Exploring the Use of 5 Different Yo-Yo Tests in Evaluating V̇O2max$$ \dot{\mathrm{V}}{\mathrm{O}}_{2\max } $$ and Fitness Profile in Team Sports: A Systematic Review and Meta-Analy…

Article *in* Scandinavian Journal of Medicine and Science in Sports · May 2025

DOI: 10.1111/sms.70054

Scandinavian Journal of Medicine & Science in Sports

WILEY

REVIEW

Exploring the Use of 5 Different Yo-Yo Tests in Evaluating $\dot{V}O_{2\,max}$ and Fitness Profile in Team Sports: A Systematic Review and Meta-Analysis

Zhiyuan Tan[1] | Carlo Castagna[2,3,4] | Peter Krustrup[4,5,6] | Del P. Wong[7] | Susana Póvoas[4,8] | Daniel Boullosa[9,10,11] | Kai Xu[12] | Ivan Cuk[1,13]

[1]Faculty of Sport and Physical Education, University of Belgrade, Belgrade, Serbia | [2]Department of Education and Sport Sciences, Pegaso Telematic University, Naples, Italy | [3]Football Training and Biomechanics Laboratory, Technical Dept FIGC, Coverciano, Italy | [4]Department of Sports Science and Clinical Biomechanics, SDU Sport and Health Sciences Cluster (SHSC), University of Southern Denmark, Odense, Denmark | [5]Danish Institute for Advanced Study (DIAS), University of Southern Denmark, Odense, Denmark | [6]Sport and Health Sciences, University of Exeter, Exeter, UK | [7]College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia | [8]Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, University of Maia, Maia, Portugal | [9]Faculty of Physical Activity and Sports Sciences, Universidad de León, León, Spain | [10]Integrated Institute of Health, Federal University of Mato Grosso Do Sul, Campo Grande, Brazil | [11]College of Healthcare Sciences, James Cook University, Townsville, Australia | [12]School of Athletic Performance, Shanghai University of Sport, Shanghai, China | [13]InterSynergy Research Center, Belgrade, Serbia

Correspondence: Carlo Castagna (castagnac@libero.it)

Received: 6 January 2025 | **Revised:** 6 March 2025 | **Accepted:** 1 April 2025

Funding: The authors received no specific funding for this work.

Keywords: aerobic capacity | athletes | correlation | endurance assessment | performance

ABSTRACT

This review aims to examine the relationship between five different Yo-Yo tests and $\dot{V}O_{2\,max}$ as well as performance in various fitness tests, to provide practical guidance for fitness assessment. Thirty-three studies ($n = 1362$) were analyzed to qualitatively score correlation magnitude (r-values) between five different Yo-Yo tests with $\dot{V}O_{2\,max}$. The association between Yo-Yo tests and physical fitness tests were examined including 36 studies ($n = 1262$ participants). Large correlations were found between $\dot{V}O_{2\,max}$ and Yo-Yo Intermittent Endurance level 1 ($r = 0.67$), and level 2 ($r = 0.67$), Yo-Yo Intermittent Recovery level 1 ($r = 0.65$), and level 2 ($r = 0.55$). Yo-Yo tests showed moderate correlations with the jump tests ($r = 0.40$). Large associations were reported between Yo-Yo tests performance and treadmill endurance test ($r = 0.59$) and the field-based endurance tests ($r = 0.66$), respectively. These findings confirm Yo-Yo tests as reliable and valid tests for estimating $\dot{V}O_{2\,max}$ and endurance performance in team sports, supporting their integration in structured testing batteries. Coaches should supplement Yo-Yo tests with specific strength and sprint assessments for a comprehensive fitness evaluation of the team sport athletes.

1 | Introduction

Team sports are characterised by short, intermittent activities such as high-speed running and sprinting, interspersed with periods of low-intensity exercise [1, 2]. Team sports athletes are required to repeatedly perform skillful actions at maximal or near-maximal effort, interspersed with brief recovery intervals [3, 4]. The physical demands are complex, requiring athletes to have highly developed speed, agility, strength, power and endurance [5]. To support athletes in achieving peak physical fitness, various tests have been developed to objectively assess players' capabilities, optimize training programs, and provide valuable feedback to both athletes and technical staff [6–8].

It is well established that assessing physical capacity of team sports athletes should include composite test batteries that target several factors in a short time [9, 10]. Indeed, laboratory testing combined with field-based testing are widely used physical fitness assessment. Field-based tests include jumping assessments, such as countermovement jumps (CMJ), to measure lower limb muscle power [11]. And sprinting tests, like 10-m and 30-m sprints, to determine maximal sprint speed [12]. Agility tests, including the Agility *T*-test and Illinois Agility Run, to evaluate change-of-direction ability in preplanned movements [13, 14]. In laboratory conditions, progressive protocols on treadmill or bicycle ergometer are used to assess maximal oxygen uptake ($\dot{V}O_{2\,max}$), for evaluating maximal aerobic power, while force platforms are employed to analyze detailed jump mechanics and muscle power with high precision [15].

Laboratory testing of $\dot{V}O_{2\,max}$ is widely regarded as the gold standard for evaluating aerobic capacity due to its high reliability. However, it has limitations, including the need for requiring specialized equipment and personnel, and being time-consuming when testing groups of athletes in team sports [16]. These constraints have driven the development of reliable and valid field-based alternatives. Over the past two decades, Yo-Yo tests has gained significant recognition in both research and practical applications for team sports [17–20]. The Yo-Yo tests are widely used for estimating $\dot{V}O_{2\,max}$ due to their simplicity, low cost (requiring only cones and speakers), and ease of implementation [17]. They have demonstrated high validity for assessing the ability of team sports athletes to perform high-intensity, repeated efforts [17, 20–23] and correlate very-largely with match performance in football and other team sports [4, 17, 24].

The Yo-Yo tests are maximal aerobic endurance fitness tests that involve running between markers placed 20 m apart at increasing speeds until exhaustion [17] (see Figure 1). The Yo-Yo tests are mainly divided into three versions (i.e., endurance, YYET, intermittent endurance, YYIE, and intermittent recovery, YYIR) each with two different levels of performance capacity (e.g., YYET1, YYET2, YYIE1, YYIE2, YYIR1, and YYIR2). The main differences between these different versions relate to different starting speeds (i.e., 8–13 km/h), and recovery times between the 2×20 m shuttle runs (i.e., 0–10 s). Moreover, two modified Yo-Yo tests have been developed for assistant football referees and futsal referees (Yo-Yo Sideways-Forwards IE2 tests) [25, 26] and for children (YYIR1 children's test) [27, 28]. However, the two modified Yo-Yo tests were not included in this analysis due to the lack of existing studies concerned on the correlation between two modified Yo-Yo tests and $\dot{V}O_{2\,max}$ as well as fitness profile.

Although the Yo-Yo tests are widely used to determine the aerobic performance of team sport athletes, there have been some disputes about its association with $\dot{V}O_{2\,max}$. There still exists a lack of evidence to sustain a cohesive link between performance in the Yo-Yo tests and $\dot{V}O_{2\,max}$, which may be related to its intermittent design. For example, Krustrup et al. [24] found a large correlation of ($r=0.56$) between YYIR1 test and $\dot{V}O_{2\,max}$ in female elite football players. Conversely, Michailidis et al. [29] found no significant correlation between the YYIR1 test and $\dot{V}O_{2\,max}$. Chuman et al. [30] found a small relationship of ($r=0.17$) between the YYIR2 test and $\dot{V}O_{2\,max}$ in young male football players. In contrast, Thomas et al. [31] found a very large correlation ($r=0.87$) between the YYIR1 test and $\dot{V}O_{2\,max}$ in team sport athletes. In addition, there have been conflicting results concerning the relationship between the Yo-Yo tests and other physical fitness components. For example, Michalsik et al. [32] found a large negative relationship ($r=-0.557$) between the YYIR1 test and sprinting performance in male team handball players. However, Medeiros et al. [33] found no significant correlation between the YYIR1 test and sprinting performance in male young football players. In addition, Castagna et al. [18] found a large positive relationship ($r=0.50$) between the YYIR1 test and jump performance in male football players. In contrast, Medeiros et al. [33] reported no correlation between the two tests in similar populations.

In the process of evaluating the physical capabilities, it is paramount to consider the relationship between different physical fitness tests to assess test interchangeability to avoid information redundancy. Furthermore, implementing a comprehensive testing battery at numerous stages during a training year is often problematic due to scheduled matches, athletes' fatigue and time requirements. A clearer synthesis of the conflicting findings regarding the Yo-Yo test's relationship with $\dot{V}O_{2\,max}$ and other physical fitness measures would enhance its rationale for use. This improved understanding could provide more accurate insights into players' physical capacities, enabling them to allocate more time to effective training and optimizing testing procedures.

This systematic review and meta-analysis aimed to synthetise evidence on the relationship between $\dot{V}O_{2\,max}$, physical fitness components and five different Yo-Yo tests in team sport players. The second aim of this study was to provide practical

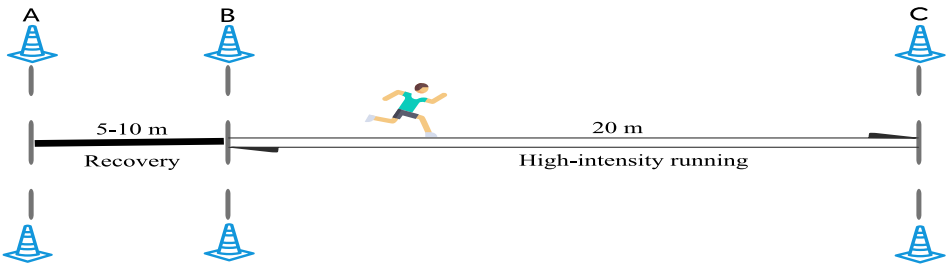

FIGURE 1 | The Yo-Yo testing diagram.

guidance to efficient physical fitness assessments for coaches and practitioners.

2 | Methods

2.1 | Design

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [34]. This review title has been registered in the Open Science Framework (osf.io/m29h7), while the protocol registration was done after study retrieval and before data analysis.

2.2 | Eligibility Criteria

We conducted a comprehensive review of the relationship between five different Yo-Yo tests, $\dot{V}O_{2\,max}$ and physical fitness components in team sports, analyzing studies that examine their correlations. Studies were included in this systematic review if they (i) were published in peer-reviewed academic journals; (ii) were published in English; (iii) were available in full-text; (iv) described the Yo-Yo tests results; (v) described the $\dot{V}O_{2\,max}$ and physical fitness tests results; (vi) described the correlation between physical fitness tests, $\dot{V}O_{2\,max}$ and Yo-Yo tests; and (vii) the subjects of study were team sport athletes with no injury. Studies on both sexes were included. In addition, meta-analysis, review, abstract only, and commentary articles were excluded, and the search was not limited to a specific year.

2.3 | Information Sources

Four electronic databases (e.g., PubMed, SPORTDiscus, Scopus, and Web of Science) were electronically searched across all publication years (i.e., from inception) up to and including September 12, 2024. In addition to the structured database search, reference lists from relevant systematic reviews were examined to identify additional eligible articles.

2.4 | Search Strategy

The search strategy combined terms related to team sports, Yo-Yo tests and fitness tests [35]. “Rugby” OR “Soccer” OR “Football” OR “Hurling” OR “Hockey” OR “Lacrosse” OR “Handball” OR “Team Sports” AND “YoYo” OR “Yo-Yo Test” OR “Yo Yo” were searched for examine the correlation between the Yo-Yo tests and $\dot{V}O_{2\,max}$ in team sports. “Endurance” OR “Change of Direction” OR “Sprint” OR “Strength” OR “Power” OR “Jump” OR “Performance” OR “Agility” OR “Repeated Sprint Ability” OR “Shuttle Run” OR “Speed” OR “Intermittent Fitness Test” AND “Rugby” OR “Soccer” OR “Football” OR “Hurling” OR “Australian rules” OR “Camogie” OR “Hockey” OR “Lacrosse” OR “Handball” OR “Team Sports” OR “Ice Hockey” OR “Volleyball” OR “Basketball” AND “YoYo” OR “Yo-Yo Test” OR “Yo Yo” were searched to examine the relationship between the Yo-Yo tests and physical fitness tests in team sports. These keywords were truncated using the quotation marks. Search strategies were modified to meet the specific requirements of each database.

2.5 | Selection Process

Two authors (Z.T. and K.X.) independently screened the titles, abstracts, and full text of these selected studies. Reference lists were also examined from the selected articles to identify any additional relevant studies. The screening process removed articles at each stage according to the inclusion and exclusion criteria, including duplicates, assessing the title, screening the abstract, and reviewing the full text. A third author (I.C.) adjudicated non-resolvable disagreements.

2.6 | Data Extraction

The database search results were imported into the bibliographic management software (Endnote X9, Clarivate INC, Philadelphia, PA 19130, USA), and duplicate copies were removed. Then, a rapid assessment was performed to analyze and discard articles based on title and abstract. Subsequently, articles that could be included were read as a whole to determine whether they were suitable for inclusion in this review. The selection process applied to the articles screened was based on the type of sport, language, result indicators, and correlation *r*-value.

Two authors (Z.T. and K.X.) independently conducted data extraction from the included articles. The studies included in this review focused on the relationship between $\dot{V}O_{2\,max}$ value, physical fitness tests results (e.g., sprint test, jump test, agility test, strength test, field-based endurance test, repeated sprint ability (RSA), treadmill endurance test, body anthropometry evaluations) and the Yo-Yo tests performance. Furthermore, the following relevant aspects were included to analyze these studies: author, year of publication, the type of sport, Yo-Yo tests result, $\dot{V}O_{2\,max}$ value, physical fitness tests result, participant characteristics (e.g., number, gender, age, height, weight, and training levels), *r*-value, main assessment methods and main findings. Where missing data could not be determined, the authors of respective studies were contacted via email to request individual and/or aggregate data.

2.7 | Data Analysis

We considered the following versions of the Yo-Yo tests: YYET1, YYET2, YYIE1, YYIE2, YYIR1, and YYIR2. Due to the limited number of studies on YYET1 and YYET2, these were combined into a single category, YYET. Physical fitness tests were categorized into sprint, jump, agility, strength, field-based endurance, RSA, treadmill endurance, body anthropometry, and team sports-specific tests. Additionally, meta-regression was performed to examine the relationship between $\dot{V}O_{2\,max}$, physical fitness tests performance and the Yo-Yo tests in the team sports. In addition, we performed subgroup analyses for the training level of the subjects. Specifically, training levels were classified into three categories: (i) recreationally active (RA),

as individuals not engaged in regular sport practice or competition; (ii) trained (TR), as individuals practicing and competing at an amateur club, university, or college, or otherwise not explicitly identified as elite or professional, and (iii) highly trained (HL), as those practicing and competing professionally or semi-professionally [36].

2.8 | Risk of Bias Assessment

Two authors (Z.T. and K.X.) independently assessed the methodological quality of each study with a third author (I.C.) to resolve discrepancies. The Agency for Healthcare Research and Quality (AHRQ) was used to determine the level of bias in all included cross-sectional studies [37]. This checklist has been shown to correlate highly with quality assessment scores [38]. AHRQ recommends using high, medium, and low-risk classification methods to assess the overall risk of bias in individual studies [39]. A scoring system was applied where yes = 1, no = 0, and unclear to determine = 0. Scores of 0 to 3, 4 to 7, and 8 tolls are generally defined as low, medium, and high quality, respectively.

2.9 | Certainty of Evidence

Evidence of validity for each study was amalgamated with quality scores to facilitate discussion of the results. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology was employed to assess the level of evidence, ranging from very low to a high level of certainty [40]. The initial level of evidence was rated as high. Evidence was downgraded to low or very low for the following limitations: sample sizes below 400 (imprecision), lack of a clear direction in combined results (imprecision), high statistical heterogeneity ($I^2 > 50\%$, inconsistency), and indications of publication bias. Each limitation resulted in a one-level downgrade in the evidence grade.

2.10 | Statistical Analysis

To address the non-normal distribution of Pearson's correlation coefficient, the r-value was transformed into Fisher's Z [41]. This transformation improves the normality of the distribution, enhancing the reliability and power of statistical tests. The z-transformed correlation coefficient was then meta-analyzed to obtain a weighted point estimate with 95% confidence intervals, and these were back-transformed into a Pearson correlation coefficient using an integral Z-to-r transformation [42].

The formula for converting Pearson's r to Fisher's Z was:

$$\text{Fisher}'\text{s } Z = 0.5 \times \text{In}\frac{1+r}{1-r}$$

Where r is Pearson's correlation coefficient. The standard error of Fisher's Z is calculated by use of the following formula [43]:

$$\text{SE} = \sqrt{1/(n-3)}$$

After pooling was completed, Fisher's Z was converted back to r using the inverse formula:

$$r = \frac{e^{-2Z}-1}{e^{-2Z}+1}$$

The Fisher's Z values were meta-analyzed using a random effects model because this approach accounts for the potential heterogeneity across studies, leading to more precise estimates [41, 43]. In the 15 analyses conducted, 13 of the analyses involved studies that reported multiple effect sizes within the same study. More importantly, multiple effect sizes from the same study may be more similar to each other than those from different studies. For instance, Alemdaroğlu et al. [44] conducted two different types of Yo-Yo tests on the same population using the same experimenters and data analysis methods. Including these effect sizes simultaneously would violate the independence assumption of traditional meta-analysis (two-level model). To address this, we employed a three-level meta-analysis (multilevel model) [45]. Multiple measurements and comparisons from the same study were nested within studies, so the variance of the observed effect sizes was decomposed into sampling variance, within-study variance (level 2), and between-study variance (level 3) [45, 46].

We employed cluster-robust variance estimation methods with small-sample adjustments to account for the within-study correlations between effect sizes [47, 48]. Assuming a within-study correlation of 0.6, sensitivity analyses with correlations of 0.4 and 0.8 showed no differences in the meta-analysis outcomes. The three-level model was selected when its goodness-of-fit indices (e.g., Akaike Information Criterion, Bayesian Information Criterion) were higher than those of the traditional two-level model; otherwise, the two-level model was used.

For the remaining two analyses without nested structures (the correlation between YYET and $\dot{V}O_{2\,max}$, and the correlation between the Yo-Yo tests and strength), we used the traditional two-level analysis model. All model parameters were assessed using the restricted maximum likelihood estimation method, with tests of individual coefficients in all models and their corresponding confidence intervals based on a t-distribution. The "optim" algorithm was applied to resolve model convergence issues by using the control setting *control = list (optimizer = "optim")* [49]. The weights were calculated using the inverse variance method based on the sample size. The analysis was performed in the *metafor* package in R studio (version 4.3.0, R Foundation for Statistical Computing) [50].

Correlations were interpreted as < 0.1 trivial; 0.1–0.29 small; 0.30–0.49 moderate; 0.5–0.69 large; 0.7–0.89 very large; 0.9–0.99 nearly perfect [41]. Cochran Q statistic was used to assess heterogeneity between studies, measuring differences in main effect between studies. The I^2 statistic was also used to assess heterogeneity, with values greater than 50% indicating high heterogeneity. Publication bias was evaluated by the visual assessment of funnel plots. When the number of included studies was ≥ 10, Egger's regression test was used to identify the presence of highly influential studies that might have biased the analysis [51]. The visualization of the funnel plot was performed using the *Metaviz* package [52].

2.11 | Subgroup Analysis and Sensitivity Analyses

Considering that the relationship between the Yo-Yo test and $\dot{V}O_{2\,max}$, sprinting, jumping, and other performance outcomes may be influenced by participants' training levels [17], we conducted subgroup analyses based on training level when the number of effect sizes per group exceeded five [53]. Leverage, outliers, and influence were assessed for all meta-analysis models by calculating hat values, Cook's distances, and studentized residuals, respectively [54–56]. Potential outliers were flagged if their hat values and Cook's distances exceeded three times the mean, and if studentized residuals were >3. After excluding these outliers, calculations were re-run. If the results remained unchanged, the original model was retained; otherwise, the model excluding the outliers was considered [57]. In addition, as per the reviewer's request, we applied Bayesian fitting methods and performed analyses using both two-level and three-level models [43]. The detailed procedure can be found at (osf.io/m29h7).

3 | Results

3.1 | Study Selection

The PRISMA flow diagram for the studies selection process is presented in Figure 2. Literature searches were performed in four databases, resulted in a total of 2922 articles, with 7 additional studies manually added after checking the references of eligible studies. Following the removal of 2770 duplicates, 159 titles and abstracts were screened. A total of 67 full-text articles were identified for retrieval and screened for eligibility. Of these, 33 studies were included in the quantitative analysis the correlation between the five Yo-Yo tests and $\dot{V}O_{2\,max}$. In addition, 42 372 articles were identified from the same four databases search for the correlation between the Yo-Yo tests and the physical fitness tests. Furthermore, five studies were manually added after reviewing the references lists. Following the removal of 24 570 duplicates, 17 807 titles and abstracts were screened. A total of 70 full-text articles were identified for retrieval and screened for eligibility, of which 36 studies were included in the quantitative analysis.

3.2 | Methodological Appraisal of the Included Studies

Overall, these studies ($n=33$) that included the relationship between the Yo-Yo tests and $\dot{V}O_{2\,max}$ were of medium quality with a mean score of 6.15. None of the included studies had means score higher than 7 out of 11 points (Table S11). Most studies addressed the following items: indicate if evaluators of subjective components of the study were masked to other aspects of the status of the participants; explain any patient exclusions from analysis; if applicable, explain how missing data were handled in the analysis; summarize patient response rates and completeness of data collection; clarify what follow-up, if any, was expected and the percentage of patients for which incomplete data or follow-up was obtained. In addition, these studies ($n=36$) that included the correlation between the Yo-Yo tests and physical fitness tests were of medium quality with a mean score of 5.83. Similarly, none of the included studies had a mean score higher than 7 out of

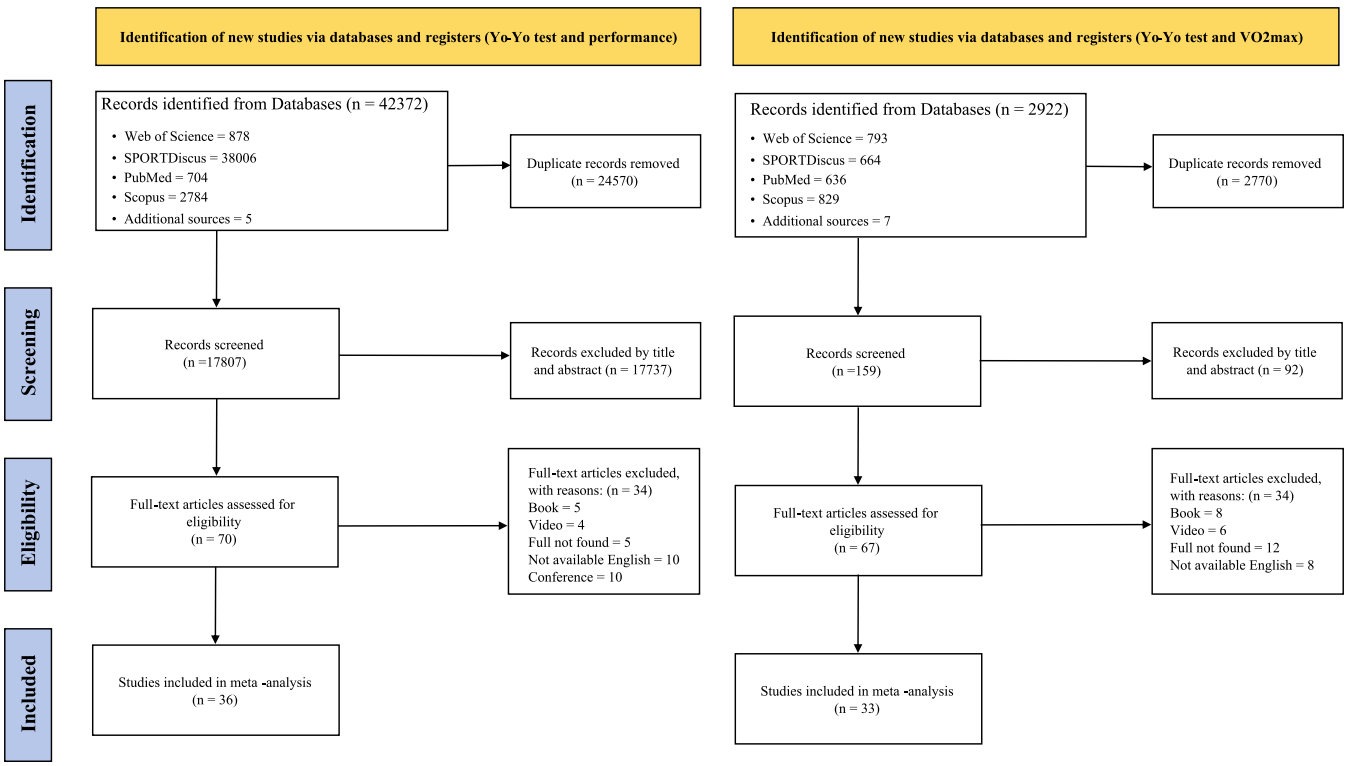

FIGURE 2 | PRISMA flow diagram of study selection for the analysis on relationships between performance, $\dot{V}O_{2\,max}$, and Yo-Yo tests. CI, Confidence interval; *K*, Number of studies; *N*, Number of participants; Yo-Yo ET, Yo-Yo Endurance test; Yo-Yo IE, Yo-Yo Intermittent Endurance test; Yo-Yo IR, Yo-Yo Intermittent Recovery test.

11 points (Table S12). Most of studies addressed the following items: if applicable, explain how missing data were handled in the analysis; summarize patient response rates and completeness of data collection; clarify what follow-up, if any, was expected and the percentage of patients for which incomplete data or follow-up was obtained.

3.3 | Study Characteristics

Studies that included the relationship between the Yo-Yo tests and $\dot{V}O_{2\,max}$ in this review were published between 2001 and 2024. In total, 1362 participants with varied training levels (e.g., high-level, trained, and recreational active), gender (1296 male and 66 female) and age (≥18 or <18 years) were considered. Additionally, the team sports mainly included were association football (soccer), football codes, basketball, hockey, cricket, rugby and futsal. The characteristics from the included studies were presented (Table S1).

Furthermore, studies that included the relationship between the Yo-Yo tests and physical fitness tests in this review were published between 2001 and 2023. In general, 1262 participants with varied training levels (e.g., high-level, trained, and recreational active), gender (1170 male and 92 female), and age (≥18 or <18 year-old) were distributed. Additionally, the team sports mainly included were association football (soccer), football codes, basketball, hockey, cricket, rugby, and handball. The characteristics of these studies included in this review were presented (Tables S2–S10).

3.4 | Risk of Bias Assessment

According to Egger's regression test analysis, which was performed for results including more than 10 effect sizes, the relationships between the two Yo-Yo tests and $\dot{V}O_{2\,max}$ were as follows: YYIR1 ($p = 0.32$) and YYIR2 ($p = 0.86$). In addition, Egger's regression test analysis showed that the relationships between the Yo-Yo tests and physical fitness tests were as follows: jump tests ($p = 0.07$) and treadmill endurance tests ($p = 0.52$). The p values from the Egger's test for all analyses were >0.05, and funnel plots for the analyses not subjected to the Egger's test were relatively symmetric. Therefore, no clear risk of publication bias was observed (Figures 3 and 4).

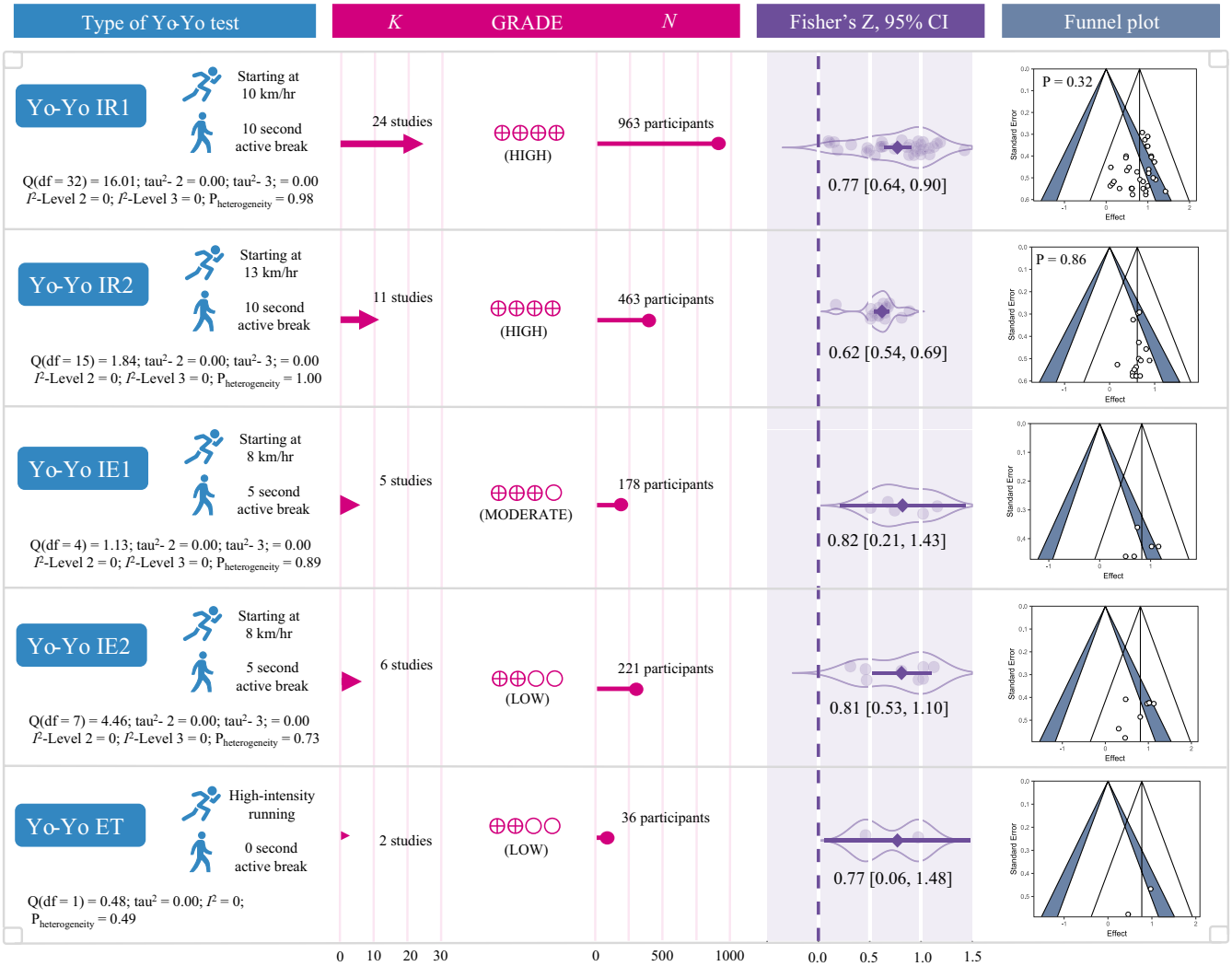

FIGURE 3 | Correlation of YYIR1, YYIR2, YYIE1, YYIE2, and YYET with $\dot{V}O_{2\,max}$. CI, Confidence interval; IR1/2, Yo-Yo Intermittent Recovery Level 1/2; *K*, Number of studies *N*, Number of participants; RSA, Repeated Sprint Ability.

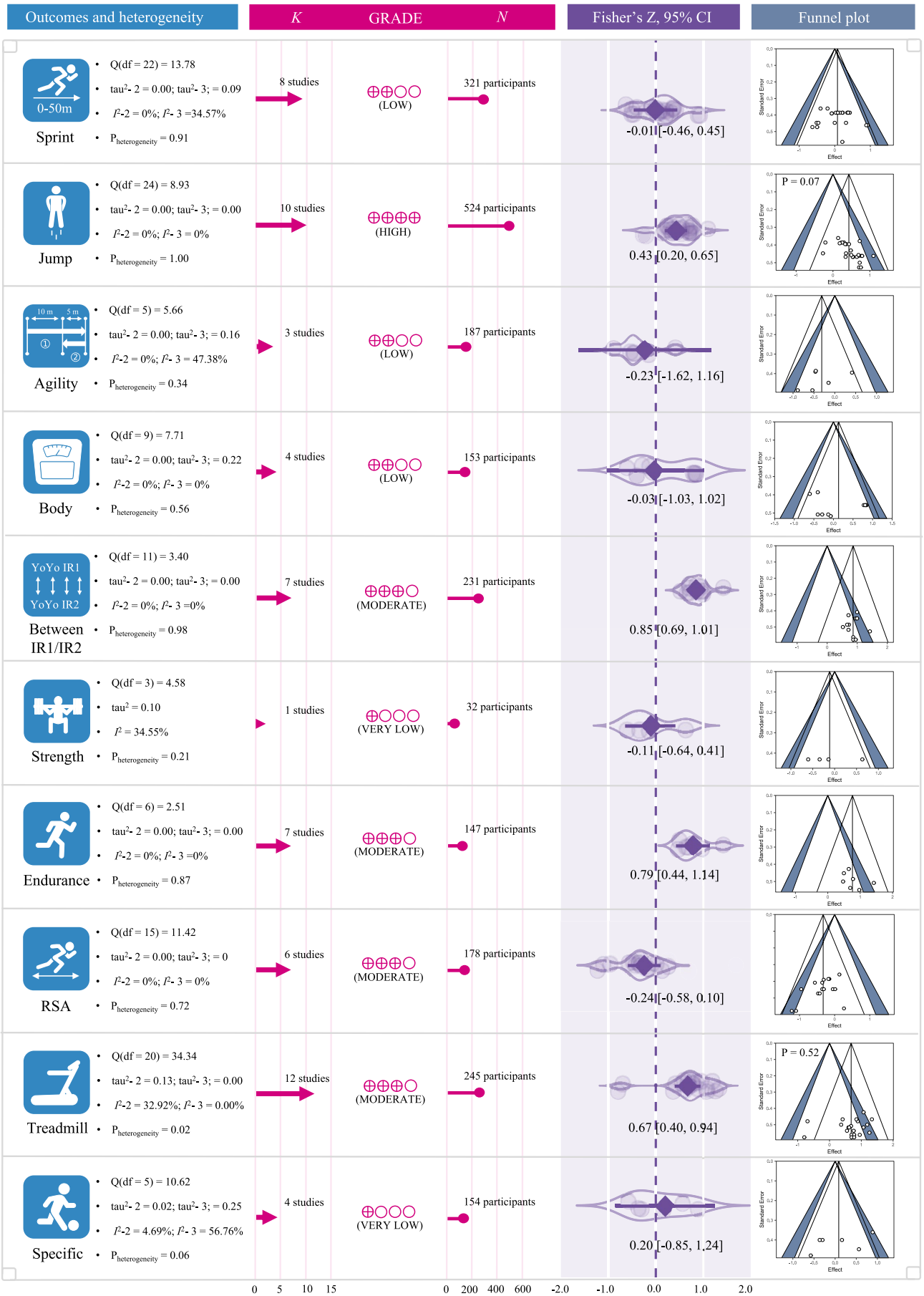

FIGURE 4 | Correlation of the various the Yo-Yo tests with physical fitness tests.

3.5 | Certainty of Evidence

In the included literature on the correlation between the five Yo-Yo tests and $\dot{V}O_{2\,max}$, the certainty of evidence ranged from low to high: YYET test (low), YYIE1 test (moderate), YYIE2 test (low), YYIR1 test (high), and YYIR2 test (high). Moreover, in the included literature on the correlation between the Yo-Yo tests and the physical fitness tests, the certainty of evidence ranged from very low to high: sprint tests (low), jump tests (high), agility tests (low), body anthropometry tests (low), and between the YYIR1and YYIR2 (moderate), strength tests (low), field-based endurance tests (moderate), RSA (moderate), treadmill endurance tests (moderate), and team sports specific tests (very low) (Figures 3 and 4).

3.6 | Meta-Analysis and Subgroup Analysis

Although the three-level model did not show a significantly better fit compared to the two-level model (Table S16), the data in our study included multiple effect sizes within each study, which likely exhibited dependencies. Therefore, we retained the results from the three-level model analysis. The meta-analysis conducted on the significant pooled results of the meta-analysis revealed that the five Yo-Yo tests were all effective methods for assessing $\dot{V}O_{2\,max}$ (Figure 3): YYET ($n=36$, $r=0.65$, large), YYIE1 ($n=178$, $r=0.67$, large), YYIE2 ($n=221$, $r=0.67$, large), YYIR1 ($n=963$, $r=0.65$, large) and YYIR2 ($n=463$, $r=0.55$, large) (Table S13). However, after performing Bayesian analysis on the YYET, the results changed, with the credible interval crossing zero, suggesting a lack of significant correlation in this model.

Furthermore, the meta-analysis of the significant pooled results demonstrated that there were correlations between Yo-Yo tests and jump performance ($n=524$, $r=0.40$, moderate), field-based endurance performance ($n=147$, $r=0.66$, large), and treadmill endurance tests performance ($n=245$, $r=0.59$, large). Additionally, the pooled results of the meta-analysis also revealed a positive correlation between the YYIR1 and YYIR2 ($n=231$, $r=0.69$, very large) (Table S14).

The results of the subgroup analysis showed that the correlation between Yo-Yo tests and $\dot{V}O_{2\,max}$ as well as physical fitness tests were presented in the Table 1. It is worth noting that, in the HL population, the correlation between the Yo-Yo tests and RSA was significantly higher than that with TR ($p=0.04$).

In all analyses, excluding outliers did not affect the results (see https://osf.io/c7nmd). Bayesian analysis showed that only the correlation between YYET and $\dot{V}O_{2\,max}$ changed, while all other results were robust (see Table S15).

4 | Discussion

To the best of our knowledge, the present study is the first systematic review and meta-analysis to examine the correlation of the five Yo-Yo tests and $\dot{V}O_{2\,max}$ as well as physical fitness components in team sports and to provide practical reference to systematic physical fitness assessments for coaches and practitioner.

Our results showed that the four different Yo-Yo tests (e.g., YYIE1, IE2, IR1, and IR2) were large correlated with $\dot{V}O_{2\,max}$, indicating they are effective tests for estimating $\dot{V}O_{2\,max}$. Among them, the YYIR2 test has the lowest correlation. This could be attributed to the rate of lactate production towards the end of the YYIR2 test which was reported to be higher than that during the YYIR1 test [20]. Indeed, the YYIR2 test was designed to examine the capacity to perform intense intermittent exercise with a large anaerobic component in combination with a significant aerobic contribution [58]. This was further confirmed by previous study showing higher peak blood lactate concentration during the YYIR2 test as compared to that during the YYIR1 test [58]. In addition, the results of sensitivity analysis showed that lack of significant correlation between YYET and $\dot{V}O_{2\,max}$ due to the limited number of studies. Therefore, practitioners should exercise caution when assessing $\dot{V}O_{2\,max}$ with YYET in practice.

Treadmill and bicycle ergometer tests for assessing $\dot{V}O_{2\,max}$ in laboratory settings are widely regarded as the gold standard for evaluating the aerobic capacity of team sports athletes [59]. Nonetheless, laboratory-based $\dot{V}O_{2\,max}$ testing is expensive, requiring advanced equipment, specialized and experienced staff [60]. Moreover, such tests are time-consuming compared to field-based alternatives, which limits their practicality for certain applications in team sports [61]. Thus, there is real interest in field-based tests to examine team sport players' aerobic power. However, conflicting results concerning the relationship between the Yo-Yo test and $\dot{V}O_{2\,max}$ in team sports have been observed in the previous studies [19, 24, 31, 62, 63], which led to uncertainty regarding the use Yo-Yo testing as a valid tool to monitor aerobic capacity of athletes in team sports. The findings of this review may contribute to clarifying the actual correlation between the Yo-Yo test and $\dot{V}O_{2\,max}$, thereby serving as a valuable resource for practitioners and researchers utilizing the Yo-Yo test to evaluate $\dot{V}O_{2\,max}$ level. Overall, the Yo-Yo tests can be recommended as an effective practical tool for practitioners, for assessing the specific aerobic capacity of team sport athletes.

This review found a correlation coefficient of 0.40 between jump ability and performance in the Yo-Yo Intermittent tests (e.g., YYIE1, IR1, and IR2), indicating a moderate relationship. The findings reflect the demands of competitive team sports, where players frequently perform activities involving turning and changes of direction at varying intensities, requiring significant muscle power and strength [64]. Additionally, studies have shown that jump performance is competitive level-dependent in team sports [65], suggesting that a player's level of maximal muscle power, expressed during muscle fiber fast stretch-shortening actions, affects performance during high-intensity intermittent running. A previous study by Wong et al. [66] implemented a 12-week strength and power training together with regular team training, resulting in a significant improvement in YYIE1 performance after intervention. Generally, both jump performance and the Yo-Yo test are essential components for success in team sports. The interaction between these two physical attributes can aid coaches in designing more effective testing and training protocols, ultimately helping athletes achieve higher levels of physical fitness. The current study revealed only a moderate correlation between explosive strength and Yo-Yo Intermittent

TABLE 1 | The results of subgroup analysis.

Test type	Studies (*k*)	ES (*n*)	Participants (*n*)	Fisher's *Z*	95% CI	Pearson *r*	*p*-values	*p*-value difference
YYIR1								
HL	6	9	255	0.73	[0.42; 1.05]	0.63	**<0.001**	Intercept
RA	7	9	292	0.84	[0.52; 1.15]	0.68	**<0.001**	−0.3256
TR	10	14	387	0.73	[0.47; 0.98]	0.62	**<0.001**	−0.3745
YYIR2								
HL	4	5	69	0.63	[0.11; 1.15]	0.56	**0.020**	Intercept
RA	1	2	30	0.73	[−0.002; 1.47]	0.63	0.051	−0.5925
TR	5	8	345	0.58	[0.25; 0.92]	0.53	**<0.001**	−0.6141
Sprint								
HL	4	12	455	0.22	[−0.32; 0.77]	0.22	0.410	Intercept
TR	3	11	424	−0.18	[−0.66; 0.30]	−0.18	0.440	0.262
Jump								
HL	3	11	402	0.51	[0.09; 0.93]	0.47	**0.020**	Intercept
RA	1	5	148	0.55	[0.06; 1.03]	0.50	**0.030**	−0.5418
TR	6	10	321	0.34	[0.02; 0.66]	0.33	**0.040**	−0.6665
IR1 and IR2								
HL	3	4	81	0.89	[0.29; 1.48]	0.71	**0.010**	Intercept
RA	1	2	33	0.81	[0.02; 1.59]	0.67	**0.040**	0.8368
TR	3	6	145	0.85	[0.38; 1.32]	0.69	**0.001**	0.9116
RSA								
HL	3	6	70	−0.64	[−1.15; −0.14]	−0.57	**0.020**	Intercept
TR	4	10	335	−0.08	[−0.46; 0.29]	−0.08	0.640	**0.0361**
Treadmill								
HL	2	3	41	0.80	[−0.02; 1.63]	0.67	0.060	Intercept
RA	1	6	130	0.89	[0.28; 1.50]	0.71	**0.010**	0.8342
TR	9	12	205	0.56	[0.14; 0.98]	0.51	**0.010**	0.5846

Note: Bold indicates the statistical significance of $p < 0.05$.

Abbreviations: CI, Confidence Interval; ES, effect size; HL, High Level; IR, Intermittent Recovery; RA, Recreationally Active; RSA, Repeated Sprint Ability; TR, Trained; YY, Yo-Yo test.

tests performance in team sports, highlighting that explosive strength and aerobic intermittent exercise performance are distinct physical qualities, both essential for athlete performance in training and match play.

A very-large ($r = 0.69$) correlation coefficient between YYIR1 and YYIR2 was found in our pooled results. This correlation magnitude can be attributed to several factors, including the fact that both tests share an identical testing process over the same distances. The crucial difference between the two tests was the stimulation of the anaerobic system, both during and towards the end of the tests. In the YYIR2, the CP levels at the end of the test were lower, and the rate of CP utilization in the last phase of the test was significantly higher compared with the YYIR1 [67]. Furthermore, the muscle lactate concentration at the end of the test was higher in the YYIR2 test than in the YYIR1, and the rate of lactate accumulation in the last phase of the test was about five times higher [58]. Accordingly, muscle pH was lower at exhaustion in the YYIR2 than in the YYIR1 [20]. In addition, in the YYIR2, the rate of lactate accumulation in the blood during the last phase of the test and the peak blood lactate concentration were higher compared to the YYIR1 [1]. In general, the rate of anaerobic energy production and lactate production rate towards the end of the YYIR2 were higher than the YYIR1 [17]. The YYIR1 focuses on the capacity to carry out intermittent exercise leading to maximal activation of the aerobic system, whereas the YYIR2 determines an individual's ability to recovery from repeated exercise with a high contribution from the anaerobic system [17]. Although there are numerous physiology differences between the YYIR1 and YYIR2, several studies

have shown a significant correlations between both [60, 68, 69]. Based on these findings, we recommend that coaches use the YYIR2 to evaluate athletes' anaerobic intermittent exercise capacity, while the YYIR1 or YYIE2 are better suited for estimating aerobic intermittent exercise capacity.

The pooled results of this review indicated large correlation coefficients between the Yo-Yo Intermittent tests (e.g., YYIE1, IR1, and IR2) and field-based endurance tests ($r = 0.66$) and the treadmill endurance test ($r = 0.59$). Commonly used to assess endurance performance, these tests include the Multistage Fitness Test (MST), 20-m Multistage Shuttle Run (MSR), and 30–15 Intermittent Fitness Test (30–15 IFT). Physiological measurements during treadmill tests are often considered the gold standard for aerobic assessment [63]. Field-based endurance tests and Yo-Yo Intermittent tests are both validated measures of maximal aerobic exertion in team sports. Their testing formats and processes, such as the MST, MSR, and 30–15 IFT, are similar, involving running between set markers at progressively increasing speeds until exhaustion [63, 70]. The key distinction between Yo-Yo Intermittent tests and treadmill endurance tests lies in their structure: the Yo-Yo tests incorporate short recovery intervals, reflecting the intermittent demands of team sports, whereas treadmill tests are continuous, with speeds increasing steadily until exhaustion. Therefore, the Yo-Yo Intermittent tests align more closely with the specific intermittent recovery characteristics of team sports. In general, the Yo-Yo Intermittent tests, field-based endurance tests, and treadmill endurance tests are all effective methods for measuring aerobic capacity in team sport athletes. However, given the unique characteristics and practical demands of team sports, we recommend that coaches and researchers prioritize the Yo-Yo Intermittent tests (particularly YYIE1, IE2, and IR1) as valid and sport-specific aerobic capacity assessment tools.

The results of the subgroup analysis showed that the correlation between YYIR1, YYIR2 as well as $\dot{V}O_{2\,max}$ were not affected by the training level of the participants. The subgroup analysis results can be attributed to the several factors. Firstly, YYIR1 primarily assesses an individual's ability to perform repeated high-intensity aerobic exercise. Its design is based on intermittent progressive speed running, and the test is completed through energy metabolism dominated by the aerobic system [17]. YYIR2 places more emphasis on the contribution of the anaerobic system, requiring subjects to complete the test at a higher initial speed (13 km/h), and therefore relies more on anaerobic capacity (such as lactate tolerance) [17, 58]. As a core indicator of aerobic capacity, $\dot{V}O_{2\,max}$ mainly reflects cardiopulmonary function and muscle oxygen utilization efficiency [71]. Since YYIR1 focuses more on aerobic metabolism, its correlation with $\dot{V}O_{2\,max}$ is significantly higher than that of YYIR2 [17]. This difference stems from the physiological demands of the test rather than the subject's training level. Secondly, The formula ($\dot{V}O_{2\,max}$ = distance × coefficient + constant) proposed by Bangsbo et al. [17] has been widely validated and is applicable to groups with different training levels. The formula is based on a large sample of data (including athletes of different training levels), and its generalizability suggests that training level has little impact on the correlation. Lastly, the strong correlation between YYIR1 and $\dot{V}O_{2\,max}$ is mainly due to its aerobic metabolism-dominated design, while the correlation of YYIR2 is weaker due to its mixed metabolic characteristics [58]. This difference is determined by the physiological demands of the test itself, not the subject's training level.

The results of the subgroup analysis of the correlation between the Yo-Yo tests and the physical fitness tests showed that the Yo-Yo tests and the sprint tests, jump tests, and treadmill endurance tests were not affected by the training level of the subjects, and there were no differences between groups. However, it is worth noting that in the HL group, the correlation between the Yo-Yo tests and the RSA was significantly higher than that in the TR group ($p = 0.04$). The results of the subgroup analysis are mainly attributed to the following factors. Firstly, the Yo-Yo tests primarily evaluate an athlete's aerobic endurance, which is crucial for sustained performance [17]. High-level athletes typically exhibit a stronger aerobic base, enabling them to recover more efficiently during high-intensity activities like the RSA [72]. This is because aerobic capacity supports the replenishment of creatine phosphate stores, which are essential for repeated sprints [73]. Secondly, high-level athletes undergo specialized training that integrates both aerobic and anaerobic exercises. This dual focus enhances their ability to perform well in both the Yo-Yo tests (aerobic) and RSA (anaerobic). For instance, aerobic training improves recovery rates, while anaerobic training directly boosts sprint performance [74]. Lastly, high-level athletes are often at their physiological peak, with well-developed aerobic and anaerobic systems. This maturity allows them to perform consistently across different types of tests, including the Yo-Yo tests and RSA.

4.1 | Practical Applications

This systematic review highlights several practical applications of Yo-Yo tests for assessing and improving athletes' physical performance. Coaches are encouraged to use Yo-Yo tests (e.g., YYIE1, IE2, IR1, and IR2) to estimate $\dot{V}O_{2\,max}$ levels and inform training design. Integrating lower limb power tests with Yo-Yo Intermittent tests (e.g., YYIE1, IR1, and IR2) can enhance the efficiency of physical fitness assessments. Specifically, YYIE1, IE2, and IR1 are effective for estimating aerobic performance, while YYIR2 is better suited for evaluating high-intensity interval recovery capacity. Yo-Yo Intermittent tests, along with field-based endurance tests (e.g., MST, MSR, 30–15 IFT) and treadmill endurance tests, are validated methods for measuring aerobic capacity in team sports.

4.2 | Limitations

Meta-analysis is a powerful tool for synthesizing and reviewing existing quantitative studies to address research questions. However, this review has several limitations that must be considered, as they could impact the results. Firstly, this systematic review and meta-analysis is a cross-sectional observational analysis and, therefore, cannot establish causality. Secondly, the limited number of female participants in the included studies restricts the generalizability of the findings. Lastly, while the funnel plot and Egger's test results indicate no significant risk of publication bias, a potential bias remains. Among the included studies, 78%

reported a significant correlation between Yo-Yo tests and $\dot{V}O_{2\,max}$ while 69% found a significant correlation between Yo-Yo tests and physical fitness tests. This suggests that studies reporting significant correlations may have been more likely to be published.

5 | Conclusion

This is the first systematic review and meta-analysis to examine the usefulness of five Yo-Yo tests in estimating $\dot{V}O_{2\,max}$ and the physical fitness profile of team sport athletes. All four Yo-Yo tests (e.g., YYIE1, IE2, IR1, and IR2) were found to estimate $\dot{V}O_{2\,max}$ though YYIR2 demonstrated the weakest correlation. Moderate correlations were observed for field-based endurance tests, but low correlations were found for explosive tests. Coaches can use Yo-Yo tests as reliable tools for estimating $\dot{V}O_{2\,max}$ and endurance performance, while explosive strength should be assessed separately. Future research should explore the correlations between $\dot{V}O_{2\,max}$, physical fitness tests, and Yo-Yo tests performance across different players' playing positions. More studies are needed to investigate testing in female athletes, players under the age of 18, and variations in correlations during different stages of the competitive season.

Author Contributions

Conception and design of study: Z.T. and K.X.; Retrieval of articles and acquisition of data: Z.T., K.X., and I.C.; Analysis of data: Z.T., K.X., and I.C.; Interpretation of data: Z.T., K.X., I.C., C.C., and P.K.; Drafting the manuscript: Z.T.; Revising the manuscript critically for important intellectual content: C.C., P.K., D.P.W., S.P., D.B., K.X., and I.C. All authors read and approved the final manuscript.

Acknowledgments

Daniel Boullosa was supported by Grant RYC202-1-031098-1 funded by MCIN/AEI/10.13039/501100011033, be "European Union NextGenerationEU/PRTR", and by a productivity research grant PQ-1-D(317126/2021-0) by CNPq (Brazil).

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets generated and/or analyzed during the current review are available in tables, Supporting Information, and at the Open Science Framework (osf.io/m29h7). All other data are available upon request.

References

1. J. Ingebrigtsen, M. Brochmann, C. Castagna, et al., "Relationships Between Field Performance Tests in High-Level Soccer Players," *Journal of Strength & Conditioning Research* 28, no. 4 (2014): 942–949, https://doi.org/10.1519/JSC.0b013e3182a1f861.

2. M. Mohr, P. Krustrup, and J. Bangsbo, "Match Performance of High-Standard Soccer Players With Special Reference to Development of Fatigue," *Journal of Sports Sciences* 21, no. 7 (2003): 519–528.

3. D. J. Bishop and O. Girard, "Determinants of Team-Sport Performance: Implications for Altitude Training by Team-Sport Athletes," *British Journal of Sports Medicine* 47, no. Suppl 1 (2013): i17–i21.

4. J. F. Vigh-Larsen and J. A. Costa, "Fatigue in Soccer: Acute and Residual Responses," in *Training Load in Professional Soccer: Guide to Monitoring Performance* (Springer, 2024), 37–57.

5. T. J. Suchomel, S. Nimphius, and M. H. Stone, "The Importance of Muscular Strength in Athletic Performance," *Sports Medicine* 46 (2016): 1419–1449.

6. F. Impellizzeri, E. Rampinini, C. Castagna, et al., "Validity of a Repeated-Sprint Test for Football," *International Journal of Sports Medicine* 29, no. 11 (2008): 899–905.

7. E. Rampinini, D. Bishop, S. Marcora, D. F. Bravo, R. Sassi, and F. Impellizzeri, "Validity of Simple Field Tests as Indicators of Match-Related Physical Performance in Top-Level Professional Soccer Players," *International Journal of Sports Medicine* 28, no. 3 (2007): 228–235.

8. M. Svensson and B. Drust, "Testing Soccer Players," *Journal of Sports Sciences* 23, no. 6 (2005): 601–618.

9. A. Turner, S. Walker, M. Stembridge, et al., "A Testing Battery for the Assessment of Fitness in Soccer Players," *Strength & Conditioning Journal* 33, no. 5 (2011): 29–39.

10. J. M. Taylor, J. L. Madden, L. P. Cunningham, and M. Wright, "Fitness Testing in Soccer Revisited: Developing a Contemporary Testing Battery," *Strength & Conditioning Journal* 44, no. 5 (2022): 10–21.

11. G. Markovic, D. Dizdar, I. Jukic, and M. Cardinale, "Reliability and Factorial Validity of Squat and Countermovement Jump Tests," *Journal of Strength & Conditioning Research* 18, no. 3 (2004): 551–555.

12. T. Haugen and M. Buchheit, "Sprint Running Performance Monitoring: Methodological and Practical Considerations," *Sports Medicine* 46, no. 5 (2016): 641–656.

13. J. Sheppard, W. B. Young, T. Doyle, T. Sheppard, and R. U. Newton, "An Evaluation of a New Test of Reactive Agility and Its Relationship to Sprint Speed and Change of Direction Speed," *Journal of Science and Medicine in Sport* 9, no. 4 (2006): 342–349.

14. Y. Hachana, H. Chaabène, M. A. Nabli, et al., "Test-Retest Reliability, Criterion-Related Validity, and Minimal Detectable Change of the Illinois Agility Test in Male Team Sport Athletes," *Journal of Strength & Conditioning Research* 27, no. 10 (2013): 2752–2759, https://doi.org/10.1519/JSC.0b013e3182890ac3.

15. M. Jemni, M. S. Prince, and J. S. Baker, "RETRACTED ARTICLE: Assessing Cardiorespiratory Fitness of Soccer Players: Is Test Specificity the Issue?—A Review," *Sports Medicine-Open* 4 (2018): 1–18.

16. N. M. Beltz, A. L. Gibson, J. M. Janot, L. Kravitz, C. M. Mermier, and L. C. Dalleck, "Graded Exercise Testing Protocols for the Determination of VO_{2max}: Historical Perspectives, Progress, and Future Considerations," *Journal of Sports Medicine* 2016 (2016): 1–12.

17. J. Bangsbo, F. M. Iaia, and P. Krustrup, "The Yo-Yo Intermittent Recovery Test: A Useful Tool for Evaluation of Physical Performance in Intermittent Sports," *Sports Medicine* 38 (2008): 37–51.

18. C. Castagna, F. M. Impellizzeri, K. Chamari, D. Carlomagno, and E. Rampinini, "Aerobic Fitness and Yo-Yo Continuous and Intermittent Tests Performances in Soccer Players: Acorrelation Study," *Journal of Strength & Conditioning Research* 20, no. 2 (2006): 320–325, https://doi.org/10.1519/00124278-200605000-00015.

19. P. Krustrup, P. S. Bradley, J. F. Christensen, et al., "The Yo-Yo IE2 Test: Physiological Response for Untrained Men Versus Trained Soccer

Players," *Medicine and Science in Sports and Exercise* 47, no. 1 (2015): 100–108.

20. P. Krustrup, M. Mohr, L. Nybo, J. M. Jensen, J. J. Nielsen, and J. Bangsbo, "The Yo-Yo IR2 Test: Physiological Response, Reliability, and Application to Elite Soccer," *Medicine & Science in Sports & Exercise* 38, no. 9 (2006): 1666–1673.

21. J. Grgic, L. Oppici, P. Mikulic, J. Bangsbo, P. Krustrup, and Z. Pedisic, "Test–Retest Reliability of the Yo-Yo Test: A Systematic Review," *Sports Medicine* 49 (2019): 1547–1557.

22. S. C. Póvoas, C. Castagna, J. M. da Costa Soares, et al., "Reliability and Construct Validity of Yo-Yo Tests in Untrained and Soccer-Trained Schoolgirls Aged 9–16," *Pediatric Exercise Science* 28, no. 2 (2016): 321–330, https://doi.org/10.1123/pes.2015-0212.

23. P.-L. Wong, A. Chaouachi, C. Castagna, P. W. Lau, K. Chamari, and U. Wisløff, "Validity of the Yo-Yo Intermittent Endurance Test in Young Soccer Players," *European Journal of Sport Science* 11, no. 5 (2011): 309–315.

24. P. Krustrup, M. Mohr, H. Ellingsgaard, and J. Bangsbo, "Physical Demands During an Elite Female Soccer Game: Importance of Training Status," *Medicine & Science in Sports & Exercise* 37, no. 7 (2005): 1242–1248.

25. P. Krustrup, M. Randers, J. Horton, J. Brito, and A. Rebelo, "Ecological Validity of the Yo-Yo SFIE2 Test," *International Journal of Sports Medicine* 33, no. 6 (2012): 432–438.

26. C. Castagna, M. Bendiksen, F. M. Impellizzeri, and P. Krustrup, "Reliability, Sensitivity and Validity of the Assistant Referee Intermittent Endurance Test (ARIET)—A Modified Yo-Yo IE2 Test for Elite Soccer Assistant Referees," *Journal of Sports Sciences* 30, no. 8 (2012): 767–775.

27. T. Ahler, M. Bendiksen, P. Krustrup, and N. Wedderkopp, "Aerobic Fitness Testing in 6-to-9-Year-Old Children: Reliability and Validity of a Modified Yo–Yo IR1 Test and the Andersen Test," *European Journal of Applied Physiology* 112 (2012): 871–876.

28. C. Castagna, P. Krustrup, S. D'Ottavio, C. Pollastro, A. Bernardini, and S. C. A. Póvoas, "Ecological Validity and Reliability of an Age-Adapted Endurance Field Test in Young Male Soccer Players," *Journal of Strength & Conditioning Research* 33, no. 12 (2019): 3400–3405.

29. Y. Michailidis, S. Chavlis, M. Mitrotasios, et al., "The Use of Yo-Yo Intermittent Recovery Test Level 1 for the Estimation of Maximal Oxygen Uptake in Youth Elite Soccer Players," *Trends in Sport Sciences* 27, no. 3 (2020): 167–173.

30. K. Chuman, Y. Hoshikawa, T. Iida, and T. Nishijima, "Relationships Between Yo-Yo Intermittent Recovery Tests and Development of Aerobic and Anaerobic Fitness in U-13 and U-17 Soccer Players," *International Journal of Sport and Health Science* 9 (2011): 91–97.

31. A. Thomas, B. Dawson, and C. Goodman, "The Yo-Yo Test: Reliability and Association With a 20-m Shuttle Run and VO_{2max}," *International Journal of Sports Physiology and Performance* 1, no. 2 (2006): 137–149.

32. L. B. Michalsik, P. Fuchs, and H. Wagner, "The Team Handball Game-Based Performance Test Is Better Than the Yo-Yo Intermittent Recovery Test to Measure Match-Related Activities in Female Adult Top-Elite Field Team Handball Players," *Applied Sciences* 11, no. 14 (2021): 6551.

33. A. I. A. Medeiros, A. D. Andrade, W. Kassiano, et al., "What Physical Performance Characteristics Are Related With Age Categories of Elite Young Soccer Players?," *Journal of Physical Education and Sport* 23, no. 4 (2023): 957–964.

34. D. V. Parums, "Review Articles, Systematic Reviews, Meta-Analysis, and the Updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 Guidelines," *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 27 (2021): e934475–e934481.

35. J. Dinet, M. Favart, and J. M. Passerault, "Searching for Information in an Online Public Access Catalogue (OPAC): The Impacts of Information Search Expertise on the Use of Boolean Operators," *Journal of Computer Assisted Learning* 20, no. 5 (2004): 338–346.

36. A. K. McKay, T. Stellingwerff, E. S. Smith, et al., "Defining Training and Performance Caliber: A Participant Classification Framework," *International Journal of Sports Physiology and Performance* 17, no. 2 (2021): 317–331.

37. L. Hartling, M. Ospina, Y. Liang, et al., "Risk of Bias Versus Quality Assessment of Randomised Controlled Trials: Cross Sectional Study," *BMJ [British Medical Journal]* 339 (2009): b4012, https://doi.org/10.1136/bmj.b4012.

38. M. Farquhar and AHRQ Quality Indicators, *Patient Safety and Quality: An Evidence-Based Handbook for Nurses* (Agency for Healthcare Research and Quality, 2008).

39. M. Viswanathan, C. D. Patnode, N. D. Berkman, et al., "Recommendations for Assessing the Risk of Bias in Systematic Reviews of Health-Care Interventions," *Journal of Clinical Epidemiology* 97 (2018): 26–34.

40. H. J. Schünemann, J. P. Higgins, G. E. Vist, et al., "Completing 'Summary of Findings' Tables and Grading the Certainty of the Evidence," in *Cochrane Handbook for Systematic Reviews of Interventions* (Wiley-Blackwell, 2019), 375–402.

41. M. Borenstein, L. V. Hedges, J. P. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis* (John Wiley & Sons, 2021).

42. A. R. Hafdahl, "Random-Effects Meta-Analysis of Correlations: Monte Carlo Evaluation of Mean Estimators," *British Journal of Mathematical and Statistical Psychology* 63, no. 1 (2010): 227–254, https://doi.org/10.1348/000711009X431914.

43. M. Harrer, P. Cuijpers, T. Furukawa, and D. Ebert, *Doing Meta-Analysis With R: A Hands-On Guide* (CRC Press, 2021).

44. U. Alemdaroğlu, "Investigation of the Performance Responses of Yo-Yo and Shuttle Run Tests With the Treadmill Run Test in Young Soccer Players," *Pamukkale Journal of Sport Sciences* 3, no. 3 (2012): 104–112.

45. M. W.-L. Cheung, "Modeling Dependent Effect Sizes With Three-Level Meta-Analyses: A Structural Equation Modeling Approach," *Psychological Methods* 19, no. 2 (2014): 211.

46. M. W.-L. Cheung, "A Guide to Conducting a Meta-Analysis With Non-Independent Effect Sizes," *Neuropsychology Review* 29, no. 4 (2019): 387–396.

47. L. V. Hedges, E. Tipton, and M. C. Johnson, "Robust Variance Estimation in Meta-Regression With Dependent Effect Size Estimates," *Research Synthesis Methods* 1, no. 1 (2010): 39–65.

48. E. Tipton and J. E. Pustejovsky, "Small-Sample Adjustments for Tests of Moderators and Model Fit Using Robust Variance Estimation in Meta-Regression," *Journal of Educational and Behavioral Statistics* 40, no. 6 (2015): 604–634.

49. J. C. Nash and R. Varadhan, "Unifying Optimization Algorithms to Aid Software System Users: Optimx for R," *Journal of Statistical Software* 43 (2011): 1–14.

50. J. E. Pustejovsky and E. Tipton, "Meta-Analysis With Robust Variance Estimation: Expanding the Range of Working Models," *Prevention Science* 23, no. 3 (2022): 425–438.

51. M. Egger and G. Smith, "Bias in Meta-Analysis Detected by a Simple, Graphical Test," *BMJ: British Medical Journal: International Edition* 315, no. 7109 (1997): 629–634.

52. M. Kossmeier, U. S. Tran, and M. Voracek, "Visual Inference for the Funnel Plot in Meta-Analysis," *Zeitschrift für Psychologie* 227, no. 1 (2019): 83–89.

53. J. J. Deeks, J. P. Higgins, D. G. Altman, and Group CSM, "Analysing Data and Undertaking Meta-Analyses," in *Cochrane Handbook for Systematic Reviews of Interventions* (Wiley-Blackwell, 2019), 241–284.

54. J. P. Stevens, "Outliers and Influential Data Points in Regression Analysis," *Psychological Bulletin* 95, no. 2 (1984): 334–344.

55. W. Viechtbauer and M. W. L. Cheung, "Outlier and Influence Diagnostics for Meta-Analysis," *Research Synthesis Methods* 1, no. 2 (2010): 112–125.

56. H. Aguinis, R. K. Gottfredson, and H. Joo, "Best-Practice Recommendations for Defining, Identifying, and Handling Outliers," *Organizational Research Methods* 16, no. 2 (2013): 270–301.

57. B. Van Hooren, I. Jukic, M. Cox, K. G. Frenken, I. Bautista, and I. S. Moore, "The Relationship Between Running Biomechanics and Running Economy: A Systematic Review and Meta-Analysis of Observational Studies," *Sports Medicine* 54, no. 5 (2024): 1269–1316.

58. P. Krustrup, M. Mohr, T. Amstrup, et al., "The Yo-Yo Intermittent Recovery Test: Physiological Response, Reliability, and Validity," *Medicine & Science in Sports & Exercise* 35, no. 4 (2003): 697–705.

59. P. J. Maud and C. Foster, *Physiological Assessment of Human Fitness* (Human Kinetics, 2006).

60. E. Rampinini, A. Sassi, A. Azzalin, et al., "Physiological Determinants of Yo-Yo Intermittent Recovery Tests in Male Soccer Players," *European Journal of Applied Physiology* 108, no. 2 (2010): 401–409.

61. P. Bradley, M. Bendiksen, A. Dellal, et al., "The Application of the Yo-Yo Intermittent Endurance Level 2 Test to Elite Female Soccer Populations," *Scandinavian Journal of Medicine & Science in Sports* 24, no. 1 (2014): 43–54, https://doi.org/10.1111/j.1600-0838.2012.01483.x.

62. Y. Michailidis, "The Relationship Between Aerobic Capacity, Anthropometric Characteristics, and Performance in the Yo-Yo Intermittent Recovery Test Among Elite Young Football Players: Differences Between Playing Positions," *Applied Sciences* 14, no. 8 (2024): 3413.

63. A. R. Aziz, F. H. Tan, and K. C. Teh, "A Pilot Study Comparing Two Field Tests With the Treadmill Run Test in Soccer Players," *Journal of Sports Science and Medicine* 4, no. 2 (2005): 105–112.

64. J. Hoff and J. Helgerud, "Endurance and Strength Training for Soccer Players: Physiological Considerations," *Sports Medicine* 34 (2004): 165–180.

65. C. Castagna and E. Castellini, "Vertical Jump Performance in Italian Male and Female National Team Soccer Players," *Journal of Strength & Conditioning Research* 27, no. 4 (2013): 1156–1161.

66. P.-l. Wong, K. Chamari, and U. Wisløff, "Effects of 12-Week On-Field Combined Strength and Power Training on Physical Performance Among U-14 Young Soccer Players," *Journal of Strength and Conditioning Research* 24, no. 3 (2010): 644–652, https://doi.org/10.1519/JSC.0b013e3181ad3349.

67. M. Mohr, P. Krustrup, J. J. Nielsen, et al., "Effect of Two Different Intense Training Regimens on Skeletal Muscle Ion Transport Proteins and Fatigue Development," *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 292, no. 4 (2007): R1594–R1602.

68. J. Ingebrigtsen, M. Bendiksen, M. B. Randers, C. Castagna, P. Krustrup, and A. Holtermann, "Yo-Yo IR2 Testing of Elite and Sub-Elite Soccer Players: Performance, Heart Rate Response and Correlations to Other Interval Tests," *Journal of Sports Sciences* 30, no. 13 (2012): 1337–1345.

69. I. Riccardo, G. Altavilla, C. Antonio, R. Gaetano, D. I. Tiziana, and G. Marco, "Performance Improvement in Yo-Yo Intermittent Recovery Test Level 2 and During Official Matches: The Role of Speed Endurance Training Production in Élite Football Players," *Sport Mont* 18 (2020): 61–66.

70. M. Buchheit and A. Rabbani, "The 30–15 Intermittent Fitness Test Versus the Yo-Yo Intermittent Recovery Test Level 1: Relationship and Sensitivity to Training," *International Journal of Sports Physiology and Performance* 9, no. 3 (2014): 522–524.

71. M. Hollenberg and I. B. Tager, "Oxygen Uptake Efficiency Slope: An Index of Exercise Performance and Cardiopulmonary Reserve Requiring Only Submaximal Exercise," *Journal of the American College of Cardiology* 36, no. 1 (2000): 194–201.

72. B. J. Peterson, "Repeated Sprint Ability: The Influence of Aerobic Capacity on Energy Pathway Response and Fatigue of Hockey Players," (2014), University of Minnesota.

73. D. Bishop and O. Girard, "Repeated Sprint Ability," in *Exercise Physiology: From a Cellular to an Integrative Approach* (IOS Press, 2010), 385–400.

74. D. L. Tomlin and H. A. Wenger, "The Relationship Between Aerobic Fitness and Recovery From High Intensity Intermittent Exercise," *Sports Medicine* 31 (2001): 1–11.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.