Heart rate responses to submaximal runs as a tool for monitoring changes in cardiorespiratory fitness in professional female soccer players

Linus Walter, [1 2] **Ludwig Ruf,** [2 3] **Martin Buchheit,** [4 5 6 7 8 9] **Alexander Woll,** [1] **Sascha Härtel,** [2] **Stefan Altmann** [1 3]

[1]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]TSG 1899 Hoffenheim, Zuzenhausen, Germany
[3]TSG ResearchLab gGmbH, Zuzenhausen, Germany
[4]Aspetar, Doha, Qatar
[5]Optimo Performance Center, Estepona, Spain
[6]Type 3.2 Performance, Montvalezan, France
[7]INSEP, Paris, France
[8]HIIT Science, Revelstoke, Canada
[9]Athletica, Revelstoke, Canada

Soccer | **Cardiorespiratory Fitness** | **Heart Rate** | **Female** | **Training** | **Testing**

Headline

While the match-running characteristics between women's and men's elite soccer players still differ, especially in higher speed zones (1), they are increasingly converging in areas such as total distance covered (2). Therefore, cardiorespiratory fitness in women's soccer is just as important as in men's soccer, and testing plays a crucial role in load management and preventing injuries (3,4). However, the gold standard multistage incremental tests with heart rate (HR) and blood lactate measures are expensive and time-consuming. Monitoring HR responses during submaximal runs offers a time-saving and inexpensive alternative to lactate testing for assessing cardiorespiratory fitness. This makes it well-suited for use during the competitive soccer season (5). While several studies have already been performed on male athletes confirming the practical utility of this simple method (6,7), research on female athletes remains limited (8).

Aim

This study aims to investigate whether the change in HR at a running speed of 12 km/h ($HR_{12km/h}$) is concomitant to the change in running speed at a lactate concentration of 4 mmol/L ($V_{4mmol/L}$) in professional female soccer players.

Methods

Design

Observational, cross-sectional.

Participants

Data was obtained from 46 professional female soccer players (age (mean ± standard deviation): 23.6 ± 3.1 years; height: 171.5 ± 6.1 cm; body mass: 64.7 ± 6.3 kg) playing for a first divisional club in Germany. Data were collected as part of the standard performance diagnostic. Therefore, ethical approval was not necessary (9). Prior to engaging in the tests, all players granted their informed consent.

Methodology

The players were tested over a period of four and a half seasons (summer 2020 to winter 2024-2025) for a minimum of 2 and a maximum of 10 times, including assessments during preseason as well as either during autumn or the winter break. The assessment consisted of an incremental treadmill test (WOODWAY GmbH, Weil am Rhein, Germany) starting at 6 km/h which increased by 2 km/h every 3 minutes until voluntary exhaustion. Between each stage (pause of 30 s), blood samples from the earlobe (C-Line Sport; EKF-diagnostic GmbH, Barleben, Germany) were collected to assess the players' blood lactate concentration. HR (H10, Polar Electro Oy, Kempele, Finland) was observed throughout the whole test.

Data Analysis

The average HR of the last 30 seconds of the 12 km/h stage ($HR_{12km/h}$) and the running velocity at a lactate concentration of 4 mmol/L (as an indicator of the fixed anaerobic threshold; $V_{4mmol/L}$) as well as the maximum achieved running speed (V_{max}) were used. The percentage changes between two testing sessions were calculated to compare the variables with $HR_{12km/h}$ being given as a percentage of the maximum HR (HR_{max}).

Based on the tests that were compared with each other, the following three time-based subgroups emerged, resulting in a total of 168 test comparisons:

- start of the season vs start of the next season (n = 41)
  - e.g. August 2022 vs August 2023
- start of the season vs within season (n = 83)
  - e.g. August 2022 vs January 2023
- within season vs start of the next season (n = 44)
  - e.g. January 2023 vs August 2023

In addition, changes in $HR_{12km/h}$ and $V_{4mmol/L}$ of a single player over a 3.5-year period (case study; n = 7 test comparisons) were also examined.

Statistical Analysis

The data were analyzed using SPSS statistical software (version 29.0.1.0; SPSS Inc, Chicago, IL) and Microsoft Excel (Microsoft, Redmond, WA).

For the entire sample, each subgroup and the case study, pearson product-moment correlations, with 95% confidence intervals (95% CI) were employed to investigate the relationships between changes in $HR_{12km/h}$ and $V_{4mmol/L}$ (expressed as percentages). According to Altmann et al. (10), the agreement between changes in $HR_{12km/h}$ and $V_{4mmol/L}$ was examined using the following categorization:

- Full agreement
- Partial agreement
- Full mismatch

To be considered as improved or impaired, both variables had to exceed the threshold for substantial changes ($HR_{12km/h}$ = 4.5% and $V_{4mmol/L}$ = 6.0%). Otherwise, they would be classified as unclear (10).

In addition, accuracy, precision, sensitivity (also referred to as recall), specificity, and the F1-Score were calculated to further evaluate the agreement in classification outcomes between $V_{4mmol/L}$ and $HR_{12km/h}$:

- Accuracy: proportion of correctly classified cases in all cases considered, irrespective of the class (e.g. $HR_{12km/h}$ and $V_{4mmol/L}$ both detected an improvement, this instance contributes to overall accuracy)
- Precision: proportion of the improvements predicted by $HR_{12km/h}$ actually confirmed by $V_{4mmol/L}$, indicating the effectiveness of positive predictions
- Sensitivity: proportion of actual impairments, as identified by $V_{4mmol/L}$, that were correctly identified by $HR_{12km/h}$
- Specificity: proportion of cases without performance impairment, as determined by $V_{4mmol/L}$, that were correctly identified as such based on the classification derived from $HR_{12km/h}$ (e.g. if $V_{4mmol/L}$ classifies an athlete as having maintained or improved performance, and $HR_{12km/h}$ likewise does not suggest impairment, this instance contributes to the specificity)
- F1-Score: balanced measure of precision and recall given the imbalanced class distribution (11,12)

To calculate these values, a confusion matrix was created (Table 1). The resulting values for true positives, true negatives, false positives, and false negatives for all three classes are shown in Table 2.

Results

Average $V_{4mmol/L}$, V_{max}, $HR_{12km/h}$, and HR_{max} were 13.0 ± 1.0 km/h, 16.2 ± 0.9 km/h, 172.6 ± 10.5 bpm, and 191.7 ± 8.8 bpm (i.e. $HR_{12km/h}$ corresponded to 90% of HR_{max} on average). The absolute average changes between two testing sessions in $V_{4mmol/L}$ and $HR_{12km/h}$ for the whole sample were 0.6 ± 0.4 km/h and 4.4 ± 3.7 bpm. Of the 168 test comparisons, 115 were full agreements, 51 were partial agreements, and 2 were full mismatches. Response classification as well as the detailed distribution of full agreements, partial agreements, or full mismatches, can be seen in Table 3.

Pearson product-moment correlations and 95% CI between changes in $V_{4mmol/L}$ and $HR_{12km/h}$ for the whole sample, each subgroup, and the case study were moderate to large and are presented in Table 4. Figures 1 (a) and 1 (b) illustrate both the absolute values and percentage changes for $HR_{12km/h}$ and $V_{4mmol/L}$ for an individual player during the period from summer 2020 to fall 2023 (i.e. case study).

The overall classification accuracy was 0.79. The micro-weighted averaged values for the classification metrics for the whole model were 0.68 for precision, 0.68 for sensitivity, 0.84 for specificity, and 0.68 for the F1-Score. Table 5 shows the class-specific values for precision, sensitivity, specificity, and the F1-Score.

Table 1. Confusion matrix comparing classifications by $HR_{12km/h}$ and $V_{4mmol/L}$.

		$V_{4mmol/L}$-based classification		
		$V_{4mmol/L}$ Impaired ↓	$V_{4mmol/L}$ Unclear ↔	$V_{4mmol/L}$ Improved ↑
$HR_{12km/h}$-based classification	$HR_{12km/h}$ Impaired ↓	4	6	1
	$HR_{12km/h}$ Unclear ↔	13	105	23
	$HR_{12km/h}$ Improved ↑	1	9	6

Table 2. Values for True Positives, True Negatives, False Positives, and False Negatives in all three classes.

	True Positives	True Negatives	False Positives	False Negatives
Impaired	4	143	7	14
Unclear	105	12	36	15
Improved	6	128	10	24

Discussion

We aimed to assess the agreement between changes in $HR_{12km/h}$ and $V_{4mmol/L}$ as indicators of changes in fitness in professional female soccer players. Using the threshold values of 4.5% for $HR_{12km/h}$ and 6% for $V_{4mmol/L}$ (10) to determine substantial changes, we observed 69% full agreement, 30% partial agreement, and 1% full mismatches across 168 test comparisons. The correlation between changes for the whole sample was r = -0.47 (95% CI -.58 to -.34) and ranged between r = -0.40 to -0.54 for the subgroups.

Those results are in line with previous studies using the same methodological approach, which reported 63% full agreement in elite male players (10) and 62% in elite youth players (13), confirming that HR responses to a submaximal run can also be implemented to assess fitness changes in elite female soccer players (8). In addition, the findings for the subgroups indicate some degree of variability of the applied measures depending on the moment during the soccer season. The reported correlations between changes in both variables in the investigations mentioned and the current study are similar (all studies r = -0.40 to -0.54)). While submaximal HR responses to a given external load in women are generally higher than in men (14), due to a lower oxygen transport capacity and a lower blood volume (15), this difference diminishes as HR approaches its maximum and is unrelated to cardiorespiratory fitness level (14). Although there exist several confounding variables of HR responses (16,17), taken together, our findings suggest that $HR_{12km/h}$ adapts similarly to a change in cardiorespiratory fitness in trained women and men.

The appearance of full mismatches in the current investigation was exceedingly rare across both the entire sample and the subgroups, occurring in only 0% to 2% of cases, which also matches the previous literature. To be precise, there were two cases out of the 168 test comparisons in which $HR_{12km/h}$ and $V_{4mmol/L}$ classified the player's fitness in opposite ways. Although full mismatches occurred very rarely, matching previous literature, these can result in contrary practical consequences by drawing wrong conclusions regarding training intensity and volume. Therefore, such cases should be examined more closely in future studies to better understand why $HR_{12km/h}$ and $V_{4mmol/L}$ rarely result in contradicting classifications.

Table 3. Classifications matrix for fitness changes for according to both $HR_{12km/h}$ and $V_{4mmol/L}$ along with full mismatches, partial agreements, and full agreements regarding the whole sample, the 3 subgroups, and the case study.

Group		↓ Fitness from $V_{4mmol/L}$	↔ Fitness from $V_{4mmol/L}$	↑ Fitness from $V_{4mmol/L}$	Agreement	Percentage
Whole sample (n = 168)	↓ Fitness from $HR_{12km/h}$	4	6	1	Full mismatch	1
	↔ Fitness from $HR_{12km/h}$	13	105	23	Partial agreement	30
	↑ Fitness from $HR_{12km/h}$	1	9	6	Full agreement	69
Start of the season vs start of the next season (n = 41)	↓ Fitness from $HR_{12km/h}$	1	1	0	Full mismatch	0
	↔ Fitness from $HR_{12km/h}$	3	25	6	Partial agreement	34
	↑ Fitness from $HR_{12km/h}$	0	4	1	Full agreement	66
Start of the season vs within season (n = 83)	↓ Fitness from $HR_{12km/h}$	0	2	1	Full mismatch	1
	↔ Fitness from $HR_{12km/h}$	1	57	14	Partial agreement	24
	↑ Fitness from $HR_{12km/h}$	0	3	5	Full agreement	75
Within season vs start of the next season (n = 44)	↓ Fitness from $HR_{12km/h}$	3	3	0	Full mismatch	2
	↔ Fitness from $HR_{12km/h}$	9	23	3	Partial agreement	39
	↑ Fitness from $HR_{12km/h}$	1	2	0	Full agreement	59
Case Study (n = 7)	↓ Fitness from $HR_{12km/h}$	1	1	0	Full mismatch	0
	↔ Fitness from $HR_{12km/h}$	2	2	0	Partial agreement	43
	↑ Fitness from $HR_{12km/h}$	0	0	1	Full agreement	57

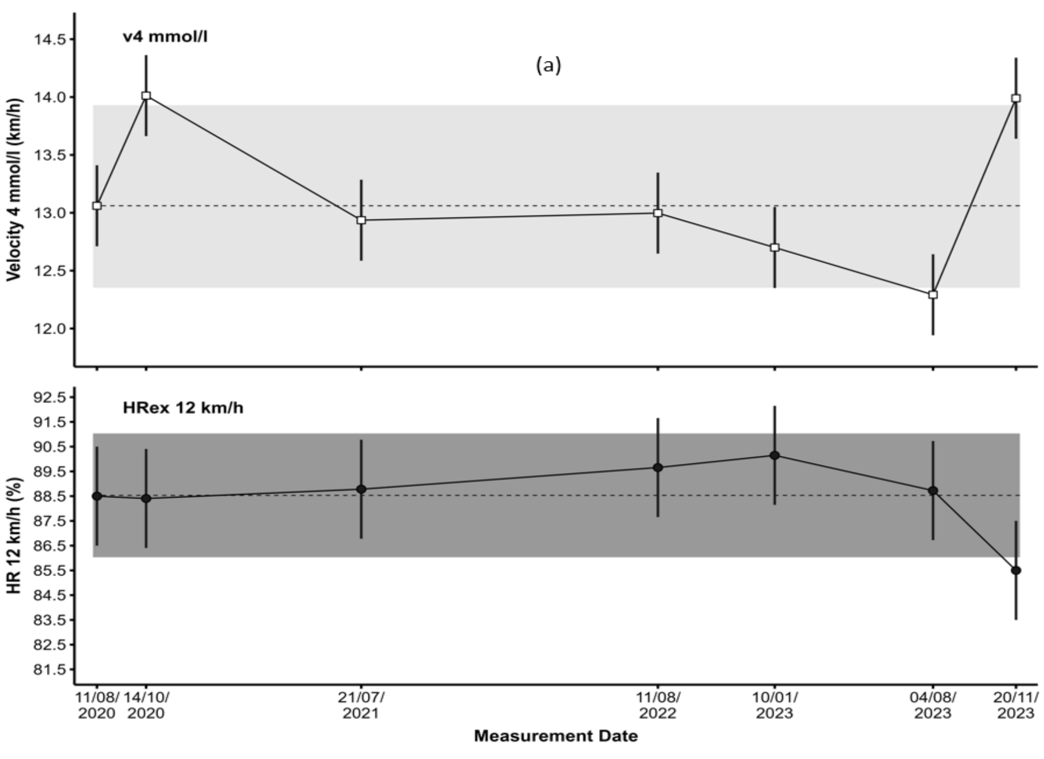
To further assess the usefulness of $HR_{12km/h}$ in relation to changes in $V_{4mmol/L}$, several additional metrics were calculated. The micro-weighted results presented for accuracy (0.79), precision (0.68), sensitivity (0.68), specificity (0.84), and the F1-Score (0.68) demonstrate that HR responses to submaximal runs perform well across categories, reflecting a generally good level of classification agreement and emphasizing their practical usefulness. On the other hand, the results for the individual classes (see Table 5) indicate reduced levels of classification agreement for certain categories. Specifically, precision, sensitivity, and the F1-Score were lower in the impaired and the improved class, while specificity was high in those classes. In the unclear class, this pattern was inverted. This suggests that submaximal HR responses perform well at detecting unclear changes but are less effective at detecting improvements and impairments.

An in-depth exploration of the changes in $HR_{12km/h}$ reveals what classifications would have been made based on this parameter in the 168 test comparisons. $HR_{12km/h}$ would have indicated impairments in 11 (7%) and an improvement in 16 (10%) cases. Assuming that $V_{4mmol/L}$ represents the actual changes in fitness, $HR_{12km/h}$ correctly identified 4 (36%) out of the 11 impairments detected, which corresponds to a precision of 0.36 in the impaired class (Table 5). Additionally, $HR_{12km/h}$ indicated impairments in 6 instances where $V_{4mmol/L}$ did not show any substantial changes, and in one case, $HR_{12km/h}$ completely misidentified the change. Similarly, regarding improvements, $HR_{12km/h}$ correctly identified 6 (38%) out of the 16 detected by $V_{4mmol/L}$, which corresponds to a precision of 0.38 in the improved class (Table 5), misidentified in 9 cases where $V_{4mmol/L}$ indicated no substantial change and was entirely incorrect in another instance. In contrast, $HR_{12km/h}$ correctly recognized 105 (74%) out of 141 test comparisons where no substantial change occurred, corresponding to a precision of 0.74 in the unclear class (Table 5). These results support that the use of HR responses to a submaximal run has difficulties recognizing clear improvements and impairments and works better at recognizing when there have been no substantial changes in cardiorespiratory fitness. In summary, of the 27 substantial changes detected by $HR_{12km/h}$, 10 (37%) were correctly identified, while 2 (7%) were misidentified. However, $HR_{12km/h}$ has failed to recognize some substantial changes altogether. Future research should focus on investigating the underlying reasons why improvements and impairments are sometimes not or incorrectly recognized or why unchanged fitness values are better detected. Additionally, as out of the 168 test comparisons, only 48 (29%) substantial changes in fitness were observed by $V_{4mmol/L}$, it would be interesting to investigate the influence of a significantly higher number of substantial fitness changes (e.g. before and after a training intervention or training camp) on the agreement distribution as well as on precision, sensitivity and specificity.

Finally, a single player was monitored from August 2020 to November 2023. During this period, there were 4 substantial changes detected by $V_{4mmol/L}$ (3 impairments and 1 improvement) and 3 substantial changes detected by $HR_{12km/h}$ (2 impairments and 1 improvement). Data points 2 and 5 in Figure 1 (b) represent the test comparisons from August 2020 to November 2020 and from January 2022 to January 2023. The player showed a substantial decline in $V_{4mmol/L}$ in both comparisons, while $HR_{12km/h}$ remained unchanged. Data points 6 and 7 represent the test comparisons from January 2023 to July 2023 and from July 2023 to August 2023. In both test comparisons, $V_{4mmol/L}$ and $HR_{12km/h}$ showed the same substantial improvement or impairment, further highlighting the practical use of submaximal HR responses on an individual level. Overall, the $HR_{12km/h}$ was effective, and changes in fitness were usually classified correctly (57% full agreements and 43% partial agreements).

Table 4. Pearson correlations and 95% confidence intervals (95% CI) between changes in $HR_{12km/h}$ and $V_{4mmol/L}$ regarding the whole sample, all subgroups, and the case study.

Group	Pearson r (95% CI)
Whole sample (n = 168)	r= -.47 (-.58 to -.34)
Start of the season vs start of the next season (n = 41)	r= -.54 (-.73 to -.29)
Start of the season vs within season (n = 83)	r= -.40 (-.57 to -.20)
Within season vs start of the next season (n = 44)	r= -.49 (-.69 to -.23)
Case Study (n =7)	r= -.49 (-.91 to .42)



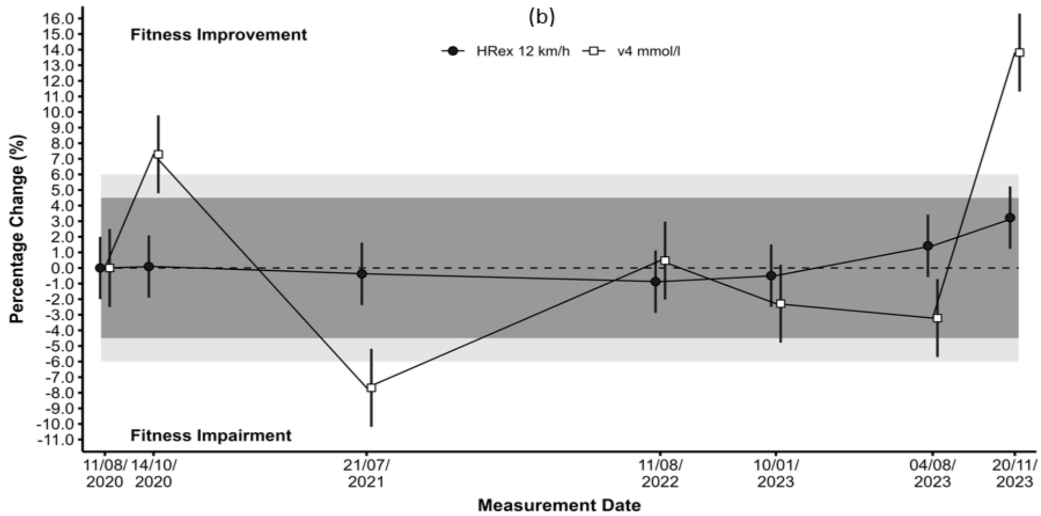


Fig. 1. **(a) Absolute value changes in $V_{4mmol/L}$ and $HR_{12km/h}$ of a single player over the period of 3.5 seasons (7 test occasions). (b) Percentage changes in $V_{4mmol/L}$ and $HR_{12km/h}$ of a single player over the period of 3.5 seasons (7 test occasions)**

Table 5. **Class-specific values for precision, sensitivity, specificity, and the F1-Score for the whole sample.**

	Impaired	Unclear	Improved
Precision	0.36	0.74	0.38
Sensitivity	0.22	0.88	0.20
Specificity	0.95	0.25	0.93
F1-Score	0.27	0.80	0.26

Limitations

This study has some limitations. We did not systematically control confounding variables that could influence $HR_{12km/h}$ and $V_{4mmol/L}$. Also, all tests were performed on a treadmill, which does not correspond to the training conditions in which submaximal HR responses would typically be used.

Practical applications

- HR responses to a submaximal 3-minute run are a practicable, time-saving, and cheap way to assess fitness changes in professional female soccer during and across the seasons.
- Agreements between changes in $HR_{12km/h}$ and $V_{4mmol/L}$ were good (69% full agreements of all test comparisons) and similar to findings in male adult and youth soccer.
- Nevertheless, certain constraints should be taken into account. It is important to understand that even if full mismatches are extremely rare, $HR_{12km/h}$ does not always detect substantial fitness changes as indicated by $V_{4mmol/L}$. This must be considered when coaches assess changes in fitness of the entire team and individual players using submaximal HR responses.

X Handle: Stefan Altmann (@StefanAltmann3), Ludwig Ruf (@RufLudwig), TSG ResearchLab gGmbH, Zuzenhausen, Germany (@TSGResearchlab), Martin Buchheit (@mart1buch)

References

1. Bradley PS, Dellal A, Mohr M, Castellano J, Wilkie A. Gender differences in match performance characteristics of soccer players competing in the UEFA Champions League. Human Movement Science. 2014;33:159–71.

2. Martínez-Lagunas V, Niessen M, Hartmann U. Women's football: Player characteristics and demands of the game. Journal of Sport and Health Science. 2014;3(4):258–72

3. Gabbett TJ. The training—injury prevention paradox: should athletes be training smarter and harder? Br J Sports Med. 2016;50(5):273–80.

4. Buchheit M, Hader K. Data everywhere, insight nowhere: a practical quadrant-based model for monitoring training load vs. response in elite football. Sport Performance & Science Reports, #258, May 2025.

5. Shushan T, McLaren SJ, Buchheit M, Scott TJ, Barrett S, Lovell R. Submaximal Fitness Tests in Team Sports: A Theoretical Framework for Evaluating Physiological State. Sports Med. 2022;52(11):2605–26.

6. Shushan T, Lovell R, Buchheit M, Scott TJ, Barrett S, Norris D, u. a. Submaximal Fitness Test in Team Sports: A Systematic Review and Meta-Analysis of Exercise Heart Rate Measurement Properties. Sports Med - Open. 2023;9(1):21.

7. Buchheit M, Simpson BM, Lacome M. Monitoring Cardiorespiratory Fitness in Professional Soccer Players: Is It Worth the Prick? International Journal of Sports Physiology and Performance. 2020;15(10):1437–41.

8. MacKenzie L, Lovell R, Atkinson L, Naumovski D, Shushan T. Validity and Sensitivity of a Submaximal Fitness Test for Assessing Cardiorespiratory Fitness and Response to Preseason Training Load in Female Football Players. International Journal of Sports Physiology and Performance. 2025;1–8.

9. Winter EM, Maughan RJ. Requirements for ethics approvals. Journal of Sports Sciences. 2009;27(10):985–985.

10. Altmann S, Neumann R, Härtel S, Woll A, Buchheit M. Using Submaximal Exercise Heart Rate for Monitoring Cardiorespiratory Fitness Changes in Professional Soccer Players: A Replication Study. International Journal of Sports Physiology and Performance. 2021;16(8):1096–102.

11. Heydarian M, Doyle TE, Samavi R. MLCM: Multi-Label Confusion Matrix. IEEE Access. 2022;10.

12. Shreffler J, Huecker MR. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025.

13. Altmann S, Ruf L, Neumann R, Härtel S, Woll A, Buchheit M. Assessing the usefulness of submaximal exercise heart rates for monitoring cardiorespiratory fitness changes in elite youth soccer players. Science and Medicine in Football. 2023;7(2):177–82.

14. Hottenrott K, Neumann G. Geschlechtsspezifische Formel für optimale Trainingsherzfrequenzen. 2012.

15. Diaz-Canestro C, Pentz B, Sehgal A, Montero D. Differences in Cardiac Output and Aerobic Capacity Between Sexes Are Explained by Blood Volume and Oxygen Carrying Capacity. Front Physiol. 2022;13:747–903.

16. Achten J, Jeukendrup AE. Heart Rate Monitoring. Sports Medicine. 2003;33(7):517–38.

17. Tschopp M. Qualitätsstandards in der Ausdauerleistungsdiagnostik. 2001.

Disclosure of interest

All contributing authors declare that they have no conflicts of interest relevant to the content of this study.