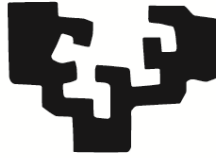


eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

International Doctorate

**EFFECTS OF DIFFERENT HIGH-INTENSITY  
INTERVAL TRAINING PRESCRIPTION MODELS ON  
PHYSIOLOGICAL RESPONSE AND PERFORMANCE  
IN TRAINED DISTANCE RUNNERS**



**Raffaele Mazzolari**

Supervisor:

**Prof. Jordan Santos-Concejero**

Vitoria-Gasteiz, 2022



eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

Department of Physical Education and Sport,  
University of the Basque Country UPV/EHU

International Doctorate

**EFFECTS OF DIFFERENT HIGH-INTENSITY INTERVAL  
TRAINING PRESCRIPTION MODELS ON PHYSIOLOGICAL  
RESPONSE AND PERFORMANCE IN TRAINED DISTANCE  
RUNNERS**

Presented by

**Raffaele Mazzolari**

Vitoria-Gasteiz, 2022



eman ta zabal zazu



Universidad del País Vasco      Euskal Herriko Unibertsitatea

Department of Physical Education and Sport,  
University of the Basque Country UPV/EHU

International Doctorate

**EFFECTS OF DIFFERENT HIGH-INTENSITY INTERVAL  
TRAINING PRESCRIPTION MODELS ON PHYSIOLOGICAL  
RESPONSE AND PERFORMANCE IN TRAINED DISTANCE  
RUNNERS**

Presented by

**Raffaele Mazzolari**

Supervised by

**Prof. Jordan Santos-Concejero**

Universidad del País Vasco/Euskal Herriko Unibertsitatea, UPV/EHU

Vitoria-Gasteiz, 2022

## ACKNOWLEDGMENTS

Testing our limits as a means to understand and surpass them is part of human nature. Competition against opponents and especially against ourselves represents an inherent behavior of our species. Common questions such as ‘why did I fail?’ and ‘how can I improve?’ represents a perfect expression of these concepts. Our mission, as men and women of science, is to learn the nature of things through observation, and their functions through experiments. Our mission, as exercise physiologists and sports scientists, is to build new knowledge and understanding within our little area of interest and apply these concepts and insights to everyday life, to help athletes and coaches, as well as sedentary and patient populations, to reach their goals in terms of performance and health. In my thesis, I contributed in this direction by answering several open questions about high-intensity interval training prescriptions in trained distance runners. Whether or not my work will be successful in improving current coaching practice through the new knowledge it provided will depend on many factors, which I cannot control for the most part. That being said, I consider the product of my work successful in achieving at least its scientific aims. If my work helps even just a few athletes to achieve their goals, I will consider it a full accomplishment.

My deepest gratitude goes to Prof. Simone Porcelli, director of the Human Integrative Physiology of Exercise (HIPE) Laboratory of the University of Pavia (Italy), where I carried out the largest part of my investigations. He provided me with the funds, the lab, and the assistance to make this research project possible. Without him, the present thesis would not exist. Thank you, Simone.

I would also like to express my appreciation to my supervisor Prof. Jordan Santos-Concejero for his guidance and support towards my project and, in particular, for his help during the draft of the research proposal, the revision, and the submission of my thesis.

I would like to thank the University of the Basque Country for having given me the possibility of accomplishing a Doctoral Degree.

Furthermore, I would like to thank all the people who participated in this research project for their time and effort.

Finally, I thank my mother for having believed in my choices and having economically and morally supported me when I needed it. I take this opportunity to thank my mother for everything she has done for me in my life. Without her, the present me would not exist.

Thank you, Liliana. For everything.



## SCIENTIFIC CONTRIBUTIONS

Part of the content of the present thesis has been considered for publication in peer-reviewed journals and/or presented in international congresses. Specifically:

- The final part of the introduction that describes the design and analysis of equivalence and non-inferiority studies in exercise physiology and sport science has been published on *Experimental Physiology* (IF: 2.969; Q3 on Physiology) (Mazzolari et al., 2022).
- I presented the results of the first family of hypotheses that I investigated in this thesis (i.e., whether individualizing high-intensity interval training prescriptions according to the physiological characteristics of the runners and their responses to exercise would have induced larger cardiorespiratory adaptations than using a standardized program) at the 26<sup>th</sup> Annual Congress of the European College of Sports Science, in which I was a candidate for the Young Investigator Award.

During the time of my doctoral program and beyond the topic of my thesis:

- I wrote a letter to the editor of *The Journal of Physiology* (IF: 4.950; Q1 on Physiology) about a study addressing the effects of endurance training dose in cardiorespiratory fitness responsiveness in healthy young males (Mazzolari, 2018).
- I was co-author in a study that was published in *Nitric Oxide: Biology and Chemistry* (IF: 4.427; Q1 on Physiology) on the time-course changes of plasma nitrate and nitrite concentrations during acclimatization to high-altitude (Rasica et al., 2021).
- I was co-signatory for a comment published in *Nature* (IF: 42.778; Q1 in Multidisciplinary) on the misuse of statistical significance (Amrhein et al., 2019).
- I was co-author in one study that was presented at the 25<sup>th</sup> Annual Congress of the European College of Sports Science about the non-invasive evaluation of the time-course changes in physiological adaptations to low-volume high-intensity interval training
- I was co-author in four studies that were presented at the 26<sup>th</sup> Annual Congress of the European College of Sports Science about:
  - 1) The effects of 8 weeks of strength training on exercise economy and race performance in triathletes



- 2) Sex differences in fatigability during an Olympic triathlon
- 3) Near infra-red spectroscopy estimation of combined skeletal muscle oxidative capacity and O<sub>2</sub> diffusion capacity in humans
- 4) Time-course recovery of central and peripheral physiological adaptations by high-intensity interval training after COVID-19

## ABSTRACT

In my thesis, I contributed to the field of exercise physiology and sport science by focusing on several open questions about the effects of different high-intensity interval training (HIIT) prescription models on physiological response and performance in trained distance runners. This aim was achieved through investigating different families of hypotheses, which focused on separated but related aspects of the topic.

In the first investigation, I tested whether a highly-individualized, physiologically-based approach to HIIT prescription was more efficacious (i.e., superior) than a standardized one that imposes the same HIIT scheme to all runners for what concerns the development of cardiorespiratory fitness, in terms of maximal oxygen consumption ( $\dot{V}O_{2max}$ ) and peak treadmill speed ( $V_{peak}$ ). Coaches often include HIIT sessions in their plan to develop specific physiological adaptations and enhance athletic performance in distance runners. The prescription is usually performed using a standardized and common empirical approach, in which runners of a similar level train together using the same HIIT scheme. However, previous studies showed that it is possible to target some physiological adaptations by individualizing HIIT prescriptions according to the physiological characteristics of the runners and their responses to exercise. This may be given that of the longest time spent at or near  $\dot{V}O_{2max}$  – and thus the stronger cardiorespiratory stimulus – that occurs with this training prescription model. Since in the standardized approach, in which the speed is the same for everyone, runners who fall outside the desired intensity range may face blunted or null gains in cardiorespiratory fitness, a smaller mean training effect coupled with a larger heterogeneity of training effect may be observed. I tested 46 distance runners [35 men and 11 women, age: 36 (8) yr, 10000-m speed: 14.4 (1.3) km · h<sup>-1</sup>] before and after 8 weeks of training, in which they replaced a similar part of their habitual training volume with two sessions per week of either individualized (IND, N = 23) or standardized (STD, N = 23) HIIT, without major modifications in the overall training structure. I assessed  $\dot{V}O_{2max}$ , running economy (RE),  $V_{peak}$ , and lactate thresholds during an incremental test performed on a treadmill, followed by a verification phase, and I measured 10000-m speed performance on a 400-m track. HIIT prescriptions were individualized for the speed and duration of each interval according to each runner's physiological values as previously described by Billat and colleagues in IND. Specifically, the speed and the duration of each interval were set at the speed halfway between the second lactate threshold and  $V_{peak}$  ( $v_{\Delta 50}$ ) and 50% of the time to exhaustion at  $v_{\Delta 50}$  ( $T_{lim}$ ), respectively, whereas the speed and the duration of each recovery period were set at 50% of the  $V_{peak}$  and 25% of  $T_{lim}$ , respectively.

The average group values were instead used for prescribing HIIT in STD. IND and STD were matched for HIIT characteristics at the group level. Relative  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  improved significantly only in IND [ $\dot{V}O_{2\max}$ :  $+2.8 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ,  $p < .001$ , 95% confidence interval (CI):  $+1.4$  to  $+4.1$ ;  $V_{\text{peak}}$ :  $+0.4 \text{ km} \cdot \text{h}^{-1}$ ,  $p < .001$ , 95% CI:  $+0.2$  to  $+0.5$ ], with a statistically significant difference of  $+3.7 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  ( $p = .01$ , 95% CI:  $+1.3$  to  $+\infty$ ),  $+0.20 \text{ L} \cdot \text{min}^{-1}$  ( $p = .02$ , 95% CI:  $+0.04$  to  $+\infty$ ), and  $+0.4 \text{ km} \cdot \text{h}^{-1}$  ( $p = .02$ , 95% CI:  $+0.1$  to  $+\infty$ ) for relative  $\dot{V}O_{2\max}$ , absolute  $\dot{V}O_{2\max}$ , and  $V_{\text{peak}}$ , respectively, between the groups. IND reduced heterogeneity of intervention effect for  $V_{\text{peak}}$  compared with STD ( $-0.4 \text{ km} \cdot \text{h}^{-1}$ ,  $p = .04$ , 95% CI:  $-\infty$  to  $-0.1$ ). RE improved significantly only in IND ( $-4.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ,  $p = .05$ , 95% CI:  $-9.1$  to  $0.0$ ), with no statistically significant difference between the groups. Fractional  $\dot{V}O_{2\max}$  at the second lactate threshold decreased significantly only in IND ( $-2.2\%$ ,  $p = .007$ , 95% CI:  $-3.8$  to  $-0.7$ ), with a statistically significant difference of  $-5\%$  ( $p = .008$ , 95% CI:  $-8.7$  to  $-1.4$ ) between the groups. No significant changes or differences were observed in the other physiological parameters investigated and 10000-m speed performance within and between the groups. These results showed that individualizing HIIT prescriptions according to the physiological responses to exercise induces superior cardiorespiratory fitness adaptations compared with standardized prescriptions and reduces the heterogeneity of intervention effects in trained distance runners. The experimental data corroborated my hypotheses, which led me to conclude that individualizing HIIT prescription according to physiological characteristics and response to exercise should be the choice for coaches and athletes in those training phases aimed to improve cardiorespiratory fitness (usually during the preparatory phase and the early part of the specific period).

Since most of the runners do not have easy access to physiology labs, medical clinics, and expert personnel to obtain valid and reliable data for individualizing HIIT prescription according to the physiological approach, I looked for alternative parameters to estimate the  $v\Delta 50$  that did not require lab tests. Informed by the previous literature on the topic, I evaluated the suitability of various treadmill-related and race pace-related measures for this purpose. Specifically, I investigated whether the relative percent of the  $V_{\text{peak}}$  at  $v\Delta 50$ , the difference between  $v\Delta 50$  and  $V_{\text{peak}}$ , the relative percent of the 10000-m speed at  $v\Delta 50$ , and the difference between  $v\Delta 50$  and 10000-m speed had sufficiently-low inter-individual variability – expressed as standard deviation (SD) – to be used as valid surrogates of  $v\Delta 50$ . The physiological and performance testing protocol was identical to the one adopted in my first investigation with no experimental intervention between the 8 weeks that separated the two testing sessions. By using a much larger sample size than what had previously been used in the literature ( $N = 75$ ) and a

repeated-measurement design (on 61 out of 75 runners for treadmill-related measures and 57 out of 75 for race pace-related measures), I was able to obtain sufficient precision in parameter estimation and to isolate the true inter-individual variability by the biological (within-individual) variability. The observed  $v\Delta 50$  corresponded to the 110% (95% CI: 109 to 111) of the 10000-m speed with a SD of 5% (95% CI: 4 to 6) and it was  $1.4 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 1.3 to 1.6) faster than the 10000-m speed with a SD of  $0.7 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.6 to 0.8). Instead, the observed  $v\Delta 50$  corresponded to the 92% (95% CI: 91 to 92) of the  $V_{\text{peak}}$  with a SD of 2% (95% CI: 0 to 2) and it was  $1.4 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 1.5 to 1.3) lower than the  $V_{\text{peak}}$  with a SD of  $0.5 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.4 to 0.8). After having accounted for the within-individual variability, the inter-individual variability of the 10000-m speed at  $v\Delta 50$  was 4% (95% CI: 3 to 6) when expressed as a percentage of the 10000-m speed and to  $0.5 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.4 to 0.8) when expressed as the difference between  $v\Delta 50$  and 10000-m speed. Instead, the inter-individual variability of the  $V_{\text{peak}}$  at  $v\Delta 50$  was 1% (95% CI: 1 to 2) when expressed as a percentage of the  $V_{\text{peak}}$  and to  $0.2 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.2 to 0.3) when expressed as the difference between  $v\Delta 50$  and  $V_{\text{peak}}$ . The data indicated that treadmill-related measures can inform about the  $v\Delta 50$  value with sufficiently-low heterogeneity between individuals to be used as valid alternatives to lab measures. On the contrary, the larger variability in the race pace-related measures at  $v\Delta 50$  does not make them suitable for prescribing HIIT according to the physiological approach. Treadmill-related measures only require a commercial (and properly calibrated) treadmill available in any gym and allow to estimate the  $v\Delta 50$  with discrete precision and reasonable effort. The use of treadmill-related measures to individualize HIIT prescriptions according to the physiological approach would make this form of HIIT more accessible to the largest part of the runner population.

Developing cardiorespiratory fitness only represents an intermediate step in training preparation, whereas maximizing performance represents the main goal when approaching competitions. As my last family of hypotheses, I investigated the effects of two different individualized HIIT prescriptions models – the physiological approach and the race pace approach – on cardiorespiratory fitness and 10000-m speed performance in trained distance runners. Following the same rationale adopted in my first investigation, I hypothesized larger gains in cardiorespiratory fitness for the physiological approach. In agreement with the modern network physiology – according to which anchoring the training intensity to a given higher percentage of the target race pace would maximize race performance through a more efficient network of physiological interactions for that task at the condition the intensity is not too high to lose the specificity for the task –, I also hypothesized larger gains in performance for the race

pace approach. Moreover, I hypothesized a larger heterogeneity of intervention effects for cardiorespiratory fitness for the race pace approach and a larger heterogeneity for 10000-m speed performance in the physiological approach. I tested 31 distance runners [20 men and 11 women, age: 38 (9) yr, 10000-m speed: 14.0 (1.4) km · h<sup>-1</sup>] before and after 8 weeks of training, in which they replaced a similar part of their habitual training volume with two sessions per week of either physiologically-based (PHY, N = 16) or race pace-based (RP, N = 15) individualized HIIT prescriptions, without major modifications in the overall training structure. The physiological and performance testing protocol was identical to the one adopted in my first investigation so was the training intervention in PHY. Race pace-based HIIT prescriptions were instead intended to mimic coaching-like practice, with the speed of each interval determined according to a given percentage of the target race pace, a fixed duration for each interval, and a given distance to cover during the recovery time. Specifically, runners were required to cover 1000 m at the 110% of the target 10000-m pace during each interval, and 300 m in 2 min during recovery, since these values are similar to the physiological-based HIIT prescriptions values at a group level in a large cohort of similar runners.  $\dot{V}O_{2max}$  and  $V_{peak}$  improved significantly in PHY ( $\dot{V}O_{2max}$ : +3.1 mL · kg<sup>-1</sup> · min<sup>-1</sup>,  $p < .001$ , 95% CI: +1.8 to +4.4;  $\dot{V}O_{2max}$ : +0.13 L · min<sup>-1</sup>,  $p = .02$ , 95% CI: +0.03 to +0.23;  $V_{peak}$ : +0.4 km · h<sup>-1</sup>,  $p < .001$ , 95% CI: +0.2 to +0.7), whereas  $\dot{V}O_{2max}$  decreased significantly in RP ( $\dot{V}O_{2max}$ : -2.0 mL · kg<sup>-1</sup> · min<sup>-1</sup>,  $p = .009$ , 95% CI: -3.4 to -0.6;  $\dot{V}O_{2max}$ : -0.11 L · min<sup>-1</sup>,  $p = .05$ , 95% CI: -0.22 to 0.00), with a statistically significant difference of +5.1 mL · kg<sup>-1</sup> · min<sup>-1</sup> ( $p < .001$ , 95% CI: +3.6 to +∞), +0.24 L · min<sup>-1</sup> ( $p = .001$ , 95% CI: +0.12 to +∞), and +0.5 km · h<sup>-1</sup> ( $p = .003$ , 95% CI: +0.2 to +∞) for relative  $\dot{V}O_{2max}$ , absolute  $\dot{V}O_{2max}$ , and  $V_{peak}$ , respectively, between the groups. On the contrary, 10000-m speed improved significantly only in RP (+0.5 km · h<sup>-1</sup>,  $p < .001$ , 95% CI: +0.2 to +0.7), with a statistically significant difference of -0.5 km · h<sup>-1</sup> ( $p = .01$ , 95% CI: -∞ to -0.1) between the groups. Fractional  $\dot{V}O_{2max}$  at the first lactate threshold improved significantly only in RP (+4.1%,  $p = .02$ , 95% CI: +0.7 to +7.5), with a statistically significant difference of -4.2% ( $p = .05$ , 95% CI: -8.3 to 0.0) between the groups. Although the experimental data were inconclusive for heterogeneity of intervention effects, they corroborated both my hypotheses about the mean training effect, which led me to conclude that the physiological approach and race pace approach may be performed within the same macrocycle, but the implementation of the former should precede the latter in terms of proximity to major competitions.

**KEYWORDS:** endurance running; exercise physiology; high-intensity interval training; sport performance

## RESUMEN

Esta tesis contribuye al campo de la fisiología del ejercicio y la ciencia del deporte centrándose en varias preguntas abiertas sobre los efectos de los diferentes modelos de prescripción de entrenamiento en intervalos de alta intensidad (HIIT) con la respuesta fisiológica y el rendimiento en corredores de fondo entrenados. Este objetivo se logró a través la investigación de diferentes familias de hipótesis, que se centraron en aspectos separados, pero relacionados al tema.

Inicialmente se estudió si un enfoque altamente individualizado y fisiológicamente basado en la prescripción de HIIT era más eficaz (es decir, superior) que uno estandarizado que impone el mismo esquema de HIIT a todos los corredores en lo que concierne al desarrollo de la aptitud cardiorrespiratoria, en términos de consumo máximo de oxígeno ( $\dot{V}O_{2max}$ ) y velocidad máxima en cinta rodante ( $V_{peak}$ ). Los entrenadores a menudo incluyen sesiones HIIT en su plan para desarrollar adaptaciones fisiológicas específicas y mejorar el rendimiento deportivo en los corredores de fondo. La prescripción generalmente se realiza usando un enfoque empírico común y estandarizado, en el que los corredores de un nivel similar entrenan juntos usando el mismo esquema HIIT. Sin embargo, estudios previos mostraron que es posible enfocarse en algunas adaptaciones fisiológicas individualizando las prescripciones de HIIT acorde con las características fisiológicas de los corredores y sus respuestas al ejercicio. Esto puede estar relacionado al mayor tiempo pasado en (o cerca del)  $\dot{V}O_{2max}$  y, por lo tanto, a un estímulo cardiorrespiratorio más fuerte que se produce con este modelo de prescripción de entrenamiento. Dado que en el enfoque estandarizado, en el que la velocidad es la misma para todos, los corredores que caen fuera del rango de intensidad deseado pueden conseguir menores ganancias (o nulas) en la aptitud cardiorrespiratoria, se puede observar un efecto promedio de grupo más pequeño junto con una mayor heterogeneidad del efecto de entrenamiento. Se analizaron a 46 corredores de fondo [35 hombres y 11 mujeres, edad: 36 (8) años, velocidad de 10000-m: 14.4 (1.3) km · h<sup>-1</sup>] antes y después de 8 semanas de entrenamiento, en las que reemplazaron una parte similar de su volumen de entrenamiento habitual con dos sesiones por semana de HIIT individualizado (IND, N = 23) o estandarizado (STD, N = 23), sin modificaciones importantes en la estructura general del entrenamiento. Se evaluó el  $\dot{V}O_{2max}$ , la economía de carrera (RE), el  $V_{peak}$  y los umbrales de lactato durante una prueba incremental realizada en una cinta de correr, seguida de una fase de verificación, y se midió el rendimiento de velocidad de 10000-m en una pista de 400 m. Las prescripciones de HIIT se individualizaron por velocidad y la duración de cada intervalo de acuerdo con los valores fisiológicos de cada

corredor, como describieron anteriormente Billat y cols. en IND. Específicamente, la velocidad y la duración de cada intervalo se establecieron a la velocidad a medio camino entre el segundo umbral de lactato y  $V_{\text{peak}}$  ( $v\Delta 50$ ) y el 50% del tiempo hasta el agotamiento en  $v\Delta 50$  ( $T_{\text{lim}}$ ), respectivamente, mientras que la velocidad y la duración de cada período de recuperación se fijó en el 50% del  $V_{\text{peak}}$  y el 25% del  $T_{\text{lim}}$ , respectivamente. En cambio, los valores medios del grupo se utilizaron para prescribir HIIT en ETS. IND y STD se emparejaron para las características HIIT a nivel de grupo. El  $\dot{V}O_{2\text{max}}$  relativo y el  $V_{\text{peak}}$  mejoraron significativamente solo en IND [ $\dot{V}O_{2\text{max}}$ :  $+2.8 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ,  $p < .001$ , intervalo de confianza (CI) del 95%:  $+1.4$  a  $+4.1$ ;  $V_{\text{peak}}$ :  $+0.4 \text{ km} \cdot \text{h}^{-1}$ ,  $p < .001$ , CI 95%:  $+0.2$  a  $+0.5$ ], con una diferencia estadísticamente significativa de  $+3.7 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  ( $p = .01$ , 95 % CI:  $+1.3$  a  $+\infty$ ),  $+0.20 \text{ L} \cdot \text{min}^{-1}$  ( $p = .02$ , 95% CI:  $+0.04$  a  $+\infty$ ), y  $+0.4 \text{ km} \cdot \text{h}^{-1}$  ( $p = .02$ , 95% CI:  $+0.1$  a  $+\infty$ ) para  $\dot{V}O_{2\text{max}}$  relativo,  $\dot{V}O_{2\text{max}}$  absoluto y  $V_{\text{peak}}$ , respectivamente, entre los grupos. IND redujo la heterogeneidad del efecto de la intervención para  $V_{\text{peak}}$  en comparación con STD ( $-0.4 \text{ km} \cdot \text{h}^{-1}$ ,  $p = .04$ , 95% CI:  $-\infty$  a  $-0.1$ ). RE mejoró significativamente sólo en IND ( $-4.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ,  $p = .05$ , 95% CI:  $-9.1$  a  $0.0$ ), sin diferencias estadísticamente significativas entre los grupos. El  $\dot{V}O_{2\text{max}}$  fraccional en el segundo umbral de lactato disminuyó significativamente solo en IND ( $-2.2\%$ ,  $p = .007$ , 95% CI:  $-3.8$  a  $-0.7$ ), con una diferencia estadísticamente significativa de  $-5\%$  ( $p = .008$ , 95% CI:  $-8.7$  a  $-1.4$ ) entre los grupos. No se observaron cambios o diferencias significativas en los otros parámetros fisiológicos investigados y en el rendimiento de velocidad de 10000-m dentro y entre los grupos. Estos resultados mostraron que la individualización de las prescripciones de HIIT, de acuerdo con las respuestas fisiológicas al ejercicio, induce adaptaciones de aptitud cardiorrespiratoria superiores en comparación con las prescripciones estandarizadas y reduce la heterogeneidad de los efectos de la intervención en corredores de fondo entrenados. Los datos experimentales corroboraron ambas hipótesis, lo que lleva a concluir que individualizar la prescripción de HIIT de acuerdo con las características fisiológicas y la respuesta al ejercicio debería ser la elección de entrenadores y deportistas en aquellas fases de entrenamiento destinadas a mejorar la aptitud cardiorrespiratoria (normalmente durante la fase preparatoria y la primera parte del período específico).

Dado que la mayoría de los corredores no tiene fácil acceso a laboratorios de fisiología, clínicas médicas y personal experto para obtener datos válidos y confiables para individualizar la prescripción de HIIT de acuerdo con el enfoque fisiológico, se buscaron parámetros alternativos para estimar el  $v\Delta 50$  que no requirieran pruebas de laboratorio. En base a la literatura previa sobre el tema, se evaluó la idoneidad de varias medidas relacionadas con el tapiz rodante y el ritmo de carrera para este propósito. Específicamente, se investigó si el

porcentaje relativo de  $V_{\text{peak}}$  en  $v\Delta 50$ , la diferencia entre  $v\Delta 50$  y  $V_{\text{peak}}$ , el porcentaje relativo de la velocidad de 10000-m en  $v\Delta 50$  y la diferencia entre  $v\Delta 50$  y 10000-m. La velocidad tenía una variabilidad interindividual – expresado como desviación estándar (SD) – suficientemente baja para ser utilizada como sustitutos válidos de  $v\Delta 50$ . El protocolo de pruebas fisiológicas y de rendimiento fue idéntico al adoptado en mi primera investigación sin intervención experimental entre las 8 semanas que separaron las dos sesiones de prueba. Mediante el uso de un tamaño de la muestra mucho mayor que el que se había utilizado previamente en la literatura ( $N = 75$ ) y un diseño de medición repetida (en 61 de 75 corredores para medidas relacionadas con la cinta de correr y 57 de 75 para medidas relacionadas con el ritmo de carrera), se pudo obtener suficiente precisión en la estimación de parámetros y aislar la verdadera variabilidad interindividual por la variabilidad biológica (intraindividual). El  $v\Delta 50$  observado correspondió al 110% (95% CI: 109 a 111) de la velocidad de 10000-m con una SD del 5% (95% CI: 4 a 6) y fue de  $1.4 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 1.3 a 1.6) más rápido que la velocidad de 10000-m con una SD de  $0.7 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.6 a 0.8). En cambio, el  $v\Delta 50$  observado correspondió al 92% (95% CI: 91 a 92) del  $V_{\text{peak}}$  con una SD del 2% (95% CI: 0 a 2) y fue de  $1.4 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 1.5 a 1.3) menor que el  $V_{\text{peak}}$  con una SD de  $0.5 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.4 a 0.8). Después de haber tenido en cuenta la variabilidad intraindividual, la variabilidad interindividual de la velocidad de 10000-m en  $v\Delta 50$ , expresada como SD, fue del 4% (95% CI: 3 a 6) cuando se expresó como un porcentaje de la velocidad de 10000-m velocidad ya  $0.5 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.4 a 0.8) cuando se expresa como la diferencia entre  $v\Delta 50$  y 10000-m de velocidad. En cambio, la variabilidad interindividual del  $V_{\text{peak}}$  en  $v\Delta 50$  fue del 1% (95% CI: 1 a 2) cuando se expresó como un porcentaje del  $V_{\text{peak}}$  y de  $0.2 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.2 a 0.3) cuando expresado como la diferencia entre  $v\Delta 50$  y  $V_{\text{peak}}$ . Los datos indicaron que las medidas relacionadas con el tapiz rodante pueden informar sobre el valor de  $v\Delta 50$  con una heterogeneidad suficientemente baja entre los individuos para usarse como alternativas válidas a las medidas de laboratorio. Por ello contrario, la mayor variabilidad en las medidas relacionadas con el ritmo de carrera en  $v\Delta 50$  no las hace adecuadas para prescribir HIIT de acuerdo con el enfoque fisiológico. Las medidas relacionadas con la tapiz rodante solo requieren una cinta de correr comercial (y debidamente calibrada) disponible en cualquier gimnasio y permiten estimar el  $v\Delta 50$  con precisión discreta y un esfuerzo razonable. El uso de medidas relacionadas con la cinta de correr para individualizar las prescripciones de HIIT de acuerdo con el enfoque fisiológico haría que esta forma de HIIT sea más accesible para la mayor parte de la población de corredores.

El desarrollo de la aptitud cardiorrespiratoria solo representa un paso intermedio en la preparación del entrenamiento, mientras que maximizar el rendimiento representa el objetivo



principal al acercarse a las competiciones. Como última familia de hipótesis, se investigaron los efectos de dos modelos de prescripción HIIT individualizados diferentes, el enfoque fisiológico y el enfoque del ritmo de carrera, sobre la aptitud cardiorrespiratoria y el rendimiento de velocidad de 10000-m en corredores de fondo entrenados. Siguiendo el mismo razonamiento adoptado en la primera investigación, se formuló la hipótesis de mayores ganancias en la aptitud cardiorrespiratoria para el enfoque fisiológico. De acuerdo con la fisiología moderna, fijar la intensidad del entrenamiento a un porcentaje mayor dado del ritmo de carrera objetivaría el rendimiento de la carrera a través de una red más eficiente de interacciones fisiológicas siempre y cuando la intensidad no sea demasiado alta. También se planteó la hipótesis de mayores ganancias en el rendimiento para el enfoque de ritmo de carrera. Además, se planteó la hipótesis de una mayor heterogeneidad de los efectos de la intervención para la aptitud cardiorrespiratoria para el enfoque del ritmo de carrera y una mayor heterogeneidad para el rendimiento de la carrera en el enfoque fisiológico. Se testaron 31 corredores de distancia [20 hombres y 11 mujeres, edad: 38 (9) años, velocidad de 10000-m: 14.0 (1.4) km · h<sup>-1</sup>] antes y después de 8 semanas de entrenamiento, en las que reemplazaron una parte similar de su volumen de entrenamiento habitual con dos sesiones por semana de prescripciones de HIIT individualizadas, ya sea fisiológicamente (PHY, N = 16) o basadas en el ritmo de carrera (RP, N = 15), sin modificaciones importantes en la estructura general de entrenamiento. El protocolo de pruebas fisiológicas y de rendimiento fue idéntico al adoptado en la primera investigación, al igual que la intervención de entrenamiento en PHY. En cambio, las prescripciones de HIIT basadas en el ritmo de carrera estaban destinadas a imitar la práctica de entrenamiento, con la velocidad de cada intervalo determinada de acuerdo con un porcentaje dado del ritmo de carrera objetivo, una duración fija para cada intervalo y una distancia determinada para cubrir durante la recuperación. Específicamente, los corredores debían cubrir 1000 m al 110% del ritmo objetivo de 10000-m durante cada intervalo, y 300 m en 2 min durante la recuperación, ya que estos valores son similares a los valores de prescripción de HIIT basados en fisiología a nivel grupal en una gran cohorte de corredores similares (Estudio 2).  $\dot{V}O_{2max}$  y  $V_{peak}$  mejoraron significativamente en PHY ( $\dot{V}O_{2max}$ : +3.1 mL · kg<sup>-1</sup> · min<sup>-1</sup>,  $p < .001$ , 95% CI: +1.8 a +4.4;  $\dot{V}O_{2max}$ : +0.13 L · min<sup>-1</sup>,  $p = .02$ , 95% CI: +0.03 a +0.23;  $V_{peak}$ : +0.4 km · h<sup>-1</sup>,  $p < .001$ , 95% CI: +0.2 a +0.7), mientras que el  $\dot{V}O_{2max}$  disminuyó significativamente en RP ( $\dot{V}O_{2max}$ : -2.0 mL · Kg<sup>-1</sup> · min<sup>-1</sup>,  $p = .009$ , 95% CI: -3.4 a -0.6;  $\dot{V}O_{2max}$ : -0.11 L · min<sup>-1</sup>,  $p = .05$ , 95% CI: -0.22 a 0.00), con una diferencia estadísticamente significativa de +5.1 mL · kg<sup>-1</sup> · min<sup>-1</sup> ( $p < .001$ , 95% CI: +3.6 a +∞), +0.24 L · min<sup>-1</sup> ( $p = .001$ , 95% CI: +0.12 a +∞) y +0.5 km · h<sup>-1</sup> ( $p = .003$ , 95% CI: +0.2 a +∞) para  $\dot{V}O_{2max}$  relativo,  $\dot{V}O_{2max}$  absoluto y  $V_{peak}$ ,

respectivamente, entre los grupos. Por el contrario, la velocidad de 10000-m mejoró significativamente solo en RP ( $+0.5 \text{ km} \cdot \text{h}^{-1}$ ,  $p < .001$ , 95% CI:  $+0.2$  a  $+0.7$ ), con una diferencia estadísticamente significativa de  $-0.5 \text{ km} \cdot \text{h}^{-1}$  ( $p = .01$ , 95% CI:  $-\infty$  a  $-0.1$ ) entre los grupos. El  $\dot{V}O_{2\text{max}}$  fraccional en el primer umbral de lactato mejoró significativamente solo en RP ( $+4.1\%$ ,  $p = .02$ , 95% CI:  $+0.7$  a  $+7.5$ ), con una diferencia estadísticamente significativa de  $-4.2\%$  ( $p = .05$ , 95% CI:  $-8.3$  a  $0.0$ ) entre los grupos. Aunque los datos experimentales no fueron concluyentes para la heterogeneidad de los efectos de la intervención, sí se corroboraron las dos hipótesis sobre el efecto promedio del grupo, lo que lleva a concluir que el enfoque fisiológico y el ritmo de carrera pueden realizarse dentro del mismo macrociclo, pero que la implementación del primero debería preceder a este último en términos de proximidad a las principales competiciones.

**PALABRAS CLAVE:** carrera de resistencia; fisiología del ejercicio; entrenamiento de intervalos de alta intensidad; rendimiento deportivo



## LIST OF SYMBOLS AND ABBREVIATIONS

$\alpha$  = Type I error rate / significance level

$\beta$  = Type II error rate

$\sigma$  = population SD

$\Delta_L$  = lower equivalence margin

$\Delta_{NI}$  = non-inferiority margin

$\Delta_{NI-C}$  = fixed-margin method for determining  $\Delta_{NI}$

$\Delta_{NI-P}$  = point-estimate method for determining  $\Delta_{NI}$

$\Delta_U$  = upper equivalence margin

$\lambda$  = preserved fraction

$\mu$  = population mean

ANOVA = analysis of variance

CI = confidence interval

CON = control group

CP = critical power

CS = critical speed

CR10 = 10-point Borg category ratio

$df$  = degrees of freedom

EXP = experimental group

GAS = General Adaptation Syndrome

GET = gas exchange threshold

HIIT = high-intensity interval training

HIPE = human exercise physiology of exercise

IND = individualized HIIT group

LT = lactate threshold

LT<sub>1</sub> = first lactate threshold

LT<sub>2</sub> = second lactate threshold

M = mean

MICT = moderate-intensity continuous training

MLSS = maximal lactate steady state

N/ $n$  = sample size

N<sub>max</sub> = the maximal number of intervals performed during the first HIIT session

PHY = physiologically-based HIIT group

$r$  = correlation coefficient / allocation ratio

RE = running economy

RP = race pace-based HIIT group

RPE = rate of perceived exertion

SD = standard deviation

SD<sub>P</sub> = pooled SD

SESOI = the smallest effect size of interest

SIT = sprint interval training

STD = standardized HIIT group

TID = training intensity distribution

T<sub>lim</sub> = time to exhaustion (at  $v\Delta 50$ )

TOST = two one-sided tests

$\dot{V}O_2$  = oxygen consumption

$\dot{V}O_{2max}$  = maximal oxygen consumption

$v\Delta 50$  = speed halfway between the second lactate threshold and  $V_{peak}$

$vLT_1$  = speed at  $LT_1$

$vLT_2$  = speed at  $LT_2$

$V_{peak}$  = peak treadmill speed

$v\dot{V}O_{2max}$  = speed at  $\dot{V}O_{2max}$

$z_\alpha$  = standardized normal deviates corresponding to the levels of  $\alpha$

$z_{\beta/2}$  = standardized normal deviates corresponding to the levels of  $\beta/2$

## **LIST OF TABLES**

TABLE 1. Values of the outcome variables before and after the training intervention.....	76
TABLE 2. Covariate-adjusted change scores and heterogeneity of intervention effects. ....	77
TABLE 3. Physiological and performance characteristics of the runners. ....	78
TABLE 4. Values of the outcome variables before and after the training intervention.....	80
TABLE 5. Covariate-adjusted change scores and heterogeneity of intervention effects. ....	81



## LIST OF FIGURES

- FIGURE 1. Overview of the main factors related to the maximal speed that can be maintained in distance races according to Bassett and Howley (1997). The maximal oxygen uptake ( $\dot{V}O_{2max}$ ), lactate threshold (LT), and running economy represent the main physiological factors determining distance running performance in any purely physiological model... 4
- FIGURE 2. The physiological factors that determine performance speed according to Joyner and Coyle (2008). This model is an expansion of the one proposed by Bassett and Howley (1997), which did not include anaerobic characteristics among the determinants..... 5
- FIGURE 3. The central governor (a) and the psychological-motivational (b) models of exercise performance by St Clair Gibson and Noakes (2004) and Marcora (2008), respectively. Despite the fundamental differences in the structure, conscious and subconscious processes represent integral parts of both models. .... 6
- FIGURE 4. Simplified representation of the integrative governor model of exercise proposed by St Clair Gibson and colleagues (2018). The model describes a continuous dynamic interplay between competing physiological and psychological drives, whose product of the relative weighting – affected by the endpoint (circled), distance, and time duration – defines the level of functional reserve maintained during a given physical task. .... 7
- FIGURE 5. Representation of four methods used to detect the maximal lactate steady-state and the first and second blood lactate threshold from the running speed-blood lactate relationship during exercise. Panel A: maximal lactate steady state (MLSS, approximated by  $LT_2$ ) – defined as the highest intensity at which the blood lactate concentration increases less than or equal to  $1.0 \text{ mmol} \cdot \text{L}^{-1}$  during the final 20 minutes of a 30-min constant-intensity test. Panel B: lactate threshold ( $LT_1$ ) – defined as the point of deflection in the log [blood lactate] versus log  $\dot{V}O_2$  transformation. Panel C: lactate threshold ( $LT_1$ ) – defined as the intensity preceding a  $0.4 \text{ mmol} \cdot \text{L}^{-1}$  increase in blood lactate above resting levels. Panel D: modified  $D_{max}$  ( $LT_2$ ) – defined as the point on running speed-blood lactate curve at maximal distance from a line connecting  $LT_1$  and the highest intensity (modified from Tanner & Gore, 2012). .... 10
- FIGURE 6. Representation of the pulmonary  $\dot{V}O_2$  (left panel) and blood lactate (right panel) response to constant-intensity exercise in the moderate-intensity (below  $LT_1$ ), heavy-intensity



(between  $LT_1$  and  $LT_2$ ), and severe-intensity (above  $LT_2$ ) domains. The shaded areas in the left panel define the slow component of  $\dot{V}O_2$ , which occurs above  $vLT_1$  (Jones et al., 2011)..... 16

FIGURE 7. The four-domain model proposed by Hill and colleagues (2002). The exercise intensities in the extreme domain are so high that exercise terminates before the runners can reach  $\dot{V}O_{2max}$ . The gas exchange threshold (GET) and the critical power (CP) reported in the figure serve the same function as the  $LT_1$  and  $LT_2$ , respectively, in separating the domains (Poole & Jones, 2012). ..... 21

FIGURE 8. The intensity ranges typically used for the different forms of HIIT. Although the inclusion of the extreme domain in the classical thresholds-based exercise intensity model may help to separate the forms of ‘aerobic’ HIIT from the forms of ‘anaerobic’ HIIT, it cannot differentiate all the forms of HIIT. On the contrary, the Delta concept (expressed as  $v\Delta 50$  in the figure) allows to precisely manipulate HIIT intensity for intensities at (or near)  $\dot{V}O_{2max}$  and, altogether with the knowledge of the boundary of the extreme domain, within the whole severe domain (modified from Buchheit & Laursen, 2013a). ..... 23

FIGURE 9. Representation of the supercompensation cycle following a single training load. The training load represents the physiological stressor imposed on the athlete, the (acute) fatigue after the load represents the alarm phase, and the supercompensation represents the (transitional) resistance phase characterized by an improved work capability. If no further stimuli occur, the work capability will return to the pre-load level (detraining). Exhaustion is not considered in this figure (from Issurin, 2010). ..... 26

FIGURE 10. The vision on network physiology of exercise is explained in a diagram: hierarchically organized physiological network levels interact both horizontally and vertically through circular causality to coordinate their functions (from Balagué et al., 2020)..... 28

FIGURE 11. Testing for superiority, equivalence, and non-inferiority within a typical parallel-group design. The error bars indicate the 95% confidence interval (CI) in relation to the traditional null-hypothesis test (Figure 11a) and non-inferiority test (Figure 11c) and the 90% CI in relation to the two one-sided test procedure (Figure 11b). The shaded areas indicate the rejection region for each hypothesis test. Figure 11a: From a traditional perspective (i.e., deciding on the presence of an effect), the superiority of the experimental group (EXP) compared with the control (CON) can be concluded in the first three examples. However, the standard of evidence to claim superiority differs between the examples. In the first example, it is only possible to reject effects that are smaller than zero. In the second example, it is also

possible to claim practical importance besides statistical significance. In the third example, it is possible to reject any effect that is not practically important – that is, an effect that is smaller than the smallest effect of interest (SESOI). Superiority cannot be concluded in the lower example, since the 95% CI extends beyond zero, which reflects in a  $p$ -value larger than  $\alpha$ . Figure 11b: It is possible to conclude equivalence between the interventions only in the second example since in the upper and lower example the 90% CI spans beyond the lower ( $\Delta_L$ ) or the upper ( $\Delta_U$ ) equivalence margin. Figure 11c: The observed data are identical to Figure 11b. Despite the wider CI, the absence of an upper margin allows concluding non-inferiority in both the second and third examples..... 36

FIGURE 12. The two-step process commonly employed to determine the non-inferiority margin ( $\Delta_{NI}$ ) in clinical research. A pooled effect estimate is calculated from a meta-analysis of hypothetical studies and the margin is determined using either the point estimate (point-estimate method;  $\Delta_{NI-P}$ ) or the lower 95% confidence limit (fixed-margin method;  $\Delta_{NI-C}$ ) of the effect size. The chosen margin ( $\Delta_{NI-C}$  in the example) is then multiplied by a pre-specified factor ( $\lambda$ ; usually 50%) to preserve a fraction of the active-control effect (shaded area). ..... 40

FIGURE 13. Testing for both equivalence and superiority. The thin error bars indicate the 95% CI in relation to the traditional null-hypothesis test, whereas the thick error bars indicate the 90% CI in relation to the two one-sided tests procedure. The solid vertical lines indicate the traditional null hypothesis, whereas the shaded area indicates the equivalence region. Conclusions for hypothesis tests are reported next to each example. .... 47

FIGURE 14. Processes of decision making for selecting the different hypothesis tests based on the research question that is being asked..... 49

FIGURE 15. Overview of research design for the first investigation. IND and STD represent the individualized and standardized group, respectively; whereas the track and the treadmill logo represent the 10000-m time trial and the incremental treadmill test followed by the constant-speed test, respectively. This pre-post parallel-group design was essentially identical for the first and the third investigation, whereas only one group and no experimental intervention was used to investigate the second family of hypotheses. .... 65

FIGURE 16. One of the participants is shown while performing the incremental test as an example..... 68

FIGURE 17. Representation of the approach used to prescribe high-intensity interval training in the individualized (IND) and standardized (STD) groups in the first investigation. The speed and the duration of each interval were set at the speed halfway between  $v_{LT_2}$  and  $V_{peak}$  ( $v_{\Delta 50}$ ) and 50% of the time to exhaustion at  $v_{\Delta 50}$  ( $T_{lim}$ ), respectively, whereas the speed and the duration of each recovery period were set at 50% of the  $V_{peak}$  and 25% of  $T_{lim}$ , respectively. HIIT prescriptions in IND were individualized according to each runner's values, whereas the average group values were used to prescribe the same HIIT session to all runners in STD. Regardless of the group, the initial number of intervals was anchored to the maximal number of intervals that each runner would have been capable to perform during the first HIIT session ( $N_{max}$ ), in such a way each runner would have alternated  $N_{max} - 1$  or  $N_{max} - 2$  intervals for the first 2 weeks of training. Other than mimicking previous research, this approach was also used to allow runners to better adapt to the new training load. From the 3<sup>rd</sup> week, the number of intervals increased on an individual basis in both groups. .... 69

FIGURE 18. Representation of the approach used to individualize high-intensity interval training prescriptions in the physiologically-based (PHY) and race pace-based (RP) group in the third investigation. The track and the treadmill logo represent the 10000-m time trial and the incremental test followed by the constant-speed treadmill test, respectively. The approach used for PHY was identical to the one used for IND in the first investigation. Runners in RP were required to cover 1000 m at 110% of the target 10000-m pace during each interval, and 300 m in 2 min during recovery. Regardless of the group, the initial number of intervals was anchored to the maximal number of intervals that each runner would have been capable to perform during the first HIIT session ( $N_{max}$ ), in such a way each runner would have alternated  $N_{max} - 1$  or  $N_{max} - 2$  intervals for the first 2 weeks of training. Other than mimicking previous research, this approach was also used to allow runners to better adapt to the new training load. From the 3<sup>rd</sup> week, the number of intervals increased on an individual basis in both groups. .... 70

FIGURE 19. Representation of the physiological rationale underlying this investigation. The much larger chance that the prescribed HIIT intensity would shift from the severe to the heavy domain in STD than IND during the training period would result in a much larger possibility than some runners may face blunted – if not null – improvements in cardiorespiratory fitness in STD than IND due to a reduced/zero time spent at (or near)  $\dot{V}O_{2max}$ . This would in turn result in an overall smaller mean training effect and larger heterogeneity of intervention effects in STD compared with IND (modified from Demarie et al., 2000). .... 87

FIGURE 20. The physiological approach to high-intensity interval training prescription may lead some runners to train at intensities that are very distant from the target speed with an

increased risk of losing the specificity for the task. The resulting inter-individual variability in the training stimulus may affect both the individual response heterogeneity and the mean training effect for 10000-m performance. .... 95



# TABLE OF CONTENTS

ACKNOWLEDGMENTS

SCIENTIFIC CONTRIBUTIONS

ABSTRACT / RESUMEN

1. INTRODUCTION .....	1
1.1 Born to run .....	3
1.2 Distance running performance: <i>E Pluribus Unum</i> .....	4
1.3 The ‘Big Three’ .....	8
1.3.1 Maximal oxygen uptake ( $\dot{V}O_{2max}$ ) .....	8
1.3.2 Lactate threshold(s).....	9
1.3.3 Running economy .....	11
1.4 Endurance training strategies to enhance the physiological determinants of distance running performance .....	13
1.4.1 One size doesn’t fit all: the role of the training level in training prescriptions .....	13
1.4.2 Lactate thresholds concept as a means to prescribe exercise intensities .....	15
1.4.3 The use of the Delta concept to prescribe high-intensity intermittent exercise .....	19
1.5 Scientific research vs training practice: how should distance runners train?.....	24
1.5.1 Physiological and race pace approach: two faces of the same medal.....	24
1.5.2 Network physiology: the forgotten piece of the puzzle .....	25
1.5.3 The right training, at the right time, in the right way.....	28
1.6 Moving beyond the traditional null-hypothesis: the use of equivalence and non-inferiority tests for interventional studies in exercise physiology and sport science .....	31
1.6.1 Investigating statistical differences (superiority).....	31
1.6.2 Investigating equivalence and non-inferiority .....	34

1.6.3 Suitability of the reference intervention .....	37
1.6.4 Determination of non-inferiority and equivalence margin(s) .....	38
1.6.5 Sample size planning for non-inferiority and equivalence studies .....	41
1.6.6 Re-imagining interventional studies using equivalence and non-inferiority tests .....	42
1.6.7 Switching between hypotheses .....	46
1.6.8 Limitations and additional considerations .....	48
1.6.9 Conclusions and recommendations for future research .....	48
2. RESEARCH GOALS .....	51
2.1 General goal .....	53
2.2 Specific goals .....	53
3. RESEARCH HYPOTHESES .....	55
4. METHODOLOGY .....	59
4.1 Participants .....	61
4.1.1 Individualizing HIIT according to the physiological characteristics .....	61
4.1.2 Testing interindividual variability in different treadmill-related measures and race pace-related measures .....	62
4.1.3 Individualizing HIIT prescriptions: physiological characteristics <i>vs</i> race pace .....	63
4.2 Experimental design .....	64
4.3 Procedures .....	65
4.3.1 Physiological and performance testing .....	65
4.3.2 Training characteristics .....	69
4.3.3 Training assessment .....	71
4.4 Statistical analysis .....	71
5. RESULTS .....	73

5.1 Individualizing HIIT according to the physiological characteristics .....	75
5.2 Testing interindividual variability in different treadmill-related measures and race pace-related measures .....	78
5.3 Individualizing HIIT prescriptions: physiological characteristics <i>vs</i> race pace .....	79
6 DISCUSSION.....	83
6.1 Individualizing HIIT according to the physiological characteristics .....	85
6.2 Testing interindividual variability in different treadmill-related measures and race pace-related measures .....	91
6.3 Individualizing HIIT prescriptions: physiological characteristics <i>vs</i> race pace .....	93
7. CONCLUSIONS .....	97
8. BIBLIOGRAPHY .....	103
9. ADDENDUMS.....	145
9.1 Addendum 1: Presentation at International Conference .....	147
9.2 Addendum 2: Letters of stay .....	150
9.3 Addendum 3: Research ethics approval .....	153
9.4 Addendum 4: Informed consent.....	154





## INTRODUCTION

*'Run often. Run long. But never outrun your joy of running.'*

*Julie Isphording – distance runner*





# 1. INTRODUCTION

## 1.1 Born to run

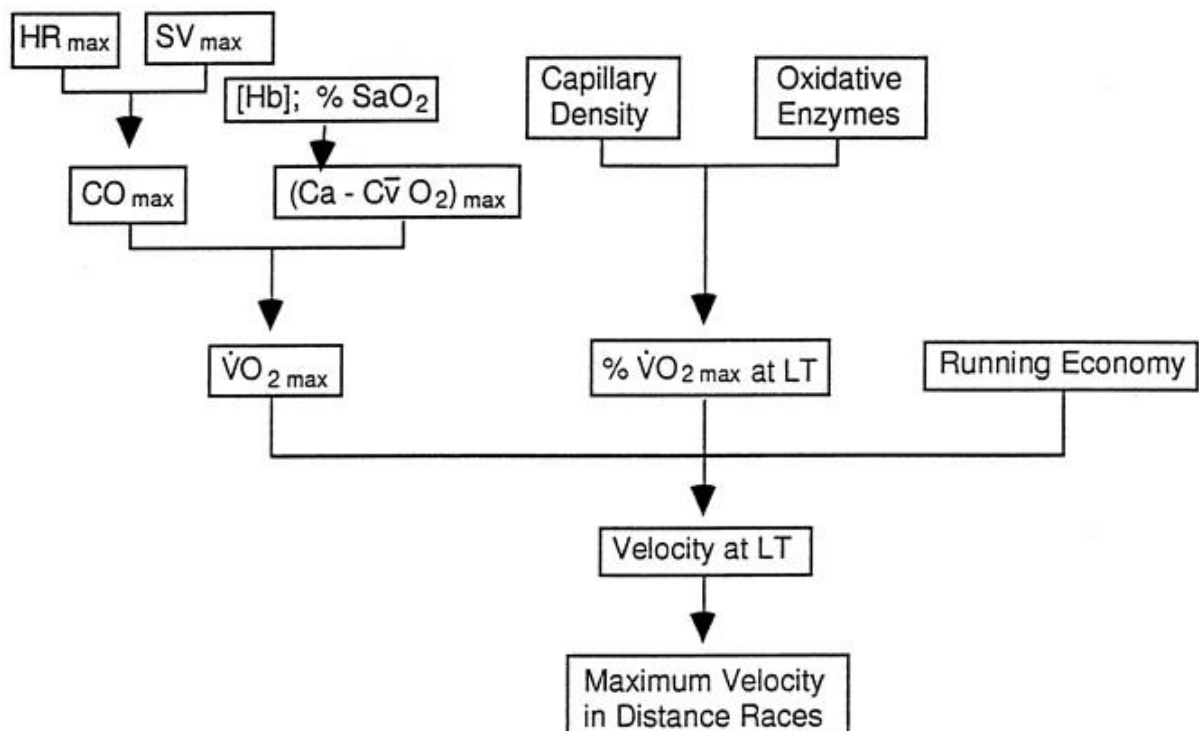
Running activities are very common in our modern society. About a tenth of the global adult population practices some form of running activity, and there are indeed more than 7.9 million people registered for running races in 2018 (Hultheen et al., 2017; Andersen, 2019). There is no surprise in these numbers. Bipedal running has accompanied the evolution of mankind for probably the last two million years (Lieberman & Bramble, 2007). For some scientists (Bramble & Lieberman, 2004), running may have shaped our body along the path of human evolution, making us unique among all primates. Running may then be responsible for what we are today.

It is then little wonder that we, as humans, require a certain amount of running, and physical activity in general, to stay healthy (Garber et al., 2011). However, cultural evolution does not pace with biological evolution (Perreault, 2012). The same adapting path that had been responsible for our survival on this planet for three hundred thousand years is now obsolete for the needs of modern human society. This ‘evolutionary mismatch’ has opened the door to most modern chronic diseases (Booth et al., 2012). Nowadays, endurance running is no longer driven by the need to survive but by the need to stay healthy in an environment that does no longer accept us as its natural parts.

Endurance running does not only represent a healthy practice, but it is also a competitive challenge to test human limits. As *Homo sapiens* species, we inherited an intraspecific competitive behavior from our ancestors (Bhattacharya, 2020). Today this behavior is no longer oriented toward ‘classical’ resources, such as food or territory, but victories in sports for many of us. As a unique ultra-cultural species (Henrich, 2011), we also have an innate desire to understand the nature of things. We wish to separate and properly characterize each piece of the often-complicated puzzle representing the topics we are interested in to have a full view and control over it. Thus, it is not surprising that, along with the natural sense of disappointment, the first questions that come after failing a goal in competition usually are ‘why did I fail?’ and ‘how can I improve?’. Exercise physiologists and sports scientists have spent a tremendous effort to answer these two common but fundamental questions. Although their conclusions did not always converge, they have provided important pieces of information to place the athletes and coaches one step closer to their goals.

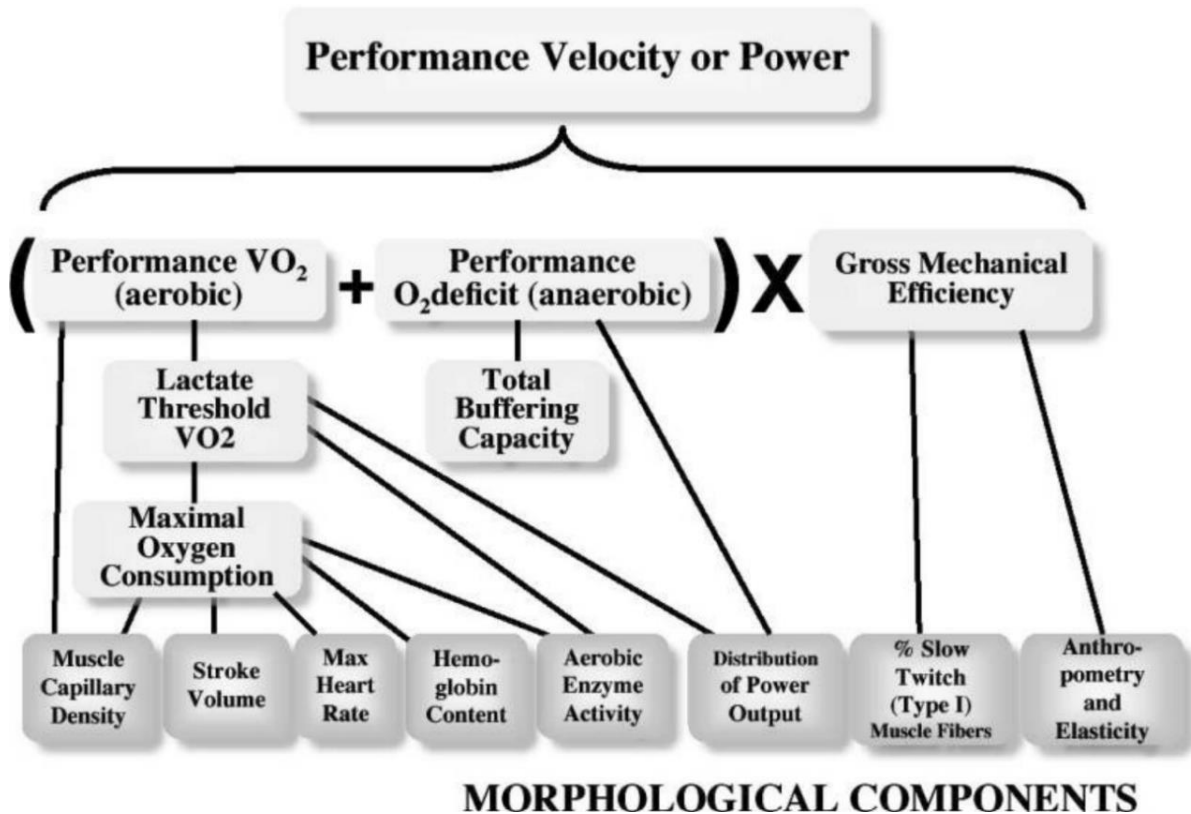
## 1.2 Distance running performance: *E Pluribus Unum*

Understanding the structure of complex phenomena is the key to reducing their perceived complexity (Mazzocchi, 2008). This also applies to human performance. Generations of exercise physiologists and sports scientists have contributed to building and continuously refining explanatory models of distance running performance. After the first pioneering studies (Hill & Lupton, 1923; Hill, 1925), the '60s and the '70s saw the growth in studies attempting to identify candidate physiological determinants and understand their linkages and interactions (Saltin & Astrand, 1967; Costill, 1967, 1970, 1972; Wyndham et al., 1969; Costill et al., 1971, 1973; Pollock, 1977; Foster et al., 1977, 1978; Davies & Thompson, 1979; Farrell et al., 1979; Conley & Krahenbuhl, 1980). However, it was not until the '80s and '90s that the first integrated physiological models were presented to the public (Sparling, 1984; Sjödín & Svedenahg, 1985; Di Prampero et al., 1986; Joyner, 1993; Coyle, 1995, 1999; Bassett & Howley, 1997, 2000) (Figure 1).



**FIGURE 1.** Overview of the main factors related to the maximal speed that can be maintained in distance races according to Bassett and Howley (1997). The maximal oxygen uptake ( $\dot{V}O_{2max}$ ), lactate threshold (LT), and running economy represent the main physiological factors determining distance running performance in any purely physiological model.

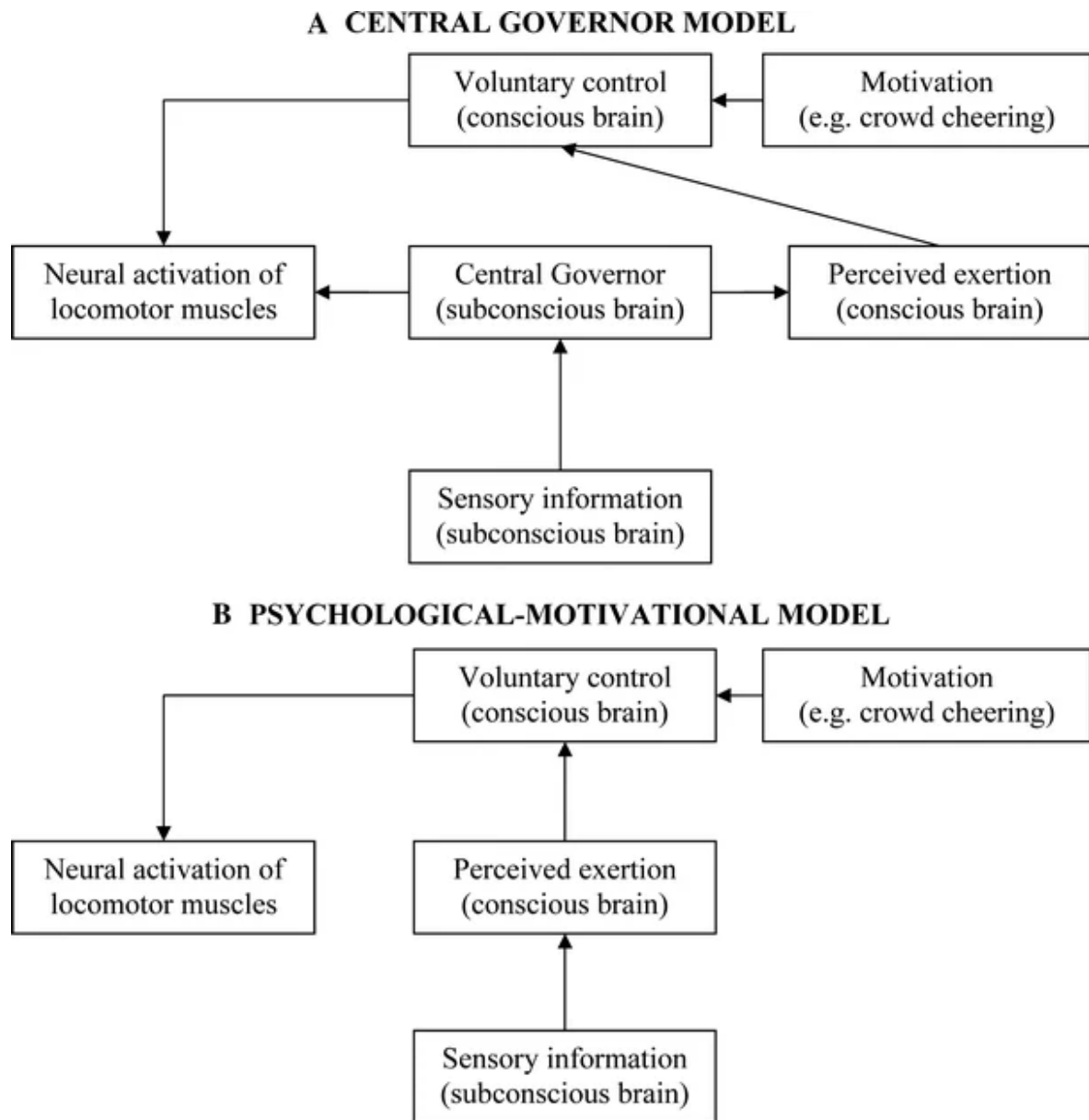
These models progressively became more and more complex by integrating other physiological components, such as neuromuscular and anaerobic characteristics, that had been overlooked by earlier scientists (Joyner & Coyle, 2008; Beattie et al., 2014) (Figure 2).



**FIGURE 2.** The physiological factors that determine performance speed according to Joyner and Coyle (2008). This model is an expansion of the one proposed by Bassett and Howley (1997), which did not include anaerobic characteristics among the determinants.

The process of expansion and refinement of physiological models of distance running – and more general endurance exercise – performance has not been free from a disagreement between scientists. Disputes about the proper physiological structure of the model and the relation among its components were raised almost immediately after the first integrated models were published (Noakes, 1988, 1996; Ulmer, 1986; Bassett & Howley, 1997; Bergh et al., 2000). The debate that arose from these disputes led to new theories that gave greater importance to the role of the brain as a regulator of effort and fatigue during exercise (Noakes, 2000; St Clair Gibson & Noakes, 2004; Lambert et al., 2005) (Figure 3a). These theories attracted, in turn, new criticisms (Marcora, 2008) (Figure 3b) that resulted, in turn, in new theories (St Clair Gibson et al., 2018; Venhorst et al., 2018a) (Figure 4). Although it is not clear whether or not science has ultimately benefited from these shortly repeated paradigm shifts (Inzlicht & Marcora, 2016; Robergs,

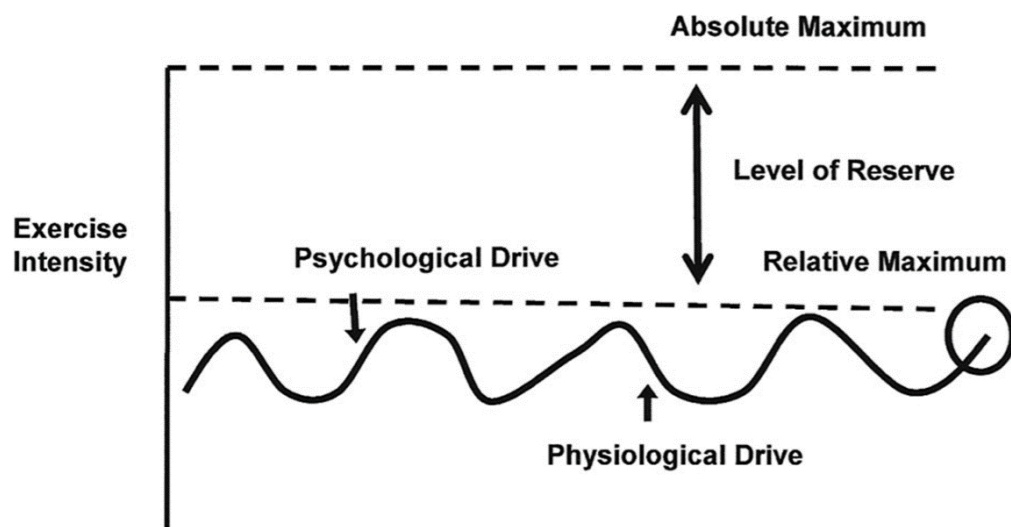
2017), we should still give credit to those scientists for having raised awareness, at least in the most recent versions of their theories, about the deep interconnection between physiological and psychological factors.



**FIGURE 3.** The central governor (a) and the psychological-motivational (b) models of exercise performance by St Clair Gibson and Noakes (2004) and Marcora (2008), respectively. Despite the fundamental differences in the structure, conscious and subconscious processes represent integral parts of both models.

Today scientists agree about the psychophysiological structure of distance running performance. However, given the complexity of human nature, it is not surprising that a single unifying model is still lacking. Recent lines of evidence define distance running performance

as a very complex phenomenon involving dynamic and non-linear interactions of several psychophysiological factors interacting at multiple levels (Millet, 2011; Knicker et al., 2011; Edwards & Polman, 2013; Pageaux, 2014; Renfree & Casado, 2018; St Clair Gibson et al., 2018; Venhorst et al., 2018a) (e.g., Figure 4). Success in competing events may then depend on the optimal dynamic interplay between physiological, psychological, technical, and tactical factors, which include interactions with other competitors (Renfree & Casado, 2018; Venhorst et al., 2018a).



**FIGURE 4.** Simplified representation of the integrative governor model of exercise proposed by St Clair Gibson and colleagues (2018). The model describes a continuous dynamic interplay between competing physiological and psychological drives, whose product of the relative weighting – affected by the endpoint (circled), distance, and time duration – defines the level of functional reserve maintained during a given physical task.

Despite the awareness of the dynamic multi-level structure of regulation of distance running performance, the knowledge about the network of linkages and interactions between the factors and levels is limited. Science is still far from precisely modeling the complex paths that exist among all the factors involved. The recent development of dynamic multiple-level psychophysiological regulatory models of endurance exercise performance and their complexity has also restricted their use in interventional studies to very late research (e.g., Venhorst et al., 2018b, c), whereas the vast majority of the literature so far focused only on a few major physiological factors (e.g., Buchheit & Laursen, 2013a, b; Beattie et al., 2014).



## 1.3 The ‘Big Three’

### 1.3.1 Maximal oxygen uptake ( $\dot{V}O_{2\max}$ )

The maximal oxygen uptake, or  $\dot{V}O_{2\max}$ , represents the maximum integrated capacity of the pulmonary, cardiovascular, and muscular systems to uptake, transport, and use  $O_2$ , respectively (Poole et al., 2008). Although there is still no consensus about the main limiting factors of the  $\dot{V}O_{2\max}$ , it can be considered synonymous with maximal aerobic power and the best indicator of cardiorespiratory fitness (Astrand, 1955; Bassett & Howley, 2000). Exercise physiologists and sports scientists typically assess  $\dot{V}O_{2\max}$  during a maximal incremental treadmill test, in which they progressively increase the speed on the treadmill until the runner is no longer able to sustain it (Bentley et al., 2007). The general criterion for  $\dot{V}O_{2\max}$  determination is the achievement of a plateau in the  $\dot{V}O_2$  (Taylor et al., 1955). While different researchers have used different criteria to define what they consider a plateau, the rationale they should have all in common is observing a larger deflection in the linear running speed- $\dot{V}O_2$  relationship than what could be explained by the measurement error alone (Midgley et al., 2007a, 2009). However, the low frequency of plateau occurrence in  $\dot{V}O_2$  – regardless of the criterion – has led scientists to adopt secondary criteria and/or supplementary procedures to validate  $\dot{V}O_{2\max}$  achievement (Midgley et al., 2007a; Midgley & Carroll, 2009). Today, it is a common practice when testing runners to require a second (usually constant-intensity) test at a similar, lower, or higher, speed than the highest speed reached during the incremental test, known as the verification phase, to confirm the attainment of  $\dot{V}O_{2\max}$  (Thoden et al., 1982; Midgley et al., 2006a, 2007a; Schaun, 2017).

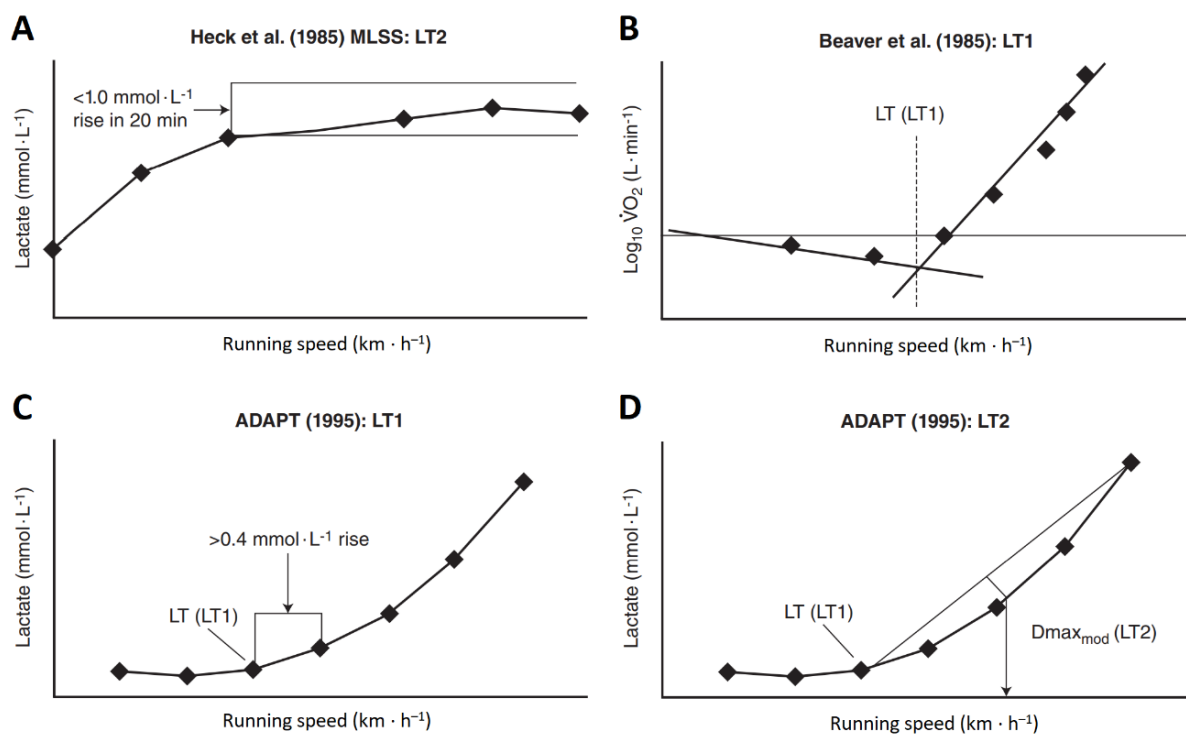
From the energetic standpoint, distance running performance is essentially an aerobic-driven activity. The aerobic metabolism indeed provides most of the energy to runners for competing in any long-distance race (Gastin, 2001). Given the essential contribution of the aerobic energy system in distance running races, it should not surprise that several scientists initially considered the  $\dot{V}O_{2\max}$  the main determinant of distance running performance (Foster et al., 1978; Davies & Thompson, 1979). Exercise physiologists and sports scientists discovered the relation between  $\dot{V}O_{2\max}$  and distance running performance almost a century years ago (Hill, 1925; Robinson et al., 1937). Numerous subsequent studies further investigated the value of the parameter in different populations and its relation with training and distance running performance (Astrand, 1955; Saltin & Astrand, 1967; Costill, 1967; Wyndham et al., 1969; Costill et al., 1971, 1973; Pollock, 1977; Daniels et al., 1978; Foster et al., 1977, 1978; Davies & Thompson, 1979; Foster, 1983). The value of the  $\dot{V}O_{2\max}$  in elite runners is usually 50–100%

greater than the one observed in untrained individuals, regardless of whether  $\dot{V}O_{2\max}$  is expressed in absolute terms ( $L \cdot \text{min}^{-1}$ ) or relative to the body mass ( $\text{mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) (Joyner & Coyle, 2008). Researchers also found that  $\dot{V}O_{2\max}$  explains distance running performance in distance runners of heterogeneous performance levels (Costill, 1967; Wyndham et al., 1969; Costill et al., 1973; Foster et al., 1977, 1978; Farrell et al., 1979; Davies & Thompson, 1979; Foster, 1983; McLaughlin et al., 2010), and the largest improvements usually take place during the first weeks of training (Daniels et al., 1978). However, the relation between  $\dot{V}O_{2\max}$  and performance is weaker in homogeneous groups of highly-trained runners (Pollock, 1977; Conley & Krahenbuhl, 1980), and the long-term trainability of  $\dot{V}O_{2\max}$  is limited in this population (Daniels et al., 1978). Today we know that, despite being a prerequisite for success in distance running,  $\dot{V}O_{2\max}$  does not represent a sensitive parameter to assess performance or track changes in performance in highly-trained runners (Legaz-Arrese et al., 2005; Joyner & Coyle, 2008).

### 1.3.2 Lactate threshold(s)

A major limitation of using  $\dot{V}O_{2\max}$  to explain distance running performance is its short sustainability.  $\dot{V}O_{2\max}$  declines after a few minutes of intense exercise mainly due to accelerated metabolic fatigue, which leads runners to reduce the intensity or interrupt the exercise (Joyner & Coyle, 2008; Billat et al., 1994; Hill & Rowell, 1997; Hill et al., 1997; Midgley & McNaughton, 2006; Black et al., 2017). Since runners cannot sustain  $\dot{V}O_{2\max}$  for more than a few minutes, most of the long-distance running events are performed at an average speed below the one corresponding to  $\dot{V}O_{2\max}$  (Costill et al., 1973; Bassett & Howley, 2000; Joyner & Coyle, 2008). The implications of this phenomenon are straightforward: for runners who are already close to their genetic potential for  $\dot{V}O_{2\max}$ , the ability to sustain high fractions of  $\dot{V}O_{2\max}$  in competition may then be a stronger determinant of performance than  $\dot{V}O_{2\max}$  itself (Wyndham et al., 1969; Costill, 1970; Costill et al., 1971, 1973; Farrell et al., 1979; Sjödín & Svedenhag, 1985; Di Prampero et al., 1986). Since the ability to sustain high fractions of  $\dot{V}O_{2\max}$  for prolonged periods finds a strong correspondence with the maximal intensity that can be sustained without a progressive rise in blood lactate levels (Sjödín & Svedenhag, 1985; Joyner, 1993; Coyle, 1995, 1999; Bassett & Howley, 1997, 2000; Joyner & Coyle, 2008), researchers have used submaximal blood lactate levels as a means to explain distance running performance (Wyndham et al., 1969; Costill, 1970; Farrell et al., 1979; Sjödín & Svedenhag, 1985). In this

regard, the concept of maximal lactate steady state (MLSS) has been utilized to identify the upper boundary of constant intensity exercise (Faude et al., 2009). However, the MLSS determination requires performing several (three to seven) prolonged ( $\geq 30$  minutes) constant-intensity tests on separate days at different running speeds, which has limited its use in the research practice (Beneke, 2003; Faude et al., 2009). Due to the difficulties in adequately estimating MLSS with a single laboratory visit, alternative methods based on the determination of physiologically meaningful ‘thresholds’ on the running speed-blood lactate curve have gained popularity among exercise physiologists and sports scientists. These thresholds consist of two typical breakpoints: the lowest intensity at which blood lactate concentration starts to increase above resting levels ( $LT_1$ ) and the highest intensity at which lactate production and clearance are in *equilibrium* ( $LT_2$ ), which is used to approximate MLSS. Various blood lactate threshold concepts have emerged [see Faude et al., (2009) for a review] (Figure 5).



**FIGURE 5.** Representation of four methods used to detect the maximal lactate steady-state and the first and second blood lactate threshold from the running speed-blood lactate relationship during exercise. Panel A: maximal lactate steady state (MLSS, approximated by  $LT_2$ ) – defined as the highest intensity at which the blood lactate concentration increases less than or equal to  $1.0 \text{ mmol} \cdot \text{L}^{-1}$  during the final 20 minutes of a 30-min constant-intensity test. Panel B: lactate threshold ( $LT_1$ ) – defined as the point of deflection in the log [blood lactate] versus log  $\dot{V}O_2$  transformation. Panel C: lactate threshold ( $LT_1$ ) – defined as the intensity preceding a  $0.4 \text{ mmol} \cdot \text{L}^{-1}$  increase in blood lactate above resting levels. Panel D: modified  $D_{\text{max}}$  ( $LT_2$ ) – defined as the point on running speed-blood lactate curve at maximal distance from a line connecting  $LT_1$  and the highest intensity (modified from Tanner & Gore, 2012).

Generally, individual threshold concepts detect the submaximal metabolic inflection points on the running speed-blood lactate curve without constraining the thresholds to any absolute blood lactate level. On the contrary, the fixed threshold concepts constrain the lactate thresholds to prespecified blood lactate levels (e.g., the running speed at 2 and/or 4 mmol · L<sup>-1</sup>). Explaining performance using fixed blood lactate thresholds was a common practice in the past (Sjodin & Jacobs, 1981; Weltman et al., 1987). However, several factors, such as the characteristics of the test, the diet, the site, the method of blood sampling, and the laboratory method may heavily influence overall blood lactate levels (Bentley et al., 2007; Faude et al., 2009). Given the large inter-individual variability in blood lactate levels at MLSS, several researchers have recommended abandoning fixed blood lactate approaches in favor of the individualized ones (Stegmann et al., 1981; Beneke et al., 2011). While different in terminology and/or the physiological background, the blood lactate threshold concepts have in common the possibility of estimating submaximal metabolic inflection points, along with  $\dot{V}O_{2\max}$  and other physiological variables, using a single incremental treadmill test (Bentley et al., 2007; Tanner & Gore, 2012).

The variety of methods available to estimate lactate thresholds has led to considerable confusion and misinterpretation (Bentley et al., 2007; Faude et al., 2009). Although there is still no agreement among researchers on the best approach(es) to estimate blood lactate thresholds, many studies found strong correlations between lactate thresholds and distance running performance, regardless of the approach adopted to estimate the thresholds (Faude et al., 2009).

### 1.3.3 Running economy

Running economy (RE) is defined as the energy demand for a given submaximal running speed, and it is typically expressed either as the submaximal  $\dot{V}O_2$  at a given running speed or the  $\dot{V}O_2$  required to cover a given distance (Sanders et al., 2004; Barnes & Kilding, 2015). While the importance of  $\dot{V}O_{2\max}$  and its sustainable fraction to determine success in the competition was already clear among exercise physiologists during the '70s (Astrand, 1955; Saltin & Astrand, 1967; Costill, 1967; Wyndham et al., 1969; Costill, 1970; Costill et al., 1971, 1973; Pollock, 1977; Daniels et al., 1978; Foster et al., 1977, 1978; Farrell et al., 1979; Davies & Thompson, 1979), RE was still an overlooked aspect of performance at that time. Although researchers have acknowledged RE as a potential factor influencing performance since their first explanatory models (Wyndham et al., 1969; Costill et al., 1971; Costill, 1972; Pollock et al.,

1977), early studies showed poor correlations between RE and performance (Costill et al., 1973; Farrell et al., 1979). The work by Conley and Krahenbuhl (1980) represented a turning point on this topic. Using a much more homogeneous group of runners than those that other researchers previously used, these researchers found that RE was indeed a major determinant of performance in highly-trained runners. This discovery should not be surprising. Since RE defines the oxygen demand at a given speed, the most economical runners may run at a higher speed than their competitors with a worse economy while sustaining a similar fraction of  $\dot{V}O_{2max}$ , all else equal (Pollock, 1977; Conley & Krahenbuhl, 1980; Noakes, 1988). Highly-trained and elite runners with a similar  $\dot{V}O_{2max}$  may vary up to ~30% in their RE (Costill et al., 1973; Conley and Krahenbuhl, 1980; Daniels, 1985; Morgan et al., 1991; Morgan & Craib, 1992). Today we know that this inter-individual variability depends on a complex interplay of different anthropometric, physiological, and biomechanical characteristics (Anderson, 1996; Arampatzis et al., 2006; Saunders et al., 2004; Nummela et al., 2007; Spurrs et al., 2003; Barnes et al., 2013). While some of these characteristics are immutable, others can change in response to specific interventions (Midgley et al., 2007b; Barnes & Kilding, 2015).

Training interventions represent the most powerful tool available to coaches to optimize the physiological characteristics of their athletes (Hawley, 2002; Joyner & Coyle, 2008). The characteristics of the training stimulus, in terms of volume (duration), intensity, and frequency, induce specific acute physiological responses which, when of sufficient magnitude and repeated over time with adequate recovery periods, lead to chronic adaptations associated with increases in athletic performance (Hawley, 2002; Midgley et al., 2007b). In the next chapter, I will review and discuss the different training approaches to prescribe exercise and manipulate the physiological determinants of distance running performance that have been investigated in the exercise physiology and sport science literature so far. The review is not meant to be exhaustive, but it intends to provide a general overview of some training aspects that will be expanded in chapter 1.5.

## 1.4 Endurance training strategies to enhance the physiological determinants of distance running performance

### 1.4.1 One size doesn't fit all: the role of the training level in training prescriptions

At the beginning of a distance running career, virtually any endurance training stimulus will induce several different physiological adaptations and improve performance (Wenger & Bell, 1986; Midgley et al., 2006b, 2007b; Beneke et al., 2011; Bangsbo, 2015). There is not much need to carefully manipulate the training variables to properly characterize, differentiate and vary the stimulus at this stage. An increase in the overall training volume is usually the first naïve but effective strategy that beginners employ when approaching distance running. However, as the runners acquire and develop physical conditioning and running techniques, the importance of the overall training volume shrinks, and other factors, such as the training intensity or its distribution across the overall training volume become predominant (Faulkner, 1968; MacDougall & Sale, 1981; Sjödín & Svedenhag, 1985; Wenger & Bell, 1986; Priest & Hagan, 1987; Pate & Branch, 1992; Laursen & Jenkins, 2002; Midgley et al., 2006b, 2007b; Bangsbo, 2015; Stögg & Sperlich, 2015).

Recreational runners can obtain very large improvements in the  $\dot{V}O_{2\max}$  just by increasing the volume of moderate-intensity runs without the need to increase the intensity (Gettman et al., 1976; Patton & Vogel, 1977; Wenger & Bell, 1986; Beneke & Hütler, 2005). On the contrary, runners who are already adapted to high training volumes may not respond to further increases in the volume and require intensities at or close  $\dot{V}O_{2\max}$  sustained for prolonged periods to further develop  $\dot{V}O_{2\max}$  adaptations (Wenger & McNab, 1975; Daniels & Scardina, 1984; Sjödín & Svedenhag, 1985; Wenger & Bell, 1986; Robinson et al. 1991; Hill & Rowell, 1997; Laursen & Jenkins, 2002; Midgley & Naughton, 2006; Midgley et al., 2006b, 2007b; Buchheit & Laursen, 2013a; Stögg & Sperlich, 2014; Wen et al., 2019). The main mediating physiological mechanism for  $\dot{V}O_{2\max}$  improvements in already trained runners seems to be the exercise-induced mechanical overload (i.e., myocardial pressure and volume overload), associated with the enhancement of the maximal stroke volume, that occurs at intensities that elicit the  $\dot{V}O_{2\max}$  (Midgley et al., 2006b, 2007b). Since elite runners have already reached their genetic potential for the  $\dot{V}O_{2\max}$ , the only changes that can be observed in these populations are generally due to seasonal fluctuations (Daniels, 1974; Daniels et al., 1978; Svedenhag & Sjödín, 1985; Berg et al., 1995; Coyle, 2005; Legaz Arrese et al., 2005; Martin et al., 2005; Jones, 2006; Sassi et al., 2008; Lacour et al., 2009).

Traditionally, coaches and athletes considered prolonged, moderate-intensity running optimally for improving lactate thresholds (MacDougall, 1977; MacDougall & Sale, 1981). However, experimental evidence did not always agree with this view (Lehmann et al., 1991). Researchers have shown that moderate-intensity running at or below the speed associated with  $LT_1$  ( $vLT_1$ ) may improve  $LT_1$  and  $LT_2$  in the short term in previously untrained individuals but not in trained individuals, which may need higher intensities [at or above  $vLT_1$  for  $LT_1$  and – probably – at or above the speed associated with  $LT_2$  ( $vLT_2$ ) for  $LT_2$ ] to develop long-term adaptations (Sjodin et al., 1982; Tanaka et al., 1986; Priest & Hagan, 1987; Yoshida et al., 1990; Weltman et al., 1992; Londeree, 1997; Billat et al., 2004; Beneke & Hütler, 2005; Philp et al., 2008). Exercise physiologists and sports scientists have proposed two possible mechanisms eliciting training adaptations in  $LT_1$  and  $LT_2$ . One is related to specific muscle adaptations in type II fibers that decrease lactate production and increase clearance at higher running speeds (MacDougall, 1977; Anderson & Rhodes, 1989; Midgley et al., 2007b). The other is related to the role of lactate as a candidate molecule for gene induction involved in lactate thresholds adaptations (Midgley et al., 2007b; Iaia & Bangsbo, 2010). Since both these mechanisms depend on reaching high lactate levels, intensities above the  $vLT_2$  should represent the optimal training stimulus to maximize adaptations in the lactate thresholds. Although this seems true for intensities at or slightly above the  $vLT_2$  (Sjodin et al., 1982; Yoshida et al., 1990; Londeree, 1997; Billat et al., 2004), intermittent exercise at or close  $\dot{V}O_{2max}$  appeared less efficacious in enhancing the lactate thresholds (Acevedo & Goldfarb, 1989; Billat et al., 1999; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003; Denadai et al., 2006), probably due to the relatively low lactate accumulation (Buchheit & Laursen, 2013b).

Despite the large inter-individual variability observed in highly-trained and elite runners, training experience seems to be a major determinant of RE (Mayhew, 1977; Sjodin & Svedenhag, 1985; Svedenhag & Sjodin, 1985; Morgan et al., 1995). Elite runners tend to have a better economy than lower-caliber runners (Sjodin & Svedenhag, 1985; Morgan et al., 1995). Data collected from world-class runners indicate that the RE improves over time (Daniels, 1974; Conley et al., 1984; Jones, 2006). This long-term improvement in RE is likely the result of continued adaptations in metabolic, biomechanical, and/or neuromuscular efficiency (Nelson & Gregor, 1976; Midgley et al., 2007b). Intensities below and above the  $vLT_1$  seem to be efficacious to improve RE in previously untrained individuals and recreational runners, respectively (Patton & Vogel, 1977; Beneke & Hütler, 2005), although it is difficult to isolate the effect of the training intensity from the training volume and time effect (Midgley et al., 2007b; Barnes & Kilding, 2015). Among the training interventions that coaches may adopt to

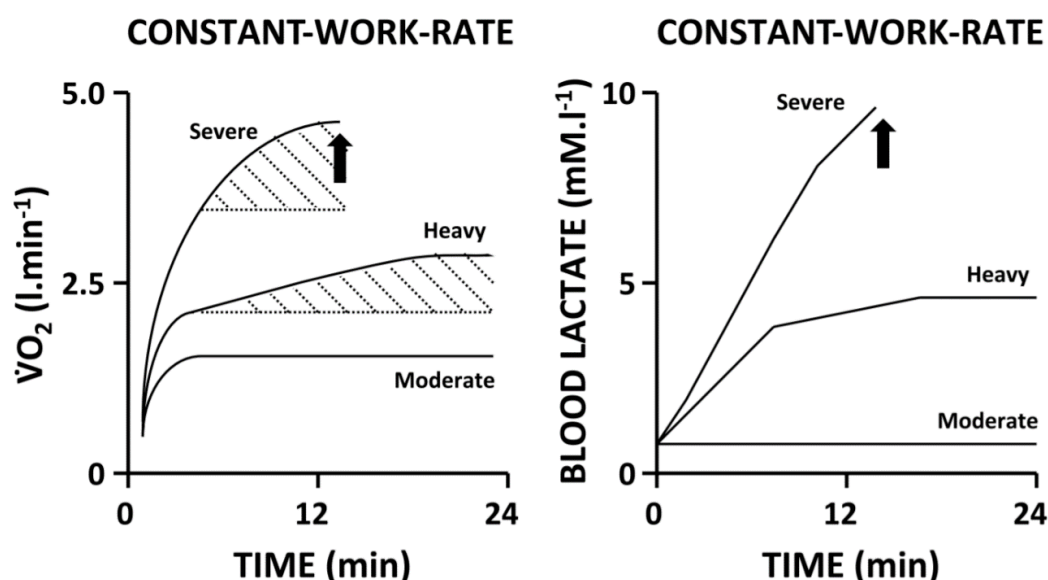
improve RE in trained distance runners, both continuous running at  $vLT_2$  (Sjodin et al., 1982) and high-intensity intermittent exercise performed at speeds around  $\dot{V}O_{2max}$  (Conley et al., 1981; Franch et al., 1998; Billat et al., 1999; Slawinski et al., 2001; Lafitte et al., 2003; Denadai et al., 2006) appear efficacious. Although some researchers speculated that very high-intensity running (e.g., equal to or larger than 130% of  $\dot{V}O_{2max}$ ) might negatively affect RE due to a deterioration in running technique or insufficient training volume achievable to induce a training effect (Zavorsky et al., 1998; Midgley et al., 2007b; Barners & Kilding, 2015) and some studies supported this view (Franch et al., 1998), other studies found this form of exercise very effective in enhancing RE in trained distance runners (Iaia et al., 2008, 2009; Bangsbo et al., 2009; Skovgaard et al., 2017).

#### 1.4.2 Lactate thresholds concept as a means to prescribe exercise intensities

Threshold or turn point concepts do not only explain performance, but they can also guide training prescriptions. Several guidelines (e.g., Garber et al., 2011; Piercy et al., 2018) often individualize training intensity prescriptions using various percentage ranges of different physiological measures (maximal heart rate,  $\dot{V}O_{2max}$ , etc.). This approach named the ‘relative percent concept’ has been criticized by many authors since it overlooks the individual variability in the physiological stimulus within a given percentage range (Katch et al., 1978; Scharhag-Rosenberger et al., 2010; Mann et al., 2013; Wolpern et al., 2015; Weatherwax et al., 2019; Meyler et al., 2021). This limitation may have important implications on the heterogeneity of the training stimulus and the magnitude of adaptations for the desired outcome(s). To prevent this risks, exercise physiologists and sports scientists usually prescribe training intensity according to specific physiological markers from incremental exercise tests (Seiler, 2010; Beneke et al., 2011). The physiological meaning of  $LT_1$  and  $LT_2$  can indeed be used to divide the overall endurance exercise intensity range into three different domains (Kindermann et al., 1979; Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Faude et al., 2009; Seiler, 2010; Jones et al., 2011, 2019) (Figure 6). At the bottom of this range, there is the ‘moderate-intensity domain’, which includes all running speeds below  $LT_1$  and marks the upper limit of a nearly exclusive aerobic metabolism. At these intensities, there is a mono-exponential increase in  $\dot{V}O_2$ , which generally reaches a steady-state within 3 min in healthy individuals (Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Seiler, 2010; Jones et al., 2011, 2019) (Figure 6). Exercise intensities within this domain allow exercising for hours and generally characterize recovery and regeneration runs (McLellan &



Skinner, 1981; Londeree, 1997; Jones, 2006). Exercise intensities between  $LT_1$  and  $LT_2$  correspond to the ‘heavy-intensity domain’ – also known as the ‘aerobic-anaerobic transition’ – and are characterized by elevated but constant blood lactate levels and delayed steady-state in  $\dot{V}O_2$  due to a short-lasting slow component, whose causes are still debated among exercise physiologists (Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Jones et al., 2011, 2019) (Figure 6). Exercise performed continuously in the heavy domain may last 45 min to approximately 4 hours, depending on the intensity (Urhausen et al., 1993; Baron et al., 2008; Fontana et al., 2009; Faude et al., 2009). Although anaerobic glycolysis is enhanced, some researchers speculated that training in the heavy domain may induce a considerable increase in the oxidative metabolism of muscle cells (Kindermann et al., 1979; Mader & Heck, 1986). Exercising at or near  $LT_2$ , commonly known as ‘threshold training’ or ‘tempo run’, is a very common training method to improve performance among coaches and athletes (Seiler, 2010; Stöggl & Sperlich, 2015; Kenneally et al., 2018, 2021a, b). Above  $LT_2$ , we have the ‘severe-intensity domain’, which is characterized by a constant increase in  $\dot{V}O_2$  and blood lactate levels. The former usually reaches its maximum within a few minutes, whereas the latter raises until exhaustion (Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Jones et al., 2011, 2019) (Figure 6). Since muscle fatigue limits exercise duration above  $LT_2$ , interval training represents the most common form of exercise in this domain (Faude et al., 2009; Bangsbo, 2015).



**FIGURE 6.** Representation of the pulmonary  $\dot{V}O_2$  (left panel) and blood lactate (right panel) response to constant-intensity exercise in the moderate-intensity (below  $LT_1$ ), heavy-intensity (between  $LT_1$  and  $LT_2$ ), and severe-intensity (above  $LT_2$ ) domains. The shaded areas in the left panel define the slow component of  $\dot{V}O_2$ , which occurs above  $vLT_1$  (Jones et al., 2011).

The specificity principle recommends training at intensities similar to those used during the competition (Hawley & Hopkins, 1995). However, the scientific evidence does not support this dogma (Hewson & Hopkins, 1996; Hawley, 2008; Beneke et al., 2011). Since runners cannot sustain exercise at  $LT_2$  for more than an hour (Urhausen et al., 1993; Baron et al., 2008; Fontana et al., 2009), many long-distance races are run at intensities that fall between the two thresholds (Costill, 1970; Costill et al. 1973; Farrell et al., 1979; Sjödin et al., 1982; Maughan & Leiper, 1983; Tanaka & Matsuura, 1984; Londeree, 1986; Bassett & Howley, 2000; Billat, 2001; Joyner & Coyle, 2008). We previously saw that training at heavy intensities is not always the best approach to develop the physiological determinants of distance running performance in already trained runners (Wenger & McNab, 1975; Daniels & Scardina, 1984; Wenger & Bell, 1986; Hill & Rowell, 1997; Midgley et al., 2006b, 2007b). Observational studies inform that successful distance runners typically perform the largest part (70–90%) of the overall training volume at moderate intensities (below  $LT_1$ ) and allocate little of their training volume around the  $LT_2$  (Billat et al., 2001b, 2003; Esteve-Lanao et al., 2005; Seiler, 2010; Enoksen et al., 2011; Stellingwerf, 2012; Tjelta et al., 2014; Stöggl & Sperlich, 2015; Kenneally et al., 2018, 2021a, b). Recreational runners also seem to spend most of their training at moderate intensities; however, they also spend a substantial part of their training at heavy intensities (Manzi et al., 2015). A possible reason for the limited training volume spent in the heavy domain by successful runners may be the negative effect that these training intensities exert on the autonomic and endocrine systems (Esteve-Lanao et al., 2007; Beneke et al., 2011; Muñoz et al., 2014; Manzi et al., 2015; Stöggl & Sperlich, 2015; Kenneally et al., 2018), other than the risk of accumulated glycogen depletion in the elite runner population (Beneke et al., 2011). While the latter may not be a serious concern for recreational runners given the much lower overall training volume, the former may explain the contradiction between the specificity principle and the training practice. In a randomized parallel study, Muñoz and colleagues (2014) did not observe statistically significant differences in 10-km race performance between two groups of recreational runners who spent the same absolute amount of time training at high intensities but gave different emphasis to moderate- and heavy-intensity training when using intention-to-treat analysis. On the contrary, Esteve-Lanao and colleagues (2007) observed significant improvements in 10.4-km cross country race performance following a reduction in the training volume performed at heavy intensities in 12 male sub-elite endurance runners, while maintaining a similar volume of severe-intensity training.

The ‘pyramidal’ approach to training intensity distribution (TID), in which runners progressively decrease the training volume from the moderate to the severe domain, and the

'polarized' approach, in which runners performed relatively high training volumes in the moderate (~80%) and severe (~20%) domains, with little or none in the heavy domain, represent the two most common TID schemes adopted by distance runners (Seiler, 2010; Stöggl & Sperlich, 2015; Kenneally et al., 2018). Both these schemes seem justified by the need of maintaining a large endurance base via moderate-intensity exercise (Sjödín & Svedenhag, 1985; Pate & Branch, 1992; Hawley, 1995; Hawley et al., 1997; Kubukeli et al., 2002), on which building the desired physiological characteristics using higher intensity exercise while minimizing the risk of overtraining (Esteve-Lanao et al., 2007; Beneke et al., 2011; Manzi et al., 2015; Stöggl & Sperlich, 2015). A recent meta-analysis conducted by Rosenblat and colleagues (2018) found that polarized TID may induce a larger improvement in performance than threshold-based training. While scientific evidence indicates both pyramidal and polarized approaches are superior to the threshold training for what concerns their ability to maximize physiological adaptations and performance (Esteve-Lanao et al., 2007; Muñoz et al., 2014; Stöggl & Sperlich, 2014; Kenneally et al., 2018), there is little consensus on which one of these two is the most efficacious (Stöggl & Sperlich, 2015; Kenneally et al., 2018). Several factors, such as the race distance, the different phases of the season, the training history of the runners, the method adopted to define the exercise intensity domains (or training zones), and the inter-individual variation in training response, may play a role in determining the efficacy of one TID scheme above the other (Muñoz et al., 2014; Sylta et al., 2014; Tjelta et al., 2014; Stöggl & Sperlich, 2015; Kenneally et al., 2018, 2021a, b; Bellinger et al., 2019). Stöggl (2018) suggested that professional endurance athletes may typically prioritize a high-volume low-intensity training during the preparation phase, a pyramidal TID during the pre-competition period, and a polarized TID during the competition phase. However, the recent findings by Kenneally and colleagues (2021a, b) partly disagree with this suggestion since they showed that, when using race pace to define the training zones, world-class long-distance runners adopt a pyramidal distribution across all phases of a 12-month season. It is worth mentioning that scientific evidence does not necessarily match training practice. Despite the superiority of pyramidal and polarized approaches to TID observed in the literature, anecdotal evidence from top-level coaches indicates that threshold-based training is a key component in structuring the world's best marathon performance when approaching competitions (Kenneally et al., 2018, 2021a, b).

Several methods exist to determine the amount of training spent in each physiological domain (or training zone) and TID across the domains/zones (Seiler, 2010; Sylta et al., 2014; Stöggl and Sperlich, 2015). These methods differ primarily in the parameter used to place the training intensity within a given zone [i.e., speed, heart rate, or rate of perceived exertion (RPE)]

(Bannister, 1991; Edwards, 1993; Foster et al., 2001; Lucia et al., 2003; Stagno et al., 2007; Manzi et al., 2009), in the metric used to quantify the training volume (i.e., time or km) (Seiler, 2010), and in the approach used to calculate the cumulative volume or dose performed in each training zone (i.e., 'time in the zone', 'session goal, or 'hybrid session-goal/time in the zone') (Sylta et al., 2014). Although there is no consensus on the best method to quantify the amount of training spent in each zone and determine TID, Manzi and colleagues (2015) observed that using the session-RPE – a method that quantifies the (internal) training load by multiplying the whole training-session RPE using the 10-point Borg category ratio (CR10) scale by its duration (Foster et al., 2001) – may overestimate the percentage of training spent in the zone 2 (corresponding to the heavy domain) and underestimate the percentage of training spent in the zone 3 (corresponding to the severe domain) compared with two different heart rate-based methods proposed by Edwards (1993) and Manzi and colleagues (2009) in recreational distance runners. However, the choice by Manzi and colleagues of using the same RPE-based thresholds from a previous study (Seiler & Kjerland, 2006) may have biased their results. Sylta and colleagues (2014) found that the 'time in the zone' approach to quantify TID, which is based on the percentage of time spent with heart rate within each intensity zone, overestimated the amount of training in zone 1 and 3, and underestimated the amount of training in the zone 2 compared with the 'hybrid session-goal/time in the zone' approach, in which the goal of the session informs in placing training time within given intensity zones, based on a combination of heart rate and speed data. Without neglecting the important distinction between internal and external training load (Borresen & Lambert, 2009; Foster, 2017; Mujika, 2017), it is important to note that the fact that heart rate slow component phenomenon may occur even during moderate-intensity exercise (Zuccarelli et al., 2018) may overestimate the amount of training performed in the higher zones when using heart rate-based methods compared with speed-based methods, regardless the approach used to quantify TID.

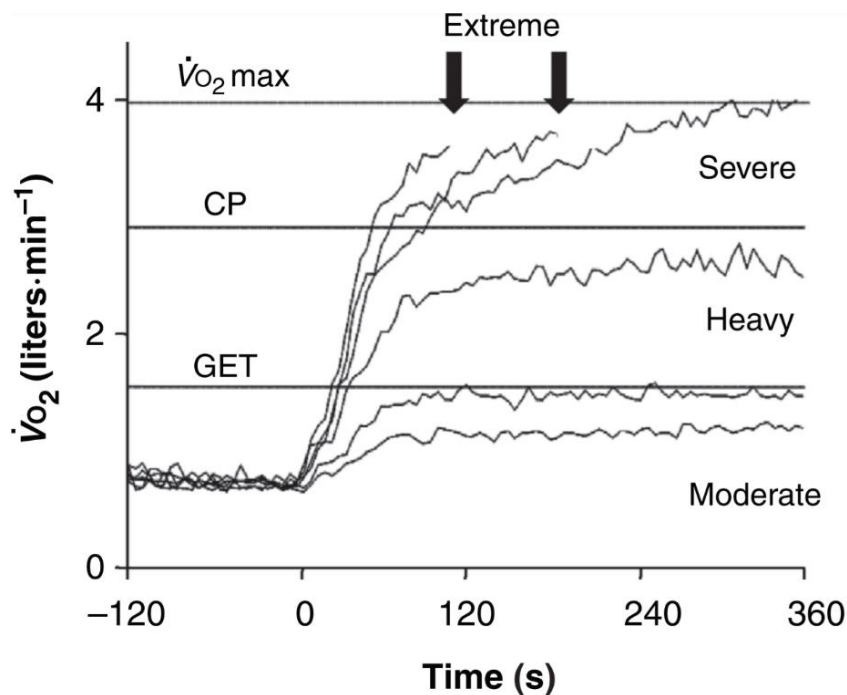
#### 1.4.3 The use of the Delta concept to prescribe high-intensity intermittent exercise

We saw in the previous sections that high-intensity running may be an efficacious approach to further stimulate different physiological adaptations in already trained individuals (MacDougall & Sale, 1981; Wenger & Bell, 1986; Priest & Hagan, 1987; Pate & Branch, 1992; Laursen & Jenkins, 2002; Midgley et al., 2006b, 2007b; Stögg & Sperlich, 2015). In this regard, a particular training method named 'high-intensity interval training (HIIT)', used by coaches and runners for almost a century has received great attention from exercise physiologists and sports

scientists in the last few decades (Billat, 2001). HIIT is generally defined as the repetition of intervals at an intensity greater than the  $vLT_2$  interspersed by recovery periods (MacDougall & Sale, 1981; Daniels & Scardina, 1984; Hawley et al., 1997; Billat, 2001; Laursen & Jenkins, 2002; Laursen, 2010; Buchheit & Laursen, 2013a, b; Tschakert & Hofmann, 2013). The rationale behind HIIT is to maximize the time spent at high intensity to stress the physiological systems used during exercise to a greater extent than required during the competition (MacDougall & Sale, 1981; Pate & Branch, 1992; Hawley et al., 1997; Billat et al., 2000; Billat, 2001; Laursen & Jenkins, 2002; Tschakert & Hofmann, 2013). Several studies supported the beneficial effect of HIIT on physiological adaptations and performance in the distance runner population (e.g., Acevedo & Goldfarb, 1989; Billat et al., 1999; Demarle et al., 2001, 2003; Slawinski et al., 2001; Lafitte et al., 2003; Smith et al., 2003; Denadai et al., 2006; Helgerud et al., 2007; Bangsbo et al., 2009; Kohn et al., 2011; Ferley et al., 2013; Muñoz et al., 2015).

The physiological adaptations induced by HIIT may depend on the intensity level (Buchheit & Laursen, 2013a, b). However, the classical three-domain/zone model described in the previous sections (Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Jones et al., 2011, 2019) cannot differentiate between the different forms of HIIT. In this regard, Hill and colleagues (2002) stressed the need for an additional, supra-severe exercise domain, in which intensity is so high that exercise will terminate before  $\dot{V}O_2$  reaches its maximal value (typically within 2 min) (Figure 7). The existence of this ‘extreme-intensity domain’ may in part explain why certain types of HIIT led to improvements in  $\dot{V}O_{2max}$  while others did not (Buchheit & Laursen, 2013a). Runners generally performed HIIT in the extreme domain as repeated maximal or near-maximal sprints. These forms of HIIT take various names – such as ‘sprint interval training’ (or ‘speed endurance production’) and ‘repeated sprint training’ (or ‘speed endurance maintenance’) – depending on the intensity and duration of the sprints and recovery periods (Iaia & Bangsbo, 2010; Buchheit & Laursen, 2013a, b; Bangsbo, 2015). Because of its very limited exercise duration, any form of HIIT in the extreme domain is unlikely to provide a substantial cardiorespiratory stimulus (Houston & Thomson, 1977; Daniels et al., 1978; Shepley et al., 1992; Billat et al., 1999; Bickham et al., 2006; Esfarjani & Laursen, 2007; Iaia et al., 2008, 2009; Bangsbo et al., 2009; Iaia & Bangsbo, 2010; Buchheit & Laursen, 2013a, b; Bangsbo, 2015; Skovgaard et al., 2017). Some researchers suggested that training in the extreme domain may still be useful to stimulate the neuromuscular and anaerobic characteristics (Buchheit & Laursen, 2013a, b; Bangsbo, 2015), which may, in turn, provide an important contribution during short distance races (Bulbulian et al., 1986; Houmard et al., 1991; Brandon, 1995; Paavolainen et al., 1999, 2000; Nummela et al., 2006; Joyner & Coyle, 2008;

Baumann et al., 2012; Bangsbo, 2015). However, only some studies observe increases in glycolytic enzyme levels after repeated sprints periods (Skovgaard et al., 2017), while others did not (Iaia et al., 2008; Bangsbo et al., 2009). That being said, a few weeks of repeated maximal or near-maximal sprints were capable to induce different muscle adaptations associated with improved fatigue resistance during intense short-term exercise (Sejersted & Sjogaard, 2000; Iaia et al., 2008; Bangsbo et al., 2009; Iaia & Bangsbo, 2010; Bangsbo, 2015; Skovgaard et al., 2017). As expectable, these findings were paralleled with an improved short-term exercise capacity (Iaia et al., 2008, Bangsbo et al., 2009; Skovgaard et al., 2017). Moreover, despite acute perturbation induced by the training session (Zavorsky et al., 1998), several studies observed improvements in RE and the speed associated with  $\dot{V}O_{2\max}$  ( $v\dot{V}O_{2\max}$ ) or the highest speed achieved at the end of an incremental treadmill test ( $V_{\text{peak}}$ ) following repeated sprint periods (Iaia et al., 2008, 2009; Bangsbo et al., 2009; Skovgaard et al., 2017; Koral et al., 2018). Finally, researchers observed improvements in 3000-m and 10000-m performance in trained distance runners after a few weeks of repeated sprints when performed together with a basic amount of aerobic training (Daniels et al., 1978; Esfarjani & Laursen, 2007; Bangsbo et al., 2009; Koral et al., 2015; Skovgaard et al., 2017).

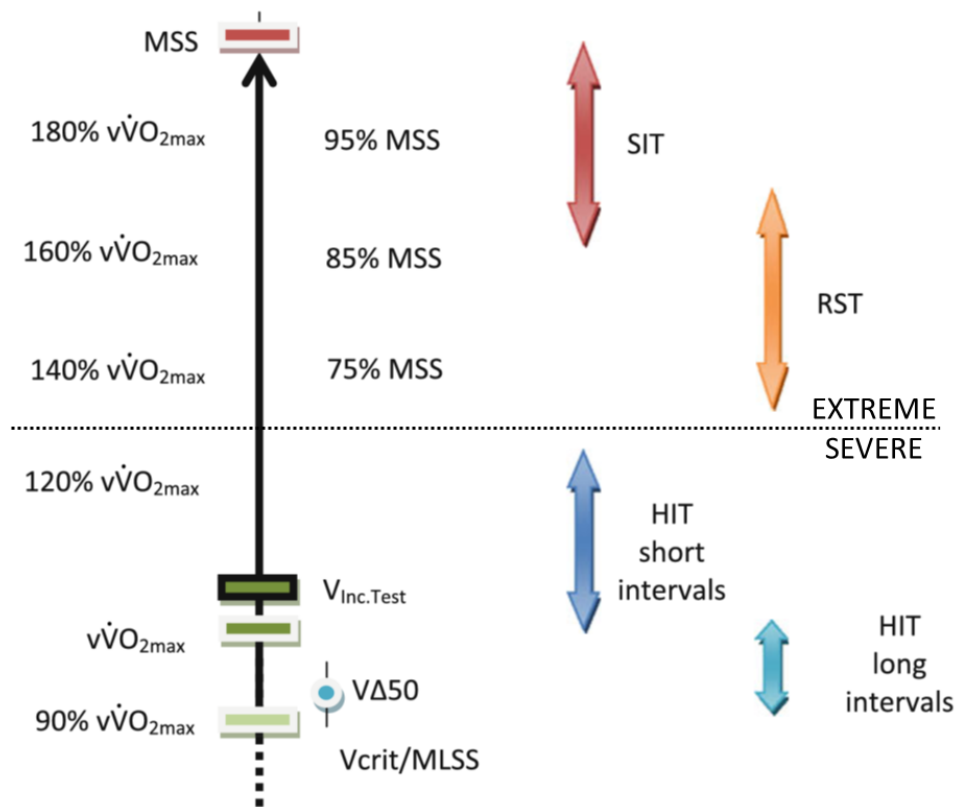


**FIGURE 7.** The four-domain model proposed by Hill and colleagues (2002). The exercise intensities in the extreme domain are so high that exercise terminates before the runners can reach  $\dot{V}O_{2\max}$ . The gas exchange threshold (GET) and the critical power (CP) reported in the figure serve the same function as the  $LT_1$  and  $LT_2$ , respectively, in separating the domains (Poole & Jones, 2012).

Although the four-domain/zone model may help to separate the forms of HIIT that potentially induce important cardiorespiratory stimuli from those that do not (Figure 8), it still fails to differentiate all the forms of HIIT. Moreover, the estimation of the lower boundary of the extreme domain requires several constant-intensity tests, therefore sharing the same feasibility issues as MLSS. These points may not be a problem for HIIT programs consisting of repeated maximal or near-maximal sprints, which can be easily prescribed without the need for exercise intensity frameworks (Buchheit & Laursen, 2013a, b). However, they may become critical when the aim is maximizing cardiorespiratory adaptations. Since spending a substantial amount of time at (or near)  $\dot{V}O_{2max}$  may play a key role in inducing  $\dot{V}O_{2max}$  development in already trained individuals (Wenger & Bell 1986; Laursen and Jenkins, 2002; Midgley & McNaughton, 2006; Midgley et al., 2006b; Buchheit & Laursen, 2013a, b), ensuring that the exercise intensity stays within the severe domain for the whole duration of the training period becomes a prerequisite for a successful HIIT program aimed to maximize cardiorespiratory adaptations during certain phases of the training season. Although using  $v\dot{V}O_{2max}$  (or  $V_{peak}$ ) may seem an intuitive way to accomplish this purpose (Billat & Koralsztein, 1996; Hill & Rowell, 1996; McLaughlin et al., 2010), this approach may not necessarily be the most effective one to maximize cardiorespiratory adaptations in trained distance runners (Hill & Rowell, 1997; Hill et al., 1997; Billat et al., 1999; Smith et al., 1999, 2003; Danadai et al., 2006). Most of the studies that used  $v\dot{V}O_{2max}$  (or  $V_{peak}$ ) for HIIT prescriptions observed significant improvements in RE and  $v\dot{V}O_{2max}$  (or  $V_{peak}$ ) in recreational and trained distance runners (Esfarjani & Laursen, 2007; Billat et al., 1999; Smith et al., 1999; Danadai et al., 2006). These studies also observed consistent improvements in 3000-m performance and less consistent improvements in 5000-m performance in these populations (Esfarjani & Laursen, 2007; Billat et al., 1999; Smith et al., 1999, 2003; Danadai et al., 2006). However, significant improvements in  $\dot{V}O_{2max}$  were observed in recreational runners (Esfarjani & Laursen, 2007) but not in trained runners (Billat et al., 1999; Smith et al., 1999, 2003; Danadai et al., 2006). Since the total time spent at (or near)  $\dot{V}O_{2max}$  is longer at submaximal speed than when running at  $v\dot{V}O_{2max}$  or  $V_{peak}$  (Hill & Rowell, 1997; Hill et al., 1997; Demarie et al., 2000), near to maximal aerobic speeds may be more efficacious than  $v\dot{V}O_{2max}$  or  $V_{peak}$  in stimulating  $\dot{V}O_{2max}$  adaptations in the distance runner population (Buchheit & Laursen, 2013a, b).

At the beginning of the century, Billat and colleagues introduced the ‘Delta concept’ as a means to differentiate HIIT within the severe domain (Demarie et al., 2000) (Figure 8). The Delta concept uses  $vLT_2$  (or MLSS) and  $v\dot{V}O_{2max}$  (or  $V_{peak}$ ) as physiological boundaries for defining near-to-maximal aerobic intensities (Demarie et al., 2000). Particularly, using  $v\Delta 50$  –

which is the speed halfway between  $vLT_2$  (or MLSS) and  $v\dot{V}O_{2max}$  (or  $V_{peak}$ ) – for HIIT prescriptions was shown to allow runners to accumulate more than 10 min at (or near)  $\dot{V}O_{2max}$  before exhaustion (Demarie et al., 2000). This duration is considered a potent stimulus for cardiorespiratory adaptations (Buchheit & Laursen, 2013a, b).



**FIGURE 8.** The intensity ranges typically used for the different forms of HIIT. Although the inclusion of the extreme domain in the classical thresholds-based exercise intensity model may help to separate the forms of ‘aerobic’ HIIT from the forms of ‘anaerobic’ HIIT, it cannot differentiate all the forms of HIIT. On the contrary, the Delta concept (expressed as  $v\Delta 50$  in the figure) allows to precisely manipulate HIIT intensity for intensities at (or near)  $\dot{V}O_{2max}$  and, altogether with the knowledge of the boundary of the extreme domain, within the whole severe domain (modified from Buchheit & Laursen, 2013a).

Similar to higher-intensity forms of HIIT, prescribing HIIT using the Delta concept showed to improve RE and  $v\dot{V}O_{2max}$  (or  $V_{peak}$ ) in trained distance runners (Demarle et al., 2001, 2003; Slawinski et al., 2001; Lafitte et al., 2003). However, despite the sound rationale for cardiorespiratory adaptations, most of the interventional studies that used the Delta concept to prescribe HIIT failed to observe improvements in  $\dot{V}O_{2max}$  in this population (Demarle et al., 2001, 2003; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003). There are several possible explanations for these results. All the studies that failed to observe changes in  $\dot{V}O_{2max}$



did not use a control group. This fact makes it very difficult to separate the net effect of the HIIT intervention from the effect of the whole training program and seasonal variation (Hecksteden et al., 2018). Moreover, the very small sample size characterizing these studies (i.e., 6 to 8 runners) may have affected the statistical power of the inferential tests. The fact that the same research group conducted all these studies also restricted the cohort of individuals from which the effect of this form of HIIT was investigated and thus the generalizability of the findings to the distance runner population. Randomized controlled studies conducted on distance runners or mixed cohorts of endurance athletes from different research groups observed significant improvements in the  $\dot{V}O_{2max}$  following near to maximal HIIT programs (Helgerud et al., 2007; Stöggl & Sperlich, 2014). Although these studies did not adopt the Delta concept to prescribe HIIT, they suggest that the Delta concept can be useful to maximize  $\dot{V}O_{2max}$ . These studies also support the use of intensities near  $\dot{V}O_{2max}$  to induce cardiorespiratory adaptations in distance runners.

## **1.5 Scientific research vs training practice: how should distance runners train?**

### **1.5.1 Physiological and race pace approach: two faces of the same medal**

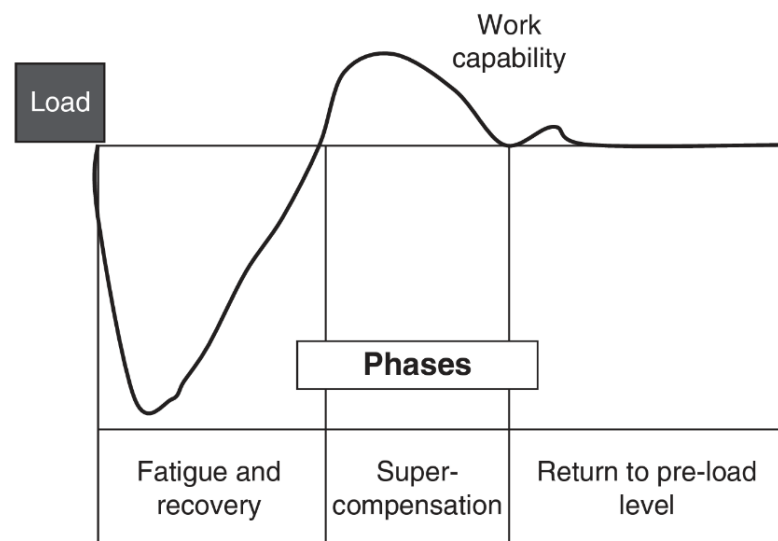
In the previous chapters, we saw that highly-trained and elite distance runners favor two specific TIDs (i.e., pyramidal and polarized), which may vary depending on the phase of the season, and that several methods exist to determine TID across the different physiological domains or training zones (Seiler, 2010; Sylta et al., 2014; Stöggl and Sperlich, 2015). An important aspect we should consider when interpreting the available evidence about the optimal TID for distance runners relates to the suitability of the underpinning model for our purpose. By taking into consideration both scientific evidence and world-class training practice, Kenneally and colleagues (2018) recently proposed a new perspective on how to approach and interpret TIDs in distance runners, whose optimization may ultimately depend on the training goal, phase of the season, race distance, and runners' characteristics, and where different approaches to training prescription may best fit different scenarios. The authors initially suggested that the anchor used to prescribe training intensity may determine the target of adaptations of a given training program. Specifically, Kenneally and colleagues speculated that the 'physiological approach' to exercise prescriptions, which defines the training zones by anchoring the training speeds to the classical exercise-intensity domain model based on a threshold or turn point

concepts (Kindermann et al., 1979; Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Faude et al., 2009; Jones et al., 2011, 2019), may be the optimal choice whenever the aim is to isolate, and thus maximize specific physiological adaptations (e.g.,  $\dot{V}O_{2\max}$ ). On the contrary, prescribing exercise according to given percentages of the athletes' event-specific target race pace – the so-called 'race pace approach' – may be more effective to improve performance by providing the perfect stimulus for the concurrent development of the physiological processes involved in that task (Kenneally et al., 2018). Kenneally and colleagues later observed that the two approaches may lead to differences in both the percentage of the overall training volume spent within each intensity zone (Kenneally et al., 2021a) and in the type of TID across the different phases of the season (Kenneally et al., 2021b) in a group of world-class long-distance runners. The authors concluded that both approaches may lack sensitivity to detect changes in TID and proposed an integrated approach, in which scientists, coaches, and/or athletes collect physiological data longitudinally, and use these data to create individual 'signature' physiological profiles that match with specific race performances (Kenneally et al., 2021a). Although the latter proposal by Kenneally and colleagues sounds appealing, the fact that the choice of the percentages of race pace used for analysis were rather arbitrary limits the power of data interpretation. Moreover, difficulties in properly isolating and modeling individual responses for single individuals (Hecksteden et al., 2015) may also limit the use of the integrated approach proposed by Kenneally and colleagues in everyday training practice. Noteworthy, the authors did not conduct any investigation to verify their initial hypothesis (i.e., prescribing exercise according to given percentages of race pace may induce different outcomes than exercise prescriptions based on the classical exercise-intensity domains), thus leaving an open question on the ability of each training prescription approach to maximize its benefits depending on the type of outcome (i.e., physiological adaptations or race performance).

### 1.5.2 Network physiology: the forgotten piece of the puzzle

Although more complex paradigms have been recently proposed (Kiely, 2018), the classical stress theory (Cannon, 1929; Selye, 1936, 1978) represents a major influencer of sports training theory to date. The General Adaptation Syndrome (GAS) (Selye, 1978), in particular, is often referred to as the starting point to understand how humans adapt to the stress of training (Virus, 2002). This conceptual framework is based on the primacy of homeostasis maintenance and defines three sequential stages of the stress response: first alarm, then resistance, and, if the stress is overwhelming, exhaustion (Selye, 1978). Sports training theory has taken the GAS as

evidence that mechanical training load (determined by the physical training parameters such as volume, intensity, and frequency) directly dictates the direction and magnitude of subsequent physiological adaptations, which take place during the (later-termed) ‘supercompensation’ phase (Virtanen, 2002) (Figure 9). On this theoretical platform – that is, the supercompensation model – many coaches have developed various training strategies to create a balance between training load and recovery to target the desired adaptations with the proper timing (Issurin, 2010, 2016).



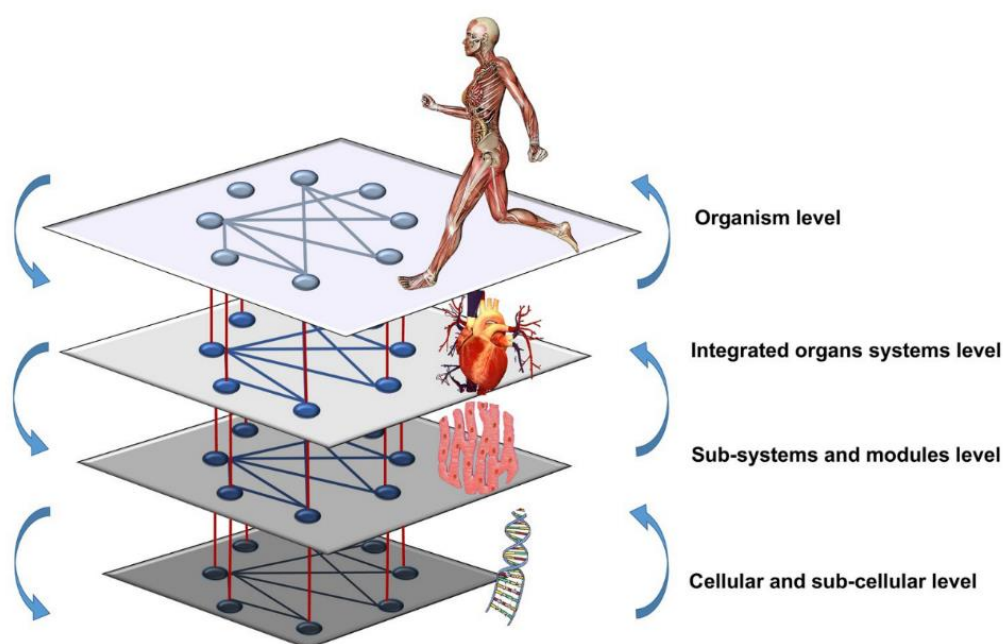
**FIGURE 9.** Representation of the supercompensation cycle following a single training load. The training load represents the physiological stressor imposed on the athlete, the (acute) fatigue after the load represents the alarm phase, and the supercompensation represents the (transitional) resistance phase characterized by an improved work capability. If no further stimuli occur, the work capability will return to the pre-load level (detraining). Exhaustion is not considered in this figure (from Issurin, 2010).

Despite the fundamental role that GAS has played in guiding exercise prescriptions as a function of maximizing physiological adaptations in the long term, the model – as it was originally conceived – is inadequate to explain and predict changes in performance in response to training prescriptions (Bannister et al., 1975). The main reason for this conceptual failure relates to the multi-level structure of human performance (Renfree & Casado, 2018; Venhorst et al., 2018a), which cannot fit in the single-level structure of the GAS model. We previously saw that success in distance running events may depend on the optimal dynamic interplay between different factors (Renfree & Casado, 2018; Venhorst et al., 2018a). Even when limiting to the solely physiological aspects, several factors are involved in distance running performance

(Bassett & Howley, 1997, 2000; Joyner & Coyle, 2008). Although the relative importance of the individual physiological predictors across the different running distances has been estimated and some integrative models have been presented (Sparling, 1984; Sjödín & Svedenhag, 1985; Di Prampero et al., 1986; Joyner, 1993; Coyle, 1995, 1999; Bassett & Howley, 1997, 2000; Joyner & Coyle, 2008; McLaughlin et al., 2010), we are still far from precisely determining the optimal combination of such determinants for any given running distance. Treating any of these factors independently without acknowledging the possible – but still largely unknown – interactions between the different factors [i.e., to what extent the development of one factor (e.g.,  $\dot{V}O_{2max}$ ) may affect the development of the others (e.g., RE) (Joyner, 1991)] may not be the best training strategy when the aim is maximizing endurance performance as a whole. There is solid evidence showing that distance running performance is better explained by the combination of the different physiological predictors than by the single factors alone (Sparling, 1984; Sjödín & Svedenhag, 1985; Di Prampero et al., 1986; Joyner, 1993; Coyle, 1995, 1999; Bassett & Howley, 1997, 2000; Joyner & Coyle, 2008; McLaughlin et al., 2010). This means that, although training programs targeting the development of specific (but isolated) physiological adaptations can be built on the GAS framework, the latter cannot be used to guide programs to maximize overall performance development, unless some framework expansion is done. It is worth mentioning that exercise physiologists and sports scientists have already developed and expanded the original GAS framework into performance-oriented multi-component models (Bannister et al., 1975, 1999; Calvert et al., 1976; Morton et al., 1990, 1991, 1997; Busso et al., 1994, 1997, 2002; Mujika et al., 1996; Avalos et al., 2003; Busso, 2003). However, these models often lack sufficient precision to explain and predict changes in performance in the athlete population (Hellard et al., 2006).

Modern physiological science has recognized the importance of the interaction(s) between the different physiological systems in successfully executing a task while ensuring a certain level of safety for the systems involved – a concept that fairly resembles the homeostasis (Bashan et al., 2012; Balagué et al., 2020; Ivanov, 2021). Particularly, the new field of network physiology acknowledges the human organism as an ‘integrated network of multi-component physiological systems, each with its regulatory mechanism, continuously interact both horizontally and vertically through circular causality to coordinate their functions’ (Balagué et al., 2020) (Figure 10). Coordinated network interactions among systems and organs are thus fundamental for determining the different physiological states and maintaining health (Balagué et al., 2020). Within this new framework, any physical task can be viewed as a whole-body stressor that activates the integrated network, where the single physiological components are

stressed only to the extent that is required to perform the task, at a given network state. The implications of applying this framework to exercise physiology and sport science scenarios are straightforward. Since a certain amount of training stress is needed to induce supercompensation and evoke adaptations, a given physical task may at the same time represent an adequate training stimulus for the bodily functions and parameters that represent the limiting factors for that task and an insufficient stimulus for those physiological components that are already sufficiently developed to sustain that task. Regardless of the impact on the single physiological components, the final result would be improving the ability of the integrated network (i.e., the human organism) to perform that task. Classical stress theory and modern network physiology provide the theoretical rationale to hypothesize that the approach used to prescribe the training can be oriented towards maximizing specific physiological determinants or performance as a whole, as proposed by Kenneally and colleagues (2018).



**FIGURE 10.** The vision on network physiology of exercise is explained in a diagram: hierarchically organized physiological network levels interact both horizontally and vertically through circular causality to coordinate their functions (from Balagué et al., 2020).

### 1.5.3 The right training, at the right time, in the right way

The points raised in the previous sections stress an intuitive but very important point, which has been well known by coaches for decades: the optimal training may ultimately depend on the given training goal for a given phase of the season. This point agrees with previous

recommendations from researchers (Stöggl and Sperlich, 2015; Stöggl, 2018) about manipulating training intensity, volume, and frequency – and thus prioritizing different TIDs – depending on the training phase. The novelty involves the possibility of optimizing training preparation by choosing the approach used for training prescriptions depending on the training goal. Specifically, coaches may prescribe based on the physiological characteristics of their athletes when they want to emphasize the development of one or more isolated physiological factors (e.g.,  $\dot{V}O_{2max}$ ) during the preparatory phase of the season. Alternatively, they may prescribe according to the race pace when they want to maximize performance when approaching competitions (Kenneally et al., 2018).

There are still several points that need to be considered and open questions, though. While essentially straightforward for the exercise prescriptions in the moderate and heavy-intensity domains, the use of the physiological approach may become complicated to prescribe HIIT sessions in the severe domain. Adopting this approach for HIIT prescriptions may indeed require multiple tests in the lab at any new training prescription (Demarle et al., 2001, 2003; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003). Reliable and valid portable versions of lab devices exist on the market (Bonaventura et al., 2015; Perez-Suarez et al., 2018); however, their use is still mostly limited to research purposes and this has possibly prevented widespread use of the delta approach among coaches and non-professional athletes. Although the critical speed (CS) can be used to set the minimum training intensity during HIIT (Gaesser & Poole, 1996; Jones & Poole, 2005; Burnley & Jones, 2007; Jones et al., 2011, 2019), several exhaustive exercise bouts are typically required to obtain an accurate estimate of this value (Jones & Vanhatalo, 2017; Jones et al., 2019). This may limit the use of the CS to adjust repeatedly HIIT intensity across subsequent mesocycles. While this may not be an issue for professional runners, adopting the physiological approach to HIIT prescription is simply not an option for the vast majority of the distance runner population, which often use extremely standardized approaches to HIIT, in which groups of runners train together. This standardized approach is often the preferred choice among many runners because it requires no equipment, little expertise, and can be prescribed quickly in any setting. Moreover, it is the only form of HIIT that can be performed together with teammates and companions (Casado et al., 2019). However, since the speed is the same for everyone, runners who perform intermittent exercise below severe intensity, either as a consequence of an overestimation of their training level when choosing the group or of an improvement in the CS during the training period, may face blunted – if not null – gains in  $\dot{V}O_{2max}$ , leading to a smaller mean training response and a larger heterogeneity of training effects. That being said, given the lower training level compared with

professionals, the overall impact of this practice may be trivial, and recreational and trained runners may still achieve similar benefits in terms of physiological adaptations when using standardized HIIT approaches as to when using individualized ones. If this were not the case, a very important missing piece of information would be the identification of valid measures, obtainable using accessible devices, that would allow most of the runners to perform HIIT according to the physiological approach without requiring repeated visits to the labs. In this regard, Billat and colleagues (Billat, 2001, Billat et al., 2002) proposed to use the speeds on given race distances to approximate physiological velocities. Alternatively, we may look for treadmill-related measures that can be obtained without the need for a metabolic chart and/or lactate analyzer. Another important point requiring investigation is whether the relation between physiological parameters and race pace across individuals is indeed sufficiently weak to make the distinct adaptations following the physiological approach and the race pace approach plausible. If so, the last point would be obtaining an empirical validation of the effects of the two approaches for HIIT prescription on the desired training outcomes.

Other than identifying meaningful research questions, it is also important to remember that data do not speak by themselves and no conclusion can be drawn without a solid statistical framework. Particular attention within this context should be posed to the different families of null hypotheses. Exercise physiology and sport science have traditionally made use of the null hypothesis of no difference to make decisions about experimental interventions. The traditional null-hypothesis test allows making informed decisions about whether an experimental intervention is superior to an inactive or reference intervention while controlling the Type I and Type II error rates. However, the traditional null hypothesis may not be the most informative choice in all research scenarios and other families of hypotheses may be more appropriate when researchers are interested in whether the experimental intervention is superior to control by more than a prespecified amount – the smallest effect size of interest (SESOI) –, two interventions are similar in efficacy, or a given intervention is not unacceptably worse than a standard one with no restriction for its maximal efficacy. These research questions acquire particular relevance whenever two training interventions substantially differ with respect to factors such as cost-effectiveness, invasiveness, or administrative procedures (Hecksteden et al., 2018). However, such methods are not widespread in exercise physiology and sport science yet, a fact which limits the tools that researchers may use to design their studies and often limits the power of data interpretation. Therefore, clarifications are needed for helping exercise physiologists and sports scientists to understand the valid statistical methods for those research questions.

## **1.6 Moving beyond the traditional null-hypothesis: the use of equivalence and non-inferiority tests for interventional studies in exercise physiology and sport science**

An often-overlooked aspect when designing and analyzing interventional studies in exercise physiology and sport science concerns the type and direction of the research hypothesis(es) (Caldwell & Chevront, 2019). Most studies use the null hypothesis of no effect when making decisions about experimental interventions. That is, researchers usually examine whether there is a statistical difference between the experimental and the control group on one or more primary outcomes. However, other hypothesis tests may be more appropriate when researchers are interested in whether two interventions are similar in efficacy but substantially differ with respect to factors such as cost-effectiveness, invasiveness, or administrative procedures (Hecksteden et al., 2018). The correct approach to designing and analyzing interventional studies in exercise physiology and sport science continues to be extensively discussed in the literature (Hopkins et al., 1999; Hecksteden et al., 2018; Mansournia & Altman, 2018; Caldwell & Chevront, 2019). Recently, several researchers have recommended complementing the traditional null hypothesis with tests of equivalence and non-inferiority, which evaluate whether two interventions or conditions are similar or do not differ by more than a given amount (Dixon et al., 2018; Caldwell & Chevront, 2019; Aisbett et al., 2020). I will review and expand the statistical toolset that can be used by exercise physiologists and sports scientists when designing and analyzing interventional studies. I will refer to the best practices as developed in biomedical, social, and behavioral research since I recognize sufficient similarities with exercise physiology and sport science regarding the design of interventional studies. To increase understanding by exercise physiologists and sports scientists, I will also provide two worked examples from exercise physiology and sport science research that highlight how typical research designs and analyses conducted using traditional null hypothesis tests could be re-imagined using equivalence or non-inferiority tests. Moreover, I will provide theoretical and practical recommendations to exercise physiologists and sports scientists who would like to apply the different hypothesis tests in future research.

### **1.6.1 Investigating statistical differences (superiority)**

Unless otherwise specified, most interventional studies in exercise physiology and sport science have the implicit aim of determining if the efficacy of a given intervention is superior, or



possibly inferior, to placebo, sham, or reference intervention. In the most common study design, researchers randomize participants to either an experimental or a control group. The observed difference in group means after the intervention period (i.e., the effect size) is used to perform a hypothesis test examining a difference in population means. Following traditional null hypothesis testing, a difference between interventions can be concluded, while controlling the Type I error rate, whenever the  $p$ -value calculated from a particular test statistic indicates the observed or more extreme data are surprising (i.e., the  $p$ -value is less than or equal to the significance level, or  $\alpha$ ), assuming there is no difference between the interventions and all other modeling assumptions are met. Alternatively, researchers can choose a confidence interval (CI) approach. Confidence intervals can be used to inspect and interpret the point estimate and the lower and upper limits of the interval in relation to effects of practical importance. Thus, a properly derived CI can also be used to evaluate superiority or any other family of null hypotheses (Bauer & Kieser, 1996). The two approaches lead to identical decisions in a hypothesis test, as  $p$  is less than or equal to .05 when a 95% CI excludes the value that is tested against (i.e., zero for the traditional null hypothesis) (Figure 11a – first example).

Regardless of the inferential approach employed, investigating differences between interventions without taking into consideration any meaningful value does not permit informed decisions regarding the practical significance of the outcome(s). From an exercise physiology and sport science perspective, testing the superiority of the experimental intervention against an effect size that is exactly zero may increase the risk of endorsing interventions, such as exercise training protocols or nutritional strategies, that are expensive, demanding, or time-consuming, but have no practical benefit – that is, they do not provide a noticeable advantage over an existing benchmark. For adequately powered tests (i.e., 80–90% power), testing data against the *nil* (zero) effect using the SESOI – which should be defined a priori and justified on sound grounds – as a target mean difference may lead to concluding efficacy for effects as low as the 60–70% of SESOI, the so-called ‘decision value’ (Chuang-Stein et al., 2010; Roychoudhury et al., 2018). Chuang-Stein et al. (2010) recommended this approach as a reasonable compromise between desirability and feasibility, stressing how this approach also acknowledges the impact of sampling variation in reducing the observed intervention effect. However, although an effect as low as 60–70% of SESOI may be observed when the true effect equals the SESOI, the opposite may not necessarily be true. By rearranging the equation used to determine the decision value, it is possible to obtain an adequate sample size that leads to rejecting the null hypothesis when the decision value equals at least the SESOI. In this way, statistical significance is ensured whenever practical relevance is observed (Figure 11a – second

example). For a deeper insight into the statistical aspects of this approach – named ‘dual-criterion designs’ – I refer the readers to Roychoudhury et al. (2018). An even more conservative criterion for assessing superiority consists of determining whether the mean difference, after having considered its uncertainty, is larger than the SESOI (Lakens, 2021) (Figure 11a – third example). This approach leads to the same conclusions as testing the shifted (non-zero) null hypothesis (Victor, 1987) or a ‘*minimum-effect test*’, whose null hypothesis assumes that the mean difference between the interventions falls within a range of practically irrelevant values (Murphy et al., 2014). However, raising the standard of evidence to claim superiority comes at a cost. Testing data against the SESOI may require sample sizes that are prohibitively large when the ‘true’ effect size is close to the SESOI unless prespecifying unrealistically large effect sizes with an attendant risk of Type II error (Gelman & Carlin, 2014). Therefore, researchers should decide very carefully what standard of evidence they want to achieve for intervention efficacy when designing their studies, taking into account the implications of their findings and their resources.

Although the definition of SESOI is self-explanatory, exercise physiologists and sports scientists should be aware that several different methods exist to determine this value, depending on data and applications (Cook et al., 2018; Lakens, 2021). The ‘anchor-based’ method, which uses the researcher’s judgment, participant’s experience, or clinical endpoint(s) to define the SESOI, provides a common approach to interpret study outcomes in clinical research. In this field, the SESOI – also known as the minimal clinically important difference – is often determined by examining the association between a certain change in an outcome variable and a meaningful change in a (hard) clinical outcome from prospective epidemiological data or randomized controlled trials. The expert panel approach, also known as the Delphi method, is an alternative (although not necessarily straightforward) way to define the SESOI based on expert consensus. Previous studies may give an indication of the expected effect sizes. However, researchers should be aware that due to publication bias published effect sizes often overestimate the true effect of interventions, and that the distribution of effect sizes observed in literature does not necessarily inform about the SESOI, whose determination needs careful consideration and justification. Cohen’s classical benchmarks (Cohen, 1988), developed for the social and behavioral sciences, are not recommended as guidance on identifying the SESOI in exercise physiology and sport science since an effect size of interest is context-dependent and should be decided based on a substantive research question (Caldwell & Vigotsky, 2020). Although some authors (Hopkins et al., 1999; Rhea, 2004) have developed scales for assessing the magnitude of effect sizes in some specific areas of exercise physiology and sport science,

researchers should be aware that determining the SESOI is not a straightforward process, and it may be challenging in many sporting and physiological contexts.

Interpreting inconclusive evidence for superiority, or interpreting failure to reject the null hypothesis, as evidence for the equality of two interventions, is a common misconception (Altman & Bland, 1995). A statistically non-significant result (e.g.,  $p > .05$ ) cannot be interpreted as the absence of an effect (Figure 11a – fourth example). To be able to conclude an effect is absent, one needs to specify the alternative hypothesis explicitly, and perform a test that statistically rejects the alternative hypothesis. The traditional null hypothesis testing only rejects the null hypothesis, and, especially in small studies, a non-significant result is not informative about whether the alternative hypothesis can be rejected. Exercise physiologists and sports scientists must keep in mind that no correct conclusions other than superiority or inferiority can be drawn using traditional hypothesis tests. Because a well-designed study is informative about both the presence and absence of an effect of interest, researchers should consider complementing traditional null hypothesis tests with equivalence and non-inferiority tests.

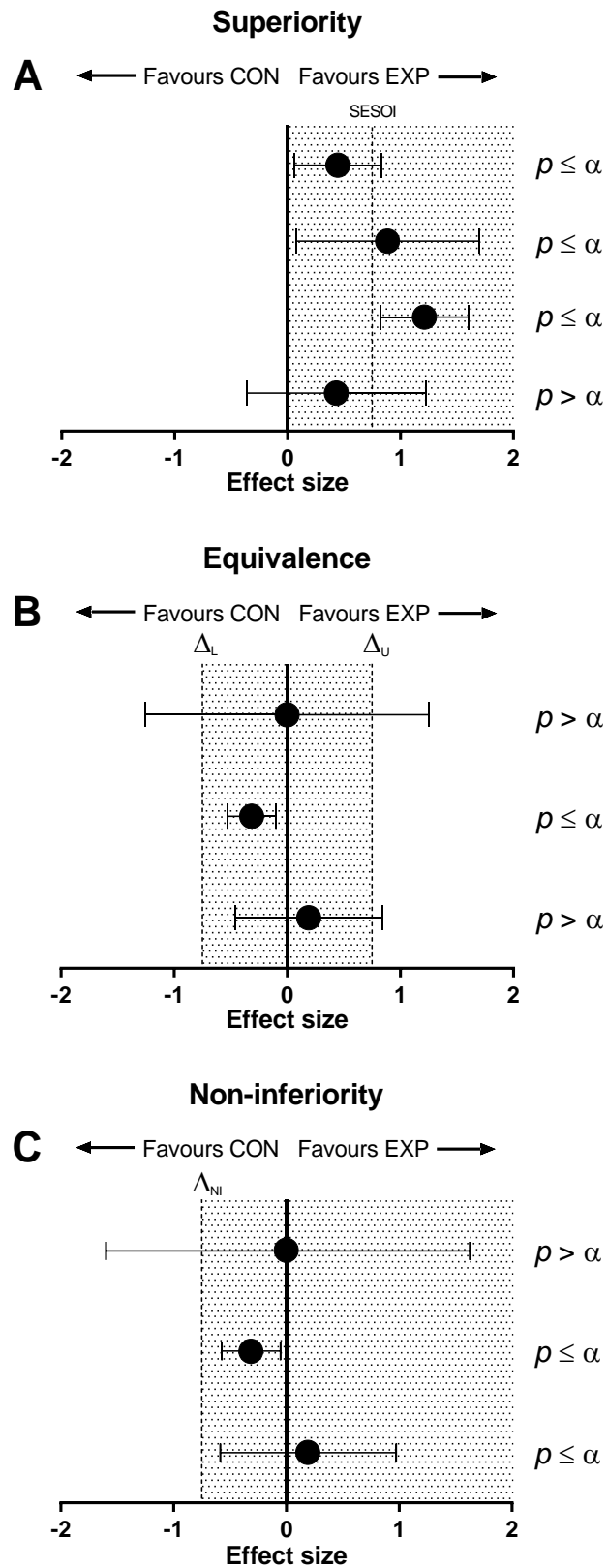
### 1.6.2 Investigating equivalence and non-inferiority

Proving that two interventions or conditions are perfectly equal in efficacy is impossible from a statistical standpoint. What is possible in a statistical test is to reject the presence of a difference that is large enough to be practically relevant, defined by the upper ( $\Delta_U$ ) and lower ( $\Delta_L$ ) equivalence margins (Hodges & Lehmann, 1954; Lakens, 2017). Although various approaches exist to perform an equivalence test (Meyners, 2012), equivalence is typically investigated via the ‘two one-sided tests’ (TOST) procedure, which is a simple variation of a traditional hypothesis test (Schuirmann, 1987). In this procedure, the null and alternative hypotheses within each set are reversed and data are tested against  $\Delta_U$  and  $\Delta_L$  in two one-sided tests, each carried out at the  $\alpha$  level (conventionally set to .05 or even to .025 in some regulated settings). Equivalence can be concluded at the  $\alpha$  level only if both tests statistically reject the presence of effects equal to or larger than the equivalence margins. It is common to report only the greater  $p$ -value of the two one-sided tests when testing for equivalence since this  $p$ -value is also the one for the overall equivalence test (Berger & Hsu, 1996). The TOST procedure is operationally identical to concluding equivalence whenever the two-sided  $100(1 - 2\alpha)\%$  CI for

the mean difference between the interventions lies entirely within the equivalence margins (Westlake, 1981; Schuirmann, 1987) (Figure 11b – second example).

Equivalence studies are very common in clinical research, in which new drug formulations or generic versions of the product are often compared to brand-name pharmaceuticals to prove bioequivalence (Senn, 2021). Moreover, this design has attracted growing interest in the social and behavioral sciences for its utility in evaluating replication results and corroborating risky predictions (Lakens, 2017; Lakens et al., 2018a). The latter application of equivalence hypotheses may also make them valuable for exercise physiology and sport science, which suffers from a shortage of replication experiments (Halperin et al., 2018). Nevertheless, until recently, investigating equivalence did not appear to be a common practice among exercise physiologists and sports scientists, who have so far restricted the use of equivalence tests mostly to measurement agreement research as an alternative or complementary approach to the Bland–Altman method (Dixon et al., 2018).

If there is an interest, along with a solid rationale, to investigate whether a given intervention is not unacceptably worse than a standard one with no restriction for its maximal efficacy, researchers can opt for a non-inferiority study. This is usually the case when the new intervention has better cost-effectiveness, is safer, is easier to implement, or is less demanding than the standard. Non-inferiority studies can also be useful to evaluate modifications to well-established interventions and extend applicability to special populations. These research questions may also apply to exercise physiology and sport science. In non-inferiority testing, the non-zero null hypothesis is shifted towards the negative side of the nil effect, favoring the standard. It follows that, when applying the CI approach, non-inferiority is conventionally concluded when the lower margin of the two-sided 95% CI for the mean difference between the interventions lies above the non-inferiority margin ( $\Delta_{NI}$ ) (Senn, 2021) (Figure 11c – second and third examples).



**Figure 11** Testing for superiority, equivalence, and non-inferiority within a typical parallel-group design. The error bars indicate the 95% confidence interval (CI) in relation to the traditional null-hypothesis test (Figure 11a) and non-inferiority test (Figure 11c) and the 90% CI in relation to the two

one-sided test procedure (Figure 11b). The shaded areas indicate the rejection region for each hypothesis test. **Figure 11a** From a traditional perspective (i.e., deciding on the presence of an effect), the superiority of the experimental group (EXP) compared with the control (CON) can be concluded in the first three examples. However, the standard of evidence to claim superiority differs between the examples. In the first example, it is only possible to reject effects that are smaller than zero. In the second example, it is also possible to claim practical importance besides statistical significance. In the third example, it is possible to reject any effect that is not practically important – that is, an effect that is smaller than the smallest effect of interest (SESOI). Superiority cannot be concluded in the lower example, since the 95% CI extends beyond zero, which reflects in a  $p$ -value larger than  $\alpha$ . **Figure 11b** It is possible to conclude equivalence between the interventions only in the second example since in the upper and lower example the 90% CI spans beyond the lower ( $\Delta_L$ ) or the upper ( $\Delta_U$ ) equivalence margin. **Figure 11c** The observed data are identical to Figure 11b. Despite the wider CI, the absence of an upper margin allows concluding non-inferiority in both the second and third examples.

Compared with classical parallel-group studies, the design and analysis of non-inferiority studies face several additional methodological challenges, which include the suitability of the reference intervention, the determination of the  $\Delta_{NI}$ , and sample size estimation. I will briefly review and discuss the main aspects of each of these challenges in the following sections. Since some of these issues also apply to equivalence studies, I will expand those parts where relevant.

### 1.6.3 Suitability of the reference intervention

From a clinical perspective, the non-inferiority of an experimental intervention can be firmly concluded only when compared to a reference intervention of well-established efficacy (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 1998, 2001; Committee for Proprietary Medicinal Products, 2000; Committee for Medicinal Products for Human Use, 2005). The design characteristics of the reference intervention (population selection, intervention protocol, primary outcome measures, etc.) should be replicated as closely as possible to reduce the risk of violating the ‘constancy assumption’, which requires consistency between the effect of the reference group in the new study and the historical effect estimated from the literature. Violating this assumption may increase the chances of incorrectly concluding non-inferiority for ineffective or even harmful interventions.

When considering the extreme paucity of replication experiments (Halperin et al., 2018), along with the small sample sizes characterizing exercise physiology and sport science research (Speed & Andersen, 2000), it becomes self-evident that satisfying the prerequisite for the choice of the comparator arm represents the first critical issue to be addressed by exercise physiologists and sports scientists interested in conducting non-inferiority studies. Even when a discrete amount of evidence is available, the large sampling variability related to studies with small sample sizes (e.g., 8–16 participants per group) makes it difficult to identify an intervention whose efficacy had been consistently demonstrated across the literature. Moreover, questionable practices such as publication bias and *p*-hacking (i.e., the manipulation of data collection and analysis to obtain statistically significant results) tend to overestimate the intervention effect in meta-analyses and thus impact the ‘assay sensitivity’ of the new investigation, which is the ability of a study to distinguish between an efficacious and less efficacious intervention. Several graphical and statistical approaches seeking to quantify or adjust for publication bias in meta-analyses have been developed (Simonsohn et al., 2014; Carter et al., 2019). However, most of these methods lack large-scale empirical validation, do not work well when there are few studies or large heterogeneity in effect sizes, and their performance and efficiency are often highly sensitive to deviations from the model assumptions. Note that the problem of publication bias and *p*-hacking would be dramatically reduced if pre-registration or Registered Reports Protocols became common practice in exercise physiology and sport science (Lakens & Evers, 2014; Caldwell et al., 2020). The aforementioned aspects highlight the importance of gaining reliable knowledge about effect sizes reported in the literature before deciding whether to adopt a non-inferiority design. This also emphasizes the need for more collaborations across exercise physiology and sport science departments to design and conduct studies with high accuracy, and the need for more transparent research practices, as stressed by several scientists in a recent call (Caldwell et al., 2020).

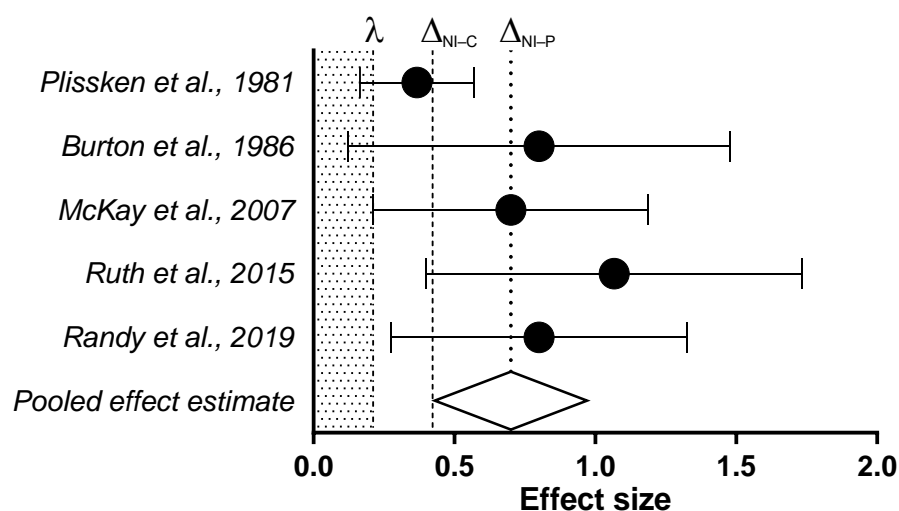
#### 1.6.4 Determination of non-inferiority and equivalence margin(s)

Once the reference intervention has been chosen, the next step in designing non-inferiority studies concerns the choice for the margin. An appropriate  $\Delta_{NI}$  should be based on a combination of statistical reasoning and domain expertise (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 1998, 2001; Committee for Proprietary Medicinal Products, 2000; Committee for Medicinal

Products for Human Use, 2005). The general principle states that the  $\Delta_{NI}$  should not be larger than the smallest effect the reference intervention would be reliably expected to have compared with a placebo. Despite more sophisticated approaches being proposed (Snapinn & Jiang, 2018a; Yu et al., 2019), the ‘point-estimate method’ and the ‘fixed-margin method’ are the most widely used for specifying the margin in clinical research (Althunian et al., 2017). In the point-estimate method, the  $\Delta_{NI}$  is based upon the pooled effect estimate of the active comparator from a meta-analysis without considering the uncertainty in the estimate ( $\Delta_{NI-P}$ ). In the fixed-margin method, the two-sided 95% CI of the meta-analytic effect size estimate that is closest to the null effect is used to determine the non-inferiority  $\Delta$  ( $\Delta_{NI-C}$ ) (Figure 12). This makes the latter approach more conservative than the former, especially when – as is often the case in exercise physiology and sport science – the precision of the individual study estimates is generally low, and the total number of studies is small. A third common approach to analyze non-inferiority trials applies the same criteria as the fixed-margin method to determine  $\Delta_{NI}$  but also adjusts the CI derived from the non-inferiority trial to account for the sampling variability in the effect of the active comparator against placebo (Holmgren, 1999; Althunian et al., 2017). This ‘synthesis method’ is slightly more efficient than the fixed-margin method but it is also more sensitive to a violation in the assumptions of assay sensitivity and constancy (Schumi & Wittes, 2011).

Regardless of the method used to determine the  $\Delta_{NI}$ , several factors such as the importance of the outcome measure, clinical or practical considerations in terms of cost-effectiveness of the active comparator, model misspecification, or violation of the constancy assumption can make putative superiority over placebo alone an insufficient criterion to establish non-inferiority and additional assurance may be needed. In this respect, pre-specifying a percentage of the historical effect of the reference intervention that must be retained by the new one (usually 50%), the so-called ‘preserved fraction’ ( $\lambda$ ), has become common practice in non-inferiority clinical trials (Figure 12) (Snapinn, 2004; Snapinn & Jiang, 2018b). Despite its widespread use in clinical research, it is important to note that there is no consensus as to whether setting the  $\Delta_{NI}$  by including a preserved fraction represents an effective discounting approach (Snapinn, 2004; Snapinn & Jiang, 2018b).





**Figure 12** The two-step process commonly employed to determine the non-inferiority margin ( $\Delta_{NI}$ ) in clinical research. A pooled effect estimate is calculated from a meta-analysis of hypothetical studies and the margin is determined using either the point estimate (point-estimate method;  $\Delta_{NI-P}$ ) or the lower 95% confidence limit (fixed-margin method;  $\Delta_{NI-C}$ ) of the effect size. The chosen margin ( $\Delta_{NI-C}$  in the example) is then multiplied by a pre-specified factor ( $\lambda$ ; usually 50%) to preserve a fraction of the active-control effect (shaded area).

Whether or not the stringency in the criteria to determine non-inferiority should be further adjusted according to the degree of magnitude of the historical effect of the comparator is a matter of debate among clinical researchers (Schumi & Wittes, 2011). Although the choice of the preserved fraction would have negligible implications on the study conclusions for small to moderate effects, considerable discrepancies may take place for largely efficacious standard interventions. In these cases, determining the fraction without any adjustment for the historical effect of the comparator may rule out a large part of the effect, eventually leading to the paradoxical situation in which non-inferiority is established although the experimental intervention is inferior compared with the standard (Schumi & Wittes, 2011; Althunian et al., 2018). A maximum margin criterion that prevents clinically important differences between the standard and the new intervention may be applied in these situations (Schumi & Wittes, 2011).

Whereas (bio)equivalence margins in clinical trials are often set by regulatory authorities (Committee for Medicinal Products for Human Use, 2010), several approaches to justify the equivalence range have been proposed in the social and behavioral sciences (Lakens, 2017, 2021; Lakens et al., 2018a). Among them, it is worth mentioning a method based on the

maximum sample size researchers are willing to collect given the available resources. This approach may be used for those situations, also common in exercise physiology and sport science, in which there are time, money, or population size constraints that limit the effect size that can be properly investigated, especially in novel lines of research. Under such conditions, determining  $\Delta_U$  and  $\Delta_L$  based on feasibility may be justified, and may represent a starting point for future studies aiming for a more precise assessment, if researchers see no way to specify the SESOI based on theoretical predictions or practical concerns.

### 1.6.5 Sample size planning for non-inferiority and equivalence studies

As I previously discussed, sample size estimation in superiority studies conventionally aims to achieve the desired level of statistical power (typically 80% or 90%) against an alternative hypothesis, expressed in terms of a target difference between interventions in the primary outcome(s), at a given  $\alpha$  (Cook et al., 2018). Since superiority and non-inferiority are logically opposite tests, sample size estimation for non-inferiority studies follows the same principles as for superiority studies. However, because the  $\Delta_{NI}$  is usually smaller than the superiority difference, larger sample size is often needed. Due to the nature of the TOST procedure, in which each one-sided test must statistically reject effects as small as the equivalence margins to prove efficacy, the power of an equivalence test equals the power to detect the smallest margin. In the light of the above, researchers should be aware that the adequate sample size for equivalence and non-inferiority tests may be prohibitively large for very small effects. For this reason, researchers should carefully consider the target or expected effect size, along with the margin(s), when planning equivalence and non-inferiority studies. Whenever there is substantial uncertainty about the mean difference between the interventions, or when it is plausible that the true effect is larger or smaller than the margin the test was powered to detect, researchers may opt for sequential analysis (Lakens et al., 2021). This efficient approach allows terminating data collection while controlling the Type I error rate as soon as there is convincing evidence to decide on the presence, or absence, of an effect.

Julious (2004) provided detailed overviews and approximations to calculate power and sample size in superiority, equivalence, and non-inferiority studies. Researchers who wish to exact solutions for power and sample size for equivalence designs may look at Shieh (2016). Moreover, there are several spreadsheets (Lakens, 2017), statistical packages (Castelloe & Watts, 2015; Lakens, 2017), and web-based applications (Magnusson, 2016; Kovacs et al.,

2021) that exercise physiologists and sports scientists can use to estimate sample sizes for equivalence and non-inferiority tests.

### 1.6.6 Re-imagining interventional studies using equivalence and non-inferiority tests

I provide two worked examples from exercise physiology and sport science research comparing sprint interval training (SIT) against moderate-intensity continuous training (MICT) to show how the statistical approaches discussed above can be applied to real-world data. I have included all the formulas used in these examples in an accompanying workbook (openly available – along with the SAS and R code used for validation – at <https://osf.io/ndqhe/>), which can also be used to perform calculations based on summary statistics or complete datasets.

*Example 1 - use of equivalence hypothesis:* In a comprehensive study investigating the effects of 4 weeks of SIT (60 min per week) or MICT (300 min per week) on cardiorespiratory, musculoskeletal, and metabolic characteristics in obese men, Cocks et al. (2016) concluded that SIT and MICT have equal benefits on aerobic capacity, as no statistical difference was observed between the two groups with respect to the changes in  $\dot{V}O_{2\max}$ . As previously stated, the absence of an effect cannot be concluded based on  $p > .05$  from the traditional null-hypothesis test. However, we wanted to determine whether the authors' conclusions concerning the absence of an effect between the groups can indeed be inferred from the observed data. Unfortunately, the authors did not report the nominal  $p$ -value for the time  $\cdot$  group interaction in the  $2 \cdot 2$  mixed analysis of variance (ANOVA) model, or any other necessary information about the differences in the changes in  $\dot{V}O_{2\max}$  between the groups. Since the authors did not make the raw data available along with the manuscript, we cannot perform a proper covariate-adjusted analysis; nonetheless, we can still appraise the between-group differences by extracting summary data from the paper. Specifically, we can estimate the standard deviation (SD) of the change score within each group by imputing different plausible correlation coefficients ( $r$ ) between pre-training and post-training scores, construct the two-sided 90% CI for the mean difference between the groups using the different SD estimates, and then perform a sensitivity analysis on the results (Higgins et al., 2019). For  $r = .5$ , the SIT – MICT 90% CI around the observed mean difference of  $-2.3 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  ranges from  $-7.1$  to  $2.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . The SDs of the change scores decrease at greater values of  $r$ , and the 90% CI narrows by  $\sim 17\%$  (ranging from  $-6.3$  to  $1.7 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) when  $r = .7$ . However, even in the optimistic scenario in which  $r = .9$ , the 90% CI for the between-group difference ranges from  $-5.2$  to  $0.6 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ , which indicates a large imprecision of the parameter estimate. Since a difference in  $\dot{V}O_{2\max}$  as

small as  $1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  has been associated with a 9% instantaneous relative risk reduction for all-cause mortality (hazard ratio 0.91) (Laukkanen et al., 2016), the mean difference between SIT and MICT that was observed by Cocks and colleagues of  $-2.3 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  is hardly trivial, let alone after having considered its uncertainty.

If we wish, we can also formally test for equivalence against symmetric margins  $\Delta_U$  and  $\Delta_L$  of  $1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  by using the TOST procedure, which is very similar to the Student's  $t$ -test when assuming equal population variances. This equivalence test examines the question of whether we can reject the presence of an effect as large, or larger than  $1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ , which we know is large enough to have practical benefits.

For  $\Delta_U$

$$t_U = \frac{M_1 - M_2 - \Delta_U}{SD_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (1)$$

where  $t_U$  is the test statistic for the one-sided  $t$ -test on  $\Delta_U$ ,  $M_1$ , and  $M_2$  are the means of the SIT and MICT group respectively,  $n_1$  and  $n_2$  are the sample size in each group, and  $SD_P$  is the pooled SD:

$$SD_P = \sqrt{\frac{(n_1 - 1) SD_1^2 + (n_2 - 1) SD_2^2}{n_1 + n_2 - 2}} \quad (2)$$

where  $SD_1$  and  $SD_2$  are the SD of the SIT and MICT group, respectively.

In this example,

$$SD_P = \sqrt{\frac{(8 - 1) 2.1^2 + (8 - 1) 4.2^2}{8 + 8 - 2}} = 3.3$$

therefore

$$t_U = \frac{2.4 - 4.7 - 1}{3.3 \sqrt{\frac{1}{8} + \frac{1}{8}}} = -2$$

which correspond to a  $p$ -value of .03 from the  $t$ -distribution with 14 degrees of freedom ( $df$ ) for a left-sided test.

For  $\Delta_L$

$$t_L = \frac{M_1 - M_2 - \Delta_L}{SD_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3)$$

being  $t_L$  the test statistic for the one-sided  $t$ -test on  $\Delta_L$ .

In this example,

$$t_L = \frac{2.4 - 4.7 - (-1)}{3.3 \sqrt{\frac{1}{8} + \frac{1}{8}}} = -0.8$$

which corresponds to a  $p$ -value of .78 from the  $t$ -distribution with 14  $df$  for a right-sided test. Since the one-sided test with the greater  $p$ -value is not statistically significant [ $t(14) = -0.8$ ,  $p = .78$ ] based on an  $\alpha = .05$ , we cannot reject differences larger than  $1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . Therefore, we cannot conclude that the difference between the two interventions is too small to matter (given a SESOI of  $1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) with respect to the changes in  $\dot{V}O_{2\text{max}}$ .

It is important to note that, unlike in traditional hypothesis tests where effects that are substantially greater than expected can compensate small sample sizes, underpowered tests inevitably increase the risk of inconclusive results in equivalence studies. If we want to estimate how many individuals Cocks and colleagues should have recruited and tested to reach an adequate level of power (e.g., 80%) for the TOST procedure at the desired  $\alpha$  level (e.g., .05), the most informative approach is to perform an *a priori* power analysis. For the sake of simplicity in calculations, we can define equivalence margins that are symmetric around a zero difference in population means ( $\mu_1 - \mu_2$ ). Moreover, we assume that the estimated pooled SD represents the true SD for the two populations ( $\sigma$ ). For simplicity, we will rely on the normal approximation of the power equation for equivalence tests (Julious, 2004) and estimate the sample size ( $n$ ) required in each group to achieve the desired power against  $\Delta_U$  and  $\Delta_L$  as:

$$n_U = \frac{(r + 1) \sigma^2 (z_\alpha + z_{\beta/2})^2}{r |\Delta_U|^2} \quad (4)$$

and

$$n_L = \frac{(r + 1) \sigma^2 (z_\alpha + z_{\beta/2})^2}{r |\Delta_L|^2} \quad (5)$$

where  $r$  is the allocation ratio ( $n_1 / n_2$ ), and  $z_\alpha$  and  $z_{\beta/2}$  are the standardized normal deviates corresponding to the levels of  $\alpha$  and  $\beta / 2$  respectively (with  $1 - \beta$  that represents the desired power). With an equal allocation (1:1 ratio), the equations 5 and 6 are reduced to:

$$n_U = n_L = \frac{2 \sigma^2 (z_\alpha + z_{\beta/2})^2}{|\Delta_U = \Delta_L|^2} \quad (6)$$

In this example,

$$n = \frac{2 \times 3.3^2 (1.6 + 1.3)^2}{1^2} = 192$$

which indicates that the *minimum* sample size that Cocks and colleagues should have recruited to have a properly powered test for equivalence was twenty-four times larger than the  $n = 8$  per group that was collected in that study. Even using a much more liberal SESOI of  $3.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ , associated with up to a 25% risk reduction in mortality (Ross et al., 2016), the smallest sample size should have been double the one collected. Note that these also represent optimistic estimations: any situation in which some inequality between interventions can be expected (i.e., the expected difference is not 0), would increase the required sample size, all else being equal.

*Example 2 - use of non-inferiority hypothesis:* Gillen et al. (2016) investigated whether 30 min per week of SIT was a time-efficient exercise strategy to improve indices of cardiometabolic health in healthy men to the same extent as 150 min per week of MICT. Although the time  $\cdot$  group interaction in the 3  $\cdot$  2 mixed ANOVA model was significant for  $\dot{V}O_{2\text{max}}$ , the authors were unable to reject a *nil* effect and conclude statistical differences between the groups after 12 weeks of training intervention. The exact  $p$ -value and the 95% CI for the between-group comparison were not reported; however, since the authors reported the 95% CI for the change scores of the two groups, as well as their sample sizes, we can obtain the information we need using statistical first principles (Higgins et al., 2019). The calculations reveal a  $p$ -value of .94 and a 95% CI ranging from  $-2.9$  to  $2.7 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  constructed around a mean difference between the interventions of  $-0.1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . From a superiority standpoint, the study is inconclusive for what concerns the ability of SIT to improve the  $\dot{V}O_{2\text{max}}$  compared with MICT. Given the rationale supporting the study, a more informative research question might be whether the improvements in the  $\dot{V}O_{2\text{max}}$  induced by SIT are not substantially lower than those induced by a standard MICT program. To answer such a question, first, we must define the  $\Delta_{\text{NI}}$  that we will use to test our hypothesis. The net effect of MICT against no-exercise control on  $\dot{V}O_{2\text{max}}$  has been estimated to be  $4.9 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  with a 95% CI ranging from  $3.5$  to  $6.3 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  (Milanović et al., 2015). If we assume the MICT protocol prescribed by Gillen and colleagues is sufficiently representative of the ‘typical’ MICT from which the average intervention effect has been estimated and we prefer a conservative approach to the margin determination without further need for a preserved fraction, we can rely on the fixed-margin method and test the SIT – MICT difference against a

$\Delta_{NI}$  of  $-3.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . The calculation of the  $t$ -statistic for the non-inferiority test is identical to those for the one-sided test against the  $\Delta_L$  in the TOST procedure.

$$t_{NI} = \frac{M_1 - M_2 - \Delta_{NI}}{SD_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (7)$$

being  $t_{NI}$  the test statistic for the non-inferiority test.

In this example,

$$t_{NI} = \frac{5.9 - 6 - (-3.5)}{2.9 \sqrt{\frac{1}{9} + \frac{1}{10}}} = 2.6$$

which corresponds to a  $p$ -value of .02 from the  $t$ -distribution with 17  $df$  for a two-sided test. If all the assumptions underlying the statistical model are correct, the non-inferiority test is significant [ $t(17) = 2.6$ ,  $p = .02$ ] for an  $\alpha = .05$ . We can then reject a loss in the efficacy of SIT compared with MICT larger than  $3.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  and conclude that SIT is non-inferior to MICT for what concerns increase in  $\dot{V}O_{2\max}$ . Unsurprisingly, given the close relationship between  $p$ -values and CIs, the CI approach leads to the same conclusion as the formal non-inferiority test since the lower 95% confidence limit of the SIT – MICT difference (i.e.,  $-2.9 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) is larger than the  $\Delta_{NI}$  (i.e.,  $-3.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ), which indicates that the entire set of plausible values for the population parameter contained in the 95% CI is consistent with the non-inferiority of SIT against MICT.

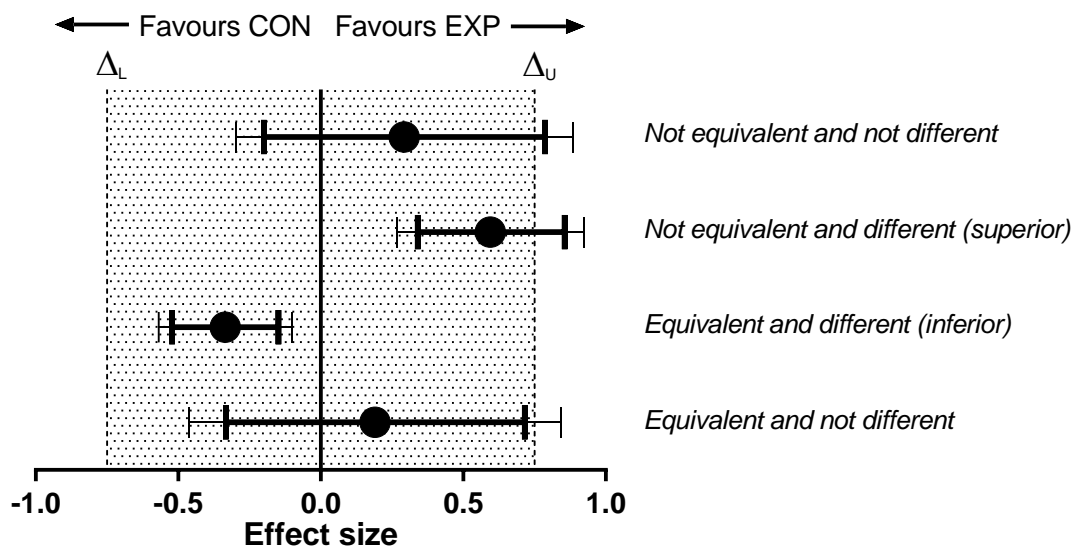
### 1.6.7 Switching between hypotheses

Switching the objective of a clinical trial from non-inferiority to superiority or vice versa may be possible at the analysis stage of the study; however, the change is not always straightforward, and several points need to be considered (Committee for Proprietary Medicinal Products, 2000; Schumi & Wittes, 2011). From a statistical perspective, testing first for non-inferiority and then for superiority, does not require a statistical penalty for multiple testing, since the closed testing procedure properly controls the overall Type I error rate of the two tests. When the  $\Delta_{NI}$  has been prespecified, and the trial design and conduct have been strict, it is also possible to test for non-inferiority after a superiority test that does not show any statistical benefit. Despite being statistically appropriate, researchers should be warned that this testing order could result in paradoxical outcomes (i.e., a new intervention that is both non-inferior and inferior to the standard), especially for largely efficacious standard interventions. As stated previously,

considering the SESOI as a criterion for the largest acceptable  $\Delta_{NI}$  may help to minimize this risk.

Departing from the initial aim of establishing equivalence does not appear to be a common practice in clinical research (Senn, 2021). Moreover, the greater value of  $\alpha$  usually adopted in such investigations would lead to an inflated Type I error rate if the researcher attempted to draw straightforward conclusions on superiority or non-inferiority. Nonetheless, various comprehensive methods to investigate equivalence along with superiority have been recently presented in the social and behavioral sciences literature (Lakens, 2017; Lakens et al., 2018a) (Figure 13). Exercise physiologists and sports scientists interested in conducting equivalence and non-inferiority studies may benefit from exploring these approaches.

It is also worth mentioning the possibility to test against both the nil effect and the SESOI in all those situations in which the researcher, after having concluded that the effect is non-zero, is interested in rejecting effects too small to be relevant.



**Figure 13** Testing for both equivalence and superiority. The thin error bars indicate the 95% CI in relation to the traditional null-hypothesis test, whereas the thick error bars indicate the 90% CI in relation to the two one-sided tests procedure. The solid vertical lines indicate the traditional null hypothesis, whereas the shaded area indicates the equivalence region. Conclusions for hypothesis tests are reported next to each example.



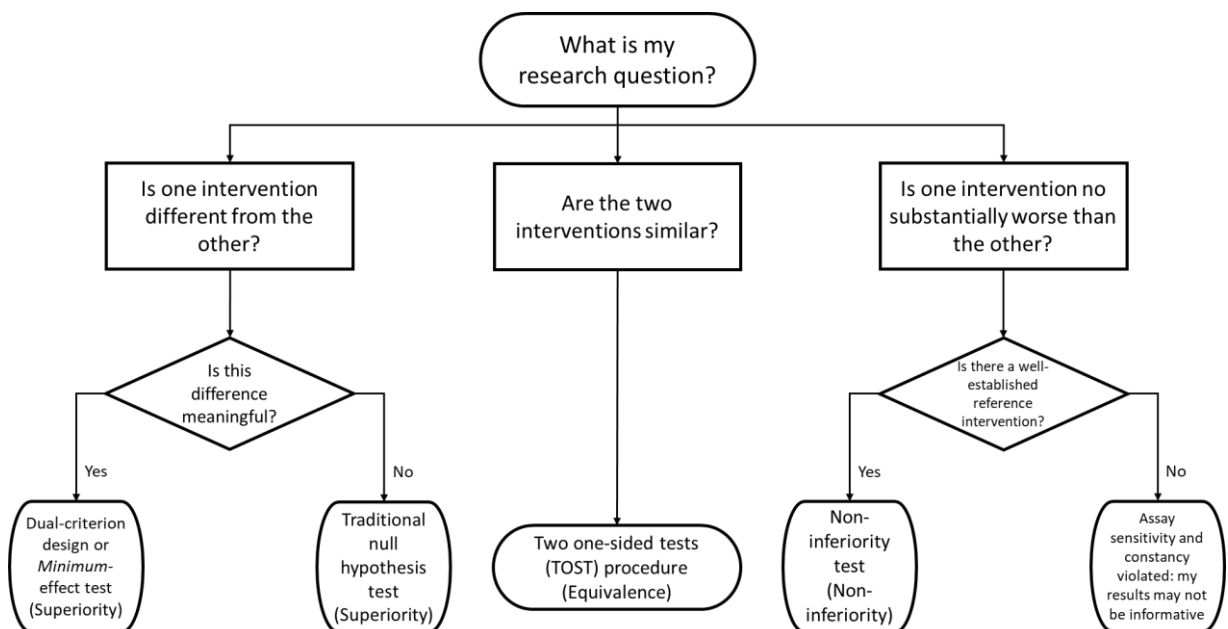
### 1.6.8 Limitations and additional considerations

I have detailed how to expand the statistical toolset used to design and analyze interventional studies in exercise physiology and sport science. To achieve clarity and brevity, I focused on parallel-group studies with means and variances determined from pairs of independent random samples of normally distributed observations. Readers must be aware that the analytical approach to other research designs or variables with different probability distributions may slightly differ from the one presented herein. When discussing the acceptable standard of evidence, I maintained consistency with the defaults commonly used in biomedical, social, and behavioral research. Nonetheless, the optimal error rates should be decided based on a cost-benefit analysis, depending on the context, goals, and resources (Lakens et al., 2018b). It is worth keeping in mind that frequentist estimation (i.e., CI) and hypothesis testing do not represent the only way to draw inferences from data. Wald's statistical decision theory provides a coherent frequentist framework to use sample data to make decisions on interventions (Manski, 2019). Compared with hypothesis testing, the Wald framework has the advantage of taking into account the magnitudes of the losses that Type I and II errors (whose probabilities are considered symmetrically) yield as an integral part of the framework. Among the alternative or complementary methods to frequentist statistics, Bayesian statistics or Likelihood approaches can also be used to answer the questions that might be of interest to researchers (Wang & Blume, 2011; van Ravenzwaaij et al., 2019; Lakens et al., 2020). These approaches have the main advantage of allowing researchers to make probabilistic statements about the (random) parameter of interest. Whenever prior data are available from other studies, Bayesian statistics also allows incorporating such information in the analysis to update the (posterior) probability of the parameter and provide the relative weight of evidence for the alternative hypothesis compared with the null. Although presenting such methods to design and analyze superiority, equivalence, and non-inferiority studies were beyond the scope of this manuscript, exercise physiologists and sports scientists should consider their use within the context of statistical inference when deciding which method(s) is the most appropriate for their research purpose(s).

### 1.6.9 Conclusions and recommendations for future research

Exercise physiology and sport science have largely relied on the traditional null hypothesis test to make informed decisions in interventional studies. This approach, combined with

underpowered tests, has often led to the misinterpretation of a non-significant test result as support for the equivalence between interventions. While it should be clear at this point that this is a statistical misconception, exercise physiologists and sports scientists should also understand that research should not be limited to investigating whether one intervention is superior or inferior to another. Equivalence and non-inferiority designs may be adopted whenever the research context, conditions, applications, researchers' interests, or reasonable beliefs justify them. Although these research hypotheses require additional methodological considerations than superiority hypotheses to be properly investigated, they may also better answer the empirical question researchers are interested in. Equivalence and non-inferiority studies may help exercise physiologists and sports scientists to answer questions that the traditional null hypothesis cannot address. Figure 14 provides a flowchart to facilitate the decision-making process about the most informative study design.



**Figure 14** Processes of decision making for selecting the different hypothesis tests based on the research question that is being asked.



## 2

**RESEARCH GOALS**

*'The aim of science is not to open the door to infinite wisdom, but to set a limit to infinite error.'*

*Bertolt Brecht – poet*





## **2. RESEARCH GOALS**

### **2.1 General goal**

The experimental part of the thesis is structured by a line of research consisting of three different families of hypotheses aimed to clarify most of the points and still open questions raised in chapter 1.5 about the effects of different HIIT prescription models on physiological response and performance in trained distance runners. The general goal of this thesis is to provide some enlightenment on the best approaches to training prescription, especially for what regards the relationship between the prescribed training and the desired outcome. The thesis is focused on HIIT since this form of training is the one that requires more careful characterization and thus the one that is potentially more sensitive to mis-prescriptions. This work may provide several important pieces of information that may serve as support for further investigations and that may help coaches to maximize the benefits of this training method in their athletes.

### **2.2 Specific goals**

- 1) Investigating whether individualizing HIIT according to the physiological characteristics and the response to exercise (i.e., the physiological approach) induces larger cardiorespiratory adaptations than a standardized HIIT program in trained distance runners.
  
- 2) Quantifying the interindividual variability that exists in different treadmill-related measures and race pace-related measures at  $v\Delta 50$ , after having accounted for the biological variability (i.e., the intra-individual variability) in the measures, to identify a valid surrogate for  $v\Delta 50$  to prescribe HIIT according to the physiological approach without the need of lab measurements.
  
- 3) Investigating whether individualizing HIIT prescriptions according to physiological characteristics (i.e., the physiological approach) or percentage of target race pace (i.e., the race pace approach) determines different physiological adaptations and performance improvement in trained distance runners.



## 3

**RESEARCH HYPOTHESES**

*'I have approximate answers and possible beliefs in different degrees of certainty about different things, but I'm not absolutely sure of anything.'*

*Richard Feynman – theoretical physicist and science communicator*







### 3. RESEARCH HYPOTHESES

- 1) The individualized, physiologically-based approach to HIIT prescription will improve  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  more than the standardized approach and it will decrease the heterogeneity of intervention effects in trained distance runners.
  
- 2) Treadmill-related measures – namely the relative percent of  $V_{\text{peak}}$  and the absolute difference between  $V_{\text{peak}}$  and  $v\Delta 50$  – can be used as valid surrogates of the delta concept to individualize HIIT prescriptions according to the physiological approach without the need for lab measurements. Please note that from a statistical standpoint, this is an estimation approach.
  
- 3) The physiological approach to HIIT prescription will improve  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  more than the race pace approach and it will lead to a smaller heterogeneity of intervention effects in trained distance runners. On the contrary, the race pace approach will improve running performance more than the physiological approach, other than leading to a smaller heterogeneity of intervention effects in this population.



## 4

**METHODOLOGY**

*'We need less research, better research, and research done for the right reasons'*

*Doug Altman – statistician*





## 4. METHODOLOGY

### 4.1 Participants

#### 4.1.1 Individualizing HIIT according to the physiological characteristics

Seventy-two distance runners (56 men, 16 women;  $\geq 2$  years of experience) were recruited for this investigation between October 2019 and January 2020. Recruitment and selection of participants were conducted in the provinces of Milan and Pavia (Italy) using flyers, social media advertising, and personal contact. Eligible participants were men and women between 18 and 50 years old that were training and competing in long-distance running events (from 5000-m to the Marathon) with a personal best achieved in the last year between 100 and 500 points and between 400 and 900 points for men and women, respectively, according to the latest edition available of the IAAF Scoring Tables for Outdoor Events (Spiriev & Spiriev, 2017). These scores correspond approximatively to times from 36 to 45 min (men) and 37 to 50 min (women) on 10000-m races and classify these individuals as ‘trained’ runners according to the classification framework recently proposed by McKay and colleagues (2022). According to the Italian national law (Law 91/1981), a valid Medical Certificate for Competitive Sports Activity was also required to participate in this investigation. Potential participants who had underlying medical problems (injury or surgery in the last 6 months, illness, infection, cardiovascular, respiratory, metabolic, and musculoskeletal diseases or dysfunctions) that may have affected their participation in the project or the results were excluded. Out of the 106 people initially contacted (80 men, 26 women), 24 (16 men, 8 women) declined participation due to lack of time or interest, 8 (7 men, 1 woman) did not meet the eligibility criteria, and 2 (1 man, 1 woman) suffered from injury before the beginning of the data collection, leaving a total of 72 runners partaking in this investigation. The sample size estimation was conducted using PROC POWER in SAS Studio 3.8 on SAS 9.4 (SAS Institute, Inc., Cary, NC, USA) to achieve an 80% power to detect a difference of at least  $3.2 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  in  $\dot{V}\text{O}_{2\text{max}}$  – corresponding to the typical within-individual variability for this population (Katch et al., 1982) – and at least  $0.5 \text{ km}^{-1} \cdot \text{h}^{-1}$  in  $V_{\text{peak}}$  – derived from the speed- $\dot{V}\text{O}_2$  linear regression model (Batliner et al., 2017) applied to a group of 22 runners from a pilot study conducted on the same population using the same incremental protocol and assuming negligible changes in RE – between the groups at a one-sided  $\alpha$  level of .05 assuming an equal number of participants in each group. The expected between-variability of  $\dot{V}\text{O}_{2\text{max}}$  ( $5.1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) and  $V_{\text{peak}}$  ( $0.8 \text{ km}^{-1} \cdot \text{h}^{-1}$ ) for this

population were instead estimated using the pooled SD extracted from previous relevant studies (Acevedo & Goldfarb, 1989; Demarle et al., 2001, 2003; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003; Denadai et al., 2006; Kohn et al., 2010). The benefit of covariate adjustment on statistical power (Thompson et al., 2015) was not considered in the calculation and the estimated sample size for the *t*-test design was considered as an ‘upper bound’ of the estimation since efficiencies are produced by the chosen design (High, 2007). Since none of these relevant studies reported information on the number of dropouts, the initial sample (33 participants per group) was inflated assuming a plausible ~8% dropout rate for this type of training intervention (Stöggl & Sperlich, 2014), and rounded up to the nearest integer.

#### 4.1.2 Testing interindividual variability in different treadmill-related measures and race pace-related measures

Seventy-five trained distance runners (58 men, 17 women;  $\geq 2$  years of experience) were recruited into this investigation for estimating the observed between-individual variability between May 2020 and May 2021. Sixty-one runners (46 men, 15 women) from this cohort underwent repeated measurements to estimate the true variability (i.e., the inter-individual variability that is left after having accounted for the biological variability). The eligibility criteria were the same as those reported when investigating the first family of hypotheses. The sample size estimation was conducted using PROC POWER in SAS. Since the main research questions involved estimation rather than hypothesis testing, sample size analysis was based on achieving a given CI precision rather than statistical power (Maxwell et al., 2008; Lakens, 2021). Specifically, the aim was to achieve a 1% and  $0.1 \text{ km}^{-1} \cdot \text{h}^{-1}$  (since this often represents the smallest speed variation that is allowed by commercial treadmills) for estimating – as two-sided 95% CI with 99% assurance – the percentage of  $V_{\text{peak}}$  at  $v\Delta 50$  and the difference between  $v\Delta 50$  and  $V_{\text{peak}}$ , respectively. Previous data from my first investigation were used to inform about the expected SD for the parameters of interest in this population (i.e., 2% and  $0.3 \text{ km}^{-1} \cdot \text{h}^{-1}$ , and 1% and  $0.2 \text{ km}^{-1} \cdot \text{h}^{-1}$  for the observed and true between-individual SD, respectively). In this regard, pre-intervention data were used to estimate the expected (gross) between-individual SD, whereas pre-intervention and post-intervention data were combined to separate the true between-individual SD from the within-individual SD via mixed modeling (using the same approach described in the statistical section). However, since the individualized and the standardized approach to HIIT prescription might have led to different effects on  $V_{\text{peak}}$  and the standardized approach might have also led to a larger heterogeneity of intervention effects, the

estimated values of the true between-individual SDs from the previous data were inflated by a fourth to compensate for any possible reasonable underestimation. The required sample size for estimating the gross between-individual SD in the percentage of  $V_{\text{peak}}$  at  $v\Delta 50$  and in the difference between  $v\Delta 50$  and  $V_{\text{peak}}$  was 23 and 55 individuals, respectively. Instead, the required sample size for estimating the true between-individual SD was 15 and 47 individuals. Since estimating the percentage of 10000-m speed at  $v\Delta 50$  and the difference between  $v\Delta 50$  and 10000-m speed was also part of the aims, sample size analysis was performed even for these parameters using the same criteria that were used for the treadmill-related measures. However, the estimated between-individual SDs (i.e., 5% and  $0.7 \text{ km}^{-1} \cdot \text{h}^{-1}$ , and 4% and  $0.5 \text{ km}^{-1} \cdot \text{h}^{-1}$  – inflated to 5% and  $0.6 \text{ km}^{-1} \cdot \text{h}^{-1}$  – for the observed and true between-individual SD, respectively) resulted in unrealistic sample size estimates (i.e., 148, 250, 141, and 200 individuals) given time and resource constraints. Therefore, a larger margin of errors was accepted for these measures (i.e., 2% for the percentage of 10000-m speed at  $v\Delta 50$  and  $0.2 \text{ km}^{-1} \cdot \text{h}^{-1}$  for the difference between  $v\Delta 50$  and 10000-m speed) to obtain feasible sample sizes of 47, 75, 45, and 61 individuals, respectively. To ensure sufficient precision for all the parameter estimations and account for the (possible) underestimated sample sizes for treadmill-related measures, the largest sample size estimates for estimating the expected (75 individuals) and true (61 individuals) between-individual SD were chosen.

#### 4.1.3 Individualizing HIIT prescriptions: physiological characteristics vs race pace

Thirty-eight distance runners (24 men, 14 women;  $\geq 2$  years of experience) were recruited for this investigation between May and September 2020. The eligibility criteria for participants and the method of recruitment were the same as those reported when investigating the first family of hypotheses. Out of the 51 people initially contacted (32 men, 19 women), 8 (5 men, 3 women) declined participation due to lack of time or interest, and 5 (3 men, 2 women) did not meet the eligibility criteria, leaving a total of 38 runners partaking in this investigation. The sample size estimation was aimed to achieve the desired error rate (i.e.,  $\alpha = .05$ ) for each specified alternative hypothesis and was conducted according to the ‘dual-criterion designs’ (Roychoudhury et al., 2018) using a customized spreadsheet. This approach provides a sample size that ensures statistical significance whenever practical relevance is observed (i.e., whenever the decision value equals at least the SESOI or the target difference) (Roychoudhury et al., 2018). Target mean differences and between-variability for  $\dot{V}O_{2\text{max}}$  and  $V_{\text{peak}}$  were the same as those reported for the first family of hypotheses [i.e.,  $3.2 (5.1) \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  and



0.5 (0.8)  $\text{km}^{-1} \cdot \text{h}^{-1}$ , respectively]. Regarding race performance, SESOI at  $0.8 \text{ km}^{-1} \cdot \text{h}^{-1}$  in 10000-m speed (corresponding to an improvement from 42:30 to 40:00) was set. This effect was considered ‘satisfactory’ according to a modified Delphi technique, which involved repeated-stage interviews with several coaches and runners. The expected between-variability of 10000-m speed ( $1.2 \text{ km}^{-1} \cdot \text{h}^{-1}$ ) for this population was instead estimated from the data available from my second investigation. The benefit of covariate adjustment on statistical power (Thompson et al., 2015) was not considered in the calculation. The largest sample size estimation for the three main outcomes (i.e., 14 participants per group) was inflated by ~36% to account for a plausible dropout rate for this type of training intervention – estimated from my first investigation – and rounded up to the nearest integer.

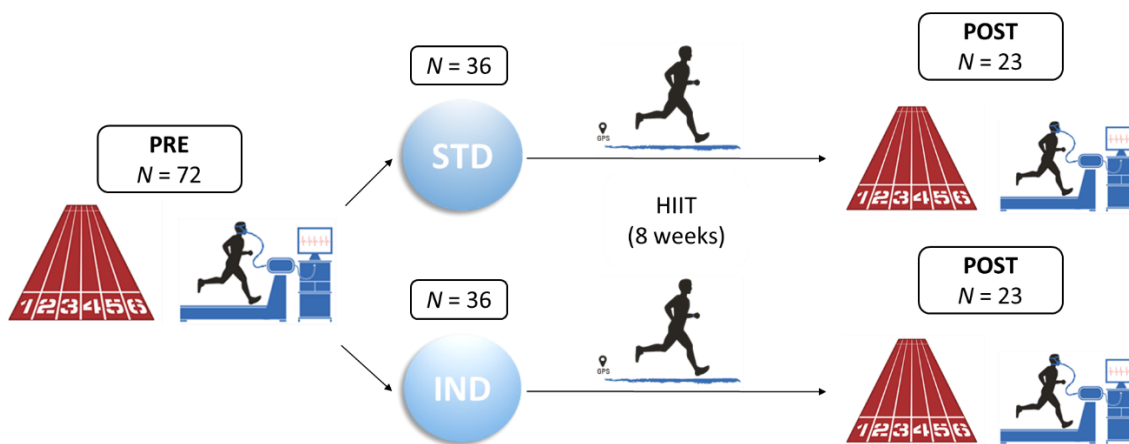
All the participants were fully informed of the aim and procedures of the study, including its risks and benefits, before signing the informed consent form. The research protocol complied with the latest revision of the Declaration of Helsinki (Fortaleza, 2013) and was approved by The Ethics Committee for Research on Human Subjects of the University of the Basque Country (CEISH-UPV/EHU 96/2018).

## 4.2 Experimental design

All the research investigations were conducted at the Human Integrative Physiology of Exercise (HIPE) Laboratory of the University of Pavia (Italy). The first investigation was carried out between January and March 2020 and it was preceded by a pilot study conducted in Madrid (Spain) between January and March 2019, which served to test the feasibility of the project and inform about the plausible direction of the effect(s). The second investigation was conducted between May 2020 and July 2021, whereas the third one took place between August and November 2020.

A pre-post parallel-group design was chosen to achieve the first and third specific aims (Figure 15).  $\dot{V}\text{O}_{2\text{max}}$  and  $V_{\text{peak}}$ , along with other physiological adaptations (RE,  $\dot{V}\text{O}_2$  at  $\text{LT}_1$  and  $\text{LT}_2$ , % $\dot{V}\text{O}_{2\text{max}}$  at  $\text{LT}_1$  and  $\text{LT}_2$ ,  $v\text{LT}_1$  and  $v\text{LT}_2$ ), were assessed during two lab tests (incremental treadmill test and constant-speed treadmill test), whereas a 10000-m time trial on a 400-m athletic track was used to determine changes in performance. I planned to perform all tests for each of these two investigations within 3 weeks (~25 runners per week), before and after an 8-week training intervention period. After the pre-intervention measurements, participants were equally allocated to two different training groups using a pseudo-random number generator

provided by PROC PLAN in SAS. The two experimental groups for each investigation [i.e., the individualized group (IND; 27 men and 9 women) and the standardized group (STD; 29 men and 7 women) for the first investigation, and the physiologically-based group (RP; 11 men and 8 women) and the race pace-based group (PHY; 13 men and 6 women) for the second investigation] replaced a similar part of the habitual training volume with two HIIT sessions per week. A longitudinal design with two blocks of repeated measures separated by 8 weeks was instead used to achieve the second specific aim. Other than the lack of a comparator and training intervention, the design was identical to the one adopted when investigating the other two families of hypotheses.



**Figure 15** Overview of research design for the first investigation. IND and STD represent the individualized and standardized group, respectively; whereas the track and the treadmill logo represent the 10000-m time trial and the incremental treadmill test followed by the constant-speed test, respectively. This pre-post parallel-group design was essentially identical for the first and the third investigation, whereas only one group and no experimental intervention was used to investigate the second family of hypotheses.

## 4.3 Procedures

### 4.3.1 Physiological and performance testing

Participants were asked to abstain from strenuous exercise and other excessive stressors that had the potential to influence performance in the 48 h before each testing session. Participants were asked to follow their usual diet during the study and maintain habitual intakes in the lead-up to the testing period. The general recommendation will be to avoid alcohol and caffeine-containing products intake during measurement periods for the 12 h preceding each

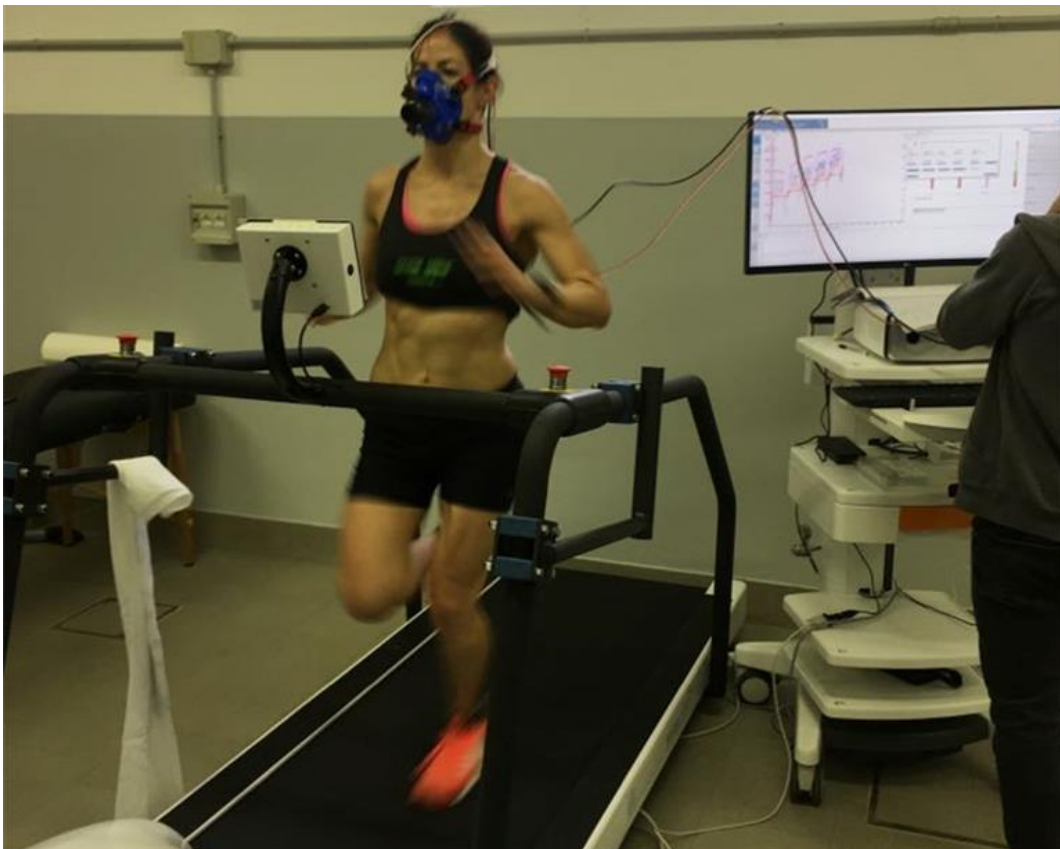
measurement. However, moderate use of caffeine was allowed for those participants with habitual caffeine practices (Jeacocke & Burke, 2010). Participants were also encouraged to get sufficient sleep the night before each session and arrive in a well-hydrated state. Careful instructions about the proper clothing and shoes to wear and about the correct way to perform the tests were provided to participants. To minimize the influence of circadian variance on the results, all tests associated with this study were performed at the same time of the day for each participant (to within  $\pm 4$  h) (Ammar et al., 2015). The incremental and constant-load treadmill tests were conducted at the HIPE Laboratory of the University of Pavia (Italy), under standard laboratory conditions ( $\sim 20$  °C ambient temperature,  $\sim 100$  kPa barometric pressure,  $\sim 50\%$  relative humidity).

The days of the 10000-m time trial were scheduled in such a way to guarantee similar weather conditions (temperature, pressure, humidity) between pre-training and post-training, although achieving a perfect matching was not possible due to seasonality and participant availability. The trial was preceded by a ‘typical’ pre-competitive resting period characterized by 3 – 4 days of low-intensity running (Muñoz et al., 2014). To avoid collective behavior influencing pacing decisions (Renfree & Casado, 2018), a maximum of 4 runners was tested at each time and each runner’s start was separated from the previous one by 60 s and performed in random order. The time on 10000-m was measured using a handheld stopwatch (Fastime 14, AST Ltd, Measham, UK) and the average speed was calculated.

Participants reported to the laboratory at least 48 h after the 10000-m time trial to perform an incremental test on a motorized treadmill (Athlete 870 C, Medisoft S.A., Sorinnes, Belgium), whose software was modified by the producer on my request to allow changes in the speed as small as  $0.1 \text{ km} \cdot \text{h}^{-1}$  (Figure 16). Accuracy and reliability in treadmill speed and the slope were assessed as recommended (Padulo et al., 2014) using a high-speed camera and digital inclinometer, respectively (iPhone 11 Pro Max, Apple Inc., Cupertino, CA, USA). Since the nominal speed of the treadmill differed from the real values, I applied a correction formula ( $y = 0.9698x - 1.2179$ , standard error of the estimate = 0.03, coefficient of determination = 1). The incremental treadmill test consisted of a discontinuous protocol preceded by a 5-min warm-up performed at the same speed as the first step of the test and separated by 3 min. To minimize the risk that  $vLT_1$  intensity was not achieved during the first step of the test, the initial running speed was set at 70% of the 10000-m speed (Dantas & Doria, 2015). The treadmill speed was then increased by  $1.5 \text{ km} \cdot \text{h}^{-1}$  every 3 min until voluntary exhaustion (Bentley et al., 2007). A gradient of 1% was maintained throughout the test to mirror the energetic cost of outdoor

running (Jones & Doust, 1996). Treadmill stages were separated by 30 s of passive rest to allow for the collection of capillary blood samples (Gullstrand et al., 1994). This protocol is appropriate to estimate  $\dot{V}O_{2\max}$  (Bentley et al., 2007; Midgley et al., 2007c, 2008), enables the collection of steady-state data for LT (Bentley et al., 2007) and RE (Shaw et al., 2013) determination, and generates a reliable and valid  $V_{\text{peak}}$  value (Machado et al., 2013; Peserico et al., 2014). Moreover, this protocol also conforms to the duration and increment rate recommended for accurate  $V_{\text{peak}}$  determination (Berthon & Fellmann, 2002). Verbal encouragement was provided on request during the test to ensure maximal effort (Halperin et al., 2015). The treadmill console displaying running speed, time, and distance were not visible to participants throughout the test. During the test, ventilatory and gas exchange data were collected breath-by-breath using a stationary metabolic cart (Vyntus CPX, Vyair Medical Inc., Mettawa, IL, USA), whose accuracy and precision had been previously reported (Groepenhoff et al., 2017; Perez-Suarez et al., 2018). The digital volume transducer of the chart was calibrated at 3 different flow rates using a 3-L syringe, whereas the  $O_2 / CO_2$  analyzer was calibrated against ambient air and a certified gas mixture of 16%  $O_2$  and 4%  $CO_2$ , according to manufacturer instructions. RE (in  $\text{mL} \cdot \text{kg}^{-1} \cdot \text{km}^{-1}$ ) at  $12 \text{ km} \cdot \text{h}^{-1}$  was determined via regression using the means of the last min of each stage performed below  $vLT_2$  under the assumption of a linear relationship between treadmill speed and  $\dot{V}O_2$  for this range of speeds (Batliner et al., 2017). The maximal  $\dot{V}O_2$  achieved during the test ( $\dot{V}O_{2\max}$ , in  $\text{mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  and  $\text{L} \cdot \text{min}^{-1}$ ) was defined as the highest 15-s average (Macfarlane, 2001; Martin-Rincon & Calbet, 2020). Before proceeding with the analysis,  $\dot{V}O_2$  data were exported and cleaned by removing test values with an externally studentized residual that was more than 3 SD away from the least-squares line of the linear regression model constructed on each time-average (Cook & Weisberg, 1982).  $V_{\text{peak}}$  was calculated as follows:  $V_{\text{peak}} = \text{speed of the last completed stage (km} \cdot \text{h}^{-1}) + [\text{running time (s) at exhaustion}] / 180 \text{ s} \times 1.5 \text{ km} \cdot \text{h}^{-1}$  (Kuipers et al., 2003). A capillary blood sample (20  $\mu\text{L}$ ) was collected from the right hyperemic earlobe after each stage of the incremental test and immediately placed in reaction tubes containing a hemolyzing solution (1000  $\mu\text{l}$ ). Lactate concentration was determined using an automated enzymatic-amperometric analyzer (Biosen C-Line, EKF-diagnostic GmbH, Barleben, Germany), which calibrated automatically before each measurement against a  $12 \text{ mmol} \cdot \text{L}^{-1}$  multi-standard solution. Treadmill speeds were then plotted against lactate values and speeds at the first lactate threshold [defined as the speed preceding an increase in lactate concentration equal to or greater than  $0.4 \text{ mmol} \cdot \text{L}^{-1}$  (Tanner & Gore, 2012)] and second lactate threshold [defined according to the modified  $D_{\max}$  method (Bishop et al., 1998)] were calculated using a lactate analysis software

(Lactate-E 2.0; Newell et al., 2007).  $\dot{V}O_2$  and fractional  $\dot{V}O_{2max}$  were derived from the speed- $\dot{V}O_2$  regression equation. After at least 20 min of recovery, participants carried out a constant-speed treadmill test to exhaustion immediately preceded by a 3-min warm-up performed at the same speed of the warm-up before the incremental test. Running speed during the test ( $v_{\Delta 50}$ ) was set halfway between  $v_{LT_2}$  and  $V_{peak}$ . The calibration of the digital volume transducer of the metabolic cart and the  $O_2 / CO_2$  analyzer was checked before the beginning of the test. The gradient level, encouragement, and  $\dot{V}O_2$  data collection were the same as the incremental test.  $T_{lim}$  (in s), and  $\dot{V}O_{2max}$  were recorded and used for training prescription and analysis. Since the characteristics of the constant-speed test make it suitable as a verification phase test for  $\dot{V}O_{2max}$  in trained runners (Demarie et al., 2000), I compared the highest  $\dot{V}O_2$  value attained in the incremental test with the one attained in the constant-speed test. However, given the lack of information about the technical error of the measurement of the Vyntus CPX metabolic cart for  $\dot{V}O_{2max}$  from at the time of the design stage of the study, I used the highest value of the two tests as  $\dot{V}O_{2max}$ .

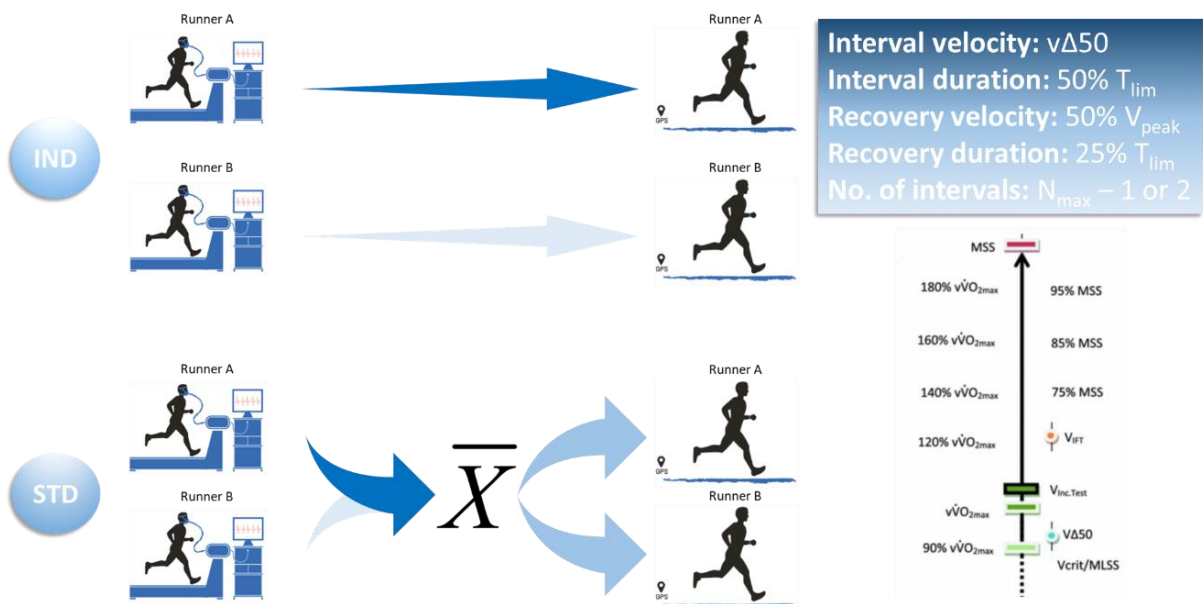


**Figure 16** One of the participants is shown while performing the incremental test as an example.

### 4.3.2 Training characteristics

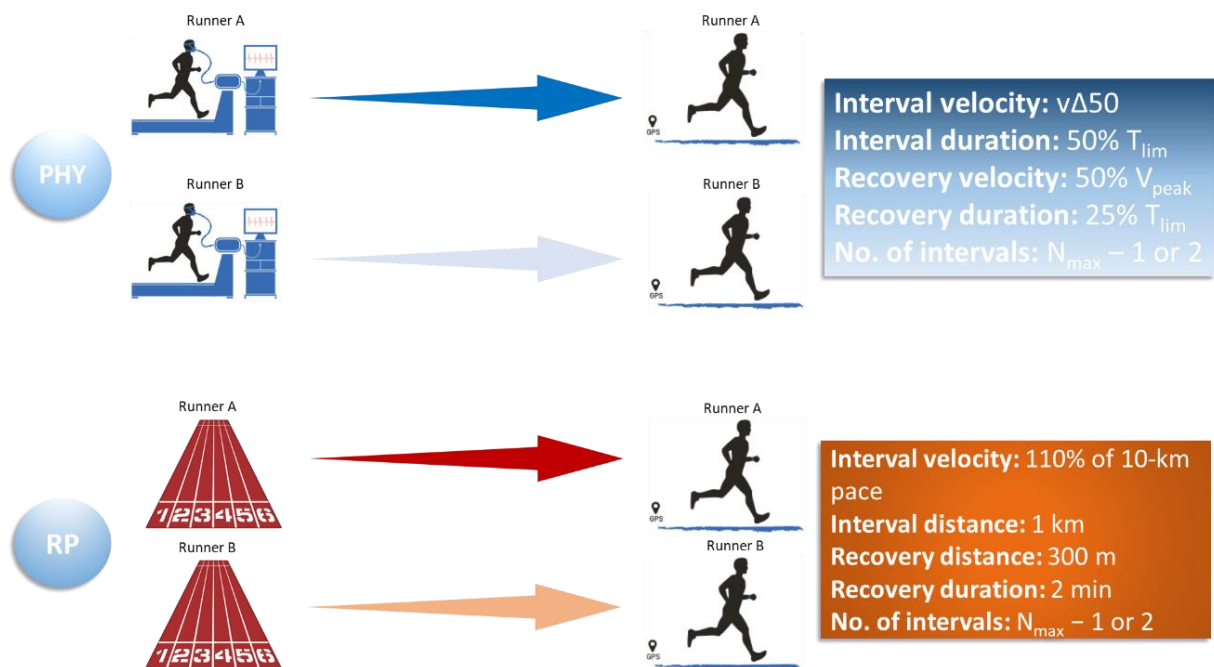
In the first investigation, HIIT characteristics were based on the protocol proposed earlier for similar cohorts of runners (Demarle et al., 2001, 2003; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003), which had shown to maximize the time spent at  $\dot{V}O_{2max}$  in this population (Demarie et al., 2000). The main difference between IND and STD was that HIIT prescriptions in IND were individualized according to each runner's values, whereas the average group values were used to prescribe the same HIIT session to all runners in STD (Figure 17). This approach would have ideally permitted to minimize the differences in the mean training stimulus between the groups while preserving the natural degree of inter-individual variability related to the training prescription method.

In the third investigation, physiological-based HIIT prescriptions in PHY were the same as those used in IND. Race pace-based HIIT prescriptions in RP were instead intended to mimic coaching-like practice, with the speed of each interval determined according to a given percentage of the target race pace, a fixed duration for each interval, and a given distance to cover during the recovery time (Figure 18). These values were extracted from the large cohort of similar runners used for the second investigation to match the intensity and distance of the intervals and recovery phases between the groups.



**FIGURE 17.** Representation of the approach used to prescribe high-intensity interval training in the individualized (IND) and standardized (STD) groups in the first investigation. The speed and the

duration of each interval were set at the speed halfway between  $v_{LT2}$  and  $V_{peak}$  ( $v_{\Delta 50}$ ) and 50% of the time to exhaustion at  $v_{\Delta 50}$  ( $T_{lim}$ ), respectively, whereas the speed and the duration of each recovery period were set at 50% of the  $V_{peak}$  and 25% of  $T_{lim}$ , respectively. HIIT prescriptions in IND were individualized according to each runner's values, whereas the average group values were used to prescribe the same HIIT session to all runners in STD. Regardless of the group, the initial number of intervals was anchored to the maximal number of intervals that each runner would have been capable to perform during the first HIIT session ( $N_{max}$ ), in such a way each runner would have alternated  $N_{max} - 1$  or  $N_{max} - 2$  intervals for the first 2 weeks of training. Other than mimicking previous research, this approach was also used to allow runners to better adapt to the new training load. From the 3<sup>rd</sup> week, the number of intervals increased on an individual basis in both groups.



**FIGURE 18.** Representation of the approach used to individualize high-intensity interval training prescriptions in the physiologically-based (PHY) and race pace-based (RP) group in the third investigation. The track and the treadmill logo represent the 10000-m time trial and the incremental test followed by the constant-speed treadmill test, respectively. The approach used for PHY was identical to the one used for IND in the first investigation. Runners in RP were required to cover 1000 m at 110% of the target 10000-m pace during each interval, and 300 m in 2 min during recovery. Regardless of the group, the initial number of intervals was anchored to the maximal number of intervals that each runner would have been capable to perform during the first HIIT session ( $N_{max}$ ), in such a way each runner would have alternated  $N_{max} - 1$  or  $N_{max} - 2$  intervals for the first 2 weeks of training. Other than mimicking previous research, this approach was also used to allow runners to better adapt to the new training load. From the 3<sup>rd</sup> week, the number of intervals increased on an individual basis in both groups.

### 4.3.3 Training assessment

Fifteen HIIT sessions were prescribed to IND and STD, and PHY and RP during the training intervention in the first and third investigation, respectively. All HIIT sessions, as well as the other training sessions, were conducted using GPS running watches [Forerunner 45, Garmin Ltd., Schaffhausen, Switzerland (accuracy: ~10 m using the combined GPS and GLONASS setting)], which monitored and recorded all the training characteristics [i.e., volume (km) and frequency (times  $\cdot$  week<sup>-1</sup>)]. All the data collected were exclusively used to characterize training, track runners' progress and adherence, but not to affect training decisions. Before the training intervention, all runners received their HIIT protocols via a proprietary app (Garmin Connect, Garmin Ltd., Schaffhausen, Switzerland). The runners were only allowed to modify the number of intervals to accommodate the expected increase in this variable throughout the 8 weeks of training. Training speed was divided into three separate exercise intensity domains as follows (Jones et al., 2019): moderate = speed < vLT<sub>1</sub>, heavy = vLT<sub>1</sub> ≤ speed ≤ vLT<sub>2</sub>, and severe = speed > vLT<sub>2</sub>, and the percentages of the overall training volume spent in each domain were calculated (Seiler, 2010).

## 4.4 Statistical analysis

Descriptive statistics are presented as mean (SD). Inferential analysis for the first and third investigations was performed using linear mixed-effects models (PROC MIXED) in SAS. Each outcome measure was modeled as a change score from baseline and percentages were treated as continuous data. Sex, age, and baseline score were identified *a priori* as putative covariates and included in each model to reduce variability and adjust for baseline imbalance between groups (Vickers, 2001; Vickers & Altman, 2001; Pocock et al., 2002). I included an interaction term between each covariate and the group predictor to allow the effect of each covariate to vary between groups. Moreover, I specified a diagonal covariance structure for the residuals to allow for unequal variance between groups. Heterogeneity of intervention effects was estimated using differences in variance between IND and STD, assuming the same (classical) measurement error in the two groups, and presented as SD (Atkinson & Batterham, 2015; Ross et al., 2019; Mills et al., 2021). Negative values of variance were presented as negative SD and interpreted as a smaller heterogeneity in IND than STD. The model assumptions were checked via residual plots, histograms, and Q-Q plots inspection; however, it was not possible to verify the normal distribution of the residuals across all levels of the predictors due to insufficient



observations in some of the subgroups. Since age was not linearly related to any of the outcomes, this covariate and the associated interaction term were dropped from the model. Models were fitted using the restricted maximum likelihood method, and  $p$ -values along with 95% CI for between-group differences, in terms of estimated marginal means, were derived using the Satterthwaite approximation for degrees of freedom. This approach maintains the actual Type I error rate close to the nominal value even with small samples (Luke, 2017). The  $\alpha$  level was set to .05 for both directional (confirmatory) and non-directional (exploratory) hypotheses. Although both within-group and between-group analyses were conducted, I based my conclusions on the between-group difference in change scores, since the latter represents the best approach to get information about intervention efficacy (Matthews & Altman, 1996; Bland & Altman, 2011, 2015). Only those participants who completed at least 12 out of 15 HIIT sessions as prescribed (i.e., a *minimum* attendance of 80%) were included in the analysis.

In the second investigation, inferential analysis for the observed between-individual variability at a single time point was performed using PROC UNIVARIATE, whereas analysis for the true between-individual variability using repeated measurements was performed via mixed modeling (PROC MIXED) in SAS. Outcome measures were modeled as single observations and percentages were treated as continuous data. To estimate the true between-individual variability, I treated participants (Subject ID) as a random effect. The linear mixed model was fitted using the restricted maximum likelihood method, and 95% CI for the overall mean, and within- and between-individual variability were derived from the model. The 95% CIs for the group mean values were calculated using the  $t$ -distribution, whereas the 95% CIs for the SD representing the inter-individual variability were calculated using the  $\chi^2$ -distribution. Normality was checked via histograms and Q-Q plots inspection.

# 5

## RESULTS

*'The good thing about science is that it's true  
whether or not you believe in it.'*

*Neil deGrasse Tyson – astrophysicist and science communicator*





## 5. RESULTS

### 5.1 Individualizing HIIT according to the physiological characteristics

Thirteen runners who belonged to IND and 13 who belonged to STD dropped from the investigation due to various injuries (3 men in IND and 1 man in STD), lack of interest, or poor training adherence (5 men and 1 woman in IND and 4 men in STD), inability to perform the prescribed training intervention (1 man and 4 women in STD), and impossibility to perform the post-training intervention measurements as a result of the restrictions imposed due to the COVID-19 pandemic by the Italian government on March 2020 (4 men in IND and 2 men and 1 woman in STD). 46 out of the 72 runners that were initially recruited completed the investigation [IND: 15 men, 8 women; age = 36 (9) years; height = 171.5 (6.7) cm, body mass = 65.1 (8.7) kg; STD: 20 men, 3 women; age = 36 (7) years; height = 175.3 (6.1) cm, body mass = 70.1 (7.0) kg]. However, gas exchange and 10000-m time data were not collected for 1 man and 1 woman in IND after the training intervention due to issues with the metabolic cart and injury, respectively.

IND ran 46.4 (15.1) km · week<sup>-1</sup> with an average frequency of 4.4 (1.0) times · week<sup>-1</sup>, while STD ran 48.7 (14.4) km · week<sup>-1</sup> with an average frequency of 4.2 (1.2) times · week<sup>-1</sup>. IND performed 70 (14)% of the overall training volume in the moderate-intensity domain, 6 (7)% in the heavy-intensity domain, and 24 (10)% in the severe-intensity domain, whereas STD performed 70 (17)% of the overall training volume in the moderate-intensity domain, 8 (12)% in the heavy-intensity domain and 23 (10)% in the severe-intensity domain. During the intervention period,  $N_{\max}$  increased from 3.8 (0.8) to 5.0 (1.0) in IND and from 4.5 (2.5) to 5.7 (2.5) in STD.

Descriptive statistics for all the outcomes are reported in Table 1, whereas inferential statistics are reported in Table 2. Relative  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  improved significantly in IND, whereas no significant changes were observed in STD. The effect of the training intervention on  $\dot{V}O_{2\max}$  (when expressed in both absolute terms and relative to the body mass) and  $V_{\text{peak}}$  was statistically larger for IND than STD. IND reduced heterogeneity of intervention effect for  $V_{\text{peak}}$  compared with STD. RE improved significantly in IND, whereas the fractional  $\dot{V}O_{2\max}$  at LT<sub>2</sub> decreased in IND and the effect of training on the fractional  $\dot{V}O_{2\max}$  at LT<sub>2</sub> was statistically larger for STD than IND.

TABLE 1. Values of the outcome variables before and after the training intervention.

	IND		STD	
	PRE	POST	PRE	POST
$\dot{V}O_{2\max}$ (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	57.8 (6.1)	60.3 (7.0)	60.7 (4.9)	61.0 (5.7)
$\dot{V}O_{2\max}$ (L · min <sup>-1</sup> )	3.78 (0.74)	3.92 (0.77)	4.21 (0.69)	4.23 (0.70)
RE (mL · kg <sup>-1</sup> · km <sup>-1</sup> )	216.4 (16.1)	212.5 (14.2)	222.7 (19.0)	218.3 (14.4)
$V_{\text{peak}}$ (km · h <sup>-1</sup> )	16.8 (1.7)	17.2 (1.8)	17.3 (0.6)	17.4 (0.9)
% $\dot{V}O_{2\max}$ at LT <sub>1</sub>	75.5 (7.8)	74.4 (5.2)	76.8 (6.8)	77.5 (6.7)
$\dot{V}O_2$ at LT <sub>1</sub> (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	43.7 (4.2)	45.0 (5.4)	46.7 (4.7)	47.1 (4.9)
vLT <sub>1</sub> (km · h <sup>-1</sup> )	12.1 (1.0)	12.6 (1.4)	12.6 (0.7)	12.9 (1.1)
% $\dot{V}O_{2\max}$ at LT <sub>2</sub>	85.9 (6.1)	84.0 (3.6)	87.4 (5.2)	87.0 (5.5)
$\dot{V}O_2$ at LT <sub>2</sub> (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	49.7 (5.2)	50.8 (5.7)	53.2 (4.4)	52.9 (4.1)
vLT <sub>2</sub> (km · h <sup>-1</sup> )	13.9 (1.5)	14.3 (1.5)	14.6 (0.6)	14.6 (0.8)
10000-m speed (km · h <sup>-1</sup> )	14.3 (1.7)	14.5 (2.2)	14.4 (1.9)	14.8 (1.0)

Data are presented as means (SD). LT<sub>1</sub>, first lactate threshold; LT<sub>2</sub>, second lactate threshold; RE, running economy; vLT<sub>1</sub>, speed at the first lactate threshold; vLT<sub>2</sub>, speed at the second lactate threshold;  $\dot{V}O_2$ , oxygen uptake;  $\dot{V}O_{2\max}$ , maximal oxygen uptake;  $V_{\text{peak}}$ , peak running speed.

TABLE 2. Covariate-adjusted change scores and heterogeneity of intervention effects.

	IND	STD	IND - STD
$VO_{2max}$ ( $mL \cdot kg^{-1} \cdot min^{-1}$ )	2.8 [1.4; 4.1]; $p < .001^*$	2.6 [2.0; 3.8]	3.7 [1.3; $\infty$ ]; $p = .01^*$
$VO_{2max}$ ( $L \cdot min^{-1}$ )	0.10 [-0.01; 0.21]; $p = .08$	0.17 [0.13; 0.26]	0.20 [0.04; $\infty$ ]; $p = .02^*$
$V_{peak}$ ( $km \cdot h^{-1}$ )	0.4 [0.2; 0.5]; $p < .001^*$	0.3 [0.3; 0.5]	0.4 [0.1; $\infty$ ]; $p = .04^*$
RE ( $mL \cdot kg^{-1} \cdot km^{-1}$ )	-4.5 [-9.1; 0.0]; $p = .05^*$	9.6 [7.3; 14.1]	-1.8 [-8.7; 5.2]; $p = .61$
% $VO_{2max}$ at LT <sub>1</sub>	-1.2 [-3.2; 0.7]; $p = .21$	4.2 [3.2; 6.1]	-4.3 [-9.1; 0.6]; $p = .08$
$VO_2$ at LT <sub>1</sub> ( $mL \cdot kg^{-1} \cdot min^{-1}$ )	0.6 [-1.8; 3.0]; $p = .62$	4.7 [3.6; 6.8]	-0.6 [-4.3; 3.2]; $p = .76$
vLT <sub>1</sub> ( $km \cdot h^{-1}$ )	0.5 [-0.1; 1.0]; $p = .08$	1.0 [0.8; 1.5]	0.0 [-0.8; 0.9]; $p = .008^*$
% $VO_{2max}$ at LT <sub>2</sub>	-2.2 [-3.8; -0.7]; $p = .007^*$	3.4 [2.6; 4.9]	-5.0 [-8.7; -1.4]; $p = .93$
$VO_2$ at LT <sub>2</sub> ( $mL \cdot kg^{-1} \cdot min^{-1}$ )	0.1 [-2.1; 2.3]; $p = .93$	4.3 [3.3; 6.3]	-0.1 [-3.0; 2.7]; $p = .40$
vLT <sub>2</sub> ( $km \cdot h^{-1}$ )	0.2 [-0.1; 0.6]; $p = .18$	0.7 [0.6; 1.1]	0.2 [-0.3; 0.8]; $p = .53$
10000-m speed ( $km \cdot h^{-1}$ )	0.2 [-0.1; 0.5]; $p = .11$	0.5 [0.4; 0.8]	-0.1 [-0.7; 0.5]; $p = .72$

Covariate-adjusted change scores are presented as least-squares means [95% CI], whereas heterogeneity of intervention effects is presented as SD [95% CI] with negative values that indicate smaller variability in IND than STD. Confirmatory hypotheses ( $VO_{2max}$ ,  $V_{peak}$ , 10000-m speed) are one-sided, whereas exploratory hypotheses are two-sided. Since the SD of the change scores within each group must be greater than zero,  $p$ -values are not reported for these metrics. The 95% CIs for the SD of the change scores within each group are calculated using the  $\chi^2$ -distribution, the 95% CIs and  $p$ -values for the SD of the difference in change scores between the groups are calculated using the normal distribution, whereas the 95% CIs and  $p$ -values for all the other metrics are calculated using the  $t$ -distribution. The abbreviations are the same as those reported in Table 1.

## 5.2 Testing interindividual variability in different treadmill-related measures and race pace-related measures

Four runners (3 men and 1 woman) did not complete the second 10000-m trial due to a close competition (1), family (1), or work commitments (2). The physiological and characteristics of the 75 runners who were recruited to estimate the observed between-individual variability [58 men, 17 women; age = 37 (8) years; height = 174.1 (6.6) cm, body mass = 69.3 (9.4) kg], the 61 who were recruited to estimate the true between-individual variability [46 men, 15 women; age = 37 (8) years; height = 173.9 (6.5) cm, body mass = 68.4 (8.6) kg] are reported in Table 3.

TABLE 3. Physiological and performance characteristics of the runners.

	75 RUNNERS	61 RUNNERS
$\dot{V}O_{2max}$ (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	58.6 (5.8)	58.3 (6.1)
$\dot{V}O_{2max}$ (L · min <sup>-1</sup> )	4.04 (0.76)	3.97 (0.76)
RE (mL · kg <sup>-1</sup> · km <sup>-1</sup> )	215.6 (18.5)	216.5 (19.0)
$V_{peak}$ (km · h <sup>-1</sup> )	17.0 (1.3)	17.1 (1.5)
$vLT_1$ (km · h <sup>-1</sup> )	12.3 (1.0)	12.3 (1.0)
$vLT_2$ (km · h <sup>-1</sup> )	14.2 (1.1)	14.3 (1.3)
10000-m speed (km · h <sup>-1</sup> )	14.2 (1.2)	14.5 (1.6)

Data are presented as means (SD).  $LT_1$ , first lactate threshold;  $LT_2$ , second lactate threshold; RE, running economy;  $vLT_1$ , speed at the first lactate threshold;  $vLT_2$ , speed at the second lactate threshold;  $\dot{V}O_2$ , oxygen uptake;  $\dot{V}O_{2max}$ , maximal oxygen uptake;  $V_{peak}$ , peak running speed.

The observed  $v\Delta 50$  corresponded to the 110% (95% CI: 109 to 111) of the 10000-m speed with a SD of 5% (95% CI: 4 to 6) and it was 1.4 km · h<sup>-1</sup> (95% CI: 1.3 to 1.6) faster than the 10000-m speed with an SD of 0.7 km · h<sup>-1</sup> (95% CI: 0.6 to 0.8). Instead, the observed  $v\Delta 50$  corresponded to the 92% (95% CI: 91 to 92) with an SD of 2% (95% CI: 0 to 2) and it was 1.4 km · h<sup>-1</sup> (95% CI: 1.5 to 1.3) slower than the  $V_{peak}$  with an SD of 0.5 km · h<sup>-1</sup> (95% CI: 0.4 to

0.8). After having accounted for the within-individual variability, the inter-individual variability in 10000-m speed at  $v\Delta 50$  was 4% (95% CI: 3 to 6) when expressed as a percentage of 10000-m speed and  $0.5 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.4 to 0.8) when expressed as the difference between  $v\Delta 50$  and 10000-m speed. Instead, the inter-individual variability in  $V_{\text{peak}}$  at  $v\Delta 50$  was 1% (95% CI: 1 to 2) when expressed as a percentage of  $V_{\text{peak}}$  and  $0.2 \text{ km} \cdot \text{h}^{-1}$  (95% CI: 0.2 to 0.3) when expressed as the difference between  $v\Delta 50$  and  $V_{\text{peak}}$ .

### 5.3 Individualizing HIIT prescriptions: physiological characteristics vs race pace

Due to COVID-19 restrictions imposed by the Italian government on November 2020, only 30 runners out of 38 (15 for each group) completed all the measurements at the end of the first recruitment sequence, while one extra runner belonging to PHY completed the laboratory measurements only [RP: 9 men, 6 women; age = 41 (8) years; height = 173.7 (7.5) cm, body mass = 65.5 (9.7) kg; PHY: 11 men, 5 women; age = 36 (10) years; height = 171.3 (6.6) cm, body mass = 65.6 (8.1) kg].

PHY ran 45.6 (17.4)  $\text{km} \cdot \text{week}^{-1}$  with an average frequency of 4.4 (1.2)  $\text{times} \cdot \text{week}^{-1}$ , while RP ran 45.9 (15.2)  $\text{km} \cdot \text{week}^{-1}$  with an average frequency of 4.0 (1.4)  $\text{times} \cdot \text{week}^{-1}$ . PHY performed 68 (16)% of the overall training volume in the moderate-intensity domain, 6 (8)% in the heavy-intensity domain, and 26 (11)% in the severe-intensity domain, whereas RP performed 64 (16)% of the overall training volume in the moderate-intensity domain, 8 (13)% in the heavy-intensity domain and 28 (14)% in the severe-intensity domain. During the intervention period,  $N_{\text{max}}$  increased from 3.6 (1.0) to 4.6 (0.9) in PHY and from 4.7 (1.3) to 5.7 (1.8) in RP.

Descriptive statistics for the 31 runners are reported in Table 4, whereas inferential statistics are reported in Table 5. The statistics that are reported for each outcome are calculated based only on the runners who completed the pre-training and post-training measurements.  $\dot{V}O_{2\text{max}}$  and  $V_{\text{peak}}$  increased significantly in PHY. However,  $\dot{V}O_{2\text{max}}$  also decreased in RP. The effect of the training intervention on  $\dot{V}O_{2\text{max}}$  and  $V_{\text{peak}}$  was statistically larger in PHY compared with RP. The 10000-m speed increased significantly in RP and the effect of the training intervention was statistically larger compared with PHY. The fractional  $\dot{V}O_{2\text{max}}$  at  $LT_1$  decreased in RP and the effect of the training intervention was statistically smaller compared with PHY.



TABLE 4. Values of the outcome variables before and after the training intervention.

	PHY		RP	
	PRE	POST	PRE	POST
$\dot{V}O_{2max}$ (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	58.0 (6.5)	60.9 (7.1)	54.8 (6.1)	52.8 (6.4)
$\dot{V}O_{2max}$ (L · min <sup>-1</sup> )	3.82 (0.74)	3.97 (0.76)	3.58 (0.65)	3.50 (0.63)
$V_{peak}$ (km · h <sup>-1</sup> )	16.9 (1.7)	17.4 (1.8)	16.6 (1.5)	16.5 (1.5)
10000-m speed (km · h <sup>-1</sup> )	14.4 (1.6)	14.6 (2.1)	13.7 (1.1)	14.1 (1.1)
RE (mL · kg <sup>-1</sup> · km <sup>-1</sup> )	212.6 (16.7)	210.2 (15.4)	200.0 (16.2)	200.5 (14.1)
% $\dot{V}O_{2max}$ at LT <sub>1</sub>	75.1 (8.0)	74.1 (5.2)	74.2 (6.3)	78.5 (6.4)
$\dot{V}O_2$ at LT <sub>1</sub> (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	43.2 (4.0)	45.4 (5.7)	40.5 (2.9)	41.2 (3.5)
vLT <sub>1</sub> (km · h <sup>-1</sup> )	12.1 (0.9)	12.7 (0.3)	12.2 (1.1)	12.3 (1.2)
% $\dot{V}O_{2max}$ at LT <sub>2</sub>	85.3 (6.5)	84.1 (3.3)	84.1 (6.6)	87.8 (5.7)
$\dot{V}O_2$ at LT <sub>2</sub> (mL · kg <sup>-1</sup> · min <sup>-1</sup> )	49.2 (4.6)	51.5 (5.9)	45.9 (4.3)	46.2 (4.9)
vLT <sub>2</sub> (km · h <sup>-1</sup> )	14.0 (1.5)	14.5 (1.5)	14.0 (1.1)	13.9 (1.3)

Data are presented as means (SD). LT<sub>1</sub>, first lactate threshold; LT<sub>2</sub>, second lactate threshold; RE, running economy; vLT<sub>1</sub>, speed at the first lactate threshold; vLT<sub>2</sub>, speed at the second lactate threshold;  $\dot{V}O_2$ , oxygen uptake;  $\dot{V}O_{2max}$ , maximal oxygen uptake;  $V_{peak}$ , peak running speed.

TABLE 5. Covariate-adjusted change scores and heterogeneity of intervention effects.

	PHY		RP		PHY - RP	
$VO_{2max}$ ( $mL \cdot kg^{-1} \cdot min^{-1}$ )	3.1 [1.8; 4.4]; $p < .001^*$	2.2 [1.6; 3.6]	-2.0 [-3.4; -0.6]; $p = .009^*$	2.4 [1.7; 3.9]	5.1 [3.6; $\infty$ ]; $p < .001^*$	-0.8 [- $\infty$ ; 2.1]; $p = .42$
$VO_{2max}$ ( $L \cdot min^{-1}$ )	0.13 [0.03; 0.23]; $p = .02^*$	0.17 [0.13; 0.28]	-0.11 [-0.22; 0.00]; $p = .05^*$	0.18 [0.13; 0.30]	0.24 [0.12; $\infty$ ]; $p = .001^*$	-0.06 [- $\infty$ ; 0.16]; $p = .43$
$V_{peak}$ ( $km \cdot h^{-1}$ )	0.4 [0.2; 0.7]; $p < .001^*$	0.4 [0.3; 0.6]	-0.1 [-0.4; 0.2]; $p = .51$	0.5 [0.4; 0.9]	0.5 [0.2; $\infty$ ]; $p = .003^*$	-0.4 [- $\infty$ ; 0.3]; $p = .13$
10000-m speed ( $km \cdot h^{-1}$ )	0.0 [-0.3; 0.4]; $p = .95$	0.6 [0.4; 0.9]	0.5 [0.2; 0.7]; $p < .001^*$	0.4 [0.3; 0.6]	-0.5 [- $\infty$ ; 0.1]; $p = .51$	0.4 [-0.6; $\infty$ ]; $p = .25$
RE ( $mL \cdot kg^{-1} \cdot km^{-1}$ )	0.3 [-5.9; 6.4]; $p = .92$	9.7 [7.0; 15.5]	-2.7 [-10.4; 4.9]; $p = .45$	12.4 [8.9; 20.5]	3.0 [-6.3; 12.3]; $p = .05^*$	-7.8 [-14.3; 9.0]; $p = .40$
% $VO_{2max}$ at LT <sub>1</sub>	0.1 [-2.8; 2.7]; $p = .96$	4.5 [3.3; 7.3]	4.1 [0.7; 7.5]; $p = .02^*$	5.9 [4.3; 9.8]	-4.2 [-8.3; 0.0]; $p = .20$	-3.8 [-6.9; 4.2]; $p = .37$
$VO_2$ at LT <sub>1</sub> ( $mL \cdot kg^{-1} \cdot min^{-1}$ )	2.3 [-0.8; 5.4]; $p = .21$	5.2 [3.8; 8.3]	0.0 [-1.9; 4.1]; $p = .44$	3.4 [2.4; 5.6]	2.3 [-1.3; 5.9]; $p = .36$	3.9 [-2.7; 6.2]; $p = .19$
$vLT_1$ ( $km \cdot h^{-1}$ )	0.5 [-0.2; 1.1]; $p = .17$	1.1 [0.8; 1.8]	0.1 [-0.3; 0.5]; $p = .57$	0.7 [0.5; 1.2]	0.3 [-0.4; 1.1]; $p = .07$	0.9 [-0.6; 1.3]; $p = .10$
% $VO_{2max}$ at LT <sub>2</sub>	-0.2 [-2.1; 1.6]; $p = .79$	3.1 [2.2; 4.9]	3.0 [-0.2; 6.3]; $p = .67$	5.6 [4.0; 9.3]	-3.3 [-6.9; 0.3]; $p = .93$	-4.7 [-7.0; 2.0]; $p = .90$
$VO_2$ at LT <sub>2</sub> ( $mL \cdot kg^{-1} \cdot min^{-1}$ )	2.3 [-0.4; 5.0]; $p = .10$	4.6 [3.3; 7.5]	-0.5 [-3.3; 2.2]; $p = .56$	4.4 [3.2; 7.3]	-0.1 [-3.0; 2.7]; $p = .93$	1.2 [-4.6; 4.9]; $p = .42$
$vLT_2$ ( $km \cdot h^{-1}$ )	0.4 [-0.1; 0.9]; $p = .10$	0.8 [0.6; 1.9]	-0.1 [-0.5; 0.3]; $p = .56$	0.7 [0.5; 1.1]	0.5 [-0.1; 1.1]; $p = .09$	0.5 [-0.6; 0.9]; $p = .42$

Covariate-adjusted change scores are presented as least-squares means [95% CI], whereas heterogeneity of intervention effects is presented as SD [95% CI] with negative values that indicate smaller variability in PHY than RP. Confirmatory hypotheses ( $VO_{2max}$ ,  $V_{peak}$ , 10000-m speed) are one-sided, whereas exploratory hypotheses are two-sided. Since the SD of the change scores within each group must be greater than zero,  $p$ -values are not reported for these metrics. The 95% CIs for the SD of the change scores within each group are calculated using the  $\chi^2$ -distribution, the 95% CIs and  $p$ -values for the SD of the difference in change scores between the groups are calculated using the normal distribution, whereas the 95% CIs and  $p$ -values for all the other metrics are calculated using the  $t$ -distribution. The abbreviations are the same as those reported in Table 4.



## 6

**DISCUSSION**

*'Without proper interpretation, data is just noise.'*

*Unknown*





## 6. DISCUSSION

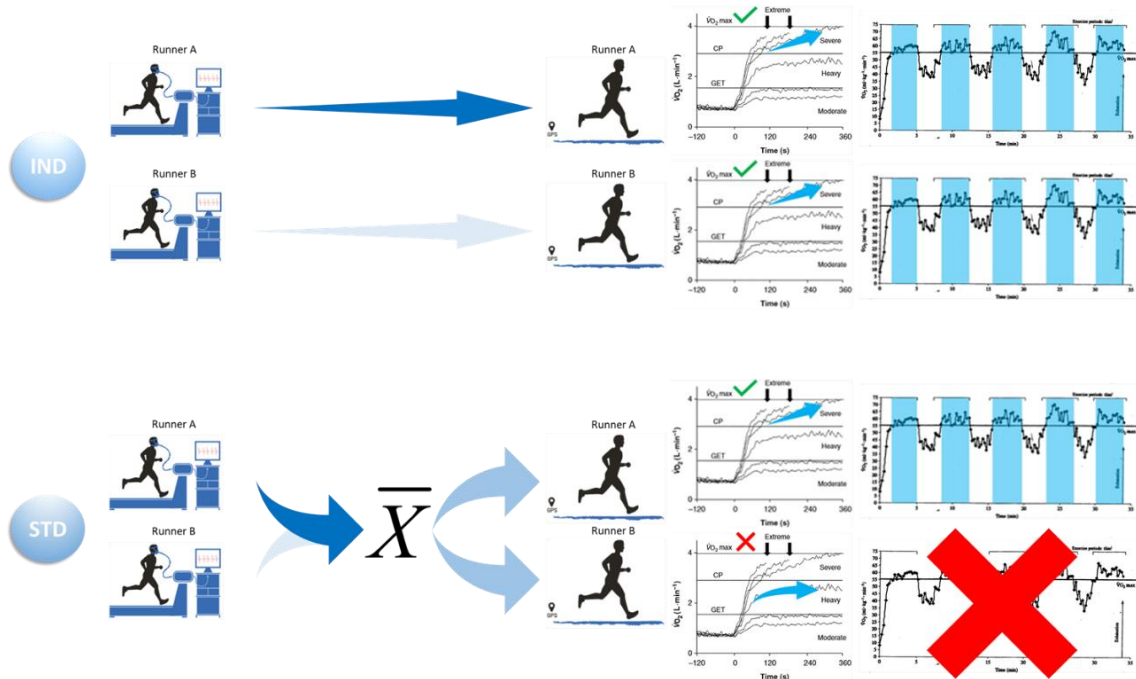
In my thesis, I aimed to clarify several still open questions about the effects of different HIIT prescription models on physiological response and performance in distance runners. This aim was achieved through three different research questions, which focused on separated but related aspects of the topic. The first question was whether a highly-individualized, physiologically-based approach to HIIT prescription was more efficacious (i.e., superior) than a standardized one that imposes the same HIIT scheme on all runners in developing cardiorespiratory fitness, in terms of  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$ , and reducing heterogeneity of training effects. The experimental data corroborated both my hypotheses, which led me to conclude that individualizing HIIT prescription according to physiological characteristics and response to exercise should be the choice for coaches and athletes in those training phases aimed to improve cardiorespiratory fitness (usually during the preparatory phase and the early part of the specific period far from major competitions). The second aim was to look for alternative parameters to estimate the  $v\Delta 50$  that did not require laboratory tests. I found that data from an incremental treadmill test conducted without any lab measurements – namely the relative percent of the  $V_{\text{peak}}$  and the absolute difference between  $V_{\text{peak}}$  and  $v\Delta 50$  – can inform about the  $v\Delta 50$  value with sufficient heterogeneity between individuals to be used as valid alternatives to the delta concept. On the contrary, the larger variability in race pace-derive measures at  $v\Delta 50$  does not make them suitable for prescribing HIIT according to the physiologically-based approach. The third question was whether the method used to prescribe HIIT intensity (i.e., physiological parameters *vs* race pace) would have affected the magnitude and the heterogeneity of cardiorespiratory adaptations and 10-km performance in moderately-trained distance runners. Although the experimental data were inconclusive for heterogeneity in the intervention effect, they corroborated both my hypotheses about the mean training effect, which led me to conclude that the two training approaches may be possible within the same macrocycle, but the implementation of the physiologically-based should precede the implementation of the race pace-based approach in terms of proximity to major competitions.

### 6.1 Individualizing HIIT according to the physiological characteristics

Previous research showed that performing prolonged intervals at intensities near  $\dot{V}O_{2\max}$  induces statistically significant improvements in  $\dot{V}O_{2\max}$  compared with lower-intensity

(domain) protocols in trained endurance runners (Helgerud et al., 2007; Stöggl & Sperlich, 2014). Several authors suggested personalizing HIIT prescriptions according to the physiological characteristics of the runners and their responses to exercise to maximize cardiorespiratory adaptations (see Buchheit & Laursen, 2013a, b for an overview). I investigated whether individualizing HIIT prescriptions using this approach would have induced higher adaptations in cardiorespiratory fitness, in terms of  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$ , and a smaller heterogeneity of training effects than standardized prescriptions, which are often adopted by distance runners because of their stronger social compatibility (Casado et al., 2019). The experimental data corroborated my first hypothesis since the mean intervention effect for  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  was statistically larger in IND than STD. Moreover, the effect size for the relative  $\dot{V}O_{2\max}$  ( $3.7 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) was similar to, or slightly lower than, the one observed by Helgerud and colleagues ( $3.7 - 3.9 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) when comparing a similar HIIT program with lower-intensity protocols. Although the smaller heterogeneity of training effects in IND than STD was not statistically significant for  $\dot{V}O_{2\max}$ , this was the case for  $V_{\text{peak}}$ . These apparent discrepancies are likely explained by the higher reliability of  $V_{\text{peak}}$  than  $\dot{V}O_{2\max}$  measures (Hopkins et al., 2001) and by the fact that the sample size for the investigation (which was also affected by the high dropout rate) was based on the mean training effect, which requires smaller sample sizes than for investigating and training response heterogeneity (Vallejo et al., 2018). These novel findings represent an important turning point for HIIT prescriptions since they highlighted the trade-off between generalizability and cardiorespiratory gains. The use of physiological anchors to prescribe exercise intensity aims to homogenize the cardiorespiratory stimulus across individuals. A major advantage of using this approach to individualize HIIT prescriptions is that reasonable errors in estimating exercise intensity, which would affect the time required to reach  $\dot{V}O_{2\max}$  (Burnley & Jones, 2007), can be compensated by an increased or reduced interval duration or the number of intervals, which would mitigate the impact of the initial misestimation on the total time spent at  $\dot{V}O_{2\max}$  (Billat et al., 2000). This mechanism relies on the presence of a continuous slow component of  $\dot{V}O_2$  kinetics that allows  $\dot{V}O_2$  to reach maximal values for sufficiently prolonged exercise, even when interspersed with brief recovery periods (Burnley & Jones, 2007; Jones et al., 2010). However, if the HIIT intensity happens to fall below the boundary of the severe domain (i.e., below  $vLT_2$ ), either as a consequence of a substantial underestimation or of an improvement in the  $vLT_2$  during the training period, any such compensatory mechanism is no longer possible, impeding to runners to reach their  $\dot{V}O_{2\max}$  during the HIIT session. This may result in a negligible cardiorespiratory stimulus in these individuals (Wenger & Bell, 1986, Midgley & McNaughton,

2006; Midgley et al., 2006b, 2007b) and consequently a large training response heterogeneity (Figure 19). The large individual variability in the exercise intensity observed in STD when running at the (same) prescribed speed [i.e., 46 (23)% of  $\Delta$  between the  $vLT_2$  and  $V_{peak}$ ] and, especially, the statistically larger heterogeneity of training effects on  $V_{peak}$  in STD than IND agree with this theory-driven rationale.



**FIGURE 19.** Representation of the physiological rationale underlying this investigation. The much larger chance that the prescribed HIIT intensity would shift from the severe to the heavy domain in STD than IND during the training period would result in a much larger possibility than some runners may face blunted – if not null – improvements in cardiorespiratory fitness in STD than IND due to a reduced/zero time spent at (or near)  $\dot{V}O_{2max}$ . This would in turn result in an overall smaller mean training effect and larger heterogeneity of intervention effects in STD compared with IND (modified from Demarie et al., 2000).

Conducting a mediation analysis on the physiological determinants of  $\dot{V}O_{2max}$  response to the different HIIT interventions was beyond the scope of my investigation. Helgerud and colleagues (2007) observed within-group improvements in the maximal stroke volume and cardiac output following eight weeks of HIIT protocol with either short (15 s) or long (4 min) intervals in trained individuals. However, no statistical differences were observed between these HIIT protocols and when compared with lower-intensity protocols by the authors. Moreover, the authors did not detect any statistical difference in the blood volume and oxygen-



carrying capacity within and between the interventions. Despite the lack of data, there is no reason to think that the observed differences in the  $\dot{V}O_{2\max}$  response between the two different HIIT protocols investigated depend on different physiological determinants according to the ‘classical’ model of  $\dot{V}O_{2\max}$  (Basset & Howley, 2000; Joyner & Coyle, 2008), since the exercise-induced mechanical overload that occurs at intensities that elicit the  $\dot{V}O_{2\max}$  represents the main mediating physiological mechanism for  $\dot{V}O_{2\max}$  improvements in already trained runners (Midgley et al., 2006b, 2007b).

The RE significantly improved in IND but not in STD and the two groups were not statistically different in their change scores. Previous uncontrolled (Demarle et al., 2001, 2003; Slawinski et al., 2001; Lafitte et al., 2003) and controlled studies comparing submaximal HIIT against lower-intensity protocols for what concerns within-group analysis (Helgerud et al., 2007; Stöggl & Sperlich, 2014) detected a significant improvement in RE following similar HIIT interventions. However, before being tempted to interpret these findings as evidence of efficacy, it is important to keep in mind the limits in terms of error rate control and informativeness of using separate analysis of changes from baseline in each parallel group when making statistical inference in randomized studies (Matthews & Altman, 1996; Bland & Altman, 2011, 2015). While the former point equally applies in all circumstances, the latter becomes particularly relevant in training intervention studies with athletes since the within-group analysis cannot isolate the effect of the experimental intervention from the rest of the training, which may be – perhaps highly – dependent on the phase of the season when the study or investigation is conducted. This makes it way more difficult to formulate and test precise hypotheses using only within-group analysis in the athlete population.

I did not observe any statistical within-group or between-group difference in the lactate thresholds when expressed in terms of  $\dot{V}O_2$  or speed. However, the fractional  $\dot{V}O_{2\max}$  at  $vLT_2$  significantly decreased in IND and it was inferior when compared with STD. Previous studies detected (Helgerud et al., 2007; Stöggl & Sperlich, 2014) and failed to detect (Demarle et al., 2001, 2003; Slawinski et al., 2001; Lafitte et al., 2003; Garcin et al., 2004; Helgerud et al., 2007; Stöggl & Sperlich, 2014) statistical changes in the lactate threshold(s) following similar submaximal HIIT protocols. Stöggl and Sperlich (2014), but not Helgerud and colleagues (2007), observed larger improvements in the speed at the lactate threshold after HIIT intervention than after moderate-intensity or heavy-intensity continuous training. Several factors, such as the choice of the measure (Basset & Howley, 2000; Midgley et al., 2007b), its method of determination (Faude et al., 2009; Poole et al., 2021), the different research designs,

and the small sample sizes characterizing these studies (Speed & Andersen, 2000), may explain these (real or apparent) discrepancies. Within the context of my investigation, I speculate that improvements in cardiorespiratory fitness and RE in IND might have contributed to reducing the fractional  $\dot{V}O_{2max}$  at  $vLT_2$  (Midgley et al., 2007b).

I did not observe any statistical change in 10000-m performance within each group nor any difference between the groups. These results may appear unexpected when considering the important role that  $\dot{V}O_{2max}$  and especially  $V_{peak}$  play in determining 10000-m performance in recreational (Machado et al., 2013) and trained (Morgan et al., 1989; Noakes et al., 1990; Evans et al., 1995) distance runners. When interpreting these findings, we should keep in mind that 10000-m performance – and more generally distance running performance – represents a rather complex phenomenon involving several psychobiological factors interacting at multiple levels (Renfree & Casado, 2018). Even when attempting to minimize the weight of the psychological and environmental components as I did when designing the time trial, several physiological factors may influence to a various extent the performance (Basset & Howley, 2000; Midgley et al., 2007b). Noteworthy, previous literature only provides a rather imprecise (and sometimes contradictory) estimation of the role of each physiological determinant, which appears to vary according to the race distance and runners' characteristics (Basset & Howley, 2000; Joyner & Coyle, 2008; Kenneally et al., 2021a). Kenneally and colleagues (2018, 2021a, b) suggested that training prescriptions according to a percentage of the (event-specific) target race pace may be more successful than prescriptions according to physiological characteristics in maximizing running performance because of their ability to stimulate the optimal combination of physiological (and possibly extra-physiological) characteristics required for that given task. Although these authors developed their reasoning from a TID perspective, the 'race pace approach' may also be extended to exercise prescription and represent an interesting candidate for future investigations on HIIT.

A missing piece of information in my investigation relates to the lack of data about the total time spent at (or near)  $\dot{V}O_{2max}$  with the two different HIIT protocols. To my knowledge, very few longitudinal HIIT studies included this measure (e.g., Turnes et al., 2015). The main reason why I decided to not include this measurement in my design related to its poor reliability (Midgley et al., 2007d), which would have required either an unrealistically larger sample or a substantially increased burden to participants due to the more testing sessions that would have been needed to achieve sufficient power to detect plausible differences between the groups. In my investigation, I manipulated the submaximal HIIT component leaving the rest of the training

substantially unchanged other than for the replacement that was necessary to maintain a similar training volume to the runners' habitual practice. This approach isolates the effect of the experimental HIIT interventions from the rest of the training when intervention allocation is randomized, thus avoiding confounding (Mansournia et al., 2018). That being said, it is worth mentioning that the implementation of my training program resulted in a polarized TID for most of the runners in IND and STD. Coaches usually adopt this TID scheme to build cardiorespiratory fitness and conditioning during the preparatory phase and the early part of the specific period far from major competitions (Kenneally et al., 2018, 2020a). Previous research found that the polarized TID led to superior improvements in physiological characteristics associated with endurance performance compared with other TIDs (Stögg & Sperlich, 2014, 2015; Rosenblat et al., 2018). Although the resulting TID in my investigation matched with the typical coaching practice, the effect of the implementation of the two HIIT protocols with different TIDs (i.e., pyramidal) remains unknown. Similarly, the impact of using different periodization models (e.g., block periodization) than my more traditional approach on the difference between the two HIIT protocols requires further investigation. From a statistical and methodological perspective, I opted for a per-protocol analysis, which refers to inclusion in the analysis of only those individuals who completed the investigation without major deviation from the experimental training protocol (Moher et al., 2010). I chose this population analysis – which is often the standard in exercise physiology and sport sciences – since I was more interested in the causal effect that would have been observed if all runners had adhered to the training protocols rather than the effect of training prescriptions *per se*. One main drawback of the per-protocol analysis relates to the increased risk of selection *bias* due to non-random adherence from participants with the assigned interventions, which may affect both average treatment estimate and heterogeneity of training effects (Mansournia et al., 2018; Bonafiglia et al., 2019). Although the overall dropout rate was similar between the groups, the dropout analysis revealed that the initial training level might have played an important role in determining more dropouts for STD than IND. By applying the same fixed HIIT characteristics to a heterogeneous group of runners in STD, I indirectly forced those runners who were not able to sustain the prescribed training protocol to withdraw from the investigation. This restricted the sample to the higher caliber runners, for which smaller gains and possibly larger heterogeneity in  $\dot{V}O_{2\max}$  response were expected. Although I partly accounted for selection bias by adjusting the analysis for pre-training scores, this approach cannot completely remove selection *bias* (Mansournia et al., 2018). An important limitation that is not related to the research design but the applicability of its findings relates to the extrinsic barriers for trained

distance runners to obtain individualized HIIT prescriptions according to physiological characteristics and response to exercise. Since my individualized approach would require at least a visit to a physiology lab in each mesocycle, many amateur runners may view the trade-off between costs of training prescriptions and the resulting benefits as non-convenient. In this regard, future research should identify valid field-based alternatives to induce a similar cardiorespiratory stimulus without the need of relying on laboratory testing.

## **6.2 Testing interindividual variability in different treadmill-related measures and race pace-related measures**

I previously showed that individualizing HIIT prescriptions using the delta concept led to superior and more homogeneous cardiorespiratory adaptations than standardized protocols and reduce the heterogeneity of training effects. Notwithstanding these proven benefits, the main limit to the implementability of the delta concept for exercise prescription is represented by the need for accessibility to physiology labs or medical clinics and expert personnel to properly conduct the measurements required to obtain valid and reliable data for exercise prescription. Therefore, I investigated whether more practical and accessible treadmill-related measures could be used as a valid surrogate of the delta concept by assessing the mean value and the individual variability occurring at  $v\Delta 50$ , an intensity that has been largely used to prescribe HIIT interventions in the literature (Billat et al., 2000; Demarie et al., 2000; Demarle et al., 2001, 2003; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003). By using the same methodological approach, I also investigated whether 10000-m speed can be used to approximate  $v\Delta 50$ , as previously suggested by different authors (Billat, 2001, Billat et al., 2002). To separate the true (inter-)individual variability from the biological (intra-individual) variability, I conducted repeated measurements on a subgroup of runners.

I identified with discrete precision the mean values of  $V_{\text{peak}}$  and 10000-m speed at which the  $v\Delta 50$  occurs. Despite the slightly different incremental treadmill protocol and the definition I used to define the  $V_{\text{peak}}$ , the point estimates for the difference between  $v\Delta 50$  and  $V_{\text{peak}}$  (i.e.,  $-1.4 \text{ km} \cdot \text{h}^{-1}$ ) and the percentage of  $v\Delta 50$  at which  $V_{\text{peak}}$  occurs (i.e., 92%) were very similar to the estimates previously observed in the literature (Demarle et al., 2000; Billat et al., 2000; Demarie et al., 2000; Demarle et al., 2001, 2003; Slawinski et al., 2001; Garcin et al., 2002; Lafitte et al., 2003). However, contrary to previous recommendations (Billat, 2001, Billat et al., 2002), the  $v\Delta 50$  was substantially ( $\sim 10\%$ ) higher than the 10000-m speed. The lower training

level of my runners compared to the elite ones investigated by Billat and colleagues (Billat, 2001, Billat et al., 2002) may explain these discrepancies.

In addition to the mean values of my candidate measures, I also quantified their inter-individual variability – expressed as SD – as a means to check whether such measures were consistent across runners. In this regard, I observed a lower variability at  $v\Delta 50$  when using  $V_{\text{peak}}$  than when using 10000-m speed expressed as both absolute difference and relative percentage. That being said, when considering the uncertainty surrounding the point estimate (i.e., the 95% CI), the maximal plausible variability at  $v\Delta 50$  expressed as the absolute difference from  $v\Delta 50$  was similar when using  $V_{\text{peak}}$  and 10000-m speed. However, after having accounted for the within-individual variability, the inter-individual variability of  $V_{\text{peak}}$  at  $v\Delta 50$  was substantially lower than the variability of 10000-m speed at  $v\Delta 50$ , regardless of the method used to express the value.

When using physiologically-based approaches to individualize HIIT prescriptions for maximizing cardiorespiratory adaptations, it is critical to determine the target intensity with high precision. This means to ensure that the training intensity is set sufficiently far from the limits of the (severe) domain to maximize the time spent at (or near)  $\dot{V}O_{2\text{max}}$ . By using a much larger sample size than what is generally used in the literature (to obtain a sufficient precision in parameter estimation) and repeated measurements (to isolate the true inter-individual variability by the biological variability), I corroborated my hypothesis that the variability in the absolute difference from  $V_{\text{peak}}$  and  $v\Delta 50$  and in the percentage of  $V_{\text{peak}}$  at  $v\Delta 50$  between runners is indeed low and can thus be used to individualize HIIT prescriptions according to the physiological approach whenever a direct measure of the  $v\Delta 50$  is not easily available (i.e., in the most of the common scenarios). On the contrary, the larger variability in the absolute difference and relative percentage of 10000-m speed at  $v\Delta 50$  does not provide sufficient precision to use this variable for individualizing this form of HIIT in this population. Since we make inferences on the population but we prescribe on individuals, estimates that have a large variability at the desired value in the population level may result in misprescriptions and suboptimal training outcomes at the individual level and therefore should be discarded in favor of less variable measures.

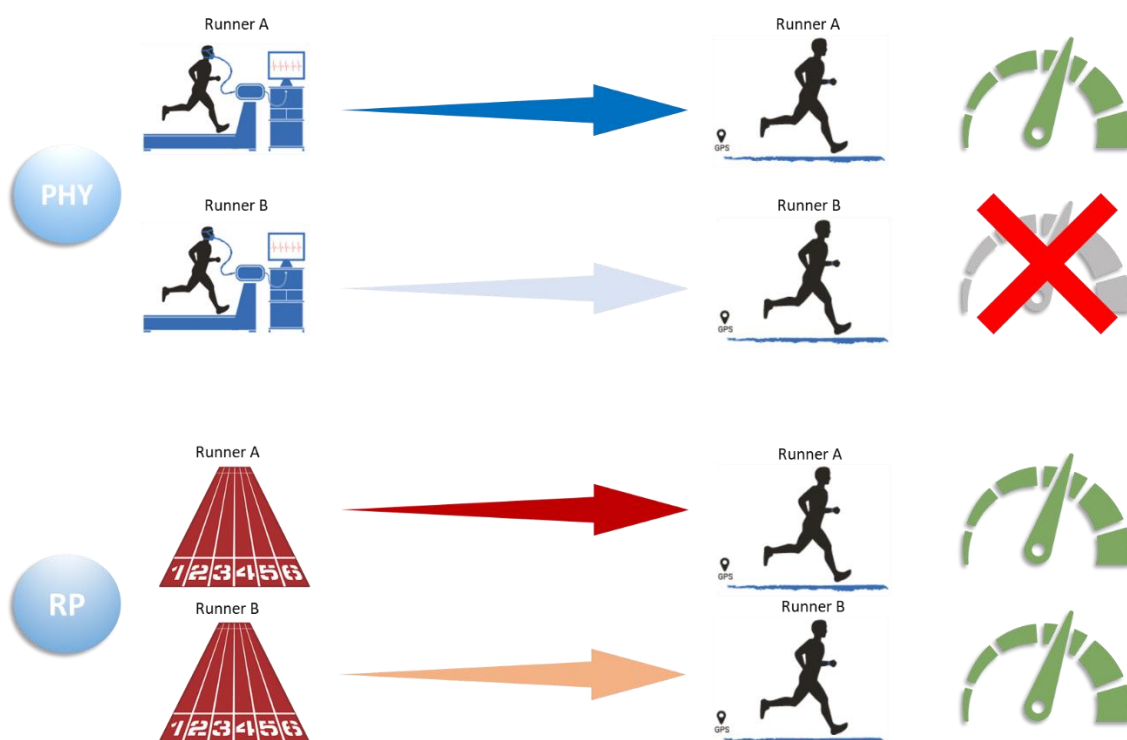
A possible limitation of this investigation regards the information provided by its main inferential model. Although the model allowed us to isolate the true inter-individual variability from the random error, it did not allow us to investigate the effect of the training level on the mean value and the variability of the measures in this population. Previous studies showed that

higher caliber runners may be able to maintain a  $\dot{V}O_2$  steady-state at higher percentages of the  $\dot{V}O_{2max}$  than what is generally expected for trained runners (Billat et al., 1998). Since  $v\Delta 50$  and the delta concept, in general, is based on two physiological parameters – namely  $vLT_2$  and  $V_{peak}$  [which is in turn determined by  $\dot{V}O_{2max}$  and RE (Billat & Koralsztein, 1996; Hill & Rowell, 1996)] – any difference in the variability in these measures compared to lower caliber runners, all else equal, would inevitably affect the variability in these measures at  $v\Delta 50$ . In this regard, it is worth mentioning that higher caliber runners may be more homogeneous than the lower caliber counterpart in some measures but not necessarily in others (Conley & Krahenbuhl, 1980). A key point when evaluating the impact of this limitation on the informativeness of this work is to not forget that, elite runners are those with the highest chances to perform physiological lab tests to obtain data for individualizing their training and therefore those who have the lesser need of surrogate measures for those variables. On the other hand, novice runners do not need near to maximal aerobic intensities to develop their  $\dot{V}O_{2max}$  (Wenger & Bell, 1986, Robinson et al. 1991; Midgley & McNaughton, 2006; Midgley et al., 2007b, 2007b) and therefore they are less affected by the need of matching and maintaining the proper physiological domain during their HIIT sessions.

### **6.3 Individualizing HIIT prescriptions: physiological characteristics vs race pace**

The lack of any statistical differences for 10000-m speed between the two training interventions in my first investigation indicates that there is no evidence about the benefits of individualizing HIIT prescriptions according to a physiologically-based approach for what concerns running performance. Kenneally and colleagues (2018, 2021a, b) suggested that individualizing training prescriptions according to a percentage of the target race pace may be more successful in maximizing running performance than physiologically-based prescriptions because of their ability to stimulate the optimal combination of determinants for that given task. This hypothesis is in agreement with modern network theory (Bashan et al., 2012; Balagué et al., 2020; Ivanov, 2021), according to which complex biological organisms, such as the human body, can adapt to a new task by developing a more efficient network of physiological interactions. Prescribing on a higher percentage of the target race pace may result in improved tolerance to the new training intensity and an improved capacity to sustain a faster pace in competition, provided that the intensity used to prescribed exercise is not too high to lose the specificity for the task. On the contrary, prescribing on a physiologically-based anchor (e.g.,  $v\Delta 50$ ) may lead some

runners to train at intensities that are very distant from the target race pace, risking losing the specificity for the task. This would result in a larger individual response heterogeneity in race performance as well as in smaller gains at the group level, even when the average training intensity is similar between the groups (Figure 20). I successfully corroborated the hypothesis of larger improvements in 10000-m performance following race pace-based HIIT prescriptions than physiologically-based HIIT prescriptions in my third investigation. However, although I observed a moderate variability in the percentage of 10000-m pace in PHY at  $v\Delta 50$  [108 (5)%], which is in agreement with the variability observed in my previous two investigations, I failed to corroborate the predicted smaller heterogeneity in race performance when prescribing according to race pace in the trained distance runner population. Moreover, other than a larger improvement in the  $\% \dot{V}O_{2\max}$  at  $LT_1$  in RP compared with PHY, I did not observe any superiority of the race pace-based approach over the physiologically-based for what concerns the physiological factors I investigated. Several reasons may explain these results. From a statistical perspective, since I performed sample size estimation for mean training response only for outcomes related to confirmatory hypotheses (i.e.,  $\dot{V}O_{2\max}$ ,  $V_{\text{peak}}$ , and 10000-m speed), the inferential test may have been underpowered to detect differences in the other outcomes and response heterogeneity between the interventions (Vallejo et al., 2018). From a psychophysiological perspective, it is important to mention that the cardiorespiratory and metabolic factors I investigated represent only a part of all the determinants of race performance (Basset & Howley, 2000; Joyner & Coyle, 2008), which is affected by several different factors interacting at multiple levels (Renfree & Casado, 2018). A more efficient network of physiological interactions for a 10000-m run may not necessarily yield to improvement in performance unless any harmful effect from the other determinants is overcome. Although I tried to minimize the weight of the psychological and environmental factors when assessing performance response to the two interventions, I cannot exclude differences in these factors that may have increased the observed heterogeneity in RP.



**FIGURE 20.** The physiological approach to high-intensity interval training prescription may lead some runners to train at intensities that are very distant from the target speed with an increased risk of losing the specificity for the task. The resulting inter-individual variability in the training stimulus may affect both the individual response heterogeneity and the mean training effect for 10000-m performance.

I also observed a larger effect on  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  following the physiologically-based HIIT prescriptions than race pace-based HIIT prescriptions, which instead led to a decrease in  $\dot{V}O_{2\max}$ . These findings corroborate my theory-driven hypothesis and agree with my previous findings showing how individualizing HIIT prescriptions according to physiological anchors should be the preferred choice when the training goal is to maximize cardiorespiratory fitness. However, I failed to detect a larger heterogeneity in  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  following the race pace approach compared to the physiological approach. Exercising at a sufficient intensity to reach the  $\dot{V}O_{2\max}$  (i.e., above the  $vLT_2$  or CS) seems to be a fundamental requirement to further develop cardiorespiratory fitness in already trained individuals (Wenger & Bell, 1986, Robinson et al. 1991; Midgley & Naughton, 2006; Midgley et al., 2006b, 2007b). I previously showed that exercise prescriptions that do not allow to train at the appropriate intensities for the whole duration of a target training period decrease mean training response in  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  and increase heterogeneity in training response in  $V_{\text{peak}}$  across individuals. I also showed that the race pace on a given running distance cannot be used to precisely estimate  $v\Delta 50$  due to the



moderate-to-large variability occurring across individuals from this population. In prescribing HIIT in RP, I individualized intervals speed at 110% of the target race pace, while maintaining intervals distance, recovery speed, and recovery distance fixed for all runners. I chose this approach to mimic coaching practice and provide an empirical comparator to investigate my main hypotheses and match the average intensity between RP and PHY. Prescribing HIIT using the race pace approach resulted in large variability in the physiological intensity [i.e., 40 (25)% of  $\Delta$  between  $vLT_2$  and  $V_{peak}$ ] that was similar to the one I observed when using standardized training prescriptions, with a high risk of falling outside the *minimum* training intensity to stimulate cardiorespiratory fitness adaptations in this population during the training period. It is important to remember that training close to or at  $\dot{V}O_{2max}$  represents a necessary but not sufficient condition to maximize the cardiorespiratory stimulus, and training intensity *per se* represents only one of the factors determining the time to exhaustion – and possibly the total time spent at (or near)  $\dot{V}O_{2max}$  – during intermittent high-intensity exercise (Jones et al., 2010; Jones & Vanhatalo, 2017). Although I did not measure the total time spent at (or near)  $\dot{V}O_{2max}$ , it may be possible that the larger degree of individualization allowed in RP than STD reduced the inter-individual variability in the training stimulus within RP and thus affected the differences in the heterogeneity of intervention effects between the groups (Midgley & Naughton, 2006; Buchheit & Laursen, 2013a, b).

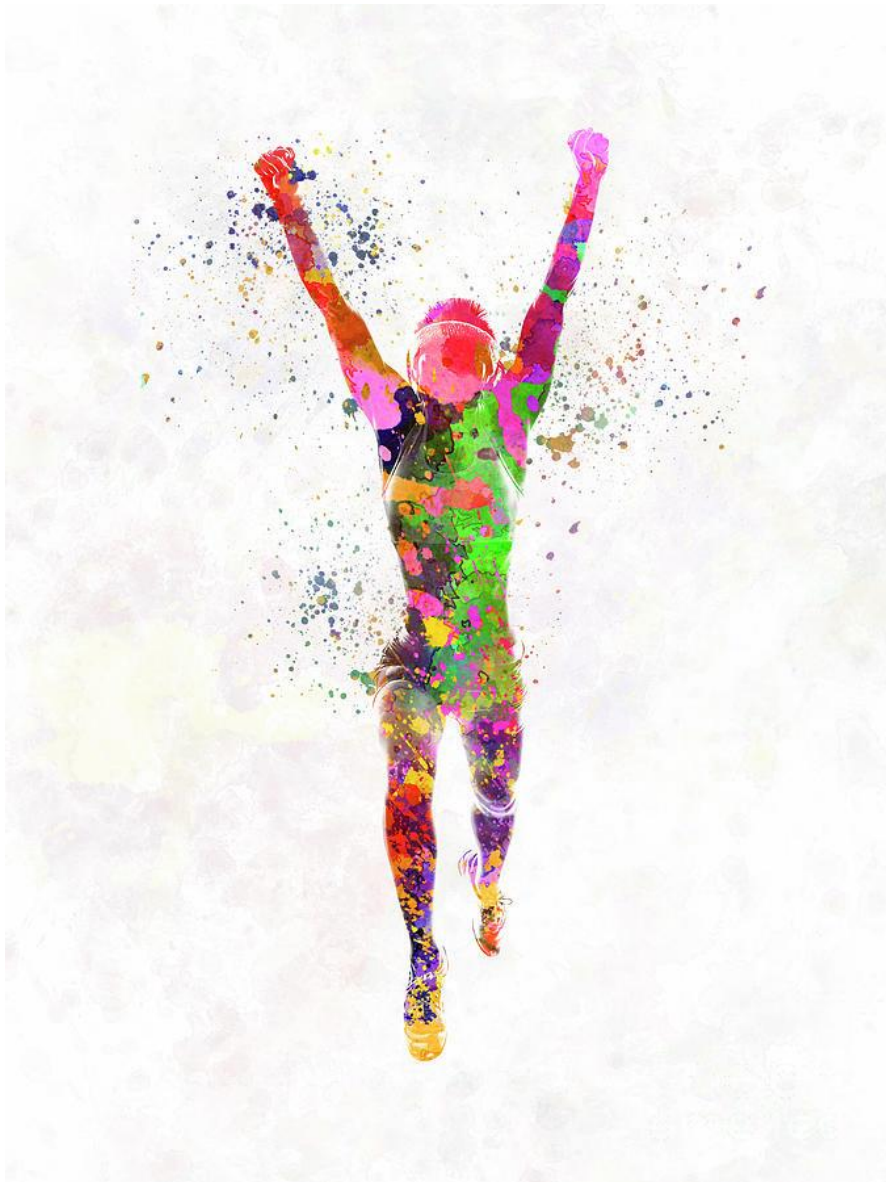
These findings may have important implications for coaches when deciding the most appropriate approach to prescribe HIIT to their athletes according to the training goal. As previously stressed (Kenneally et al., 2018a, b), maximizing  $\dot{V}O_{2max}$  development during a particular training phase may help in sustaining higher intensities during later performance-oriented stages of training preparation. Therefore, the use of the physiologically-based approach and the one based on race pace may be possible within the same macrocycle, but the implementation of the former should precede the implementation of the latter in terms of proximity to major competitions. I want to stress that, despite the promising findings, there is no evidence that the 110% of the race pace represents the optimal HIIT intensity for maximizing performance on 10000-m or any other running distance. As I previously discussed, I chose this value to match the training intensity between the two groups I investigated, but the percentage value *per se* is not less arbitrary than those previously proposed by Kenneally and colleagues (2021a, b) to characterize the TID of athletes. Future research should identify the optimal percentage value or range to prescribe HIIT for any given race distance (i.e., how much above the target race pace we can be prescribed HIIT intensity to stimulate adaptations without losing the specificity for the task).

# 7

## CONCLUSIONS

*'Everything is theoretically impossible, until it is done.'*

*Robert A. Heinlein – science fiction author, engineer, and naval officer*





## 7. CONCLUSIONS

**1)** Individualizing HIIT prescriptions according to physiological characteristics and response to exercise induce larger improvements in  $\dot{V}O_{2\max}$  and  $V_{\text{peak}}$  and reduce response heterogeneity for  $V_{\text{peak}}$  compared with standardized prescriptions in distance runners. Although no conclusion can be drawn about the efficacy of the standardized approach *per se*, these results stressed the importance of favoring individualized prescriptions whenever improving these physiological attributes represents a priority for coaches and athletes.

**2)** I identified valid surrogates of the delta concept that permit individualizing HIIT prescriptions according to the physiological approach making this form of HIIT more accessible to the largest part of the runner population. These measures, namely the relative percent of the  $V_{\text{peak}}$  at  $v\Delta 50$  and the difference between  $v\Delta 50$  and  $V_{\text{peak}}$ , only require a commercial (and properly calibrated) treadmill available in any commercial gym and also in some private houses and allow to estimate  $v\Delta 50$  with discrete precision and reasonably low effort.

**3)** The approach adopted to individualize HIIT prescriptions (i.e., physiological *vs* race pace approach) impacts the type of training response, in terms of cardiorespiratory fitness and running performance. These findings complement those from the first family of hypotheses in helping coaches and athletes to choose the most appropriate approach to HIIT prescription according to the desired goal of any given training phase.



## 7. CONCLUSIONES

**1)** La individualización de las prescripciones de HIIT de acuerdo con las características fisiológicas y la respuesta al ejercicio induce mayores mejoras en el  $\dot{V}O_{2max}$  y la  $V_{peak}$  y reduce la heterogeneidad de la respuesta para la  $V_{peak}$  en comparación con las prescripciones estandarizadas en corredores de fondo. Aunque no se puede sacar ninguna conclusión sobre la eficacia del enfoque estandarizado per se, estos resultados enfatizan la importancia de favorecer las prescripciones individualizadas siempre que mejorar estos atributos fisiológicos represente una prioridad para los entrenadores y atletas.

**2)** Identifiqué sustitutos válidos del concepto delta que permiten individualizar las prescripciones de HIIT de acuerdo con el enfoque fisiológico, haciendo que esta forma de HIIT sea más accesible para la mayor parte de la población de corredores. Estas medidas, a saber, el porcentaje relativo de  $V_{peak}$  en  $v\Delta 50$  y la diferencia entre  $v\Delta 50$  y  $V_{peak}$ , solo requieren una cinta de correr comercial (y debidamente calibrada) disponible en cualquier gimnasio comercial y también en algunas casas particulares y permiten estimar  $v\Delta 50$  con una precisión discreta y un esfuerzo razonablemente bajo.

**3)** El enfoque adoptado para individualizar las prescripciones de HIIT (es decir, enfoque fisiológico versus ritmo de carrera) impacta el tipo de respuesta al entrenamiento, en términos de aptitud cardiorrespiratoria y rendimiento de carrera. Estos hallazgos complementan los de la primera familia de hipótesis para ayudar a los entrenadores y atletas a elegir el enfoque más apropiado para la prescripción de HIIT de acuerdo con el objetivo deseado de cualquier fase de entrenamiento determinada.



## BIBLIOGRAPHY

*'Without data, you're just another person with an opinion'*

*W. Edwards Deming – engineer, statistician, professor, and much more*







## 8. BIBLIOGRAPHY

Acevedo, E. O., & Goldfarb, A. H. (1989). Increased training intensity effects on plasma lactate, ventilatory threshold, and endurance. *Medicine and Science in Sports and Exercise*, *21*, 563–568.

Aisbett, J., Lakens, D., & Sainani, K. L. (2020). *Magnitude based inference in relation to one-sided hypotheses testing procedures*. SportRxiv. <https://doi.org/10.31236/osf.io/pn9s3>

Althunian, T. A., de Boer, A., Groenwold, R. H. H., & Klungel, O. H. (2017). Defining the noninferiority margin and analysing noninferiority: an overview. *British Journal of Clinical Pharmacology*, *83*, 1636–1642. <https://doi.org/10.1111/bcp.13280>

Althunian, T. A., de Boer, A., Groenwold, R. H. H., & Klungel, O. H. (2018). Using a single noninferiority margin or preserved fraction for an entire pharmacological class was found to be inappropriate. *Journal of Clinical Epidemiology*, *104*, 15–23. <https://doi.org/10.1016/j.jclinepi.2018.07.004>

Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal*, *311*, 485. <https://doi.org/10.1136/bmj.311.7003.485>

Ammar, A., Chtourou, H., Trabelsi, K., Padulo, J., Turki, M., El Abed, K., Hoekelmann, A., & Hakim, A. (2015). Temporal specificity of training: intra-day effects on biochemical responses and Olympic-Weightlifting performances. *Journal of Sports Sciences*, *33*, 358–368. <https://doi.org/10.1080/02640414.2014.944559>

Andersen, J. J. (2021, December 12). *The State of Running 2019*. RunRepeat. <https://runrepeat.com/state-of-running>

Anderson T. (1996). Biomechanics and running economy. *Sports Medicine*, *22*, 76–89. <https://doi.org/10.2165/00007256-199622020-00003>

Anderson, G. S., & Rhodes, E. C. (1989). A review of blood lactate and ventilatory methods of detecting transition thresholds. *Sports Medicine*, *8*, 43–55. <https://doi.org/10.2165/00007256-198908010-00005>

Arampatzis, A., De Monte, G., Karamanidis, K., Morey-Klapsing, G., Stafilidis, S., & Brüggemann, G. P. (2006). Influence of the muscle-tendon unit's mechanical and

morphological properties on running economy. *The Journal of Experimental Biology*, *209*, 3345–3357. <https://doi.org/10.1242/jeb.02340>

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*, 305–307. <https://doi.org/10.1038/d41586-019-00857-9>

Åstrand, P. O. (1995). New Records in Human Power. *Nature* *176*, 922–923. <https://doi.org/10.1038/176922a0>

Atkinson, G., & Batterham, A. M. (2015). True and false interindividual differences in the physiological response to an intervention. *Experimental Physiology*, *100*, 577–588. <https://doi.org/10.1113/EP085070>

Avalos, M., Hellard, P., & Chatard, J. C. (2003). Modeling the training-performance relationship using a mixed model in elite swimmers. *Medicine and Science in Sports and Exercise*, *35*, 838–846. <https://doi.org/10.1249/01.MSS.0000065004.05033.42>

Balagué, N., Hristovski, R., Almarcha, M., Garcia-Retortillo, S., & Ivanov, P. C. (2020). Network Physiology of Exercise: Vision and Perspectives. *Frontiers in Physiology*, *11*, 611550. <https://doi.org/10.3389/fphys.2020.611550>

Bangsbo, J., Gunnarsson, T. P., Wendell, J., Nybo, L., & Thomassen, M. (2009). Reduced volume and increased training intensity elevate muscle Na<sup>+</sup>-K<sup>+</sup> pump  $\alpha_2$ -subunit expression as well as short- and long-term work capacity in humans. *Journal of Applied Physiology*, *107*, 1771–1780. <https://doi.org/10.1152/jappphysiol.00358.2009>

Bangsbo, J. (2015). Performance in sports – With specific emphasis on the effect of intensified training. *Scandinavian Journal of Medicine & Science in Sports*, *25*, 88–99. <https://doi.org/10.1111/sms.12605>

Banister, E. W., Calvert, T. W., Savage, M. V., & Bach, T. (1975). A systems model of training for athletic performance. *Australian Journal of Sports Medicine*, *7*, 57–61.

Banister, E. W. (1991). Modeling Elite Athletic Performance. In J. D. Macdougall, H. A. Wenger, & H. J. Green (Eds.), *Physiological Testing Of Elite Athletes* (2<sup>nd</sup> ed., pp. 403–424). Champaign, IL: Human Kinetics.

Banister, E. W., Carter, J. B., & Zarkadas, P. C. (1999). Training theory and taper: validation in triathlon athletes. *European Journal of Applied Physiology and Occupational Physiology*, *79*, 182–191. <https://doi.org/10.1007/s004210050493>

- Barnes, K. R., Hopkins, W. G., McGuigan, M. R., Northuis, M. E., & Kilding, A. E. (2013). Effects of resistance training on running economy and cross-country performance. *Medicine and Science in Sports and Exercise*, *45*, 2322–2331. <https://doi.org/10.1249/MSS.0b013e31829af603>
- Barnes, K. R., & Kilding, A. E. (2015). Running economy: measurement, norms, and determining factors. *Sports Medicine - Open*, *1*, 8. <https://doi.org/10.1186/s40798-015-0007-y>
- Baron, B., Noakes, T. D., Dekerle, J., Moullan, F., Robin, S., Matran, R., & Pelayo, P. (2008). Why does exercise terminate at the maximal lactate steady state intensity? *British Journal of Sports Medicine*, *42*, 828–833. <https://doi.org/10.1136/bjism.2007.040444>
- Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S., & Ivanov, P. (2012). Network physiology reveals relations between network topology and physiological function. *Nature Communications*, *3*, 702. <https://doi.org/10.1038/ncomms1705>
- Bassett, D. R., & Howley, E. T. (1997). Maximal oxygen uptake: "classical" versus "contemporary" viewpoints. *Medicine and Science in Sports and Exercise*, *29*, 591–603. <https://doi.org/10.1097/00005768-199705000-00002>
- Bassett, D. R., Jr, & Howley, E. T. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine and Science in Sports and Exercise*, *32*, 70–84. <https://doi.org/10.1097/00005768-200001000-00012>
- Batliner, M. E., Kipp, S., Grabowski, A. M., Kram, R., & Byrnes, W. C. (2017). Does Metabolic Rate Increase Linearly with Running Speed in all Distance Runners? *Sports Medicine International Open*, *2*, E1–E8. <https://doi.org/10.1055/s-0043-122068>
- Bauer, P., & Kieser, M. (1996). A Unifying Approach for Confidence Intervals and Testing of Equivalence and Difference. *Biometrika*, *83*, 934–937. <https://doi.org/10.1093/biomet/83.4.934>
- Baumann, C. W., Rupp, J. C., Ingalls, C. P., & Doyle, J. A. (2012). Anaerobic work capacity's contribution to 5-km-race performance in female runners. *International Journal of Sports Physiology and Performance*, *7*, 170–174. <https://doi.org/10.1123/ijsp.7.2.170>
- Beattie, K., Kenny, I. C., Lyons, M., & Carson, B. P. (2014). The effect of strength training on performance in endurance athletes. *Sports Medicine*, *44*, 845–865. <https://doi.org/10.1007/s40279-014-0157-y>

- Bellinger, P., Arnold, B., & Minahan, C. (2019). Quantifying the Training-Intensity Distribution in Middle-Distance Runners: The Influence of Different Methods of Training-Intensity Quantification. *International Journal of Sports Physiology and Performance*, 1–5. Advance online publication. <https://doi.org/10.1123/ijsp.2019-0298>
- Beneke, R. (2003). Methodological aspects of maximal lactate steady state-implications for performance testing. *European Journal of Applied Physiology*, 89, 95–99. <https://doi.org/10.1007/s00421-002-0783-1>
- Beneke, R., & Hütler, M. (2005). The effect of training on running economy and performance in recreational athletes. *Medicine and Science in Sports and Exercise*, 37, 1794–1799. <https://doi.org/10.1249/01.mss.0000176399.67121.02>
- Beneke, R., Leithäuser, R. M., & Ochentel, O. (2011). Blood lactate diagnostics in exercise testing and training. *International Journal of Sports Physiology and Performance*, 6, 8–24. <https://doi.org/10.1123/ijsp.6.1.8>
- Bentley, D. J., Newell, J., & Bishop, D. (2007). Incremental exercise test design and analysis: implications for performance diagnostics in endurance athletes. *Sports Medicine*, 37, 575–586. <https://doi.org/10.2165/00007256-200737070-00002>
- Berg, K., Latin, R. W., & Hendricks, T. (1995). Physiological and physical performance changes in female runners during one year of training. *Sports Medicine, Training and Rehabilitation*, 5, 311–319. <https://doi.org/10.1080/15438629509512027>
- Berger, R. L., & Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11, 283–319. <https://doi.org/10.1214/ss/1032280304>
- Bergh, U., Ekblom, B., & Astrand, P. O. (2000). Maximal oxygen uptake "classical" versus "contemporary" viewpoints. *Medicine and Science in Sports and Exercise*, 32, 85–88. <https://doi.org/10.1097/00005768-200001000-00013>
- Berthon, P., & Fellmann, N. (2002). General review of maximal aerobic velocity measurement at laboratory. Proposition of a new simplified protocol for maximal aerobic velocity assessment. *The Journal of Sports Medicine and Physical Fitness*, 42, 257–266.

- Bhattacharya, N. (2018). Intraspecific Competition and Natural Selection in Homo Sapience Sapience: A Critical Review. *International Journal of Humanities and Social Science*, 3, 14–22.
- Bickham, D. C., Bentley, D. J., Le Rossignol, P. F., & Cameron-Smith, D. (2006). The effects of short-term sprint training on MCT expression in moderately endurance-trained runners. *European Journal of Applied Physiology*, 96, 636–643. <https://doi.org/10.1007/s00421-005-0100-x>
- Billat, L. V. (2001). Interval training for performance: a scientific and empirical practice. Special recommendations for middle- and long-distance running. Part I: aerobic interval training. *Sports Medicine*, 31, 13–31. <https://doi.org/10.2165/00007256-200131010-00002>
- Billat, V., Renoux, J. C., Pinoteau, J., Petit, B., & Koralsztein, J. P. (1994). Times to exhaustion at 100% of velocity at  $\dot{V}O_{2max}$  and modelling of the time-limit/velocity relationship in elite long-distance runners. *European Journal of Applied Physiology and Occupational Physiology*, 69, 271–273. <https://doi.org/10.1007/BF01094801>
- Billat, L. V., & Koralsztein, J. P. (1996). Significance of the velocity at  $\dot{V}O_{2max}$  and time to exhaustion at this velocity. *Sports Medicine*, 22, 90–108. <https://doi.org/10.2165/00007256-199622020-00004>
- Billat, V., Binsse, V., Petit, B., & Koralsztein, J. P. (1998). High level runners are able to maintain a  $\dot{V}O_2$  steady-state below  $\dot{V}O_{2max}$  in an all-out run over their critical velocity. *Archives of Physiology and Biochemistry*, 106, 38–45. <https://doi.org/10.1076/apab.106.1.38.4396>
- Billat, V. L., Flechet, B., Petit, B., Muriaux, G., & Koralsztein, J. P. (1999). Interval training at  $VO_{2max}$ : effects on aerobic performance and overtraining markers. *Medicine and Science in Sports and Exercise*, 31, 156–163. <https://doi.org/10.1097/00005768-199901000-00024>
- Billat, V. L., Slawinski, J., Bocquet, V., Demarle, A., Lafitte, L., Chassaing, P., & Koralsztein, J. P. (2000). Intermittent runs at the velocity associated with maximal oxygen uptake enables subjects to remain at maximal oxygen uptake for a longer time than intense but submaximal runs. *European Journal of Applied Physiology*, 81, 188–196. <https://doi.org/10.1007/s004210050029>

- Billat, V. L., Demarle, A., Slawinski, J., Paiva, M., & Koralsztein, J. P. (2001). Physical and training characteristics of top-class marathon runners. *Medicine and Science in Sports and Exercise*, *33*, 2089–2097. <https://doi.org/10.1097/00005768-200112000-00018>
- Billat, V., Demarle, A., Paiva, M., & Koralsztein, J. P. (2002). Effect of training on the physiological factors of performance in elite marathon runners (males and females). *International Journal of Sports Medicine*, *23*, 336–341. <https://doi.org/10.1055/s-2002-33265>
- Billat, V., Lepretre, P. M., Heugas, A. M., Laurence, M. H., Salim, D., & Koralsztein, J. P. (2003). Training and bioenergetic characteristics in elite male and female Kenyan runners. *Medicine and Science in Sports and Exercise*, *35*, 297–306. <https://doi.org/10.1249/01.MSS.0000053556.59992.A9>
- Billat, V., Sirvent, P., Lepretre, P. M., & Koralsztein, J. P. (2004). Training effect on performance, substrate balance and blood lactate concentration at maximal lactate steady state in master endurance-runners. *European Journal of Physiology*, *447*, 875–883. <https://doi.org/10.1007/s00424-003-1215-8>
- Bishop, D., Jenkins, D. G., & Mackinnon, L. T. (1998). The relationship between plasma lactate parameters, Wpeak and 1-h cycling performance in women. *Medicine and Science in Sports and Exercise*, *30*, 1270–1275. <https://doi.org/10.1097/00005768-199808000-00014>
- Black, M. I., Jones, A. M., Blackwell, J. R., Bailey, S. J., Wylie, L. J., McDonagh, S. T., Thompson, C., Kelly, J., Sumners, P., Mileva, K. N., Bowtell, J. L., & Vanhatalo, A. (2017). Muscle metabolic and neuromuscular determinants of fatigue during cycling in different exercise intensity domains. *Journal of Applied Physiology*, *122*, 446–459. <https://doi.org/10.1152/jappphysiol.00942.2016>
- Bland, J. M., & Altman, D. G. (2011). Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials*, *12*, 264. <https://doi.org/10.1186/1745-6215-12-264>
- Bland, J. M., & Altman, D. G. (2015). Best (but oft forgotten) practices: testing for treatment effects in randomized trials by separate analyses of changes from baseline in each group is a misleading approach. *The American Journal of Clinical Nutrition*, *102*, 991–994.
- Bonafiglia, J. T., Brennan, A. M., Ross, R., & Gurd, B. J. (2019). An appraisal of the SDIR as an estimate of true individual differences in training responsiveness in parallel-arm exercise

randomized controlled trials. *Physiological Reports*, 7, e14163.  
<https://doi.org/10.14814/phy2.14163>

Bonaventura, J. M., Sharpe, K., Knight, E., Fuller, K. L., Tanner, R. K., & Gore, C. J. (2015). Reliability and accuracy of six hand-held blood lactate analysers. *Journal of Sports Science & Medicine*, 14, 203–214.

Booth, F. W., Roberts, C. K., & Laye, M. J. (2012). Lack of exercise is a major cause of chronic diseases. *Comprehensive Physiology*, 2, 1143–1211. <https://doi.org/10.1002/cphy.c110025>

Borresen, J., & Lambert, M. I. (2009). The quantification of training load, the training response and the effect on performance. *Sports Medicine*, 39, 779–795.  
<https://doi.org/10.2165/11317780-000000000-00000>

Bramble, D. M., & Lieberman, D. E. (2004). Endurance running and the evolution of Homo. *Nature*, 432, 345–352. <https://doi.org/10.1038/nature03052>

Brandon L. J. (1995). Physiological factors associated with middle distance running performance. *Sports Medicine*, 19, 268–277. <https://doi.org/10.2165/00007256-199519040-00004>

Buchheit, M., & Laursen, P. B. (2013). High-intensity interval training, solutions to the programming puzzle: Part I: cardiopulmonary emphasis. *Sports Medicine*, 43, 313–338.  
<https://doi.org/10.1007/s40279-013-0029-x>

Buchheit, M., & Laursen, P. B. (2013). High-intensity interval training, solutions to the programming puzzle. Part II: anaerobic energy, neuromuscular load and practical applications. *Sports Medicine*, 43, 927–954. <https://doi.org/10.1007/s40279-013-0066-5>

Bulbulian, R., Wilcox, A. R., & Darabos, B. L. (1986). Anaerobic contribution to distance running performance of trained cross-country athletes. *Medicine and Science in Sports and Exercise*, 18, 107–113.

Burnley, M., & Jones, A. M. (2007). Oxygen uptake kinetics as a determinant of sports performance. *European Journal of Sport Science*, 7, 63–79.  
<https://doi.org/10.1080/17461390701456148>

Busso, T. (2003). Variable dose-response relationship between exercise training and performance. *Medicine and Science in Sports and Exercise*, 35, 1188–1195.  
<https://doi.org/10.1249/01.MSS.0000074465.13621.37>



- Busso, T., Candau, R., & Lacour, J. R. (1994). Fatigue and fitness modelled from the effects of training on performance. *European Journal of Applied Physiology*, *69*, 50–54. <https://doi.org/10.1007/BF00867927>
- Busso, T., Denis, C., Bonnefoy, R., Geysant, A., & Lacour, J. R. (1997). Modeling of adaptations to physical-training by using a recursive least-squares algorithm. *Journal of Applied Physiology*, *82*, 1685–1693. <https://doi.org/10.1152/jappl.1997.82.5.1685>
- Busso, T., Benoit, H., Bonnefoy, R., Feasson, L., & Lacour, J. R. (2002). Effects of training frequency on the dynamics of performance response to a single training bout. *Journal of Applied Physiology*, *92*, 572–580. <https://doi.org/10.1152/japplphysiol.00429.2001>
- Caldwell, A. R., & Chevront, S. N. (2019). Basic statistical considerations for physiology: the journal *Temperature* toolbox. *Temperature (Austin)*, *6*, 181–210. <https://doi.org/10.1080/23328940.2019.1624131>
- Caldwell, A., & Vigotsky, A. D. (2020). A case against default effect sizes in sport and exercise science. *PeerJ*, *8*, e10314. <https://doi.org/10.7717/peerj.10314>
- Caldwell, A. R., Vigotsky, A. D., Tenan, M. S., Radel, R., Mellor, D. T., Kreutzer, A., Lahart, I. M., Mills, J. P., Boisgontier, M. P., & Consortium for Transparency in Exercise Science (COTES) Collaborators. (2020). Moving sport and exercise science forward: a call for the adoption of more transparent research practices. *Sports Medicine*, *50*, 449–459. <https://doi.org/10.1007/s40279-019-01227-1>
- Calvert, T. W., Banister, E. W., & Savage, M. V. (1976). A systems model of the effects of training on physical performance. *SMC Systems, Man and Cybernetics*, *2*, 94–102.
- Cannon, W. B. (1929). The control of homeostasis by the sympathetic system. *Transactions of the Association of American Physicians*, *41*.
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: a comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*, 115–144. <https://doi.org/10.1177/2515245919847196>
- Casado, A., Moreno-Pérez, D., Larrosa, M., & Renfree, A. (2019). Different psychophysiological responses to a high-intensity repetition session performed alone or in a group by elite middle-distance runners. *European Journal of Sport Science*, *19*, 1045–1052. <https://doi.org/10.1080/17461391.2019.1593510>

Castelloe, J., & Watts, D. (2015). Equivalence and Noninferiority Testing Using SAS/STAT® Software (Paper SAS1911-2015). *Proceedings of the SAS Global Forum 2015 Conference*. Cary, NC: SAS Institute Inc. <https://support.sas.com/resources/papers/proceedings15/SAS1911-2015.pdf>

Chuang-Stein, C., Kirby, S., Hirsch, I., & Atkinson, G. (2011). The role of the minimum clinically important difference and its impact on designing a trial. *Pharmaceutical statistics*, 10, 250–256. <https://doi.org/10.1002/pst.459>

Cocks, M., Shaw, C. S., Shepherd, S. O., Fisher, J. P., Ranasinghe, A., Barker, T. A., & Wagenmakers, A. J. (2016). Sprint interval and moderate-intensity continuous training have equal benefits on aerobic capacity, insulin sensitivity, muscle capillarisation and endothelial eNOS/NAD(P)H oxidase protein ratio in obese men. *Journal of Physiology*, 594, 2307–2321. <https://doi.org/10.1113/jphysiol.2014.285254>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>

Committee for Medicinal Products for Human Use. (2005). *Guideline on the choice of the non-inferiority margin*. European Medicines Agency. [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf)

Committee for Medicinal Products for Human Use. (2010). *Guideline on the investigation of bioequivalence*. European Medicines Agency. [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf)

Committee for Proprietary Medicinal Products. (2000). *Points to consider on switching between superiority and non-inferiority*. European Medicines Agency. [https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-switching-between-superiority-non-inferiority\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-switching-between-superiority-non-inferiority_en.pdf)

Conley, D. L., & Krahenbuhl, G. S. (1980). Running economy and distance running performance of highly trained athletes. *Medicine and Science in Sports and Exercise*, 12, 357–360.

- Conley, D. L., Krahenbuhl, G. S., & Burkett, L. N. (1981). Training for Aerobic Capacity and Running Economy. *The Physician and sportsmedicine*, 9, 107–146. <https://doi.org/10.1080/00913847.1981.11711060>
- Conley, D. K., Burkett, L. N., & Millar, A. L. (1984). Following Steve Scott: physiological changes accompanying training. *The Physician and Sportsmedicine*, 12, 103–106. <https://doi.org/10.1080/00913847.1984.11701746>
- Cook, J. A., Julious, S. A., Sones, W., Hampson, L. V., Hewitt, C., Berlin, J. A., Ashby, D., Emsley, R., Fergusson, D. A., Walters, S. J., Wilson, E. C. F., Maclennan, G., Stallard, N., Rothwell, J. C., Bland, M., Brown, L., Ramsay, C. R., Cook, A., Armstrong, D., ... Vale, L. D. (2018). DELTA<sup>2</sup> guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *British Medical Journal*, 363, k3750. <https://doi.org/10.1186/s13063-018-2884-0>
- Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Springer.
- Costill, D. L. (1967). The relationship between selected physiological variables and distance running performance. *J Sports Med Phys Fitness*, 7, 61–66.
- Costill, D. L. (1970). Metabolic responses during distance running. *Journal of Applied Physiology*, 28, 251–255. <https://doi.org/10.1152/jap.1970.28.3.251>
- Costill, D. L. (1972). Physiology of marathon running. *JAMA*, 221, 1024–1029.
- Costill, D. L., Branam, G., Eddy, D., & Sparks, K. (1971). Determinants of Marathon running success. *Internationale Zeitschrift für Angewandte Physiologie, Einschliesslich Arbeitsphysiologie*, 29, 249–254. <https://doi.org/10.1007/BF01100536>
- Costill, D. L., Thomason, H., & Roberts, E. (1973). Fractional utilization of the aerobic capacity during distance running. *Medicine and Science in Sports*, 5, 248–252.
- Coyle, E. F. (1995). Integration of the physiological factors determining endurance performance ability. *Exercise and Sport Sciences Reviews*, 23, 25–63.
- Coyle, E. F. (1999). Physiological determinants of endurance exercise performance. *Journal of Science and Medicine in Sport*, 2, 181–189. [https://doi.org/10.1016/s1440-2440\(99\)80172-8](https://doi.org/10.1016/s1440-2440(99)80172-8)

- Coyle, E. F. (2005). Improved muscular efficiency displayed as Tour de France champion matures. *Journal of Applied Physiology*, *98*, 2191–2196. <https://doi.org/10.1152/jappphysiol.00216.2005>
- Daniels, J. (1974). Physiological characteristics of champion male athletes. *Research Quarterly*, *45*, 342–348.
- Daniels, J. T., Yarbrough, R. A., & Foster, C. (1978). Changes in  $\dot{V}O_{2\max}$  and running performance with training. *European Journal of Applied Physiology and Occupational Physiology*, *39*, 249–254. <https://doi.org/10.1007/BF00421448>
- Daniels J. T. (1985). A physiologist's view of running economy. *Medicine and Science in Sports and Exercise*, *17*, 332–338.
- Daniels, J., & Scardina, N. (1984). Interval training and performance. *Sports Medicine*, *1*, 327–334. <https://doi.org/10.2165/00007256-198401040-00006>
- Dantas, J. L., & Doria, C. (2015). Detection of the Lactate Threshold in Runners: What is the Ideal Speed to Start an Incremental Test? *Journal of Human Kinetics*, *45*, 217–224. <https://doi.org/10.1515/hukin-2015-0022>
- Davies, C. T., & Thompson, M. W. (1979). Aerobic performance of female marathon and male ultramarathon athletes. *European Journal of Applied Physiology and Occupational Physiology*, *41*, 233–245. <https://doi.org/10.1007/BF00429740>
- Demarie, S., Koralsztein, J. P., & Billat, V. (2000). Time limit and time at  $VO_{2\max}$  during a continuous and an intermittent run. *The Journal of Sports Medicine and Physical Fitness*, *40*, 96–102.
- Demarle, A. P., Slawinski, J. J., Laffite, L. P., Bocquet, V. G., Koralsztein, J. P., & Billat, V. L. (2001). Decrease of  $O_2$  deficit is a potential factor in increased time to exhaustion after specific endurance training. *Journal of Applied Physiology*, *90*, 947–953. <https://doi.org/10.1152/jappl.2001.90.3.947>
- Demarle, A. P., Heugas, A. M., Slawinski, J. J., Tricot, V. M., Koralsztein, J. P., & Billat, V. L. (2003). Whichever the initial training status, any increase in velocity at lactate threshold appears as a major factor in improved time to exhaustion at the same severe velocity after training. *Archives of Physiology and Biochemistry*, *111*, 167–176. <https://doi.org/10.1076/apab.111.2.167.14003>

- Denadai, B. S., Ortiz, M. J., Greco, C. C., & de Mello, M. T. (2006). Interval training at 95% and 100% of the velocity at  $\dot{V}O_{2max}$ : effects on aerobic physiological indexes and running performance. *Applied Physiology, Nutrition, and Metabolism*, *3*, 737–743. <https://doi.org/10.1139/h06-080>
- Di Prampero, P. E., Atchou, G., Brückner, J. C., & Moia, C. (1986). The energetics of endurance running. *European Journal of Applied Physiology and Occupational Physiology*, *55*, 259–266. <https://doi.org/10.1007/BF02343797>
- Dixon, P. M., Saint-Maurice, P. F., Kim, Y., Hibbing, P., Bai, Y., & Welk, G. J. (2018). A primer on the use of equivalence testing for evaluating measurement agreement. *Medicine & Science in Sports & Exercise*, *50*, 837–845. <https://doi.org/10.1249/MSS.0000000000001481>
- Edwards, A. M., & Polman, R. C. (2013). Pacing and awareness: brain regulation of physical activity. *Sports Medicine*, *43*, 1057–1064. <https://doi.org/10.1007/s40279-013-0091-4>
- Edwards, S. (1993). High performance training and racing. In S. Edwards (Eds.). *The heart rate monitor book* (pp. 113–123). Sacramento, CA: Feet Fleet Press.
- Enoksen, E., Shalfawi, S. A., & Tønnessen, E. (2011). The effect of high- vs. low-intensity training on aerobic capacity in well-trained male middle-distance runners. *Journal of Strength and Conditioning Research*, *25*, 812–818. <https://doi.org/10.1519/JSC.0b013e3181cc2291>
- Esfarjani, F., & Laursen, P. B. (2007). Manipulating high-intensity interval training: effects on  $\dot{V}O_{2max}$ , the lactate threshold and 3000 m running performance in moderately trained males. *Journal of Science and Medicine in Sport*, *10*, 27–35. <https://doi.org/10.1016/j.jsams.2006.05.014>
- Esteve-Lanao, J., San Juan, A. F., Earnest, C. P., Foster, C., & Lucia, A. (2005). How do endurance runners actually train? Relationship with competition performance. *Medicine and Science in Sports and Exercise*, *37*, 496–504. <https://doi.org/10.1249/01.mss.0000155393.78744.86>
- Esteve-Lanao, J., Foster, C., Seiler, S., & Lucia, A. (2007). Impact of training intensity distribution on performance in endurance athletes. *Journal of Strength and Conditioning Research*, *21*, 943–949. <https://doi.org/10.1519/R-19725.1>

- Evans, S. L., Davy, K. P., Stevenson, E. T., & Seals, D. R. (1995). Physiological determinants of 10-km performance in highly trained female runners of different ages. *Journal of Applied Physiology*, *78*, 1931–1941. <https://doi.org/10.1152/jappl.1995.78.5.1931>
- Farrell, P. A., Wilmore, J. H., Coyle, E. F., Billing, J. E., & Costill, D. L. (1979). Plasma lactate accumulation and distance running performance. *Medicine and Science in Sports*, *11*, 338–344.
- Faude, O., Kindermann, W., & Meyer, T. (2009). Lactate threshold concepts: how valid are they? *Sports Medicine*, *39*, 469–490. <https://doi.org/10.2165/00007256-200939060-00003>
- Faulkner, J. A. (1968). New perspectives in training for maximum performance. *JAMA*, *205*, 741–746. doi: 10.1001/jama.1968.03140370043009
- Ferley, D. D., Osborn, R. W., & Vukovich, M. D. (2013). The effects of uphill vs. level-grade high-intensity interval training on  $\dot{V}O_{2\max}$ ,  $V_{\max}$ ,  $V_{LT}$ , and  $T_{\max}$  in well-trained distance runners. *Journal of Strength and Conditioning Research*, *27*, 1549–1559. <https://doi.org/10.1519/JSC.0b013e3182736923>
- Fontana, P., Boutellier, U., & Knöpfli-Lenzin, C. (2009). Time to exhaustion at maximal lactate steady state is similar for cycling and running in moderately trained subjects. *European Journal of Applied Physiology*, *107*, 187–192. <https://doi.org/10.1007/s00421-009-1111-9>
- Foster, C. (1983).  $\dot{V}O_{2\max}$  and training indices as determinants of competitive running performance. *Journal of Sports Sciences* *1*, 13–22. <https://doi.org/10.1080/02640418308729657>
- Foster, C., Daniels, J. T., & Yarbrough, R. A. (1977). Physiological and Training Correlates of Marathon Running Performance. *Australian Journal Of Sports Medicine*, *9*, 58–61.
- Foster, C., Costill, D. L., Daniels, J. T., & Fink, W. J. (1978). Skeletal muscle enzyme activity, fiber composition and  $\dot{V}O_{2\max}$  in relation to distance running performance. *European Journal of Applied Physiology and Occupational Physiology*, *39*, 73–80. <https://doi.org/10.1007/BF00421711>
- Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. *Journal of Strength and Conditioning Research*, *15*, 109–115.

- Foster, C., Rodriguez-Marroyo, J. A., & de Koning, J. J. (2017). Monitoring Training Loads: The Past, the Present, and the Future. *International Journal of Sports Physiology and Performance*, *12*, S22–S28. <https://doi.org/10.1123/ijsp.2016-0388>
- Franch, J., Madsen, K., Djurhuus, M. S., & Pedersen, P. K. (1998). Improved running economy following intensified training correlates with reduced ventilatory demands. *Medicine and Science in Sports and Exercise*, *30*, 1250–1256. <https://doi.org/10.1097/00005768-199808000-00011>
- Garber, C. E., Blissmer, B., Deschenes, M. R., Franklin, B. A., Lamonte, M. J., Lee, I. M., Nieman, D. C., Swain, D. P., & American College of Sports Medicine (2011). American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Medicine and Science in Sports and Exercise*, *43*, 1334–1359. <https://doi.org/10.1249/MSS.0b013e318213febf>
- Gaesser, G. A., & Poole, D. C. (1996). The slow component of oxygen uptake kinetics in humans. *Exercise and Sport Sciences Reviews*, *24*, 35–71.
- Garcin, M., Fleury, A., & Billat, V. (2002). The ratio HLa : RPE as a tool to appreciate overreaching in young high-level middle-distance runners. *International Journal of Sports Medicine*, *23*, 16–21. <https://doi.org/10.1055/s-2002-19275>
- Gastin, P. B. (2001). Energy system interaction and relative contribution during maximal exercise. *Sports Medicine*, *31*, 725–741. <https://doi.org/10.2165/00007256-200131100-00003>
- Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on psychological science: a journal of the Association for Psychological Science*, *9*, 641–651. <https://doi.org/10.1177/1745691614551642>
- Gettman, L. R., Pollock, M. L., Durstine, J. L., Ward, A., Ayres, J., & Linnerud, A. C. (1976). Physiological responses of men to 1, 3, and 5 day per week training programs. *Research Quarterly*, *47*, 638–646.
- Gillen, J. B., Martin, B. J., MacInnis, M. J., Skelly, L. E., Tarnopolsky, M. A., & Gibala, M. J. (2016). Twelve weeks of sprint interval training improves indices of cardiometabolic health

similar to traditional endurance training despite a five-fold lower exercise volume and time commitment. *PLoS One*, *11*, e0154075. <http://doi.org/10.1371/journal.pone.0154075>

Groepenhoff, H., de Jeu, R. C., & Schot, R. (2017). Vyntus CPX compared to Oxycon pro shows equal gas-exchange and ventilation during exercise. *European Respiratory Journal*, *50*, PA3002. <https://doi.org/10.1183/1393003.congress-2017.PA3002>

Gullstrand, L., Sjödin, B., & Svedenhag, J. (1994). Blood sampling during continuous running and 30-second intervals on a treadmill. *Scandinavian Journal of Medicine & Science in Sports*, *4*, 239–242.

Halperin, I., Pyne, D. B., & Martin, D. T. (2015). Threats to internal validity in exercise science: a review of overlooked confounding variables. *International Journal of Sports Physiology and Performance*, *10*, 823–829. <https://doi.org/10.1123/ijsp.2014-0566>

Halperin, I., Vigotsky, A. D., Foster, C., & Pyne, D. B. (2018). Strengthening the practice of exercise and sport-science research. *International Journal of Sports Physiology and Performance*, *13*, 127–134. <https://doi.org/10.1123/ijsp.2017-0322>

Hawley, J. A. (1995). State of the art training guidelines for endurance performance. *South African Journal of Sports Medicine*, *2*, 7-12.

Hawley J. A. (2002). Adaptations of skeletal muscle to prolonged, intense endurance training. *Clinical and Experimental Pharmacology & Physiology*, *29*, 218–222. <https://doi.org/10.1046/j.1440-1681.2002.03623.x>

Hawley J. A. (2008). Specificity of training adaptation: time for a rethink? *The Journal of Physiology*, *586*, 1–2. <https://doi.org/10.1113/jphysiol.2007.147397>

Hawley, J. A., & Hopkins, W. G. (1995). Aerobic glycolytic and aerobic lipolytic power systems. A new paradigm with implications for endurance and ultraendurance events. *Sports Medicine*, *19*, 240–250. <https://doi.org/10.2165/00007256-199519040-00002>

Hawley, J. A., Myburgh, K. H., Noakes, T. D., & Dennis, S. C. (1997). Training techniques to improve fatigue resistance and enhance endurance performance. *Journal of Sports Sciences*, *15*, 325–333. <https://doi.org/10.1080/026404197367335>

Hecksteden, A., Kraushaar, J., Scharhag-Rosenberger, F., Theisen, D., Senn, S., & Meyer, T. (2015). Individual response to exercise training - a statistical perspective. *Journal of Applied Physiology*, *118*, 1450–1459. <https://doi.org/10.1152/jappphysiol.00714.2014>



- Hecksteden, A., Faude, O., Meyer, T., & Donath, L. (2018). How to construct, conduct and analyze an exercise training study? *Frontiers in Physiology*, *9*, 1007. <https://doi.org/10.3389/fphys.2018.01007>
- Helgerud, J., Høydal, K., Wang, E., Karlsen, T., Berg, P., Bjerkaas, M., Simonsen, T., Helgesen, C., Hjørth, N., Bach, R., & Hoff, J. (2007). Aerobic high-intensity intervals improve  $\dot{V}O_{2\max}$  more than moderate training. *Medicine and Science in Sports and Exercise*, *39*, 665–671. <https://doi.org/10.1249/mss.0b013e3180304570>
- Hellard, P., Avalos, M., Lacoste, L., Barale, F., Chatard, J. C., & Millet, G. P. (2006). Assessing the limitations of the Banister model in monitoring training. *Journal of Sports Sciences*, *24*, 509–520. <https://doi.org/10.1080/02640410500244697>
- Henrich, J. (2021, December 12). *A cultural species: How culture drove human evolution*. Psychological Science Agenda. <http://www.apa.org/science/about/psa/2011/11/human-evolution>
- Hewson, D. J., & Hopkins, W. G. (1996). Specificity of training and its relation to the performance of distance runners. *International Journal of Sports Medicine*, *17*, 199–204. <https://doi.org/10.1055/s-2007-972832>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions* (2<sup>nd</sup> ed.). Chichester: Wiley. <https://doi.org/10.1002/9781119536604>
- High, R. (2021, December 12). *An Introduction to Statistical Power Calculations for Linear Models with SAS 9.1*. <https://www.lexjansen.com/pnwsug/2007/Robin%20High%20-%20Statistical%20Power%20Calculations%20for%20Linear%20Models.pdf>
- Hill, A. V. (1925). The Physiological Basis of Athletic Records. *Nature*, *116*, 544–548. <https://doi.org/10.1038/116544a0>
- Hill, A. V., Lupton, H. (1923). Muscular Exercise, Lactic Acid, and the Supply and Utilization of Oxygen. *QJM: An International Journal of Medicine*, *16*, 135–171, <https://doi.org/10.1093/qjmed/os-16.62.135>
- Hill, D. W., & Rowell, A. L. (1996). Running velocity at  $\dot{V}O_{2\max}$ . *Medicine and Science in Sports and Exercise*, *28*, 114–119. <https://doi.org/10.1097/00005768-199601000-00022>

- Hill, D. W., & Rowell, A. L. (1997). Responses to exercise at the velocity associated with  $\dot{V}O_{2\max}$ . *Medicine and Science in Sports and Exercise*, 29, 113–116. <https://doi.org/10.1097/00005768-199701000-00016>
- Hill, D. W., Williams, C. S., & Burt, S. E. (1997). Responses to exercise at 92% and 100% of the velocity associated with  $\dot{V}O_{2\max}$ . *International Journal of Sports Medicine*, 18, 325–329. <https://doi.org/10.1055/s-2007-972641>
- Hill, D. W., Poole, D. C., & Smith, J. C. (2002). The relationship between power and the time to achieve  $\dot{V}O_{2\max}$ . *Medicine and Science in Sports and Exercise*, 34, 709–714. <https://doi.org/10.1097/00005768-200204000-00023>
- Hodges, J. L., & Lehmann, E. L. (1954). Testing the approximate validity of statistical hypotheses. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 16, 261–268. <https://doi.org/10.1111/j.2517-6161.1954.tb00169.x>
- Holmgren, E. B. (1999). Establishing equivalence by showing that a specified percentage of the effect of the active control over placebo is maintained. *Journal of Biopharmaceutical Statistics*, 9, 651–659. <https://doi.org/10.1081/bip-100101201>
- Hopkins, W. G., Hawley, J. A., & Burke, L. M. (1999). Design and analysis of research on sport performance enhancement. *Medicine & Science in Sports & Exercise*, 31, 472–485. <https://doi.org/10.1097/00005768-199903000-00018>
- Hopkins, W. G., Schabert, E. J., & Hawley, J. A. (2001). Reliability of power in physical performance tests. *Sports Medicine*, 31, 211–234. <https://doi.org/10.2165/00007256-200131030-00005>
- Houmard, J. A. (1991). Impact of reduced training on performance in endurance athletes. *Sports Medicine*, 12, 380–393. <https://doi.org/10.2165/00007256-199112060-00004>
- Houston, M. E., & Thomson, J. A. (1977). The response of endurance-adapted adults to intense anaerobic training. *European Journal of Applied Physiology and Occupational Physiology*, 36, 207–213. <https://doi.org/10.1007/BF00421751>
- Hulteen, R. M., Smith, J. J., Morgan, P. J., Barnett, L. M., Hallal, P. C., Colyvas, K., & Lubans, D. R. (2017). Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis. *Preventive medicine*, 95, 14–25. <https://doi.org/10.1016/j.ypmed.2016.11.027>

Iaia, F. M., Thomassen, M., Kolding, H., Gunnarsson, T., Wendell, J., Rostgaard, T., Nordsborg, N., Krstrup, P., Nybo, L., Hellsten, Y., & Bangsbo, J. (2008). Reduced volume but increased training intensity elevates muscle Na<sup>+</sup>-K<sup>+</sup> pump alpha1-subunit and NHE1 expression as well as short-term work capacity in humans. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, 294, R966–R974. <https://doi.org/10.1152/ajpregu.00666.2007>

Iaia, F. M., Hellsten, Y., Nielsen, J. J., Fernström, M., Sahlin, K., & Bangsbo, J. (2009). Four weeks of speed endurance training reduces energy expenditure during exercise and maintains muscle oxidative capacity despite a reduction in training volume. *Journal of Applied Physiology*, 106, 73–80. <https://doi.org/10.1152/jappphysiol.90676.2008>

Iaia, F. M., & Bangsbo, J. (2010). Speed endurance training is a powerful stimulus for physiological adaptations and performance improvements of athletes. *Scandinavian Journal of Medicine & Science in Sports*, 20, 11–23. <https://doi.org/10.1111/j.1600-0838.2010.01193.x>

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1998). *ICH E9: statistical principles for clinical trials*. European Medicines Agency. [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf)

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2001). *ICH E10: Choice of control group in clinical trials*. European Medicines Agency. [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf)

Inzlicht, M., & Marcora, S. M. (2016). The Central Governor Model of Exercise Regulation Teaches Us Precious Little about the Nature of Mental Fatigue and Self-Control Failure. *Frontiers in Psychology*, 7, 656. <https://doi.org/10.3389/fpsyg.2016.00656>

Issurin, V. B. (2010). New horizons for the methodology and physiology of training periodization. *Sports Medicine*, 40, 189–206. <https://doi.org/10.2165/11319770-000000000-00000>

Issurin, V. B. (2016). Benefits and Limitations of Block Periodized Training Approaches to Athletes' Preparation: A Review. *Sports Medicine*, 46, 329–338. <https://doi.org/10.1007/s40279-015-0425-5>

- Ivanov, P. C. (2021). The New Field of Network Physiology: Building the Human Physiome. *Frontiers in Network Physiology, 1*, 711778. <https://doi.org/10.3389/fnetp.2021.711778>
- Jeacocke, N. A., & Burke, L. M. (2010). Methods to standardize dietary intake before performance testing. *International Journal of Sport Nutrition and Exercise Metabolism, 20*, 87–103. <https://doi.org/10.1123/ijsnem.20.2.87>
- Jones, A. M. (2006). The Physiology of the World Record Holder for the Women's Marathon. *International Journal of Sports Science & Coaching, 1*, 101–116. <https://doi.org/10.1260/174795406777641258>
- Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences, 14*, 321–327. <https://doi.org/10.1080/02640419608727717>
- Jones, A. M., & Poole, D. C. (2005). Oxygen uptake dynamics: from muscle to mouth--an introduction to the symposium. *Medicine and Science in Sports and Exercise, 37*, 1542–1550. <https://doi.org/10.1249/01.mss.0000177466.01232.7e>
- Jones, A. M., & Vanhatalo, A. (2017). The 'Critical Power' Concept: Applications to Sports Performance with a Focus on Intermittent High-Intensity Exercise. *Sports Medicine, 47*, 65–78. <https://doi.org/10.1007/s40279-017-0688-0>
- Jones, A. M., Vanhatalo, A., Burnley, M., Morton, R. H., & Poole, D. C. (2010). Critical power: implications for determination of  $\dot{V}O_2$ max and exercise tolerance. *Medicine and Science in Sports and Exercise, 42*, 1876–1890. <https://doi.org/10.1249/MSS.0b013e3181d9cf7f>
- Jones, A. M., Grassi, B., Christensen, P. M., Krstrup, P., Bangsbo, J., & Poole, D. C. (2011). Slow component of  $\dot{V}O_2$  kinetics: mechanistic bases and practical applications. *Medicine and Science in Sports and Exercise, 43*, 2046–2062. <https://doi.org/10.1249/MSS.0b013e31821fcfc1>
- Jones, A. M., Burnley, M., Black, M. I., Poole, D. C., & Vanhatalo, A. (2019). The maximal metabolic steady state: redefining the 'gold standard'. *Physiological reports, 7*, e14098. <https://doi.org/10.14814/phy2.14098>
- Joyner, M. J. (1993). Physiological limiting factors and distance running: influence of gender and age on record performances. *Exercise and Sport Sciences Reviews, 21*, 103–133.

- Joyner, M. J., & Coyle, E. F. (2008). Endurance exercise performance: the physiology of champions. *The Journal of Physiology*, *586*, 35–44. <https://doi.org/10.1113/jphysiol.2007.143834>
- Julious, S. A. (2004). Sample sizes for clinical trials with normal data. *Statistics in Medicine*, *23*, 1921–1986. <https://doi.org/10.1002/sim.1783>
- Katch, V., Weltman, A., Sady, S., & Freedson, P. (1978). Validity of the relative percent concept for equating training intensity. *European Journal of Applied Physiology and Occupational Physiology*, *39*, 219–227. <https://doi.org/10.1007/BF00421445>
- Kenneally, M., Casado, A., & Santos-Concejero, J. (2018). The Effect of Periodization and Training Intensity Distribution on Middle- and Long-Distance Running Performance: A Systematic Review. *International Journal of Sports Physiology and Performance*, *13*, 1114–1121. <https://doi.org/10.1123/ijsp.2017-0327>
- Kenneally, M., Casado, A., Gomez-Ezeiza, J., & Santos-Concejero, J. (2021). Training intensity distribution analysis by race pace vs. physiological approach in world-class middle- and long-distance runners. *European Journal of Sport Science*, *21*, 819–826. <https://doi.org/10.1080/17461391.2020.1773934>
- Kenneally, M., Casado, A., Gomez-Ezeiza, J., & Santos-Concejero, J. (2021). Training Characteristics of a World Championship 5000-m Finalist and Multiple Continental Record Holder Over the Year Leading to a World Championship Final. *International Journal of Sports Physiology and Performance*, 1–5. Advance online publication. <https://doi.org/10.1123/ijsp.2021-0114>
- Kiely, J. (2018). Periodization Theory: Confronting an Inconvenient Truth. *Sports Medicine*, *48*, 753–764. <https://doi.org/10.1007/s40279-017-0823-y>
- Kindermann, W., Simon, G., & Keul, J. (1979). The significance of the aerobic-anaerobic transition for the determination of work load intensities during endurance training. *European Journal of Applied Physiology and Occupational Physiology*, *42*, 25–34. <https://doi.org/10.1007/BF00421101>
- Knicker, A. J., Renshaw, I., Oldham, A. R., & Cairns, S. P. (2011). Interactive processes link the multiple symptoms of fatigue in sport competition. *Sports Medicine*, *41*, 307–328. <https://doi.org/10.2165/11586070-000000000-00000>

- Kohn, T. A., Essén-Gustavsson, B., & Myburgh, K. H. (2011). Specific muscle adaptations in type II fibers after high-intensity interval training of well-trained runners. *Scandinavian Journal of Medicine & Science in Sports*, *21*, 765–772. <https://doi.org/10.1111/j.1600-0838.2010.01136.x>
- Koral, J., Oranchuk, D. J., Herrera, R., & Millet, G. Y. (2018). Six Sessions of Sprint Interval Training Improves Running Performance in Trained Athletes. *Journal of Strength and Conditioning Research*, *32*, 617–623. <https://doi.org/10.1519/JSC.0000000000002286>
- Kovacs, M., van Ravenzwaaij, D., Hoekstra, R., & Aczel, B. (2021). *SampleSizePlanner: A Tool to Estimate and Justify Sample Size for Two-Group Studies*. MetaArXiv. <https://doi.org/10.31222/osf.io/rm9dn>
- Kubukeli, Z. N., Noakes, T. D., & Dennis, S. C. (2002). Training techniques to improve endurance exercise performances. *Sports Medicine*, *32*, 489–509. <https://doi.org/10.2165/00007256-200232080-00002>
- Kuipers, H., Rietjens, G., Verstappen, F., Schoenmakers, H., & Hofman, G. (2003). Effects of stage duration in incremental running tests on physiological variables. *International Journal of Sports Medicine*, *24*, 486–491. <https://doi.org/10.1055/s-2003-42020>
- Lacour, J. R., Messonnier, L., & Bourdin, M. (2009). Physiological correlates of performance. Case study of a world-class rower. *European Journal of Applied Physiology*, *106*, 407–413. <https://doi.org/10.1007/s00421-009-1028-3>
- Laffite, L. P., Mille-Hamard, L., Koralsztejn, J. P., & Billat, V. L. (2003). The effects of interval training on oxygen pulse and performance in supra-threshold runs. *Archives of Physiology and Biochemistry*, *111*, 202–210. <https://doi.org/10.1076/apab.111.3.202.23455>
- Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*, 355–362. <https://doi.org/10.1177/1948550617697177>
- Lakens, D. (2021). *Sample size justification*. PsyArXiv. <https://doi.org/10.31234/osf.io/9d3yf>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: a tutorial. *Advances in Methods and Practices in Psychological Science*, *1*, 259–269. <https://doi.org/10.1177/2515245918770963>

Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S-C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., ... Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, 2, 168–171. <https://doi.org/10.1038/s41562-018-0311-x>

Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridor of stability: practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, 9, 278–292. <https://doi.org/10.1177/1745691614528520>

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *J The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 75, 45–57. <https://doi.org/10.1093/geronb/gby065>

Lakens, D., Pahlke, F., & Wassmer, G. (2021). *Group sequential designs: a tutorial*. PsyArXiv. <https://doi.org/10.31234/osf.io/x4azm>

Lambert, E. V., St Clair Gibson, A., & Noakes, T. D. (2005). Complex systems model of fatigue: integrative homeostatic control of peripheral physiological systems during exercise in humans. *British Journal of Sports Medicine*, 39, 52–62. <https://doi.org/10.1136/bjism.2003.011247>

Laukkanen, J. A., Zaccardi, F., Khan, H., Kurl, S., Jae, S. Y., & Rauramaa, R. (2016). Long-term Change in Cardiorespiratory Fitness and All-Cause Mortality: A Population-Based Follow-up Study. *Mayo Clinic proceedings*, 91, 1183–1188. <https://doi.org/10.1016/j.mayocp.2016.05.014>

Laursen, P. B., & Jenkins, D. G. (2002). The scientific basis for high-intensity interval training: optimising training programmes and maximising performance in highly trained endurance athletes. *Sports Medicine*, 32, 53–73. <https://doi.org/10.2165/00007256-200232010-00003>

Legaz Arrese, A., Serrano Ostáriz, E., Jcasajús Mallén, J. A., & Munguía Izquierdo, D. (2005). The changes in running performance and maximal oxygen uptake after long-term training in elite athletes. *The Journal of Sports Medicine and Physical Fitness*, 45, 435–440.

Lehmann, M., Dickhuth, H. H., Gendrisch, G., Lazar, W., Thum, M., Kaminski, R., Aramendi, J. F., Peterke, E., Wieland, W., & Keul, J. (1991). Training-overtraining. A prospective,

experimental study with experienced middle- and long-distance runners. *International Journal of Sports Medicine*, 12, 444–452. <https://doi.org/10.1055/s-2007-1024711>

Lieberman, D. E., & Bramble, D. M. (2007). The evolution of marathon running: capabilities in humans. *Sports Medicine*, 37, 288–290. <https://doi.org/10.2165/00007256-200737040-00004>

Londeree, B. R. (1997). Effect of training on lactate/ventilatory thresholds: a meta-analysis. *Medicine and Science in Sports and Exercise*, 29, 837–843. <https://doi.org/10.1097/00005768-199706000-00016>

Lucia, A., Hoyos, J., Santalla, A., Earnest, C. & Chicharro, J. L. (2003). Tour De France Versus Vuelta A Espana: Which Is Harder? *Medicine and Science in Sports and Exercise*, 35, 872–878. <https://doi.org/10.1249/01.MSS.0000064999.82036.B4>

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49, 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>

MacDougall, J. D. (1977). The anaerobic threshold: its significance for the endurance athlete. *Canadian Journal of Applied Sport Sciences*, 2, 137–40.

MacDougall, D., & Sale, D. (1981). Continuous vs. interval training: a review for the athlete and the coach. *Canadian Journal of Applied Sport Sciences*, 6, 93–97.

Macfarlane, D. J. (2001). Automated metabolic gas analysis systems: a review. *Sports Medicine*, 31, 841–861. <https://doi.org/10.2165/00007256-200131120-00002>

Machado, F. A., Kravchychyn, A. C., Peserico, C. S., da Silva, D. F., & Mezzaroba, P. V. (2013). Incremental test design, peak 'aerobic' running speed and endurance performance in runners. *Journal of Science and Medicine in Sport*, 16, 577–582. <https://doi.org/10.1016/j.jsams.2012.12.009>

Mader, A., & Heck, H. (1986). A theory of the metabolic origin of "anaerobic threshold". *International Journal of Sports Medicine*, 7, 45–65.

Magnusson, K. (2021, December 12). *Equivalence, non-inferiority and superiority testing – an interactive visualization*. R Psychologist. <https://rpsychologist.com/d3/equivalence/>



- Manski, C. F. (2019). Treatment Choice With Trial Data: Statistical Decision Theory Should Supplant Hypothesis Testing. *The American Statistician*, *73*, 296–304. <https://doi.org/10.1080/00031305.2020.1717621>
- Mann, T., Lamberts, R. P., & Lambert, M. I. (2013). Methods of prescribing relative exercise intensity: physiological and practical considerations. *Sports Medicine*, *43*, 613–625. <https://doi.org/10.1007/s40279-013-0045-x>
- Mansournia, M. A., Higgins, J. P., Sterne, J. A., & Hernán, M. A. (2017). Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists. *Epidemiology*, *28*, 54–59. <https://doi.org/10.1097/EDE.0000000000000564>
- Mansournia, M. A., & Altman, D. G. (2018). Invited commentary: methodological issues in the design and analysis of randomised trials. *British Journal of Sports Medicine*, *52*, 553–555. <https://doi.org/10.1136/bjsports-2017-09824515>
- Manzi, V., Iellamo, F., Impellizzeri, F., D'Ottavio, S., & Castagna, C. (2009). Relation between individualized training impulses and performance in distance runners. *Medicine and Science in Sports and Exercise*, *41*, 2090–2096. <https://doi.org/10.1249/MSS.0b013e3181a6a959>
- Manzi, V., Bovenzi, A., Castagna, C., Sinibaldi Salimei, P., Volterrani, M., & Iellamo, F. (2015). Training-Load Distribution in Endurance Runners: Objective Versus Subjective Assessment. *International Journal of Sports Physiology and Performance*, *10*, 1023–1028. <https://doi.org/10.1123/ijsp.2014-0557>
- Marcora S. M. (2008). Do we really need a central governor to explain brain regulation of exercise performance? *European Journal of Applied Physiology*, *104*, 929–935. <https://doi.org/10.1007/s00421-008-0818-3>
- Martin, D. T., Quod, M. J., Gore, C. J., & Coyle, E. F. (2005). Has Armstrong's cycle efficiency improved? *Journal of Applied Physiology*, *99*, 1628–1629. <https://doi.org/10.1152/jappphysiol.00507.2005>
- Martin-Rincon, M., & Calbet, J. (2020). Progress Update and Challenges on  $\dot{V}O_{2\max}$  Testing and Interpretation. *Frontiers in Physiology*, *11*, 1070. <https://doi.org/10.3389/fphys.2020.01070>
- Matthews, J. N., & Altman, D. G. (1996). Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ*, *313*, 808. <https://doi.org/10.1136/bmj.313.7060.808>

- Maughan, R. J., & Leiper, J. B. (1983). Aerobic capacity and fractional utilisation of aerobic capacity in elite and non-elite male and female marathon runners. *European Journal of Applied Physiology and Occupational Physiology*, *52*, 80–87. <https://doi.org/10.1007/BF00429030>
- Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, *59*, 537–563. <https://doi.org/10.1146/annurev.psych.59.103006.093735>
- Mayhew, J. L. (1977). Oxygen cost and energy expenditure of running in trained runners. *British Journal of Sports Medicine*, *11*, 116–121. <https://doi.org/10.1136/bjism.11.3.116>
- Mazzocchi, F. (2008). Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. *EMBO reports*, *9*, 10–14. <https://doi.org/10.1038/sj.embor.7401147>
- Mazzolari, R. (2018). Exercise dose and individual response of healthy adults: is it time to re-evaluate exercise responsiveness and training recommendations? *The Journal of Physiology*, *596*, 3807–3808. <https://doi.org/10.1113/JP276141>
- Mazzolari, R., Porcelli, S., Bishop, D. J., & Lakens, D. (2022). Myths and Methodologies: The use of equivalence and non-inferiority tests for interventional studies in exercise physiology and sport science. *Experimental physiology*, 10.1113/EP090171. Advance online publication. <https://doi.org/10.1113/EP090171>
- McLaughlin, J. E., Howley, E. T., Bassett, D. R., Jr, Thompson, D. L., & Fitzhugh, E. C. (2010). Test of the classic model for predicting endurance running performance. *Medicine and Science in Sports and Exercise*, *42*, 991–997. <https://doi.org/10.1249/MSS.0b013e3181c0669d>
- McLellan, T. M., & Skinner, J. S. (1981). The use of the aerobic threshold as a basis for training. *Canadian Journal of Applied Sport Sciences*, *6*, 197–201.
- McKay, A., Stellingwerff, T., Smith, E. S., Martin, D. T., Mujika, I., Goosey-Tolfrey, V. L., Sheppard, J., & Burke, L. M. (2022). Defining Training and Performance Caliber: A Participant Classification Framework. *International Journal of Sports Physiology and Performance*, 1–15. Advance online publication. <https://doi.org/10.1123/ijsp.2021-0451>
- Meyler, S., Bottoms, L., & Muniz-Pumares, D. (2021). Biological and methodological factors affecting  $\dot{V}O_{2\max}$  response variability to endurance training and the influence of exercise

intensity prescription. *Experimental Physiology*, 106, 1410–1424.  
<https://doi.org/10.1113/EP089565>

Meyners, M. (2012). Equivalence tests – a review. *Food Quality and Preference*, 26, 231–245.  
<https://doi.org/10.1016/j.foodqual.2012.05.003>

Midgley, A. W., & McNaughton, L. R. (2006). Time at or near  $\dot{V}O_{2\max}$  during continuous and intermittent running. A review with special reference to considerations for the optimisation of training protocols to elicit the longest time at or near  $\dot{V}O_{2\max}$ . *The Journal of Sports Medicine and Physical Fitness*, 46, 1–14.

Midgley, A. W., McNaughton, L. R., & Carroll, S. (2006). Verification phase as a useful tool in the determination of the maximal oxygen uptake of distance runners. *Applied Physiology, Nutrition, and Metabolism*, 31, 541–548. <https://doi.org/10.1139/h06-023>

Midgley, A. W., McNaughton, L. R., & Wilkinson, M. (2006). Is there an optimal training intensity for enhancing the maximal oxygen uptake of distance runners?: empirical research findings, current opinions, physiological rationale and practical recommendations. *Sports Medicine*, 36, 117–132. <https://doi.org/10.2165/00007256-200636020-00003>

Midgley, A. W., McNaughton, L. R., Polman, R., & Marchant, D. (2007). Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Medicine*, 37, 1019–1028. <https://doi.org/10.2165/00007256-200737120-00002>

Midgley, A. W., McNaughton, L. R., & Jones, A. M. (2007). Training to enhance the physiological determinants of long-distance running performance: can valid recommendations be given to runners and coaches based on current scientific knowledge? *Sports Medicine*, 37, 857–880. <https://doi.org/10.2165/00007256-200737100-00003>

Midgley, A. W., McNaughton, L. R., & Carroll, S. (2007). Reproducibility of time at or near  $\dot{V}O_{2\max}$  during intermittent treadmill running. *International Journal of Sports Medicine*, 28, 40–47. <https://doi.org/10.1055/s-2006-923856>

Midgley, A. W., McNaughton, L. R., & Carroll, S. (2007). Time at  $\dot{V}O_{2\max}$  during intermittent treadmill running: test protocol dependent or methodological artefact? *International Journal of Sports Medicine*, 28, 934–939. <https://doi.org/10.1055/s-2007-964972>

- Midgley, A. W., Bentley, D. J., Luttikholt, H., McNaughton, L. R., & Millet, G. P. (2008). Challenging a dogma of exercise physiology: does an incremental exercise test for valid  $\dot{V}O_{2\max}$  determination really need to last between 8 and 12 minutes? *Sports Medicine*, *38*, 441–447. <https://doi.org/10.2165/00007256-200838060-00001>
- Midgley, A. W., & Carroll, S. (2009). Emergence of the verification phase procedure for confirming 'true'  $\dot{V}O_{2\max}$ . *Scandinavian Journal of Medicine & Science in Sports*, *19*, 313–322. <https://doi.org/10.1111/j.1600-0838.2009.00898.x>
- Midgley, A. W., Carroll, S., Marchant, D., McNaughton, L. R., & Siegler, J. (2009). Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Applied Physiology, Nutrition, and Metabolism*, *34*, 115–123. <https://doi.org/10.1139/H08-146>
- Milanović, Z., Sporiš, G., & Weston, M. (2015). Effectiveness of high-intensity interval training (HIT) and continuous endurance training for  $\dot{V}O_{2\max}$  improvements: a systematic review and meta-analysis of controlled trials. *Sports Medicine*, *45*, 1469–1481. <http://doi.org/10.1007/s40279-015-0365-0>
- Millet, G. Y. (2011). Can neuromuscular fatigue explain running strategies and performance in ultra-marathons?: the flush model. *Sports Medicine*, *41*, 489–506. <https://doi.org/10.2165/11588760-000000000-00000>
- Mills, H., Higgins, J. P., Morris, R., Kessler, D., Heron, J., Wiles, N., et al. (2020). *Detecting heterogeneity of intervention effects using analysis and meta-analysis of differences in variance between arms of a trial*. MedRxiv. <https://doi.org/10.1101/2020.03.07.20032516>
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*, c869. <https://doi.org/10.1136/bmj.c869>
- Morgan, D. W., Baldini, F. D., Martin, P. E., & Kohrt, W. M. (1989). Ten kilometer performance and predicted velocity at  $\dot{V}O_{2\max}$  among well-trained male runners. *Medicine and Science in Sports and Exercise*, *21*, 78–83. <https://doi.org/10.1249/00005768-198902000-00014>

- Morgan, D. W., Martin, P. E., Krahenbuhl, G. S., & Baldini, F. D. (1991). Variability in running economy and mechanics among trained male runners. *Medicine and Science in Sports and Exercise*, *23*, 378–383.
- Morgan, D. W., & Craib, M. (1992). Physiological aspects of running economy. *Medicine and Science in Sports and Exercise*, *24*, 456–461.
- Morgan, D. W., Bransford, D. R., Costill, D. L., Daniels, J. T., Howley, E. T., & Krahenbuhl, G. S. (1995). Variation in the aerobic demand of running among trained and untrained subjects. *Medicine and Science in Sports and Exercise*, *27*, 404–409.
- Morton, R. H., Fitz-Clarke, J. R., & Banister, E. W. (1990). Modeling human performance in running. *Journal of Applied Physiology*, *69*, 1171–1177. <https://doi.org/10.1152/jappl.1990.69.3.1171>
- Morton, R. H. (1991). The quantitative periodization of athletic training: a model study. *Sports Medicine Training and Rehabilitation*, *3*, 19–28.
- Morton, R. H. (1997). Modeling training and overtraining. *Journal of Sports Sciences*, *15*, 335–340. <https://doi.org/10.1080/026404197367344>
- Mujika, I., Busso, T., Lacoste, L., Barale, F., Geysant, A., & Chatard, J. C. (1996). Modeled responses to training and taper in competitive swimmers. *Medicine and Science in Sports and Exercise*, *28*, 251–258. <https://doi.org/10.1097/00005768-199602000-00015>
- Mujika, I. (2017). Quantification of Training and Competition Loads in Endurance Sports: Methods and Applications. *International Journal of Sports Physiology and Performance*, *12*, S29–S217. <https://doi.org/10.1123/ijsp.2016-0403>
- Muñoz, I., Seiler, S., Bautista, J., España, J., Larumbe, E., & Esteve-Lanao, J. (2014). Does polarized training improve performance in recreational runners? *International Journal of Sports Physiology and Performance*, *9*, 265–272. <https://doi.org/10.1123/ijsp.2012-0350>
- Muñoz, I., Seiler, S., Alcocer, A., Carr, N., & Esteve-Lanao, J. (2015). Specific Intensity for Peaking: Is Race Pace the Best Option? *Asian Journal of Sports Medicine*, *6*, e24900. <https://doi.org/10.5812/asj.24900>
- Murphy, K. R., Myers, B., & Wolach, A. (2014). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (4<sup>th</sup> ed.). New York, NY: Routledge. <https://doi.org/10.4324/9781315773155>

- Nelson, R. C., & Gregor, R. J. (1976). Biomechanics of distance running: a longitudinal study. *Research quarterly*, 47, 417–428.
- Newell, J., Higgins, D., Madden, N., Cruickshank, J., Einbeck, J., McMillan, K., & McDonald, R. (2007). Software for calculating blood lactate endurance markers. *Journal of Sports Sciences*, 25, 1403–1409. <https://doi.org/10.1080/02640410601128922>
- Noakes, T. D. (1988). Implications of exercise testing for prediction of athletic performance: a contemporary perspective. *Medicine and Science in Sports and Exercise*, 20, 319–330. <https://doi.org/10.1249/00005768-198808000-00001>
- Noakes, T. D., Myburgh, K. H., & Schall, R. (1990). Peak treadmill running velocity during the  $\dot{V}O_{2\max}$  test predicts running performance. *Journal of Sports Sciences*, 8, 35–45. <https://doi.org/10.1080/02640419008732129>
- Noakes, T. D. (1997). 1996 J.B. Wolffe Memorial Lecture. Challenging beliefs: *ex Africa semper aliquid novi*. *Medicine and Science in Sports and Exercise*, 29, 571–590. <https://doi.org/10.1097/00005768-199705000-00001>
- Noakes, T. D. (2000). Physiological models to understand exercise fatigue and the adaptations that predict or enhance athletic performance. *Scandinavian Journal of Medicine & Science in Sports*, 10, 123–145. <https://doi.org/10.1034/j.1600-0838.2000.010003123.x>
- Nummela, A. T., Paavolainen, L. M., Sharwood, K. A., Lambert, M. I., Noakes, T. D., & Rusko, H. K. (2006). Neuromuscular factors determining 5 km running performance and running economy in well-trained athletes. *European Journal of Applied Physiology*, 97, 1–8. <https://doi.org/10.1007/s00421-006-0147-3>
- Nummela, A., Keränen, T., & Mikkelsen, L. O. (2007). Factors related to top running speed and economy. *International Journal of Sports Medicine*, 28, 655–661. <https://doi.org/10.1055/s-2007-964896>
- Paavolainen, L. M., Nummela, A. T., & Rusko, H. K. (1999). Neuromuscular characteristics and muscle power as determinants of 5-km running performance. *Medicine and Science in Sports and Exercise*, 31, 124–130. <https://doi.org/10.1097/00005768-199901000-00020>
- Paavolainen, L., Nummela, A., & Rusko, H. (2000). Muscle power factors and  $\dot{V}O_{2\max}$  as determinants of horizontal and uphill running performance. *Scandinavian Journal of Medicine & Science in Sports*, 10, 286–291. <https://doi.org/10.1034/j.1600-0838.2000.010005286.x>

- Padulo, J., Chamari, K., & Ardigò, L. P. (2014). Walking and running on treadmill: the standard criteria for kinematics studies. *Muscles, Ligaments and Tendons Journal*, *4*, 159–162.
- Pageaux, B. (2014). The psychobiological model of endurance performance: an effort-based decision-making theory to explain self-paced endurance performance. *Sports Medicine*, *44*, 1319–1320. <https://doi.org/10.1007/s40279-014-0198-2>
- Patton, J. F., & Vogel, J. A. (1977). Cross-sectional and longitudinal evaluations of an endurance training program. *Medicine and Science in Sports*, *9*, 100–103.
- Pate, R. R., & Branch, J. D. (1992). Training for endurance sport. *Medicine and Science in Sports and Exercise*, *24*, S340–S343.
- Perez-Suarez, I., Martin-Rincon, M., Gonzalez-Henriquez, J. J., Fezzardi, C., Perez-Regalado, S., Galvan-Alvarez, V., Juan-Habib, J. W., Morales-Alamo, D., & Calbet, J. (2018). Accuracy and Precision of the COSMED K5 Portable Analyser. *Frontiers in Physiology*, *9*, 1764. <https://doi.org/10.3389/fphys.2018.01764>
- Perreault C. (2012). The pace of cultural evolution. *PloS One*, *7*, e45150. <https://doi.org/10.1371/journal.pone.0045150>
- Peserico, C. S., Zagatto, A. M., & Machado, F. A. (2014). Reliability of peak running speeds obtained from different incremental treadmill protocols. *Journal of Sports Sciences*, *32*, 993–1000. <https://doi.org/10.1080/02640414.2013.876087>
- Philp, A., Macdonald, A. L., Carter, H., Watt, P. W., & Pringle, J. S. (2008). Maximal lactate steady state as a training stimulus. *International Journal of Sports Medicine*, *29*, 475–479. <https://doi.org/10.1055/s-2007-965320>
- Piercy, K. L., Troiano, R. P., Ballard, R. M., Carlson, S. A., Fulton, J. E., Galuska, D. A., George, S. M., & Olson, R. D. (2018). The Physical Activity Guidelines for Americans. *JAMA*, *320*, 2020–2028. <https://doi.org/10.1001/jama.2018.14854>
- Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine*, *21*, 2917–2930. <https://doi.org/10.1002/sim.1296>
- Pollock M. L. (1977). Submaximal and maximal working capacity of elite distance runners. Part I: Cardiorespiratory aspects. *Annals of the New York Academy of Sciences*, *301*, 310–322. <https://doi.org/10.1111/j.1749-6632.1977.tb38209.x>

- Poole, D. C., Wilkerson, D. P., & Jones, A. M. (2008). Validity of criteria for establishing maximal O<sub>2</sub> uptake during ramp exercise tests. *European Journal of Applied Physiology*, *102*, 403–410. <https://doi.org/10.1007/s00421-007-0596-3>
- Priest, J. W., & Hagan, R. D. (1987). The effects of maximum steady state pace training on running performance. *British Journal of Sports Medicine*, *21*, 18–21. <https://doi.org/10.1136/bjism.21.1.18>
- Rasica, L., Porcelli, S., Limper, U., Mrakic-Sposta, S., Mazzolari, R., Gelmini, F., Beretta, G., & Marzorati, M. (2021). Beet on Alps: Time-course changes of plasma nitrate and nitrite concentrations during acclimatization to high-altitude. *Nitric Oxide: Biology and Chemistry*, *107*, 66–72. <https://doi.org/10.1016/j.niox.2020.12.004>
- Renfree, A., & Casado, A. (2018). Athletic Races Represent Complex Systems, and Pacing Behavior Should Be Viewed as an Emergent Phenomenon. *Frontiers in Physiology*, *9*, 1432. <https://doi.org/10.3389/fphys.2018.01432>
- Rhea, M. R. (2004). Determining the magnitude of treatment effects in strength training research through the use of the effect size. *The Journal of Strength & Conditioning Research*, *18*, 918–920. <https://doi.org/10.1519/14403.1>
- Robergs, R. A. (2017). Lessons from Popper for science, paradigm shifts, scientific revolutions and exercise physiology. *BMJ Open Sport & Exercise Medicine*, *3*, e000226. <https://doi.org/10.1136/bmjsem-2017-000226>
- Robinson, S., Edwards, H. T., & Dill, D. B. (1937). New records in human power. *Science*, *85*, 409–410. <https://doi.org/10.1126/science.85.2208.409>
- Robinson, D. M., Robinson, S. M., Hume, P. A., & Hopkins, W. G. (1991). Training intensity of elite male distance runners. *Medicine and Science in Sports and Exercise*, *23*, 1078–1082.
- Rosenblat, M. A., Perrotta, A. S., & Vicenzino, B. (2019). Polarized vs. Threshold Training Intensity Distribution on Endurance Sport Performance: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Journal of Strength and Conditioning Research*, *33*, 3491–3500. <https://doi.org/10.1519/JSC.0000000000002618>
- Ross, R., Blair, S. N., Arena, R., Church, T. S., Després, J. P., Franklin, B. A., Haskell, W. L., Kaminsky, L. A., Levine, B. D., Lavie, C. J., Myers, J., Niebauer, J., Sallis, R., Sawada, S. S., Sui, X., Wisløff, U., American Heart Association Physical Activity Committee of the Council



on Lifestyle and Cardiometabolic Health, Council on Clinical Cardiology, Council on Epidemiology and Prevention, ... Stroke Council. (2016). Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. *Circulation*, *134*, e653–e699. <https://doi.org/10.1161/CIR.0000000000000461>

Ross, R., Goodpaster, B. H., Koch, L. G., Sarzynski, M. A., Kohrt, W. M., Johannsen, N. M., Skinner, J. S., Castro, A., Irving, B. A., Noland, R. C., Sparks, L. M., Spielmann, G., Day, A. G., Pitsch, W., Hopkins, W. G., & Bouchard, C. (2019). Precision exercise medicine: understanding exercise response variability. *British Journal of Sports Medicine*, *53*, 1141–1153. <https://doi.org/10.1136/bjsports-2018-100328>

Roychoudhury, S., Scheuer, N., & Neuenschwander, B. (2018). Beyond *p*-values: A phase II dual-criterion design with statistical significance and clinical relevance. *Clinical Trials*, *15*, 452–461. <https://doi.org/10.1177/1740774518770661>

Saltin, B., & Astrand, P. O. (1967). Maximal oxygen uptake in athletes. *Journal of Applied Physiology*, *23*, 353–358. <https://doi.org/10.1152/jappl.1967.23.3.353>

Sassi, A., Impellizzeri, F. M., Morelli, A., Menaspà, P., & Rampinini, E. (2008). Seasonal changes in aerobic fitness indices in elite cyclists. *Applied Physiology, Nutrition, and Metabolism*, *33*, 735–742. <https://doi.org/10.1139/H08-046>

Saunders, P. U., Pyne, D. B., Telford, R. D., & Hawley, J. A. (2004). Factors affecting running economy in trained distance runners. *Sports Medicine*, *34*, 465–485. <https://doi.org/10.2165/00007256-200434070-00005>

Scharhag-Rosenberger, F., Meyer, T., Gässler, N., Faude, O., & Kindermann, W. (2010). Exercise at given percentages of  $VO_{2max}$ : heterogeneous metabolic responses between individuals. *Journal of Science and Medicine in Sport*, *13*, 74–79. <https://doi.org/10.1016/j.jsams.2008.12.626>

Schaun G. Z. (2017). The Maximal Oxygen Uptake Verification Phase: a Light at the End of the Tunnel?. *Sports Medicine - Open*, *3*, 44. <https://doi.org/10.1186/s40798-017-0112-1>

Schuirman, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Pharmacodynamics*, *15*, 657–680. <https://doi.org/10.1007/BF01068419>

- Schumi, J., & Wittes, J. T. (2011). Through the looking glass: understanding non-inferiority. *Trials*, *12*, 106. <https://doi.org/10.1186/1745-6215-12-106>
- Seiler, S. (2010). What is best practice for training intensity and duration distribution in endurance athletes? *International Journal of Sports Physiology and Performance*, *5*, 276–291. <https://doi.org/10.1123/ijsp.5.3.276>
- Seiler, K. S., & Kjerland, G. Ø. (2006). Quantifying training intensity distribution in elite endurance athletes: is there evidence for an "optimal" distribution? *Scandinavian Journal of Medicine & Science in Sports*, *16*, 49–56. <https://doi.org/10.1111/j.1600-0838.2004.00418.x>
- Sejersted, O. M., & Sjøgaard, G. (2000). Dynamics and consequences of potassium shifts in skeletal muscle and heart during exercise. *Physiological Reviews*, *80*, 1411–1481. <https://doi.org/10.1152/physrev.2000.80.4.1411>
- Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature*, *138*, 32. <https://doi.org/10.1038/138032a0>
- Selye, H. (1978). *The Stress of Life* (2<sup>nd</sup> ed.). New York, NY: McGraw-Hill.
- Senn, S. (2021). *Statistical issues in drug development* (3rd ed.). Hoboken, NJ: Wiley.
- Shaw, A. J., Ingham, S. A., Fudge, B. W., & Folland, J. P. (2013). The reliability of running economy expressed as oxygen cost and energy cost in trained distance runners. *Applied Physiology, Nutrition, and Metabolism*, *38*, 1268–1272. <https://doi.org/10.1139/apnm-2013-0055>
- Shepley, B., MacDougall, J. D., Cipriano, N., Sutton, J. R., Tarnopolsky, M. A., & Coates, G. (1992). Physiological effects of tapering in highly trained athletes. *Journal of Applied Physiology*, *72*, 706–711. <https://doi.org/10.1152/jappl.1992.72.2.706>
- Shieh G. (2016). Exact Power and Sample Size Calculations for the Two One-Sided Tests of Equivalence. *PloS one*, *11*, e0162093. <https://doi.org/10.1371/journal.pone.0162093>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-curve and effect size: correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*, 666–681. <https://doi.org/10.1177/1745691614553988>

- Sjödin, B., Jacobs, I., & Svedenhag, J. (1982). Changes in onset of blood lactate accumulation (OBLA) and muscle enzymes after training at OBLA. *European Journal of Applied Physiology and Occupational Physiology*, *49*, 45–57. <https://doi.org/10.1007/BF00428962>
- Sjödin, B., & Jacobs, I. (1981). Onset of blood lactate accumulation and marathon running performance. *International Journal of Sports Medicine*, *2*, 23–26. <https://doi.org/10.1055/s-2008-1034579>
- Sjödin, B., & Svedenhag, J. (1985). Applied physiology of marathon running. *Sports Medicine*, *2*, 83–99. <https://doi.org/10.2165/00007256-198502020-00002>
- Skovgaard, C., Almquist, N. W., & Bangsbo, J. (2017). Effect of increased and maintained frequency of speed endurance training on performance and muscle adaptations in runners. *Journal of Applied Physiology*, *122*, 48–59. <https://doi.org/10.1152/jappphysiol.00537.2016>
- Slawinski, J., Demarle, A., Koralsztein, J. P., & Billat, V. (2001). Effect of supra-lactate threshold training on the relationship between mechanical stride descriptors and aerobic energy cost in trained runners. *Archives of Physiology and Biochemistry*, *109*, 110–116. <https://doi.org/10.1076/apab.109.2.110.4270>
- Smith, T. P., McNaughton, L. R., & Marshall, K. J. (1999). Effects of 4-wk training using  $V_{\max}/T_{\max}$  on  $\dot{V}O_{2\max}$  and performance in athletes. *Medicine and Science in Sports and Exercise*, *31*, 892–896. <https://doi.org/10.1097/00005768-199906000-00019>
- Smith, T. P., Coombes, J. S., & Geraghty, D. P. (2003). Optimising high-intensity treadmill training using the running speed at maximal  $O_2$  uptake and the time for which this can be maintained. *European Journal of Applied Physiology*, *89*, 337–343. <https://doi.org/10.1007/s00421-003-0806-6>
- Snapinn, S. M. (2004). Alternatives for discounting in the analysis of noninferiority trials. *Journal of Biopharmaceutical Statistics*, *14*, 263–273. <https://doi.org/10.1081/BIP-120037178>
- Snapinn, S., & Jiang, Q. (2018). Controlling the type 1 error rate in non-inferiority trials. *Statistics in Medicine*, *27*, 371–381. <https://doi.org/10.1002/sim.3072>
- Snapinn, S., & Jiang, Q. (2018). Preservation of effect and the regulatory approval of new treatments on the basis of non-inferiority trials. *Statistics in Medicine*, *27*, 382–391. <https://doi.org/10.1002/sim.3073>

- Sparling, P. B. (1984). Physiological Determinants of Distance Running Performance. *The Physician and Sportsmedicine*, 12, 68–77. <https://doi.org/10.1080/00913847.1984.11701795>
- Speed, H. D., & Andersen, M. B. (2000). What exercise and sport scientists don't understand. *Journal of Science and Medicine in Sport*, 3, 84–92. [https://doi.org/10.1016/s1440-2440\(00\)80051-1](https://doi.org/10.1016/s1440-2440(00)80051-1)
- Spiriev, B., & Spiriev, A. (2021, December 12). *IAAF Scoring Tables Of Athletics – 2017 Revised Edition*. World Athletics. <https://www.worldathletics.org/news/iaaf-news/scoring-tables-2017>
- Spurrs, R. W., Murphy, A. J., & Watsford, M. L. (2003). The effect of plyometric training on distance running performance. *European Journal of Applied Physiology*, 89, 1–7. <https://doi.org/10.1007/s00421-002-0741-y>
- Stagno, K. M., Thatcher, R., & van Someren, K. A. (2007). A modified TRIMP to quantify the in-season training load of team sport players. *Journal of Sports Sciences*, 25, 629–634. <https://doi.org/10.1080/02640410600811817>
- St Clair Gibson, A., & Noakes, T. D. (2004). Evidence for complex system integration and dynamic neural regulation of skeletal muscle recruitment during exercise in humans. *British Journal of Sports Medicine*, 38, 797–806. <https://doi.org/10.1136/bjism.2003.009852>
- St Clair Gibson, A., Swart, J., & Tucker, R. (2018). The interaction of psychological and physiological homeostatic drives and role of general control principles in the regulation of physiological systems, exercise and the fatigue process - The Integrative Governor theory. *European Journal of Sport Science*, 18, 25–36. <https://doi.org/10.1080/17461391.2017.1321688>
- Stegmann, H., & Kindermann, W. (1982). Comparison of prolonged exercise tests at the individual anaerobic threshold and the fixed anaerobic threshold of 4 mmol · l<sup>-1</sup> lactate. *International Journal of Sports Medicine*, 3, 105–110. <https://doi.org/10.1055/s-2008-1026072>
- Stöggl, T. (2021, December 12). *What is the Best Way to Train to Become a Star Endurance Athlete?* Frontiers for Young Minds. <https://kids.frontiersin.org/articles/10.3389/frym.2018.00017#ref1>

- Stöggl, T., & Sperlich, B. (2014). Polarized training has greater impact on key endurance variables than threshold, high intensity, or high volume training. *Frontiers in Physiology, 5*, 33. <https://doi.org/10.3389/fphys.2014.00033>
- Stöggl, T. L., & Sperlich, B. (2015). The training intensity distribution among well-trained and elite endurance athletes. *Frontiers in Physiology, 6*, 295. <https://doi.org/10.3389/fphys.2015.00295>
- Stellingwerf, T. (2012). Case study: Nutrition and training periodization in three elite marathon runners. *International Journal of Sport Nutrition and Exercise Metabolism, 22*, 392–400. <https://doi.org/10.1123/ijsnem.22.5.392>
- Svedenhag, J., & Sjödin, B. (1985). Physiological characteristics of elite male runners in and off-season. *Canadian Journal of Applied Sport Sciences, 10*, 127–133.
- Sylta, O., Tønnessen, E., & Seiler, S. (2014). From heart-rate data to training quantification: a comparison of 3 methods of training-intensity analysis. *International Journal of Sports Physiology and Performance, 9*, 100–107. <https://doi.org/10.1123/IJSPP.2013-0298>
- Tanaka, K., Watanabe, H., Konishi, Y., Mitsuzono, R., Sumida, S., Tanaka, S., Fukuda, T., & Nakadomo, F. (1986). Longitudinal associations between anaerobic threshold and distance running performance. *European Journal of Applied Physiology and Occupational Physiology, 55*, 248–252. <https://doi.org/10.1007/BF02343795>
- Tanaka, K., & Matsuura, Y. (1984). Marathon performance, anaerobic threshold, and onset of blood lactate accumulation. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology, 57*, 640–643. <https://doi.org/10.1152/jappl.1984.57.3.640>
- Tanner, R., & Gore, C. (2012). *Physiological Tests for Elite Athletes* (2<sup>nd</sup> ed.). Champaign, IL: Human Kinetics.
- Taylor, H. L., Buskirk, E., & Henschel, A. (1955). Maximal oxygen intake as an objective measure of cardio-respiratory performance. *Journal of Applied Physiology, 8*, 73–80. <https://doi.org/10.1152/jappl.1955.8.1.73>
- Thoden, J. S., Wilson, B. A., & MacDougall, J. D. (1982) Testing aerobic power. In J. D. MacDougall, H. A. Wenger, & H. J. Green (Eds.), *Physiological testing of the elite athlete* (pp. 39–54). Canadian Association of Sports Sciences.

- Thompson, D. D., Lingsma, H. F., Whiteley, W. N., Murray, G. D., & Steyerberg, E. W. (2015). Covariate adjustment had similar benefits in small and large randomized controlled trials. *Journal of Clinical Epidemiology*, *68*, 1068–1075. <https://doi.org/10.1016/j.jclinepi.2014.11.001>
- Tjelta, L. I. (2016). The training of international level distance runners. *International Journal of Sports Science & Coaching*, *11*, 122–134. <https://doi.org/10.1177/1747954115624813>
- Tjelta, L. I., Tønnessen, E., & Enoksen, E. (2014). A Case Study of the Training of Nine Times New York Marathon Winner Grete Waitz. *International Journal of Sports Science & Coaching*, *9*, 139–158. <https://doi.org/10.1260/1747-9541.9.1.139>
- Tschakert, G., & Hofmann, P. (2013). High-intensity intermittent exercise: methodological and physiological aspects. *International Journal of Sports Physiology and Performance*, *8*, 600–610. <https://doi.org/10.1123/ijsp.8.6.600>
- Turnes, T., de Aguiar, R. A., Cruz, R. S., & Caputo, F. (2016). Interval training in the boundaries of severe domain: effects on aerobic parameters. *European Journal of Applied Physiology*, *116*, 161–169. <https://doi.org/10.1007/s00421-015-3263-0>
- Ulmer, H. V. (1996). Concept of an extracellular regulation of muscular metabolic rate during heavy exercise in humans by psychophysiological feedback. *Experientia*, *52*, 416–420. <https://doi.org/10.1007/BF01919309>
- Urhausen, A., Coen, B., Weiler, B., & Kindermann, W. (1993). Individual anaerobic threshold and maximum lactate steady state. *International Journal of Sports Medicine*, *14*, 134–139. <https://doi.org/10.1055/s-2007-1021157>
- Vallejo, G., Ato, M., Fernández, M. P., & Livacic-Rojas, P. E. (2019). Sample size estimation for heterogeneous growth curve models with attrition. *Behavior Research Methods*, *51*, 1216–1243. <https://doi.org/10.3758/s13428-018-1059-y>
- Van Ravenzwaaij, D., Monden, R., Tendeiro, J. N., Ioannidis, J. P. A. (2019). Bayes factors for superiority, non-inferiority, and equivalence designs. *BMC Medical Research Methodology*, *19*, 71. <https://doi.org/10.1186/s12874-019-0699-7>
- Venhorst, A., Micklewright, D., & Noakes, T. D. (2018). Towards a three-dimensional framework of centrally regulated and goal-directed exercise behaviour: a narrative review. *British Journal of Sports Medicine*, *52*, 957–966. <https://doi.org/10.1136/bjsports-2016-096907>

Venhorst, A., Micklewright, D. P., & Noakes, T. D. (2018). The Psychophysiological Regulation of Pacing Behaviour and Performance Fatigability During Long-Distance Running with Locomotor Muscle Fatigue and Exercise-Induced Muscle Damage in Highly Trained Runners. *Sports Medicine - Open*, 4, 29. <https://doi.org/10.1186/s40798-018-0143-2>

Venhorst, A., Micklewright, D. P., & Noakes, T. D. (2018). Modelling perception-action coupling in the phenomenological experience of "hitting the wall" during long-distance running with exercise induced muscle damage in highly trained runners. *Sports Medicine - Open*, 4, 30. <https://doi.org/10.1186/s40798-018-0144-1>

Vickers, A. J. (2001). The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Medical Research Methodology*, 1, 6. <https://doi.org/10.1186/1471-2288-1-6>

Vickers, A. J., & Altman, D. G. (2001). Statistics notes: Analysing controlled trials with baseline and follow up measurements. *BMJ*, 323, 1123–1124. <https://doi.org/10.1136/bmj.323.7321.1123>

Victor, N. (1987). On clinically relevant differences and shifted null hypotheses. *Methods of Information in Medicine*, 26, 109–116. <https://doi.org/10.1055/s-0038-1635499>

Viru, A. (2002). Early contributions of Russian stress and exercise physiologists. *Journal of Applied Physiology*, 92, 1378–1382. <https://doi.org/10.1152/jappphysiol.00435.2001>

Wang, S. J., & Blume, J. D. (2011). An evidential approach to non-inferiority clinical trials. *Pharmaceutical Statistics*, 10, 440–447. <https://doi.org/10.1002/pst.513>

Weatherwax, R. M., Harris, N. K., Kilding, A. E., & Dalleck, L. C. (2019). Incidence of  $\dot{V}O_{2max}$  Responders to Personalized versus Standardized Exercise Prescription. *Medicine and Science in Sports and Exercise*, 51, 681–691.

Weltman, A., Snead, D., Seip, R., Schurrer, R., Levine, S., Rutt, R., Reilly, T., Weltman, J., & Rogol, A. (1987). Prediction of lactate threshold and fixed blood lactate concentrations from 3200-m running performance in male runners. *International Journal of Sports Medicine*, 8, 401–406. <https://doi.org/10.1055/s-2008-1025694>

Weltman, A., Seip, R. L., Snead, D., Weltman, J. Y., Haskvitz, E. M., Evans, W. S., Veldhuis, J. D., & Rogol, A. D. (1992). Exercise training at and above the lactate threshold in previously

untrained women. *International Journal of Sports Medicine*, *13*, 257–263. <https://doi.org/10.1055/s-2007-1021263>

Wen, D., Utesch, T., Wu, J., Robertson, S., Liu, J., Hu, G., & Chen, H. (2019). Effects of different protocols of high intensity interval training for  $\dot{V}O_{2\max}$  improvements in adults: A meta-analysis of randomised controlled trials. *Journal of Science and Medicine in Sport*, *22*, 941–947. <https://doi.org/10.1016/j.jsams.2019.01.013>

Wenger, H. A., & Macnab, R. B. (1975). Endurance training: the effects of intensity, total work, duration and initial fitness. *The Journal of Sports Medicine and Physical Fitness*, *15*, 199–211.

Wenger, H. A., & Bell, G. J. (1986). The interactions of intensity, frequency and duration of exercise training in altering cardiorespiratory fitness. *Sports Medicine*, *3*, 346–356. <https://doi.org/10.2165/00007256-198603050-00004>

Westlake, W. J. (1981). Response to T.B.L. Kirkwood: Bioequivalence testing – a need to rethink. *Biometrics*, *37*, 589–594. <https://doi.org/10.2307/2530573>

Wolpern, A. E., Burgos, D. J., Janot, J. M., & Dalleck, L. C. (2015). Is a threshold-based model a superior method to the relative percent concept for establishing individual exercise intensity? a randomized controlled trial. *BMC Sports Science, Medicine & Rehabilitation*, *7*, 16. <https://doi.org/10.1186/s13102-015-0011-z>

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*, *310*, 2191–2194. <https://doi.org/10.1001/jama.2013.281053>

Wyndham, C. H., Strydom, N. B., van Rensburg, A. J., & Benade, A. J. (1969). Physiological requirements for world-class performances in endurance running. *South African Medical Journal*, *43*, 996–1002.

Yoshida, T., Udo, M., Chida, M., Ichioka, M., Makiguchi, K., & Yamaguchi, T. (1990). Specificity of physiological adaptation to endurance training in distance runners and competitive walkers. *European Journal of Applied Physiology and Occupational Physiology*, *61*, 197–201. <https://doi.org/10.1007/BF00357599>

Yu, B., Yang, H., & Sabin, B. (2019). A note on the determination of non-inferiority margins with application in oncology clinical trials. *Contemporary Clinical Trials Communications*, *16*, 100454. <https://doi.org/10.1016/j.conctc.2019.100454>



Zavorsky, G. S., Montgomery, D. L., & Pearsall, D. J. (1998). Effect of intense interval workouts on running economy using three recovery durations. *European Journal of Applied Physiology and Occupational Physiology*, 77, 224–230.

Zuccarelli, L., Porcelli, S., Rasica, L., Marzorati, M., & Grassi, B. (2018). Comparison between Slow Components of HR and  $\dot{V}O_2$  Kinetics: Functional Significance. *Medicine and Science in Sports and Exercise*, 50, 1649–1657. <https://doi.org/10.1249/MSS.0000000000001612>

Zavorsky, G. S., Montgomery, D. L., & Pearsall, D. J. (1998). Effect of intense interval workouts on running economy using three recovery durations. *European Journal of Applied Physiology and Occupational Physiology*, 77, 224–230. <https://doi.org/10.1007/s004210050326>

## 9

**ADDENDUMS**

*'Above all, don't fear difficult moments. The best comes from them.'*

*Rita Levi-Montalcini – neurologist*





## 9. ADDENDUMS

### 9.1 Addendum 1: Presentation at International Conference



**EUROPEAN COLLEGE OF  
SPORT SCIENCE**

#### EUROPEAN COLLEGE OF SPORT SCIENCE

Aachener Str. 1053 -1055

50858 Cologne

GERMANY

VAT-ID: DE251715668 - St.Nr.: 223/5905/0216

register of associations: VR12508

Cologne, 11.09.2021 - 13:57:13

### Confirmation of Presentation

This is to certify that the following title has been presented at the 26th Annual Congress of the European College of Sport Science between 8 - 10 September 2021.

#### Raffaele Mazzolari

University of the Basque Country (UPV/EHU)

Portal de Lasarte 71

01007 Vitoria-Gasteiz, Spain

Abstr.-ID: 418, Presentation format: Oral , Session name: OP-API2 - Analyses of Workload and performance

Title: Individualized high-intensity interval training improves some physiological determinants of performance more than standardized training in distance runners

Authors: Mazzolari, R.1,2, Villanova, S.2, Santos Concejero, J.1, Porcelli, S.2,3

Institution: 1University of the Basque Country ; 2University of Pavia; 3National Research Council

Presentation date: 08.09.2021, 16:30, Lecture room: Track 3, No: 2

European College of Sport Science

This document has been created digitally and is valid without a signature

---

Privacy Policy (<http://sport-science.org/index.php/privacy-policy>) - Terms & Conditions (<https://sport-science.org/index.php/privacy-policy?id=78>)

Copyright © 2021 European College of Sport Science, All Rights Reserved.  
The ECSS is a non profit organisation, dedicated to Sport Science.

compared with passive heating and threshold running. Such findings provide further support for the application of passive hot water immersion and high intensity exercise for maintaining cerebral and peripheral vasculature health in young healthy individuals.

#### ECCENTRIC VERSUS CONCENTRIC CYCLING EFFECTS ON CEREBRAL BLOOD FLOW IN HEALTHY ADULTS

BOLAM, L.M., CARTER, H.H., NOSAKA, K., GREEN, D.J., NAYLOR, L.H.

THE UNIVERSITY OF WESTERN AUSTRALIA

##### INTRODUCTION:

INTRODUCTION: Eccentric cycling (ECC) allows for a greater muscular workload at a reduced cardiovascular burden, prolongs the onset of the ventilatory threshold, and is associated with greater neuronal activation compared to concentric cycling (CON). The present study investigated if cerebrovascular responses would be different between ECC and CON.

##### METHODS:

METHODS: Healthy adults (8♂, 8♀) were randomized to a within-subject cross-over study, using ramped CON and ECC cycling protocols with workload being matched at 50-65% heart rate (HR) reserve. Middle (MCAV) and posterior (PCAV) cerebral artery velocities, and femoral artery blood flow during and after cycling were measured via B-mode ultrasound. Blood pressure (MAP), HR, rate pressure product (RPP), cardiorespiratory measures and rate of perceived exertion (RPE) were also compared between exercise modalities.

##### RESULTS:

RESULTS: Cerebral blood flow increased to the same extent during ECC and CON, despite lower respiratory ( $\dot{V}O_2$   $14 \pm 3$  vs  $17 \pm 3$  ml·kg<sup>-1</sup>·min<sup>-1</sup>,  $P=0.019$ ; PETCO<sub>2</sub>  $27 \pm 3$  vs  $31 \pm 3$  mmHg,  $P=0.031$ ) and cardiovascular responses (SBP  $151 \pm 18$  vs  $166 \pm 18$  mmHg,  $P=0.032$ , RPP  $18359 \pm 643$  vs  $20308 \pm 660$  mmHg·bpm), with a hyperventilatory response (RR  $29 \pm 5$  vs  $22 \pm 5$  BPM,  $P=0.025$ ;  $\dot{V}_t$   $1 \pm 0.2$  vs  $2 \pm 0.2$  L,  $P<0.001$ ) during ECC than CON. Similar physical RPE ( $14 \pm 2$  vs  $13 \pm 2$ ,  $P=1.00$ ) and workload ( $146 \pm 51$ W vs  $89 \pm 51$ W,  $P=0.16$ ) were reported, with higher reported mental RPE ( $14 \pm 2$  vs  $11 \pm 3$ ,  $P=0.035$ ) in ECC.

##### CONCLUSION:

CONCLUSION: When matched for HR, ECC cycling conferred similar cerebral blood flow velocity despite significantly lower physiological regulators such as PETCO<sub>2</sub>,  $\dot{V}O_2$  and SBP. This highlights the potential for ECC cycling as a beneficial mode of exercise to optimise cerebrovascular blood flow for individuals with reduced  $\dot{V}O_2$ , such as those with advancing age and clinical conditions, including cardiovascular disease.

#### OP-AP12 Analyses of Workload and performance

#### LONGITUDINAL EXAMINATION OF IN-SEASON EXTERNAL WORKLOAD IN WOMEN COLLEGIATE BASKETBALL ATHLETES

BROWN, F.S.A., FIELDS, J.B., BAKER, R.E., JONES, M.T.

GEORGE MASON UNIVERSITY, USA; SPRINGFIELD COLLEGE, USA

##### INTRODUCTION:

Quantifying external workload provides insight into physical stress imposed upon athletes during training and games. By design, pre-season training should prepare athletes for the rigors of the competitive season, yet athletes often experience higher workloads in games than practice (1). Further, lower pre-season workloads are associated with a higher incidence of in-season injury (2). Substantial differences in workloads of primary and secondary players have been reported in sports other than women's collegiate basketball (3). Therefore, the purpose was to compare game external workloads of primary and secondary players, and categorize pre-season practices relative to game load.

##### METHODS:

National Collegiate Athletic Association Division I women collegiate basketball athletes ( $n=16$ , mean±SD age:  $20.46 \pm 1.56$  yr; body mass:  $83.01 \pm 42.94$  kg.; height:  $176.95 \pm 3.17$ cm; bodyfat:  $24.07 \pm 9.68$ %;  $\dot{V}O_{2max}$ :  $45.45 \pm 7.56$ ml·kg<sup>-1</sup>·min<sup>-1</sup>) participated. External load metrics included player load (PL), PL/min, high inertial movement analysis (IMA;  $>3.5$  m/s<sup>2</sup>), and total jumps (TJ). Metrics were collected for two seasons of conference games ( $n=32$ ), and pre-season practices ( $n=57$ ) utilizing 10 Hz GPS/GNSS technology. Athletes were classified based upon game minutes played as primary (played  $\geq 15$  minutes/game) or secondary (played  $< 15$  minutes/game). Multivariate analyses of variance assessed differences in game metrics between primary and secondary players ( $p<0.05$ ). Practice metrics were classified relative to game metrics as: high ( $>1$  SD above the mean), medium (1 SD below the mean), low (2 SD below the mean), and very low (3 SD below the mean).

##### RESULTS:

During gameplay, primary players experienced greater PL ( $F=470.37$ ,  $p<0.001$ ,  $612.50 \pm 144.69$  vs.  $295.32 \pm 210.64$  AU), PL/min ( $F=453.91$ ,  $p<0.001$ ,  $5.49 \pm 1.28$  vs.  $2.68 \pm 1.22$  AU/min), IMA ( $F=334.17$ ,  $p<0.001$ ,  $40.27 \pm 15.76$  vs.  $13.53 \pm 11.40$  #), and TJ ( $F=61.20$ ,  $p<0.001$ ,  $52.49 \pm 20.70$  vs.  $36.52 \pm 17.80$  #) compared to secondary players. Compared to games, pre-season practice metrics were classified as: a) PL: 7% high, 44% medium, 42% low, 7% very low; b) PL/min: 12% high, 40% medium, 37% low, 10% very low; c) IMA: 2% high, 47% medium, 49% low, 2% very low; d) TJ: 42% high, 40% medium, 16% low, 2% very low.

##### CONCLUSION:

Findings indicate primary players experienced greater external game workloads than secondary players. Further, the majority of pre-season practices were of medium loads for PL and PL/min, low loads for IMA, and high loads for TJ, indicating game loads were not achieved consistently. It is recommended practitioners strategically plan pre-season practices in order to ensure athletes are prepared to sustain game workloads. In-season, supplemental training for secondary players may be useful to avoid detraining adaptations.

#### INDIVIDUALIZED HIGH-INTENSITY INTERVAL TRAINING IMPROVES SOME PHYSIOLOGICAL DETERMINANTS OF PERFORMANCE MORE THAN STANDARDIZED TRAINING IN DISTANCE RUNNERS

MAZZOLARI, R.1,2, VILLANOVA, S.2, SANTOS CONCEJERO, J.1, PORCELLI, S.2,3

1UNIVERSITY OF THE BASQUE COUNTRY; 2UNIVERSITY OF PAVIA; 3NATIONAL RESEARCH COUNCIL

##### INTRODUCTION:

Coaches often include high-intensity interval training (HIIT) sessions in their plan to develop specific physiological adaptations and enhance athletic performance in distance runners. The prescription is usually performed by a standardized and common empirical approach, in which runners of a similar level train together using the same exercise intensities. However, previous studies [1-5] showed that it is possible to target some physiological adaptations by individualizing HIIT prescriptions according to the physiological characteristics of the runners and their responses to exercise. As far as we know, there has been no investigation comparing the effects of an evidence-based individualized HIIT against a standardized HIIT in a large group of runners. Therefore, we compared the effects of these two different training prescription models on physiological adaptations and performance in distance runners.

#### METHODS:

Forty-five distance runners (35 men and 10 women, age: 36±8 yr, 10000-m performance: 44:06±5:01) were tested before and after 8 weeks of training. Maximal oxygen uptake (VO<sub>2</sub>max), running economy (RE), peak treadmill running speed (V<sub>peak</sub>), and lactate threshold (D<sub>mod</sub>) were determined during an incremental test performed on a treadmill, followed by a verification test. 10000-m performance was measured on a 400-m track. The intervention consisted of replacing ~20% of the runners' habitual weekly training volume with 2 sessions of either individualized (IND; n=23) or standardized (STD; n=22) HIIT. Repeated intervals were performed at the speed halfway between V<sub>peak</sub> and D<sub>mod</sub>, as described by [1-5]. HIIT prescriptions were individualized according to each runner's physiological values in IND, whereas the average group values were used for STD. IND and STD were matched for HIIT characteristics at the group level.

#### RESULTS:

VO<sub>2</sub>max significantly improved only in IND (+2.7 mL/kg/min, p=.003, 95% confidence interval [CI]: +1.0 to +4.5), with a statistical difference of 3.4 mL/kg/min (p=.038, 95%CI: 0.2 to 6.5) between the groups. RE significantly improved only in IND (-4.5 mL/kg/km, p=.030, 95%CI: -8.6 to -0.5), with no significant difference between the groups. V<sub>peak</sub> significantly improved only in IND (+0.4 km/h, p<.001, 95%CI: +0.2 to +0.6), with a statistical difference of 0.5 km/h (p=.022, 95%CI: 0.1 to 0.9) between the groups. No significant changes or differences were observed in D<sub>mod</sub> and 1000-m performance.

#### CONCLUSION:

The present findings show that individualizing HIIT prescriptions according to the physiological responses to exercise induces superior adaptations for some physiological variables related to performance (VO<sub>2</sub>max and V<sub>peak</sub>) compared with standardized prescriptions in distance runners. These results may be helpful for coaches aiming to improve specific physiological attributes of their athletes during certain training phases.

1. Demarle et al. (2001) 2. Slawinski et al. (2001) 3. Demarle et al. (2003) 4. Garcin et al. (2002) 5. Lafitte et al. (2004)

### TRAINING LOAD MONITORING IN TEAM SPORTS: A PRACTICAL APPROACH TO ADDRESSING MISSING DATA

GRIFFIN, A., KENNY, I.C., COMYNS, T.M., PURTILL, H., TIERNAN, C., O SHAUGHNESSY, E., LYONS, M.

UNIVERSITY OF LIMERICK

#### INTRODUCTION:

Training load (TL) is a modifiable risk factor that may provide practitioners with opportunities to mitigate injury risk and increase sports performance. A regular problem encountered by practitioners however, is the issue of missing TL data. The purpose of this study was to examine the impact of missing TL data in team sports and to offer a practical and effective method of missing value imputation (MVI) to address this.

#### METHODS:

Session rating of perceived exertion (sRPE) data from 10 male professional soccer players (age, 24.8 ± 5.0 years; height, 181.2 ± 5.1 cm; mass, 78.7 ± 6.4 kg) were collected over a 32-week season. Data were randomly removed from the complete dataset at a range of 5 – 50% in increments of 5%. Data were then imputed using twelve MVI methods. These methods were theorized and examined based on their structure and their practicality (i.e. can be feasibly used by a practitioner working with a team). Additional metrics derived from sRPE data were calculated, and illustrated, to demonstrate the impact missing data has on their computation. The Normalized Root Mean Squared Error (NRMSE) and the Mean of Absolute Deviations (MAD), standard statistical metrics to measure model performance, were used to compare the performance of the MVI methods across the levels of missingness.

#### RESULTS:

The best-fitting MVI method was Daily Team Mean (DTMean) with NRMSE ranging from 0.42 – 0.46 across all levels of missingness. The MVI methods of Daily Team Median (DTMedian), Mixed Mean, Mixed Median, Mixed R4W Mean and Mixed R4W Median produced similar NRMSE values of 0.44 – 0.49. The methods of Player Mean (PM), Player Median, R4W Player Mean, R4W Player Median, Day of Week Mean and Day of Week Median were each a poorer fit with NRMSE values of 0.48 – 0.68. For all imputation methods, as the percentage of missing data increases so too does the MAD in a linear manner. At 5% missingness, the mean MAD value of all MVI methods is 9.2 while at 50% missingness this increased to 90.8. Additionally, not addressing missing sRPE data may lead to more inaccurate calculations of other TL metrics (e.g. acute chronic workload ratio, training monotony, training strain).

#### CONCLUSION:

Practitioners should strive to keep missing data at a minimum. However, when collecting TL data longitudinally, missing data is inevitable due to many factors. The DTMean MVI method offers practitioners, working with teams, a practical and effective method of addressing the negative consequences of missing TL data. The DTMean MVI method appears to be easy to use, quick to calculate, does not require specialist software or previous data collection. As the monitoring of TL is becoming more and more prevalent, the findings of this study address an ever increasing challenge for practitioners working with team sport athletes.

### ASSUMING THE STANDARD 1 MET VALUE OF 3.5 ML/KG/MIN IN OLDER ADULTS MISCLASSIFIES TIME SPEND IN SB AND PA INTENSITIES.

LEAL MARTÍN, J., MAÑAS, A., ALFARO ACHA, A., MUÑOZ MUÑOZ, M., ALEGRE, L.M., GARCÍA GARCÍA, F.J., ARA, I.

UNIVERSITY OF CASTILLA-LA MANCHA

#### INTRODUCTION:

The metabolic equivalent of task (MET) provides a feasible approach for classifying physical activity (PA) intensity as a multiple of the resting metabolic rate (RMR). The RMR standard value of 1 MET is generally assumed as 3.5 ml/kg/min, a value that has been criticized for being inappropriate in the older adult population. This fact could represent an important bias when classifying the activity intensity and derived scientific conclusions. However, to our knowledge, it has hardly been investigated. Therefore, the aim of this work was to assess

## 9.2 Addendum 2: Letters of stay



September 4, 2019

To Whom It May Concern,

I am pleased to write on behalf of Raffaele Mazzolari who is currently undertaking the Doctoral Programme in Physical Activity and Sport at the University of the Basque Country (UPV/EHU) in Vitoria-Gasteiz, Spain.

Raffaele is currently attending the Physiopathology of Exercise Laboratory "Rodolfo Margaria" at the National Research Council in Segrate (Milan), Italy, where he has been collaborating on various ongoing projects since the time of his arrival (April 24, 2019), with a particular interest in an ongoing line of research about trainability and inter-individual response to exercise. During the time he spent with us, Raffaele has been actively involved in recruiting volunteers and organizing the experimental setup. Moreover, he provided assistance during several physiological evaluations aimed to characterize pulmonary, cardiovascular and metabolic response to exercise of participants volunteering for other research projects running in the lab.

Sincerely,

Simone Porcelli, MD PhD  
Senior Lecturer  
Physiopathology of Exercise Lab "Rodolfo Margaria"  
Institute of Biomedical Technologies  
LITA Building, Via Fratelli Cervi 93, Segrate, Italy  
Email: [simone.porcelli@itb.cnr.it](mailto:simone.porcelli@itb.cnr.it)

*Sede Istituto*  
Via F.lli Cervi, 93  
20090 Segrate (MI)  
Tel: +39 02 26422702  
Fax: +39 02 26422770

*UOS di Bari*  
Via G. Amendola 122/D  
70126 Bari  
Tel: +39 080 5929680  
Fax: +39 080 5929690

*UOS di Roma*  
C.ne Nomentana, 496  
00162 Roma  
Tel: +39 06 441622207  
Fax: +39 06 44254397

*UO di Pisa*  
Via G. Moruzzi, 1  
56100 Pisa  
Tel: +39 050 3152792  
Fax: +39 050 3153973



September 3, 2020

To Whom It May Concern,

I am pleased to write on behalf of Raffaele Mazzolari who is currently undertaking the Doctoral Programme in Physical Activity and Sport at the University of the Basque Country (UPV/EHU) in Vitoria-Gasteiz, Spain.

During the last year spent in Italy, Raffaele has moved the experimental phase of his PhD Project from the Institute of Biomedical Technologies at the National Research Council in Segrate (Milan) to the Department of Molecular Medicine of the University of Pavia, in accordance to my new appointment as Assistant Professor in that Institution. Other than working on his PhD research project, Raffaele collaborated with other members of our Research Group, being involved in several new studies about environmental and exercise physiology (including animal models), and sports performance. Raffaele worked with great diligence and attitude, demonstrating interest for extending his current knowledge and competences.

Sincerely,



Simone Porcelli, MD PhD  
Assistant Professor in Human Physiology  
Institute of Physiology  
Department of Molecular Medicine  
University of Pavia, Italy  
Via Forlanini 6, 27100 Pavia - Italy





September 3, 2021

To Whom It May Concern,

I am pleased to write on behalf of Raffaele Mazzolari who is currently undertaking the Doctoral Programme in Physical Activity and Sport at the University of the Basque Country (UPV/EHU) in Vitoria Gasteiz, Spain.

During the last year (from September 2020 to September 2021), Raffaele spent his time working with my research group at the Department of Molecular Medicine of the University of Pavia (Italy). Raffaele's work was mainly focused on collecting data for his PhD project but he also collaborated to other projects of our Research Group. Raffaele has kept working with great diligence and attitude, extending his previous knowledge and competencies.

Sincerely,



Simone Porcelli, MD PhD  
Assistant Professor in Human Physiology  
Institute of Physiology  
Department of Molecular Medicine  
University of Pavia, Italy  
Via Forlanini 6, 27100 Pavia - Italy

## 9.3 Addendum 3: Research ethics approval



NAZIOARTEKO  
BIKAINASUN  
CAMPUSA  
CAMPUS DE  
EXCELENCIA  
INTERNACIONAL

IKERKETA SAILEKO ERREKTOREORDETZA  
VICERRECTORADO DE INVESTIGACIÓN

GIZAKIEKIN ETA HAUEN LAGIN ETA DATUEKIN  
EGINDAKO IKERKETEI BURUZKO ETIKA  
BATZORDEAREN (GIEB-UPV/EHU) TXOSTENA

M<sup>a</sup> Jesús Marcos Muñoz andreak, Universidad del País Vasco/Euskal Herriko Unibertsitateko (UPV/EHU) GIEBeko idazkari gisa,

### ZIURTATZEN DU

Ezen gizakiek in egindako ikerkuntzaren etika batzorde honek, GIEB-UPV/EHU, (2014/2/17ko 32. EHAA)

Balioetsi duela ondoko ikertzailearen proposamen hau:

Jordan Santos Concejero andreak, M10\_2018\_012, honako ikerketa proiektu hau egiteko:

"Respuesta individual al entrenamiento de alta intensidad en corredores de resistencia de alto nivel"

Eta aintzat hartuta ezen

1. Ikerketa justifikatuta dago, bere helburuei esker jakintza areagotu eta gizarteari onura ekarriko baitio, ikerlanak lekartzakeen eragozpen eta arriskuak arazoizko izanik.
2. Ikertzaile taldearen gaitasuna eta erabilgarri dituzten baliabideak aproposak dira proiektua gauzatzeko.
3. Ikerketaren planteamendua bat dator era honetako ikerkuntza egin ahal izateko baldintza metodologiko eta etikoekin, ikerkuntza zientifikoaren praktika egokien irizpideei jarraiki.
4. Indarreko arauak betetzen ditu, ikerketa egin ahal izateko baimenak, akordioak edo hitzarmenak bame.

Aldeko Txostena eman du 2018ko urtarrilaren 25ean egin duen bileran (96/2018akta) aipatutako ikerketa proiektua ondoko ikertzaileek osatutako taldeak egin dezan:

Jordan Santos Concejero  
Raffaele Mazzolari

Eta halaxe sinatu du Leioan, 2018ko otsailaren 12an

M<sup>a</sup> Jesús Marcos Muñoz  
GIEB-UPV/EHUko idazkari teknikoa  
Secretaria Técnica del CEISH-UPV/EHU

INFORME DEL COMITÉ DE ÉTICA PARA LAS  
INVESTIGACIONES CON SERES HUMANOS, SUS  
MUESTRAS Y SUS DATOS (CEISH-UPV/EHU)

M<sup>a</sup> Jesús Marcos Muñoz como Secretaria del CEISH de la Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)

### CERTIFICA

Que este Comité de Ética para la Investigación con Seres Humanos, CEISH-UPV/EHU, BOPV 32, 17/2/2014, Ha evaluado la propuesta del investigador:

D. Jordan Santos Concejero, M10\_2018\_012, para la realización del proyecto de investigación: "*Respuesta individual al entrenamiento de alta intensidad en corredores de resistencia de alto nivel*"

Y considerando que,

1. La investigación está justificada porque sus objetivos permitirán generar un aumento del conocimiento y un beneficio para la sociedad que hace asumibles las molestias y riesgos previsibles.
2. La capacidad del equipo investigador y los recursos disponibles son los adecuados para realizarla.
3. Se plantea según los requisitos metodológicos y éticos necesarios para su ejecución, según los criterios de buenas prácticas de la investigación científica.
4. Se cumple la normativa vigente, incluidas las autorizaciones, acuerdos o convenios necesarios para llevarla a cabo.

Ha emitido en la reunión celebrada el 25 de enero de 2018 (acta 96/2018), INFORME FAVORABLE a que dicho proyecto de investigación sea realizado, por el equipo investigador:

Jordan Santos Concejero  
Raffaele Mazzolari

Lo que firmo en Leioa, a 12 de febrero de 2018



## 9.4 Addendum 4: Informed consent

**Titolo del progetto:**

Effetti di diversi modelli prescrittivi dell'allenamento intervallato ad alta intensità sugli adattamenti fisiologici, performance e sul passo in corridori di endurance moderatamente allenati.

**Nome dello sperimentatore:**

Raffaele Mazzolari; Dipartimento di Educazione fisica e sport, Facoltà di attività fisica e scienze dello sport, Università dei Paesi Baschi (UPV / EHU), Vitoria-Gasteiz, Spagna.

**Nome del supervisore:**

Prof. Jordan Santos-Concejero; Dipartimento di Educazione fisica e sport, Facoltà di attività fisica e scienze dello sport, Università dei Paesi Baschi (UPV / EHU), Vitoria-Gasteiz, Spagna

**Livello di studio:**

Progetto di ricerca di dottorato

**Indirizzo email:**

[rmazzolari001@ikasle.ehu.eus](mailto:rmazzolari001@ikasle.ehu.eus)

**Introduzione:**

Periodi prolungati di corsa a moderata e ad alta intensità vengono tradizionalmente raccomandati tra gli approcci più efficaci per migliorare gli attributi fisiologici relativi alla performance negli atleti di endurance. Ciononostante, è stato dimostrato che stimoli allenanti più brevi, intermittenti ad alta intensità, comunemente noti come "allenamento intervallato ad alta intensità (HIIT)", possano essere altrettanto efficaci, soprattutto in corridori già condizionati alle altre strategie di allenamento. Sebbene la ricerca abbia confermato gli effetti positivi dell'HIIT su corridori di endurance, importanti differenze in termini di adattamenti fisiologici e performance sono state osservate in letteratura. Seppur tali discrepanze possano essere in parte spiegate dalle differenze nelle caratteristiche fisiologiche dei corridori, nelle apparecchiature e nei protocolli di test tra i vari studi, un fattore importante è verosimilmente legato ai diversi protocolli HIIT adottati in letteratura. A tal proposito, numerosi autori hanno privilegiato prescrizioni altamente individualizzate, dove ogni variabile allenante è adattata alle caratteristiche fisiologiche di ogni soggetto, rispetto ad approcci più standardizzati, dove lo stesso tipo di protocollo è prescritto in corridori di simile livello, o ad approcci più da campo, dove il passo in gara è comunemente usato per prescrivere l'HIIT ("ripetute"). Tuttavia, nonostante sussista una solida logica per favorire un allenamento individualizzato rispetto ad uno standardizzato, non vi è alcuna chiara evidenza scientifica circa differenze sostanziali in termini di adattamenti fisiologici e performance tra i due diversi metodi di prescrizione dell'HIIT. Dati i requisiti in termini di tempo, costi e sforzo più elevati per i programmi HIIT personalizzati rispetto a programmi standardizzati o alle prescrizioni con metodi da campo, il primo approccio può

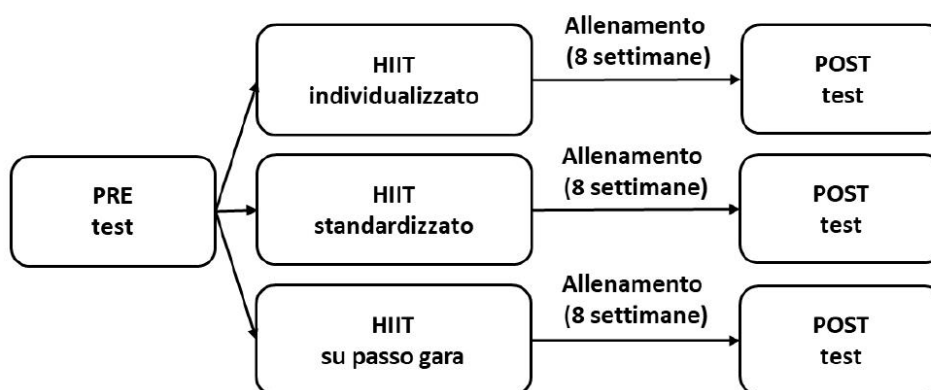
essere giustificato solo una sostanziale inferiorità degli altri in termini di adattamenti fisiologici e performance.

#### Scopo:

Lo scopo primario del presente studio sarà quello di investigare gli effetti di tre diversi metodi di prescrizione HIIT (individualizzato, standardizzato e basato sul passo di gara) sugli adattamenti fisiologici, sulla performance e sul passo in corridori di endurance moderatamente allenati, nonché quello di identificare valide alternative 'da campo' per individualizzare l'allenamento senza bisogno di ricorrere a test di laboratorio.

#### Design e Metodi:

La durata dello studio sarà di circa 10 settimane, di cui 8 di allenamento sperimentale per 3 gruppi e 8 di allenamento abituale per il quarto gruppo (Figura 1).



**Figura 1.** Riepilogo del disegno dello studio. Il gruppo che manterrà l'allenamento abituale non è riportato.

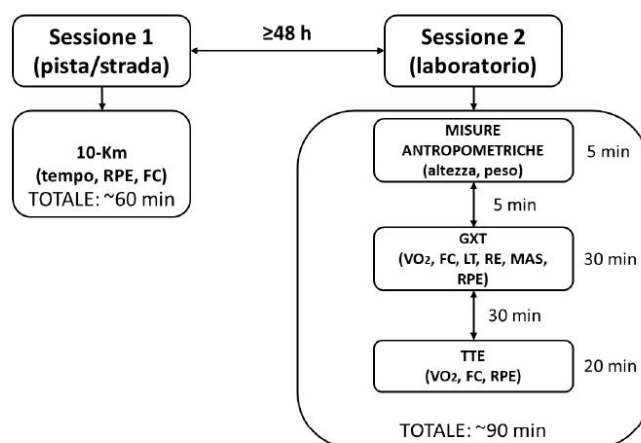
Gli atleti che parteciperanno a questo lavoro di tesi di dottorato saranno corridori di fondo – uomini e donne – di età superiore ai 18 anni. Ulteriori specifiche circa i criteri di selezione (tempi gara sulle varie distanze, eventuali patologie ed infortuni pregressi, ecc.) verranno discusse al primo incontro con lo sperimentatore ed il medico. Diversi test e misurazioni verranno eseguiti nel corso dello studio (Figura 2):

#### Sessione 1)

- Prova cronometrata di 10 km in pista (o strada). Verrà registrato il tempo necessario per completare i 10 km. Inoltre, verrà registrata registrato lo sforzo percepito (RPE) secondo la scala di Borg ogni 2 giri. Quando possibile, verrà registrata la frequenza cardiaca di ogni atleta durante l'intera durata della prova.

**Sessione 2)**

- Misurazione di variabili antropometriche (altezza, peso).
- Test incrementale massimale su tapis roulant (GXT) con misurazione dei parametri respiratori e di lattato [consumo massimo di ossigeno ( $VO_{2max}$ ), frequenza cardiaca (FC), soglia del lattato (LT), economia di corsa (RE), massima velocità aerobica (MAS)] oltre allo sforzo percepito (RPE).
- Test di massima durata a carico costante su tapis roulant (TTE). Verrà misurato il tempo massimo che si è in grado di sostenere ad un'intensità di circa il 95% del  $VO_{2max}$ .



**Figura 2.** Test effettuati prima (PRE) e dopo (POST) il periodo di allenamento.

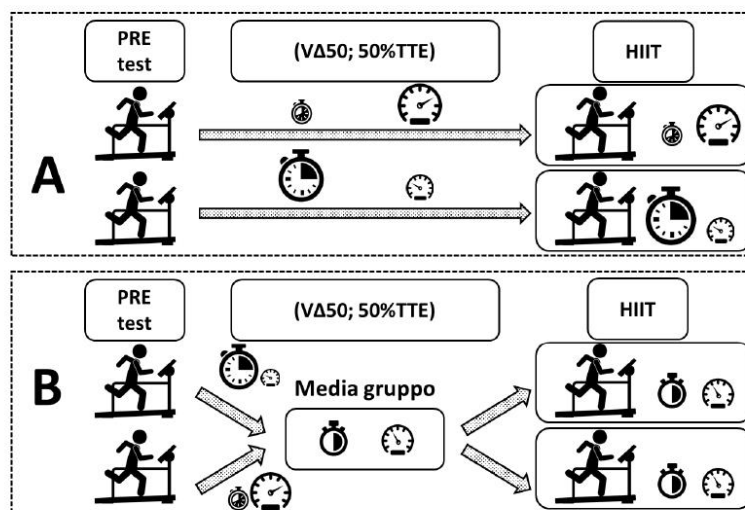
Una volta completati i test, i corridori verranno assegnati a caso ad uno dei seguenti gruppi:

### 1) Allenamento HIIT personalizzato

Continuerà l'altrimenti previsto programma di allenamento, ma sostituendo 2 sessioni per settimana con sessioni di HIIT le cui caratteristiche in termini di durata, intensità e recupero verranno determinate su base individuale a seconda dei dati ottenuti durante i test fatti ad inizio studio (Figura 3). Specificatamente, l'intensità di ciascun intervallo corrisponderà alla velocità media tra la LT e il  $VO_{2max}$  ( $V_{\Delta 50}$ ), mentre la durata degli intervalli sarà la metà del TTE. L'intensità nel recupero sarà il 50% del MAS mentre la durata del recupero tra le sessioni sarà il 25% del TTE (rapporto 2:1). Il numero di ripetizioni per sessione verrà stabilito in base al numero massimo di ripetizioni ottenute durante la prima sessione di allenamento HIIT ( $N_{max}$ ) e corrisponderà a 1-2 sessioni in meno del numero massimo sostenibile ( $N_{max} - 1$  or  $N_{max} - 2$ ); numero che verrà aumentato su base individuale durante il corso dello studio secondo la capacità individuale del corridore.

## 2) Allenamento HIIT standardizzato

Continuerà l'altrimenti previsto programma di allenamento, ma sostituendo 2 sessioni per settimana con sessioni di HIIT le cui caratteristiche in termini di durata, intensità e recupero verranno determinate a seconda dei dati ottenuti durante i test fatti ad inizio studio (Figura 3), calcolando però i valori medi per tutti i soggetti del gruppo. Il numero di ripetizioni per sessione verrà stabilito in base al numero massimo di ripetizioni ottenute durante la prima sessione di allenamento HIIT ( $N_{max}$ ) e corrisponderà a 1-2 sessioni in meno del numero massimo sostenibile ( $N_{max} - 1$  or  $N_{max} - 2$ ); numero che verrà aumentato su base individuale durante il corso dello studio secondo la capacità individuale del corridore.



**Figura 3.** Approccio alla prescrizione delle sessioni di HIIT. "A" rappresenta una tipica sessione HIIT prescritta usando il metodo individualizzato, mentre "B" rappresenta una sessione HIIT prescritta con il metodo standardizzato. Le dimensioni delle icone rappresentano l'entità dei valori dei parametri correlati.

## 3) Allenamento HIIT su passo gara

Continuerà l'altrimenti previsto programma di allenamento, ma sostituendo 2 sessioni per settimana con sessioni di HIIT le cui caratteristiche in termini di durata, intensità e recupero verranno determinate a seconda del passo durante la prova dei 10 km. Specificatamente, l'intensità di ciascun intervallo corrisponderà al 110% della velocità media nella prova dei 10 km ( $V_{10km}$ ), mentre la durata degli intervalli sarà simile alla durata media negli altri due gruppi (circa 4 min). Il recupero consisterà nel coprire circa 300 metri in circa 2 minuti (rapporto 2:1). Il numero di ripetizioni per sessione verrà stabilito in base al numero massimo di ripetizioni ottenute durante la prima sessione di allenamento HIIT ( $N_{max}$ ) e corrisponderà a 1-2 sessioni in meno del numero massimo sostenibile ( $N_{max} - 1$  or  $N_{max} - 2$ ); numero che verrà aumentato su base individuale durante il corso dello studio secondo la capacità individuale del corridore.

La mia partecipazione allo studio non comporta alcuna spesa.

- Comprendo che la mia partecipazione è volontaria.
- Comprendo che posso ritirarmi dallo studio:
  1. In qualsiasi momento
  2. Senza dover spiegare.
  3. Senza nessuna conseguenza economica.
- Partecipo liberamente allo studio e do il mio consenso per l'accesso e l'uso dei miei dati nelle condizioni dettagliate nella scheda informativa.

Io, ....., maggiorenne, e con documento d'identità No. ....,

DICHIARO CHE:

- Ho letto il foglio informativo che mi è stato dato.
- Sono stato in grado di porre domande sullo studio.
- Ho parlato con Raffaele Mazzolari
- Ho ricevuto abbastanza informazioni sullo studio.

Consapevole di quanto sopra scritto, firmo il presente documento in ..... il .....,

AUTORIZZO IL RESTO DELLE MISURAZIONI

Firma del partecipante:  Nome:  Documento d'identità No.:
---

Firma dell'investigatore:  Nome:  Documento d'identità No.:
---

\*\* Se hai bisogno di maggiori informazioni o hai domande, contatta Raffaele Mazzolari, tel. 3468754234  
email: [rmazzolari001@ikasle.ehu.eus](mailto:rmazzolari001@ikasle.ehu.eus).