot s^{-1}$ for this trial.

Based on these differences in metabolic rate estimations between level and uphill running at similar Stryd running power, we then decided to estimate the metabolic efficiency for each trial by dividing the estimated metabolic power by the recorded running power. For level running, the estimated metabolic efficiency was around 24%, independent of body mass (Fig. 5a). For uphill running, the estimated metabolic efficiency decreased as a function of body mass from 29% for the least heavy participant (43.1 kg) down to 24% for the heaviest participants (over 90 kg) (Fig. 5a). To further assess the metabolic efficiency during level running, relative to the estimated metabolic rate, we created a Bland-Altman plot (Fig. 5b), which confirmed the consistency of this efficiency estimate for level running.

We could not measure Coros Power during the TTEs as the Coros Watch does not estimate power for treadmill running. Therefore, we cannot compare the running power of the TTEs between devices. During the 3MT, Coros Power visually fluctuated more than Stryd Power (Fig. 1). This might be due to Coros' power estimation derived from a more variable GPS-based speed. Also, the recorded total distance during the 3MT using the Coros watch in track mode in lane 1 did not reliably match the actual covered distance on the track, typically overestimating the total distance.

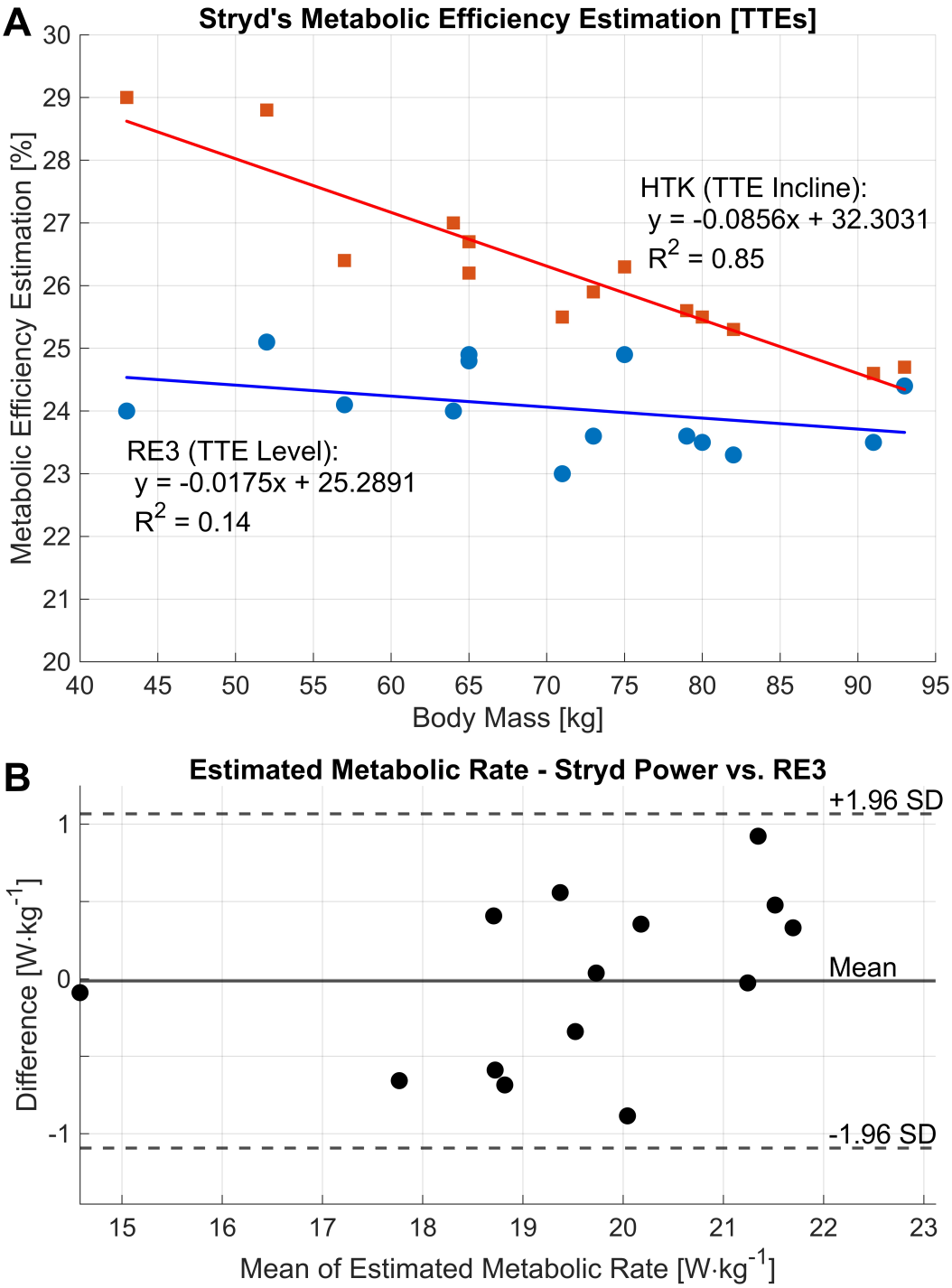

Fig. 5 Metabolic efficiency estimation using linear regression for the level time-to-exhaustion (TTE_{Level} in blue) showed a weak but high correlation for the uphill time-to-exhaustion test ($TTE_{Incline}$ in red) for body mass (Fig. 5a). Linear regression equations and R^2 are displayed next to the regression curves. For the TTE_{Level}, we used the RE3 equation (Looney et al. 2023), and for the $TTE_{Incline}$, we used the HTK equation (Hoogkamer et al. 2014) and corresponding running power of the time-to-exhaustion test ($Power_{TTE}$). The estimated metabolic rate of the level time-to-exhaustion test (TTE_{Level}) was similar between Stryd-derived Power vs. Speed from the TTE_{Level} using the RE3 equation and a 24% metabolic efficiency (Fig. 5b)

As detailed in the method section, we used CS for the TTE_{Level} and matching power output for the $TTE_{Incline}$ instead of CP because of differences between overground (track) and treadmill running, which affect the runner's speed-power ratio. During the TTE_{Level}, we observed a minor power drift of 2.19 ± 1.82 % from the start of the TTE (average over 0:30-1:00 min) to the end of the TTE (average over 0:40 to 0:10 min before the first NaN reading). For the $TTE_{Incline}$, the drift was not significant (0.17 ± 0.90 %). We chose these time windows to capture the power output while running at CS without the initial acceleration and deceleration phase of the TTE. Even though this power drift is small compared to accuracy numbers reported for a range of cycling power meters (1-5%) (Maier et al. 2017), we wanted to see if it was related to changes in cadence during the TTEs. We detected cadence drift of -2.99 ± 1.64% for the TTE_{Level} and -1.96 + 2.29% for the $TTE_{Incline}$ ($p < 0.05$), yet the power drift was not correlated to the cadence drift using linear regression analysis ($R^2 = 0.02$). Participants averaged 182.0 ± 10.1 steps per minute for the TTE_{Level} and 173.1 ± 9.5 steps per minute for the $TTE_{Incline}$.

We assessed muscle oxygen demand during the TTEs and observed significantly higher oxygen depletion for the TTE_{Level} over time using the delayed slope concept (Kirby et al. 2021). The oxygen level decreased by 2.48 ± 1.32 $\%\cdot min^{-1}$ for the TTE_{Level} and by 1.62 ± 2.11 $\%\cdot min^{-1}$ for the $TTE_{Incline}$. Kirby et al. (2021) reported a delayed slope of -2.65 ± 0.77 $\%\cdot min^{-1}$ during exercise within the severe intensity domain and a consistently positive slope for exercise in the heavy intensity domain of 0.53 ± 0.14 $\%\cdot min^{-1}$. These findings suggest that our participants likely performed within the severe intensity domain, especially for the TTE_{Level}. This might explain the relatively short $Time_{TTE}$ we observed for both trials running at critical intensity, the intensity which theoretically can be sustained indefinitely, but in reality for approximately 16 to 30 minutes (Pepper et al. 1992; Bull et al. 2008; Figueiredo et al. 2021). The non-competitive nature of the lab setting and the unawareness of participants of elapsed time and running speed might have contributed to the relatively short $Time_{TTE}$ we observed, although our $Time_{TTE}$ was still shorter than reported by Pepper et al. (1992) in similar lab and awareness conditions (986 ± 365 seconds) who determined CS from a series of 5 TTEs on a level treadmill instead of using a 3MT.

Numerous papers have compared the 3MT-derived CS or CP estimations to the current gold standard of time trial efforts in running (e.g., Broxterman et al. 2013; Gama et al. 2018; Hunter et al. 2023) and have demonstrated similar CS and CP estimations on a group level. However, critical intensity estimations can substantially differ on an individual level, as shown by Broxterman et al. (2013). On a group level, the difference between methods averaged out at only 2.32% for Broxterman's cohort, but importantly, the average absolute intra-individual difference was 6.58%, ranging from 0.52% up to 19.20% among participants. Such an over- and underestimation of the 3MT-derived critical intensity will substantially impact a runner's time-to-exhaustion

because of the hyperbolic relationship between running speed to time-to-exhaustion (Daniels et al. 1978; Poole et al. 2016). For example, the current men's world records in 5000 m (12:35.36 min) and 10.000m (26:11.00 min) set by Joshua Cheptegei in 2020 differ only 3.8% in running speed, but this small speed difference enables an effort more than twice as long in duration. For our participants, the estimated metabolic rate was ~9% lower during the $TTE_{Incline}$, and they managed to run 42% longer on average, in line with this hyperbolic relationship.

Importantly the TTE durations varied substantially between individuals (167 s to 730 s), which could have been related to over- or underestimations of their 3MT-derived critical intensity. Therefore, we evaluated the execution of 3MT using existing criteria (Muniz-Pumares et al. 2019). The 3MT assumes that towards the end of the test, a runner can only sustain their CS or CP as their finite capacity above CS or CP - called D-Prime (D') or W-Prime (W') (proxies for anaerobic working capacity) - has been fully depleted (Pettitt 2016) while the mean speed or power of the last 30 seconds often serves as CS or CP approximation (Pettitt 2016). Muniz-Pumares et al. (2019) developed criteria for the 3MT in cycling to assess whether a participant has properly executed their test. Hunter et al. (2023) used these criteria to establish running CS and CP values based on Stryd's foot pod. These criteria include: 1) a plateau in power output in the last 30 seconds of the test; 2) the attainment of peak power output within the first 10 seconds of the test; 3) rapid initial decrease of power output to deplenish >90% of the work above end power, a proxy for anaerobic working capacity; 4) no decrease in power output of >5% of end power for >5s during the test; 5) an end test cycling cadence within 10 rpm of preferred cadence; 6) the attainment of $\dot{V}O_2max$, and 7) a blood lactate concentration >8mmol·L^{-1}. We have assessed 4 out of the 7 criteria (criteria 1 - 4) for the 3MT using the running power of the Stryd foot pod and observed 8.71 ± 4.34 s (Peak Power: 546 ± 122 W; Peak Speed: 6.95 ± 0.58 m·s^{-1}) for criterion 2, 92.37 ± 4.54 % for criterion 3, and 3.50 ± 5.26 s for criterion 4. Additionally, as there is currently no standardized criterion for evaluating the plateau during end power (criterion 1, Muniz-Pumares et al. 2019), we developed a standard to assess steady state during end power that assesses if any of the data points exceed 2 standard deviations above or below the mean. For our 3MT, 4 out of 14 participants met all criteria, while criterion 1) was met by 11 out of 14 participants, criterion 2) by 11 out of 14 participants, criteria 3) by 10 out of 14 participants, and criteria 4) by 11 out of 14 participants. Herein, 12 out of 14 participants missed only one criterion, 1 out of 14 missed two criteria, and 1 out of 14 missed three criteria. Interestingly, the participant who missed three criteria ran the second longest TTE_{Level} of all the participants with 727 seconds. We cannot analyze criterion 5 as it is specific to cycling and cannot be applied directly to running, and criteria 6 and 7 require additional equipment. Compared to Hunter et al. (2023) who assessed two 3MTs, our participant cohort reached a higher peak speed (6.95 ± 0.58 m·s^{-1} vs. 5.92 ± 1.09 and 5.69 ± 1.04), and a higher peak power (546 ± 122 W vs. 451 ± 101 W and 454 ± 106 W), which they also reached after a shorter time (8.71 ± 4.34 s vs. 12.6 ±

3.6 s and 17.0 ± 10.8 s) (criterion 2). Moreover, a higher percentage of our participants met criterion 3 (71.4 % vs. 21.7 % and 30.4 %) and depleted their W' by more than >90 % within the first 90 seconds. However, 3 out of our 14 participants missed criterion 4 and dipped below 5 % of their end power for more than 5 seconds compared to none of the participants in Hunter's group. Still, these participants did not run shorter than the other participants in the TTE_{Level}. These findings suggest that our participants may have depleted their anaerobic working capacity at a faster rate than Hunter's group, as demonstrated by criterion 3, but dipped below their end power and might have, therefore, recuperated some of their finite working capacity that would elevate their end power. We conclude that standardized criteria for the 3MT, as developed by Muniz-Pumares et al. (2019) for cycling, would enhance the comparability of the 3MT for running across studies. However, we need to acknowledge the differences between stationary bike cycling and overground running, particularly in the acceleration phase.

Limitations

We used Stryd's foot pod on the track using the Stryd app's Outdoor Mode and the treadmill using the Stryd app's Indoor Mode. Our results are mainly based on using Stryd's Indoor Mode on the treadmill. We must acknowledge the differences between Stryd's Indoor Mode and Stryd's Outdoor Mode in the way they consider elevation changes to estimate running power. Therefore, conclusions on the usability of Stryd's food pod for running at critical intensity can only be made for treadmill uphill running. Further, we performed the 3MT in track lane 1, and as such, participants were aware of how much distance they had covered, which could have affected the 3MT results. We decided to use lane 1 to enhance the Coros GPS watch accuracy since, during pilot testing, we identified substantial differences in the participant's running distance with the Coros Watch captured distance when using lane 5 or 6 (input in the Coros' track mode setting). In hindsight, this anomaly seems to be independent of the running lane. Lastly, we did not measure metabolic rate but estimated it using literature-based equations (HTK; Hoogkamer et al. 2014 and RE3; Looney et al. 2023), and those are likely not to be entirely accurate at the individual level.

Conclusion

Our data suggests that athletes should be careful when using running power at critical intensity to guide their racing and training in non-level conditions. Further, our findings suggest that running power significantly differs between foot- vs. wrist-based wearables.

Declarations

Disclosures

The authors have no relevant financial or non-financial interests to disclose.

Funding:

The authors did not receive support from any organization for the submitted work.

Ethical Approval

Approval was obtained from the Institutional Review Board of the University of Massachusetts Amherst (Protocol Nr.: 5639). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Informed Consent

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

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Apte S, Falbriard M, Meyer F, Millet GP, Gremeaux V, Aminian K (2023) Estimation of horizontal running power using foot-worn inertial measurement units. Front. Bioeng. Biotechnol. 11:1167816. doi: 10.3389/fbioe.2023.1167816

Aubry RL, Power GA, Burr JF (2018) An Assessment of Running Power as a Training Metric for Elite and Recreational Runners. J. Strength Cond. Res. 32(8):2258–2264. doi: 10.1519/JSC.0000000000002650

Austin CL, Hokanson JF, McGinnis PM, Patrick S (2018) The Relationship between Running Power and Running Economy in Well-Trained Distance Runners. Sports (Basel, Switzerland) 6(4). doi: 10.3390/sports6040142

Borg G, Dahlstrom H (1962) A pilot study of perceived exertion and physical working capacity. Acta Soc. Med. Upsal. 67:21–27

Brill JW, Kram R (2021) Does the preferred walk-run transition speed on steep inclines minimize energetic cost, heart rate or neither? J. Exp. Biol. 224(Pt 3). doi: 10.1242/jeb.233056

Broxterman RM, Ade CJ, Poole DC, Harms CA, Barstow TJ (2013) A single test for the determination of parameters of the speed-time relationship for running. Respir. Physiol. Neurobiol. 185(2):380–385. doi: 10.1016/j.resp.2012.08.024

Bull AJ, Housh TJ, Johnson GO, Rana SR (2008) Physiological responses at five estimates of critical velocity. Eur. J. Appl. Physiol. 102(6):711–720. doi: 10.1007/s00421-007-0649-7

Cavagna GA, Kaneko M (1977) Mechanical work and efficiency in level walking and running. J. Physiol. 268(2):467--81. doi: 10.1113/jphysiol.1977.sp011866

Cavagna GA, Saibene FP, Margaria R (1964) Mechanical work in running. J. Appl. Physiol. 19:249–256. doi: 10.1152/jappl.1964.19.2.249

Cerezuela-Espejo V, Hernández-Belmonte A, Courel-Ibáñez J, Conesa-Ros E, Martínez-Cava A, Pallarés JG (2020) Running power meters and theoretical models based on laws of physics: effects of environments and running conditions. Physiol. Behav. 223:112972. doi: 10.1016/j.physbeh.2020.11297

Chisholm DM, Collis ML, Kulak LL, Davenport W, Gruber N (1975). Physical activity readiness. B. C. Med. J. 17, 375-378.

Clark IE, West BM, Reynolds SK, Murray SR, Pettitt RW (2013) Applying the critical velocity model for an off-season interval training program. J. Strength Cond. Res. 27(12):3335–3341. doi: 10.1519/JSC.0b013e31828f9d87

Daniels J, Fitts RK, Sheehan G (1978) Conditioning for distance running, 2. pr. American college of sports medicine series. Wiley, New York

Dearing CG, Paton CD (2023) Is Stryd critical power a meaningful parameter for runners? Biol. Sport. 40(3):657–664. doi: 10.5114/biolsport.2023.118025

Feldmann A, Schmitz R, Erlacher D (2019) Near-infrared spectroscopy-derived muscle oxygen saturation on a 0% to 100% scale: reliability and validity of the Moxy Monitor. J. Biomed. Opt. 24(11):1–11. doi: 10.1117/1.JBO.24.11.115001

Fenn WO (1930) Frictional and kinetic factors in the work of sprint running. Am. J. Physiol.-Legac. Content 92(3):583–611. doi: 10.1152/ajplegacy.1930.92.3.583

Figueiredo DH, Figueiredo DH, Manoel FA, Machado FA (2023) Peak Running Velocity vs. Critical Speed: Which One Is Better to Prescribe Endurance Training to Recreational Runners? J. Strength Cond. Res. 37(9):1783–1788. doi: 10.1519/JSC.0000000000004452

Figueiredo DH, Figueiredo DH, Manoel FdA, Machado FA (2021) Peak Running Velocity or Critical Speed Under Field Conditions: Which Best Predicts 5-km Running Performance in Recreational Runners? Front. Physiol. 12:680790. doi: 10.3389/fphys.2021.680790

Gaesser GA, Brooks GA (1975) Muscular efficiency during steady-rate exercise: effects of speed and work rate. J. Appl. Physiol. 38(6):1132–1139. doi: 10.1152/jappl.1975.38.6.1132

Gama MCT, Dos Reis, Ivan Gustavo Masselli, Sousa, Filipe Antônio de Barros, Gobatto CA (2018) The 3-min all-out test is valid for determining critical power but not anaerobic work capacity in tethered running. PloS one 13(2):e0192552. doi: 10.1371/journal.pone.0192552

Heglund NC, Cavagna GA, Taylor CR (1982) Energetics and mechanics of terrestrial locomotion. III. Energy changes of the centre of mass as a function of speed and body size in birds and mammals. J. Exp. Biol. 97:41–56. doi: 10.1242/jeb.97.1.41

Hoogkamer W, Taboga P, Kram R (2014) Applying the cost of generating force hypothesis to uphill running. PeerJ 2:e482. doi: 10.7717/peerj.482

Hunter B, Ledger A, Muniz-Pumares D (2023) Remote Determination of Critical Speed and Critical Power in Recreational Runners. Int. J. Sports Physiol. Perform. 18(12):1449–1456. doi: 10.1123/ijspp.2023-0276

Imbach F, Candau R, Chailan R, Perrey S (2020) Validity of the Stryd Power Meter in measuring running parameters at submaximal speeds. Sports 8(7):103. doi: 10.3390/sports8070103

Kirby BS, Clark DA, Bradley EM, Wilkins BW (2021) The balance of muscle oxygen supply and demand reveals critical metabolic rate and predicts time to exhaustion. J. Appl. Physiol. (Bethesda, Md. : 1985) 130(6):1915–1927. doi: 10.1152/japplphysiol.00058.2021

Kram R, Taylor CR (1990) Energetics of running: a new perspective. Nature 346(6281):265–267. doi: 10.1038/346265a0

Lai A, Schache AG, Lin Y-C, Pandy MG (2014) Tendon elastic strain energy in the human ankle plantar-flexors and its role with increased running speed. J. Exp. Biol. 217(Pt 17):3159–3168. doi: 10.1242/jeb.100826

Looney DP, Hoogkamer W, Kram R, Arellano CJ, Spiering BA (2023) Estimating Metabolic Energy Expenditure During Level Running in Healthy, Military-Age Women and Men. J. Strength Cond. Res 37(12):2496–2503. doi: 10.1519/JSC.0000000000004626

Maier T, Schmid L, Müller B, Steiner T, Wehrlin JP (2017) Accuracy of Cycling Power Meters against a Mathematical Model of Treadmill Cycling. Int. J. Sports Med. 38(6):456–461. doi: 10.1055/s-0043-102945

Margaria R (1968) Positive and negative work performances and their efficiencies in human locomotion. Int. Z. Angew. Physiol. Einschl. Arbeitsphysiol. 25(4):339–351. doi: 10.1007/BF00699624

Margaria R, Cerretelli P, Aghemo P, Sassi G (1963) Energy cost of running. J. Appl. Physiol. 18:367–370. doi: 10.1152/jappl.1963.18.2.367

Matthews IR, Heenan LJ, Fisher KG, Flood EF, Wehrman LW, Kirby BS, Wilkins BW (2023) Identification of maximal steady-state metabolic rate by the change in muscle oxygen saturation. J. Appl. Physiol. (Bethesda, Md. : 1985) 134(6):1349–1358. doi: 10.1152/japplphysiol.00706.2022

Montalvo-Pérez A, Alejo LB, Valenzuela PL, Castellanos M, Gil-Cabrera J, Talavera E, Lucia A, Barranco-Gil D (2021) Validity of the Favero Assioma Duo Power Pedal System for Measuring Power Output and Cadence. Sensors (Basel, Switzerland) 21(7). doi: 10.3390/s21072277

Monte A, Maganaris C, Baltzopoulos V, Zamparo P (2020) The influence of Achilles tendon mechanical behaviour on "apparent" efficiency during running at different speeds. Eur. J. Appl. Physiol. 120(11):2495–2505. doi: 10.1007/s00421-020-04472-9

Muniz-Pumares D, Karsten B, Triska C, Glaister M (2019) Methodological Approaches and Related Challenges Associated With the Determination of Critical Power and Curvature Constant. J. Strength Cond. Res. 33(2):584–596. doi: 10.1519/JSC.0000000000002977

Passfield L, Hopker JG, Jobson S, Friel D, Zabala M (2017) Knowledge is power: Issues of measuring training and performance in cycling. J. Sports Sci. 35(14):1426–1434. doi: 10.1080/02640414.2016.1215504

Pepper ML, Housh TJ, Johnson GO (1992) The accuracy of the critical velocity test for predicting time to exhaustion during treadmill running. Int. J. Sports Med. 13(2):121–124. doi: 10.1055/s-2007-1021242

Pettitt RW (2016) Applying the Critical Speed Concept to Racing Strategy and Interval Training Prescription. Int. J. Sports Physiol. Perform. 11(7):842–847. doi: 10.1123/ijspp.2016-0001

Pinedo-Jauregi A, Garcia-Tabar I, Carrier B, Navalta JW, Cámara J (2022) Reliability and validity of the Stryd Power Meter during different walking conditions. Gait Posture 92:277–283. doi: 10.1016/j.gaitpost.2021.11.041

Poole DC, Burnley M, Vanhatalo A, Rossiter HB, Jones AM (2016) Critical Power: An Important Fatigue Threshold in Exercise Physiology. Med. Sci. Sports Exerc. 48(11):2320–2334. doi: 10.1249/MSS.0000000000000939

Ruiz-Alias SA, Olaya-Cuartero J, Ñancupil-Andrade AA, García-Pinillos F (2022) 9/3-Minute Running Critical Power Test: Mechanical Threshold Location With Respect to Ventilatory Thresholds and Maximum Oxygen Uptake. Int. J. Sports Physiol. Perform. 17(7):1111–1118. doi: 10.1123/ijspp.2022-0069

Scherr J, Wolfarth B, Christle JW, Pressler A, Wagenpfeil S, Halle M (2013) Associations between Borg's rating of perceived exertion and physiological measures of exercise intensity. Eur. J. Appl. Physiol. 113(1):147–155. doi: 10.1007/s00421-012-2421-x

Snyder KL, Mohrman WP, Williamson JA, Li K (2018) Methodological Flaws in Aubry, RL, Power, GA, and Burr, JF. An Assessment of Running Power as a Training Metric for Elite and Recreational Runners. J. Strength Cond. Res 32: 2258-2264. doi: 10.1519/JSC.0000000000002894

SRM GmbH (2024, September 20). History of Power Meters | SRM. https://www.srm.de/the-history-of-power-meters/

Stadheim HK, Stensrud T, Brage S, Jensen J (2021) Caffeine Increases Exercise Performance, Maximal Oxygen Uptake, and Oxygen Deficit in Elite Male Endurance Athletes. Med. Sci. Sports Exerc. 53(11):2264–2273. doi: 10.1249/MSS.0000000000002704

Stryd Team (2017, December 7) How to Lead the Pack: Running Power Meters & Quality Data | White paper. Stryd Blog. https://blog.stryd.com/2017/12/07/how-to-lead-the-pack-running-power-meters-quality-data/

Stryd Team (2020, January 8) Adding air resistance costs to Power Estimates in Running | White paper. Stryd Blog. https://blog.stryd.com/2019/08/17/adding-air-resistance-costs-to-power-estimates-in-running-white-paper/

Swinnen W, Hoogkamer W, Delabastita T, Aeles J, Groote F de, Vanwanseele B (2019) Effect of habitual foot-strike pattern on the gastrocnemius medialis muscle-tendon interaction and muscle force production during running. J. Appl. Physiol. (Bethesda, Md. : 1985) 126(3):708–716. doi: 10.1152/japplphysiol.00768.2018

Taboga P, Giovanelli N, Spinazzè E, Cuzzolin F, Fedele G, Zanuso S, Lazzer S (2022) Running power: lab based vs. portable devices measurements and its relationship with aerobic power. Eur. J. Sport Sci. 22(10):1555–1568. doi: 10.1080/17461391.2021.1966104

Turley KR (2016) Effects of Caffeine on Exercise Responses and Performance in Children and Youth. Am. J. Lifestyle Med. 10(6):417–421. doi: 10.1177/1559827614554593

van Rassel CR, Ajayi OO, Sales KM, Griffiths JK, Fletcher JR, Edwards WB, MacInnis MJ (2023) Is Running Power a Useful Metric? Quantifying Training Intensity and Aerobic Fitness Using Stryd Running Power Near the Maximal Lactate Steady State. Sensors (Basel, Switzerland) 23(21). doi: 10.3390/s23218729

van Rassel CR, Sales KM, Ajayi OO, Nagai K, MacInnis MJ (2024) A Comparison of Critical Speed and Critical Power in Runners Using Stryd Running Power. Int. J. Sports Physiol. Perform. 19(1):84–87. doi: 10.1123/ijspp.2023-0260