SYSTEMATIC REVIEW

Open Access

Submaximal Fitness Test in Team Sports: A Systematic Review and Meta-Analysis of Exercise Heart Rate Measurement Properties

Tzlil Shushan^{1*}[®], Ric Lovell^{1,2}[®], Martin Buchheit^{3,4,5,6}, Tannath J. Scott^{7,8}[®], Steve Barrett⁹[®], Dean Norris¹[®] and Shaun J. McLaren^{10,11}[®]

Abstract

Background Submaximal fitness tests (SMFT) are a pragmatic approach for evaluating athlete's physiological state, due to their time-efficient nature, low physiological burden and relative ease of administration in team sports settings. While a variety of outcome measures can be collected during SMFT, exercise heart rate (HRex) is the most popular. Understanding the measurement properties of HRex can support the interpretation of data and assist in decision making regarding athlete's current physiological state and training effects.

Objectives The aims of our systematic review and meta-analysis were to: (1) establish meta-analytic estimates of SMFT HRex reliability and convergent validity and (2) examine the moderating influence of athlete and protocol characteristics on the magnitude of these measurement properties.

Methods We conducted a systematic literature search with MEDLINE, Scopus and Web of Science databases for studies published up until January 2022 since records began. Studies were considered for inclusion when they included team sports athletes and the reliability and/or convergent validity of SMFT HRex was investigated. Reliability statistics included the group mean difference (MD), typical error of measurement (TE) and intraclass correlation coefficient (ICC) derived from test–retest(s) designs. Pearson's correlation coefficient (*r*) describing the relationship between SMFT HRex and a criterion measure of endurance performance was used as the statistic for convergent validity. Qualitative assessment was conducted using risk of bias assessment tool for non-randomised studies. Mixed-effects, multilevel hierarchical models combined with robust variance estimate tests were performed to obtain pooled measurement property estimates, effect heterogeneity, and meta-regression of modifying effects.

Results The electronic search yielded 21 reliability (29 samples) and 20 convergent validity (29 samples) studies that met the inclusion criteria. Reliability meta-analysis indicated good absolute (MD = 0.5 [95% CI 0.1 to 0.9] and TE = 1.6 [95% CI 1.4 to 1.9] % points), and high relative (ICC = 0.88 [95% CI 0.84 to 0.91]) reliability. Convergent validity meta-analysis indicated an inverse, large relationship (r = -0.58 [95% CI -0.62 to -0.54]) between SMFT HRex and endurance tests performance. Meta-regression analyses suggested no meaningful influence of SMFT protocol or athlete characteristics on reliability or convergent validity estimates.

Conclusions Submaximal fitness test HRex is a reliable and valid proxy indicator of endurance performance in team sport athletes. Athlete and SMFT protocol characteristics do not appear to have a meaningful effect on these measurement properties. Practitioners may implement SMFT HRex for monitoring athlete's physiological state by using our

*Correspondence: Tzlil Shushan

Tzlil21092@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

applied implications to guide the interpretation of data in practice. Future research should examine the utility of SMFT HRex to track within-athlete changes in aerobic capacity, as well as any further possible effects of SMFT protocols design elements or HRex analytical methods on measurement properties.

Registration Protocol registration can be found in Open Science Framework and available through https://doi.org/10. 17605/OSF.IO/9C2JV.

Keywords Submaximal fitness tests, Exercise heart rate, Meta-analysis, Measurement properties

Key Points

- Exercise heart rate (HRex) during submaximal fitness tests is a reliable and valid proxy measure of endurance performance in team sports.
- Athlete and test protocol-related characteristics do not appear to meaningfully affect HRex measurement properties.
- Our findings provide implications for test protocol selection, as well as conceptual and statistical data interpretations when using HRex for evaluating athlete's physiological state in team sports.

Introduction

Quantifying athlete's responses to training programmes is an integral part of the training process of team sports [1]. Assessing aerobic-oriented training effects is preferably made via maximal exhaustive tests administered in field-based or laboratory conditions, with higher physiological or performance outcomes indicative of improved aerobic capacity or adaptation to a training programme [2, 3]. However, given the limited viability of repeated maximal endurance performance tests in team sports, there is a growing interest in collecting proxy (surrogate) outcome measures of physiological capacities during alternative, time-efficient and non-exhaustive exercise assessments [4, 5].

Submaximal Fitness Tests (SMFT) provide a pragmatic approach for evaluating physiological state by assessing an athlete's internal load or responses to a standardised physical stimulus [6, 7]. We recently undertook a review of the literature on SMFT in team sports [8] and identified many variations in protocols, outcome measures and monitoring purposes. Exercise heart rate (HRex) is the most utilised SMFT outcome measure, perhaps due to the strong relationship to oxygen uptake during a continuous, (intended steady-state) exercise [4], supporting its use as a surrogate marker of aerobic capacity of cardiovascular fitness [4, 9]. Although various resting, exercise and recovery HR-derived measures may provide an insight into the cardiac autonomic nervous system state [10], HRex has also been proposed as a marker of shorter-term physiological stress given its sensitivity to factors such as an exposure to extreme environments (e.g., heat, altitude) [11, 12] or a training-induced overreaching state [10, 13]. Therefore, SMFT HRex might

provide practitioners with a valuable insight into athlete monitoring in team sports [4, 9].

Understanding the measurement properties of an outcome measure is a fundamental aspect in sports performance testing and monitoring [14]. Reliability refers to the consistency and reproducibility of an outcome measure [15, 16]. This is typically assessed using test-retest measurements [17] and quantifies the expected 'noise' (biological or measurement error) of the outcome measure [15, 17], which can subsequently assist in establishing meaningful changes for decision making [4]. Validity refers to the extent to which a measure represents the variable or construct it is intended to and includes several dimensions. For example, construct validity is the extent to which variables are reflective of theoretical constructs described by constitutive definitions and operationalizations. A sub-domain of construct validity is convergent validity-the degree to which two measures of a theoretical construct that should be related, are in fact related [18]. Within the SMFT framework, HRex should, theoretically at least, be associated with both positive [6] and negative [19] training effects.

To better understand the utility of SMFT HRex as a monitoring tool, a thorough knowledge of its measurement properties is required. There have been a considerable number of studies examining the reliability and convergent validity of SMFT HRex in team sport athletes; however, the overall (pooled) estimates of these measurement properties, the associated between- and within-study variance, and the influence of modifying effects have yet to be established. Therefore, the aims of our current paper are: (1) to provide the first meta-analysis of SMFT HRex reliability (absolute and relative) and convergent validity; and (2) to examine the influence of athlete's and SMFT characteristics on these measurement properties.

Methods

Registration and Search Strategy

Our systematic review and meta-analysis was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2020 guidelines [20] (PRISMA checklist items are available in the Additional file 1), and was prospectively registered in Open Science Framework (available through https://doi. org/10.17605/OSF.IO/9C2JV). Details on the registration process and changes made between the original to final protocol are available in Additional file 2. The research question for this review was developed as part of an iterative process originated in our previous review [8]. For this, we updated the systematic searches with an extension of the inclusion and exclusion criteria (Table 1). The electronic databases MEDLINE, Scopus and Web of Science were searched on multiple occasions, commencing in January 2020 and finalised in January 2022, denoting the earliest search date of the original review and the most updated searches for the meta-analysis, respectively. Detailed descriptions presenting the search strategy and results are provided in the Additional file 2.

Screening and Study Selection

The updated aspects in the inclusion-exclusion criteria (Table 1) for this review were related to study design and SMFT outcome measure, and the searching and screening processes are presented in Fig. 1. We sought to include studies examining SMFT HRex reliability using test-retest(s) designs and reporting absolute (group mean difference [MD], typical error of measurement [TE]) and/or relative (intraclass correlation coefficient [ICC]) effect estimates. Convergent validity studies included correlational study designs examining the relationship (correlation coefficient [r]) between SMFT HRex and an established team sport endurance performance test administered in a cross-sectional manner. In team sports, assessing endurance performance has been undertaken using via a variety of maximal effort and exhaustive tests administered in laboratory (typically, maximal oxygen uptake) or field-based (e.g., intermittent endurance capacity) conditions. While laboratory tests measuring maximal oxygen uptake are considered as a criterion measure of aerobic capacity, field-based tests are more pronounced in team sports (perhaps due to their practicality) and have been shown to deliver a satisfactory degree of reliability and validity [2, 21, 22]. These tests generally involve external intensity measures (either total distance covered or final running speed achieved) that are used as a proxy indicator of aerobic

Table 1 Meta-analysis inclusion and exclusion criteria

Criteria Inclusion

1	Original research published in peer- reviewed journal	Excluded non-investigation studies such as reviews, book chapters, abstracts, theses, opinion piece, surveys, letters to editor, etc.
2	Studies published until January 2022	In accordance with our most updated searches, studies published from the inception up until January 2022 were considered for inclusion, including studies published in the <i>Epub ahead of print</i> format
3	Available in English language	Full text is in English
4	Population	Team sport athletes with no restriction of age, level and sex
5	Study design	Test-retest designs
		Studies examining the reliability of SMFT HRex in controlled settings (e.g., scheduling, training context) and reporting absolute (group mean difference [MD], typical error of measurement [TE]), relative (intraclass correlation coefficient [ICC]) effect estimates, or both
		Correlational designs
		Studies examining the convergent validity of SMFT HRex with reference to a maximal endurance per- formance test and reporting correlational statistic (correlation coefficient [r]) of this relationship
6	Settings	SMFT were administered either in laboratory, indoor/outdoor field-based formats or combination of two
7	SMFT	Any SMFT including cycling, running, or standardised drills and games (refer to Fig. 2, 'SMFT protocol categories')
	Intensity	Excluded prolonged 'all-out' maximal intensities, intensities that cause a voluntary cessation, and intensities that elicit excessive training stimulus beyond that originally intended (for more details refer to a previously published review [8]; 'Submaximal Fitness Tests Definition')
	Duration	The duration was \leq 15 min of <i>exercise</i> for all SMFT, while <i>exercise</i> refers to the time when HRex was monitored. Therefore, the duration of rest intervals was excluded (for example, recovery between sets)
8	Outcome Measure	HRex (expressed as %HR _{max})

Description or reasons for exclusion

HRex exercise heart rate, %HR_{max} percentage of HR maximum, SMFT submaximal fitness tests



Fig. 1 PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow chart. The total number of studies included in the systematic review and meta-analysis was 30. 10 and 9 studies reported reliability and convergent validity effect estimates, respectively, with an additional 11 studies reporting effect estimates of both measurement properties. *n* number of studies, *WoS* Web of Science

capacity, albeit we acknowledge other, non-aerobic qualities (anaerobic, neuromuscular) may contribute towards performance result [2, 21].

Data Extraction and Coding

We extracted and coded the following study characteristics (Tables 2 and 3):

meta-analysis
he reliability i
included in t
if the studies
aracteristics o
and SMFT ch
2 Sample a
Table

Study	Year	Samp	le Characteri	stics			SMFT						Multiple Effects	RoB (L/U/H)	Reliab Prope	ility ties
		z	Mean Age	Sex	Sport	Level	Category	Protocol	Environment	Exe Dur. (min)	Exe Int. (HRex)	Analysis			L DM	UC E
Veugelers et al. [46]	2016	25	23.0	Male	ARF	Elite	=	Mod Yo-YoIR2	Indoor AM	2-4	74.4–82.4	30 s	Vol./Int Test-retest	5/1/0	`` ``	(A) 🖌 (3,1)
Scott et al. [3]	2018	25	23.2	Male	Rugby	Elite	CF	Ind. Shuttle	Outdoor AM 20–24 °C	4	81.1	30 s		6/0/0	``	(A) 🖌 (3,1)
Buchheit et al. [34]	2010	15	13.9 15.7	Male	Soccer	Elite	CF	Track	Outdoor 21.6–23 °C	5	81.8 76.0	30 s	Sub-groups	3/1/2	>	
Bradley et al. [35]	2011	32		Male	Soccer	Elite	=	Yo-YolR2	Indoor	Q	NS	Fixed		1/2/3		🗸 (_{NS)}
Buchheit et al. [<mark>37</mark>]	2011	12	17.0	Male	lce Hockey	Elite	=	30-15 _{IFT}	Indoor	~3&6	74.6 & 88.6	20 s	Vol./Int	4/1/1	``	(A) 🖌 (3,1)
Fanchini et al. [7]	2014	24	17.0	Male	Soccer	Sub-elite	=	Yo-YolR1	NS	9~	91.3	Fixed		3/1/2	``	(A) 🖌 (2,1)
Hulin et al. [6] ^a	2019	17	24.1	Male	Rugby	Elite	=	Yo-YolR1	Outdoor 19.0 ± 2.6 °C 55 ± 13% RH	₹ 4	89.0 & 91.5	10 s	Test-retest	3/2/1	``````````````````````````````````````	(A) 🖌 (3,1)
Hulka et al. [83]	2015	25	17.7	Male	Soccer	NS	≥	4v4 SSG	Indoor	12	79.7–81.8	AII		3/3/0	``	(A) 🗸 (NS)
Hulse et al. [26]	2013	13 15 18	U9-11 U12-14 U15-18	Male	Soccer	NS	=	MSFT	Indoor PM 10.5–12.4 °C	~ 5.5	95.7 88.7 84.3	~ 60 s	Sub-groups	4/1/1	``````````````````````````````````````	(B) 🖌 (2,1)
Ingebrigt- sen et al. [28]	2014	10-37	. 22.0	Male	Soccer	Mixed	=	Yo-YolR 1/2	Indoor	2,4&6	NS	Fixed	Vol./Int	1/4/1		✓ _(3,1)
Rabbani et al. [53]	2018	12	25.1	Male	Soccer	Elite	CF	100 m shuttle	AM 24–26 °C	4	76.7	30 s	Test-retest	5/1/0	``	(A) 🖌 (3,1)
Owen et al. [29]	2017	10	17.8	Male	Soccer	Elite	=	Yo-YolR1	Same time during the day	9	82.1	Fixed		2/4/0	``	(B) 🗸 (2,1)
Buchheit et al. [23]	2008	15	15.6	Male	Handball	Sub-elite	CF	Ind. Track	Indoor 18–22 °C	9	80.5	AII		3/2/1	>	(B) 🖌 (NS)
Deprez et al. [25]	2014	9–27 8–26 4–19	12.5 14.0 16.2	Male	Soccer	Mixed	=	Yo-YolR1	Indoor Similar °C	~ 2.5, 3.5 & 6.5	91.8–97.0 91.5–96.7 91.0–94.6	Fixed	Vol./Int Sub-groups	4/2/0	``	(B) × (3,1)
Owen et al. [82]	2020	23	25.3	Male	Soccer	Elite	≥	5V5 SSG	Outdoor AM	6	87.0	AII		2/2/2	,	(A) 🖌 (3,1)

Study	Year	Sam	iple Characteri	istics			SMFT						Multiple Effects	RoB (L/U/H)	Relia Prop	bility erties
		z	Mean Age	Sex	Sport	Level	Category	Protocol	Environment	Exe Dur. (min)	Exe Int. (HRex)	Analysis			MD	LE ICC
Doncaster et al. [54]	2018	œ	12.9	Male	Soccer	Elite	=	Yo-YolR1	Outdoor 11–13.2 °C	3&6	88.3–93.8	15 s	Vol./Int	5/1/0	>	V _(A) V _(3,1)
									63-72%RH							
Deprez	2015	22	13.9	Male	Soccer	Elite	=	Yo-YolR1	Outdoor PM	~ 6.5	95.1–95.4	Fixed	Sub-groups	5/1/0	>	🖌 _(A) 🖌 _(NS)
et al. [24]		10 4	16.2 18.1						~ 10 °C		91.8–92.8 88.1–90.0		Test-retest			
Dell lacono	2019	10	18.5	Male	Soccer	Elite	\geq	Pass. Drill	Outdoor PM	12	85.1-87.2	AII	Sub-groups	5/1/0	>	🗸 _(A) 🖌 _(3.1)
et al. [<mark>27</mark>] ^b		10	18.7	Male	Soccer	Elite	\geq	5V5 SSG	22±2.5 °C	12	86.2-87.1					
									65 ± 3.8%RH							
Younesi	2021a	16	27.2	Male	Soccer	Elite	≥	3V3-6V6	Outdoor PM	9	80.8-81.8	AII	Protocol	5/1/0	>	✓ (A) ✓ (3,1)
et al. [33]								SSG	25-32 °C				Test-retest			
									54-76%8RH							
Ryan et al. [31]	2019	16	24.6	Male	ARF	Elite	CF	80 m shut- tle	Outdoor	2		30 s		4/2/0	>	🖌 _(A) 🖌 _(3,1)
2								2	∩ <u>,</u> כ.כ∠−0∠							
Scott et al. [36] ^c	2022	17	22.0	Male	Rugby	Elite	G	Ind. Shuttle	21–23 °C	4	84.9	30 s		5/1/0	>	V _(A) V _(3,1)
<i>30-15</i> _{IFT} 30-1 <i>II</i> Intermitter <i>RH</i> relative h Yo-Yo interm variance betu	5 intermin it increme umidity, <i>F</i> ittent recover	ttent fit ental, <i>Ir.</i> <i>RoB</i> Risl overy tr ividuals	ness test, AM mo d. Individualised, c of Bias, SSG sma est level 2, mod. Y and ICC (Eq. 3). It	vrning hc <i>Jut</i> inter ill-sided (<i>o-YolR2</i> CC (_{2,1}) al	burs, <i>ARF</i> Autrasity, <i>IV</i> inter games, <i>TE</i> ty modified (18 nd ICC _(3,1) re	stralian footb rmittent varia pical error of 3 m shuttle) Y ifer to ICC typ	iall, <i>All</i> overall able, <i>L</i> low, <i>m</i> ^f measuremer 'o-Yo intermit	exercise, <i>Fixed</i> f metres, <i>MD</i> mea nt, <i>U</i> unclear, s so tent recovery te	Txed time point(s) an difference, <i>MSF</i> econd, <i>SMFT</i> subr ist level 2. TE _(A) ; TE	. °C temperatu T multistage fi laximal fitness derived from t	re in Celsius, CF tness test, N san test, Vol volume the mean varian	continuous fi) nple size, NS n 2, Yo-YolR1 Yo- ce of individu	ked, <i>Exe</i> . Exercis ot specified, <i>PA</i> Yo intermittent als score chang	e, <i>H</i> high, <i>ICC</i> <i>A</i> afternoon h recovery test le (Eq. 2), TE _(B) :	intraclas ours, <i>Pas</i> level 1,) : TE deriv	s correlation, s passing, fo-YoIR2 ed from the

^b Test-retest reliability results of the first two weeks which met the inclusion criteria of total duration were established using the raw data provided by the authors

^c Test-retest reliability results were established using the raw data provided by the authors

.5
ž
g
Ē
Ľ,
ţ
é
≥
÷
≝
Ś
÷
E.
Ř
5
Š
Ē
8
<u>م</u>
Ĕ
t
.⊆
σ
Ð
9
-
Ĕ
ű
÷
\Box
st
ē
÷
÷
0
8
Ξ.
·£
ē
IJ
Ľ
g
÷
Ē
Ľ,
\geq
5
2
ar
Ð
õ
Ê
ar
Ś
m
-
-
lde

Study	Year	Sam	ple Characte	ristics			SMFT						Multiple	RoB	Performance
		z	Mean age	Sex	Sport	Level	Category	Protocol	Season phase	Exe Dur. (min)	Exe Int. (HRex)	Analysis	ellects		Criterion
Veugelers et al. [46]	2016	38	23.0	Male	ARF	Elite	=	Yo-YolR2 Mod Yo-YolR2	Pre-season	2-8	76.4–92.8	30 s	Vol./Int	5/1/0	Yo-YolR2 (m)
Mohr & Krustrup [45]	2014	172	25.8	Male	Soccer	Sub-elite	=	Yo-YolR1	Pre & in- season	Q	85.9-90.9	Fixed	Repeated Measures	2/3/1	Yo-YolR1 & 2 (m)
Bradley et al. [<mark>35</mark>]	2011	32 14		Male	Soccer	Elite	=	Yo-YolR2	NS	4&6	87.0 & 91.5	Fixed	Vol./Int Sub-groups	1/4/1	Yo-YolR2 (m)
Bradley et al. [74]	2014	28	23.0	Female	Soccer	Elite	=	Yo-YolR2	NS	2 & 4	0.06-0.06	Fixed	Vol./Int	2/4/0	Yo-YolR2 (m)
Fanchini et al. [7]	2014	24	17.0	Male	Soccer	Sub-elite	=	Yo-YolR1	Pre & in- season	9	90.3–93.9	Fixed	Repeated Measures	2/3/1	Yo-YolR1 (m)
Ingebrigt- sen et al. [79]	2012	39 12	20.0 26.0	Male	Soccer	Sub-elite Elite	=	Yo-YolR1/2	In-season	2 & 4	91.1–97.8 80.8	Fixed	Vol./Int Protocol Sub-groups	2/2/2	Yo-YolR1 and 2 (m)
Ingebrigt- sen et al. [28]	2014	57	22.0	Male	Soccer	Mixed	=	Yo-YolR1	In-season	2 & 4	NS	Fixed	Vol./Int Protocol	1/4/1	Yo-YolR1 and 2 (m)
Lignell et al. [47]	2018	18	28.0	Male	lce Hockey	Elite	=	Yo-YolR1	In-season	9	78.9	Fixed		3/3/0	VO ₂ max (ml·min ⁻¹ ·kg ⁻¹)
Rabbani et al. [<mark>53</mark>]	2018	4	26.7	Male	Soccer	Elite	CF	100 m shuttle	In-season	4	76.7	30 s		5/1/0	30-15 _{IFT} (km·h ¹)
Vigh-Larsen et al. [39]	2019	245	21.7	Male	Ice Hockey	Mixed	=	Yo-YolR1	In-season	9	83.3	60 s		3/3/0	Yo-YolR1 (m)
Hulin et al. [6]	2019	32	24.1	Male	Rugby	Elite	=	Yo-YolR1	Pre-season	~4&6	89.8 & 92.8	10 s	Vol./Int	5/1/0	Yo-YolR1 (m)
Francini et al. [40]	2019	68	14.5	Male	Soccer	Sub-elite	= CF	Mognomi's Yo-YolR1	In-season	9	95.4 94.4	60 s Fixed	Protocol	3/3/0	Yo-YolR1 (m)
Pereira et al. [30]	2019	40	11.5	Male	Soccer	NS	ſ	5'–5' 20 m rectangle	NS	2	78.4	30 s		5/1/0	Yo-YolR1 (m)
Buchheit et al. [34]	2010	33	14.7	Male	Soccer	Elite	CF	Track	In-season	2	78.4	30 s		3/3/0	Vam-Eval (km·h ⁻¹)
Owen et al. [82]	2020	23	25.3	Male	Soccer	Elite	≥	5V5 SSG	In-season	6	87.0	All		3/2/1	Yo-YolR1 (m)

Study	Year	Sam	ple Charact	eristics			SMFT						Multiple	RoB	Performance
		z	Mean age	Sex	Sport	Level	Category	Protocol	Season phase	Exe Dur. (min)	Exe Int. (HRex)	Analysis	effects	(L/U/H)	Criterion
Araújo et al.	2019	16	24.0	Male	Soccer	Elite	=	ISRT	In-season	2, 4, 6 and 9	74.0–91.0	Fixed	Vol./Int	4/2/0	ISRT (m)
[78]		14	24.0	Male		Sub-elite					77.0-92.0		Sub-groups		
		16	21.0	Female		Elite					83.0-97.0				
		15	25.0	Female		Sub-elite					88.0-97.0				
Deprez et al.	2014	19	12.5	Male	Soccer	Mixed	=	Yo-YolR1	In-season	~ 6.5	97.0	Fixed	Sub-groups	2/2/2	Yo-YolR1 (m)
[25]		18	14.0								96.7				
		4	16.2								94.6				
Deprez et al.	2015	22	13.9	Male	Soccer	Elite	=	Yo-YolR1	NS	~ 6.5	95.3	Fixed	Sub-groups	4/1/1	Yo-YolR1 (m)
[24]		10	16.2	Male	Soccer	Elite					92.3				
		4	18.1	Male	Soccer	Elite					89.2				
Younesi et al. [<mark>32</mark>] ^a	2021b	16	27.2	Male	Soccer	Elite	≥	3V3 SSG	Pre-season	9	80.1-80.8	All	Repeated Measures	3/2/1	30-15 _{IFT} (km·h ⁻¹)
Scott et al. [36] ^b	2022	12	24.0	Male	Rugby	Elite	CF	Ind. Shuttle	Pre-season	4	83.2	30 s		5/1/0	30-15 _{IFT} (km·h ⁻¹)
<u>30-15_{IFT} 30-15</u> variable, <i>L</i> low	N sample	ent fitr e size,	NS not specifie	Australian f ed, <i>RoB</i> Risł	football, All on k of Bias, U un	verall exercise Iclear, s secon	e, <i>Fixed</i> fixed t. Id, <i>SMFT</i> subm	ime point(s), <i>CF</i> c naximal fitness te	ontinuous fixed sst, <i>Vol</i> volume,	d, <i>H</i> high, <i>ll</i> Inter <i>Yo-YolR1</i> Yo-Yo i	mittent increm ntermittent rec	ental, <i>Int</i> inte overy test lev	nsity, <i>ISRT</i> inter rel 1, Yo-YolR2 Yo	val shuttle run o-Yo intermitte	test, // intermittent nt recovery test

level 2, SSG small-sided games

^a Correlation coefficient of each independent session were provided by the authors

 $^{\mathrm{b}}$ Exercise intensity value was established using the raw data provided by the authors

Table 3 (continued)

- Cohort mean age (years) and age category (youth [<18 years] or senior [≥18 years]).
- 3. Sex.
- 4. Sport.
- 5. Competition level (elite, sub-elite and mixed). Based on the information provided by the authors, elite players compete in the highest level of their age group in the country. Sub-elite players do not compete in the highest level of their age group, whereas mixed represents a group of players from both levels.
- 6. SMFT protocol category. We earlier proposed an operational taxonomy for unique SMFT protocols based on two levels of classification, including the exercise regimen (continuous or intermittent) and rate of changes in exercise intensity (fixed, incremental or variable). Figure 2 presents the characteristics of protocols and their utilisation in the included studies.
- 7. SMFT duration (minutes). Based upon the details provided in the text, the total duration was rounded to the nearest 30 s.
- SMFT internal exercise intensity, expressed as mean HRex. In our analyses, we considered HRex values analysed as percentage points of heart rate maximum (%HR_{max}). Accordingly, when a study reported statistical results for both beats·min⁻¹

and %HR_{max}, we selected the results presented in %HRmax only. In addition, if a study reported findings solely in beats·min⁻¹, we used additional information to convert the results into %HR_{max} (see "Handling Missing Data" section for details).

- 9. SMFT HRex collection methodology (fixed time point, mean range, or mean overall). HRex was recorded at a specific fixed time point(s) throughout the test, calculated as the mean HRex during the last during 10–60 s of the test, or calculated as the mean HRex during the overall test.
- 10. Reliability statistics: methodology used to calculate TE and ICC type.
- 11. Convergent validity (*r*): endurance performance indicator and season phase (pre-season or in-season period).

Handling Missing Data

We used direct contact details of the corresponding author(s), along with social networks (e.g., ResearchGate, Twitter) to attain missing information including: (1) sample characteristics (e.g., mean age, sample size); (2) SMFT characteristics (duration, exercise or drill/game configuration, etc.); (3) outcome measures, including HRex values recorded during the SMFT, or sample HR_{max;} and (4) other relevant reliability and/or convergent validity statistical calculation methods or results. Overall, we

SMFT Category	Description	Illustration	References
Continuous-Fixed	SMFT involving continuous exercises performed at a fixed intensity that remains constant for the entire test		3 23 30 31 34 36 40 53
	Example: 4 minutes running at 12 km·h ⁻¹		0,20,00,01,04,00,40,00
Continuous-Incremental	SMFT characterised by a progressive increase in exercise intensity within (single) or between (multiple) exercise bouts, with each bout lasting for several minutes.		NA
	<u>Examples</u> : 1) one running bout for 1, 1, and 3 minutes at 9, 10 and 12 km h^{-1} , respectively, or 2) 3 sets \times 3 minutes bouts at 10, 11 and 12 km h^{-1}		-
Intermittent-Fixed	SMFT including reoccurring activities performed at a constant intensity and rest intervals	11111	
	Example: 4 running bouts × 50 meters at 18 km·h ⁻¹ , separated by 30 seconds of recovery		NA
Intermittent-Incremental	SMFT that predominantly involves fixed rest periods, while exercise intensity is increased between activity bouts		6,7,24,25,26,28,29,35,37,39,
	<u>Example</u> : shorter version (e.g., 6 minutes) of the 30–15 $_{\rm IFT}$ or Yo-YoIR1		40,45,46,47,54,74,78,79
Intermittent-Variable	Drill-based: may include specific and non-specific standardised drills which do not involve competition features		27
	Example: passing drills formats of the sport	- Malah Malah Mahah	
	Game-based: may include specific and non-specific standardised games which include competition features		27,32,33,82,83
	Example: small-sided games of the sport		

Fig. 2 SMFT protocol categories and their utilisation in the included studies. NA not available

contacted 12 authors (15 studies) [3, 6, 23–35] for additional information, and 9 (12 studies) of them replied [3, 6, 23–27, 29, 31–34], with most providing explanations [23–25, 29, 31, 32, 34] or the entire raw data [3, 6, 27, 36]. The remaining information was either excluded from pooled effect estimates analyses, obtained from alternative details provided within the text, or omitted when a meta-regression analysis was performed. All data estimations employed were considered in the qualitative assessment of individual studies (see "Study Qualitative Assessment" section.

In some cases, the original sample size reported in the study was modified when reliability or convergent validity experiments were administered. When the new characteristics (as an example, mean age) relating to the modified sample were not reported, the original information provided was used. With regards to the reliability dataset, two studies [7, 37] that collected SMFT HRex in beats min⁻¹ also reported HR maximum derived from the maximal version of the same test, meaning the conversion into relative values was straight-forward. One study [26] analysed SMFT HRex similarly (beats \cdot min⁻¹) and did not report HR_{max}. In addition, ages were reported as U9–U11, U12–U14 and U15–U18. In this case, $\mathrm{HR}_{\mathrm{max}}$ was calculated using an estimated equation $(HR_{max} = 208)$ $-0.7 \times$ mean age [38]), and mean ages were assumed as 9, 12 and 16 years, respectively. Similar HR_{max} estimation was employed in two previous studies [29, 33]. Graph digitizer software (WebPlotDigitizer, https://apps.autom eris.io/wpd/) was used to extract the first two SMFT HRex group means and standard deviations of a testretest study that administered repeated measures [34].

With respect to the convergent validity dataset, in two studies [7, 30] using SMFT HRex as beats·min⁻¹, one [7] reported HR_{max} , and for the other [30] we estimated HR_{max} as described above. One study [35] reported age as youth U19, the estimated mean age was 17, and the same graph digitizer software was used to extract the mean SMFT exercise intensity (HRex). Further, three studies [34, 39, 40] combined sub-groups when convergent validity was analysed. In that case, the weighted mean SMFT HRex and age of the sub-groups were calculated for exercise intensity and mean age, respectively.

Effect Estimates Mean Difference

The Mean Difference (MD) is an absolute reliability estimate which indicates the degree of variability relating to systematic, group-level bias (i.e., total sample) [15]. As discussed in previous sections, we considered MD effect estimates presented raw units (%HR_{max}), which were calculated as the change in group means across test–retest [41].

Typical Error of Measurement

Another absolute reliability estimate is Typical Error of Measurement (TE) [15, 16]. The TE is an indicator of random, within-subject variability and is sometimes presented in the form of coefficient of variation (CV%) [15]. This approach is appealing for normalization and interpretation of variability, especially across different outcome measures. However, since the primary outcome measure in our meta-analysis is already analysed in relative values ($%HR_{max}$), it is not appropriate (or interpretable) to present TE as a CV (i.e., percentage of percentage). Therefore, when a study presented the TE as a CV value, we estimated the former using the following equation:

$$\left(\frac{\text{CV}}{100}\right) \times \text{grand mean HRex}$$
 (1)

where grand mean is the mean HRex across test 1 and test 2.

Two common methods for calculating TE are cited in the sport science literature [15, 16], and were identified in our review. The first approach proposed by Atkinson and Nevill [16] uses the ICC in combination with the pooled between-athlete variance (Eq. 2). The second method proposed by Hopkins [42] uses the mean variance of within-athlete change scores (Eq. 3). Since the former approach requires the sample standard deviation and ICC, it is usually easier to extract from an article. However, after performing different simulations with our previous (unpublished) work and the available data in our included studies, and generating fictitious data with similar values, we found that the difference in TE calculated using both methods negligible to almost non-existent (available in the Additional file 3), suggesting that it is appropriate to meta-analyse estimates calculated from either method without any transformation. If the exact method was not specified nor provided after personal communication to the author, or TE was not disclosed in the analysis, we used the Atkinson and Nevill [16] method (i.e., Eq. 2), where possible [35].

$$\Gamma E = SD \times \sqrt{1 - ICC}$$
⁽²⁾

$$TE = \frac{SD_{\Delta}}{\sqrt{2}}$$
(3)

where SD_{Δ} is the standard deviation of individual change scores (test 2–test 1% point) and SD is the pooled between-athlete standard deviation observed in test 1 and test 2.

Intraclass Correlation Coefficient

The ICC represents reproducibility in the rank order of athletes over repeated measures (i.e., relative reliability) [43]. In most of our eligible studies, two types of ICC were identified: those describing absolute agreement in a single measure from a two-way random effects model $(ICC_{2,1})$, and those describing consistency in a single measure from a two-way mixed-effects model $(ICC_{3,1})$ [43], whereas in other studies ICC type was not clearly specified or attainable (Table 2). Based on the comparisons used in the TE data synthesis, the differences between ICC types were negligible (Additional file 3), and we thereby treated all data the same. We acknowledge that there are conceptual differences between ICC types [43]. However, it was not possible to examine this effect due to the low number of estimates in some levels (e.g., ICC_{2.1}, n = 3 studies).

Pearson's Correlation Coefficient (r)

We elected to extract Pearson's product-moment correlation coefficient (*r*) as the principal statistic estimate to describe the magnitude of association between SMFT HRex and maximal endurance performance [44]. As demonstrated in Table 3, most reference tests included fieldbased, intermittent incremental protocols with change of direction (COD) such as the Yo-Yo Intermittent Recovery Test Level 1 (Yo-YoIR1) [45] or 2 (Yo-YoIR2) [46], while other individual studies involved continuous incremental protocols, including the Vam-Eval test [34] and maximal oxygen uptake in laboratory conditions [47].

Statistical Methods and Data Analysis Synthesis of Effect Estimates Prior and Post Analysis

An overview of calculations for effect estimates and their sampling variance is presented in Table 4. The MD estimate and its sampling variance were calculated from the test–retest statistics of sample size (n), group means and standard deviations while assuming equal sampling variances within groups [41]. The ICC and r estimates and their sampling variances were converted to Fisher's z values to approximate normally distributed data [41, 48], and subsequently back-transformed for analyses interpretations of the pooled estimates (intercept-only model). To our knowledge, there is no documented approach for a meta-analysis of TE. Therefore, we used the method proposed by Nakagawa et al. [49] for meta-analysis of variation. This method uses a log-transformed standard deviation (adjusted for sample size), along with its sampling variance for each individual group [49]. For obtaining the pooled estimate, the results were back-converted to their original values, including the bias correction for sample size. It is important to note that considering that the transformation of meta-regression coefficients (fullmodel) back into their original values is not straight-forward and may result in erroneous interpretations of the modifying effects, coefficients are presented in Fisher's z-scale (ICC and r) and log-transformed (TE) values. To facilitate models' interpretations, we constructed metaregression bubble plots, presenting the predicted values and their corresponding confidence and prediction interval bounds.

Overall Meta-Analysis

All data analyses were conducted using the *metafor* [50] and *clubSandwich* [51] packages in the R studio environment (version 1.4.1106) [52]. The datasets and analysis codes are openly available in the Additional files (https://osf.io/mqnt9/). In most of the included studies, we were able to extract more than a single effect size. Multiple effect sizes were derived from different sub-groups within studies, but more frequently from a variety of SMFT characteristics (e.g., SMFT protocols, duration or intensities) within sub-groups. In addition, some of the reliability studies [6, 24, 46, 53, 54] included more than two

Primary outcome	Effect estimate		Sampling variance
MD	Raw values	$MD = M_1 - M_2$	$V_{\rm MD} = \frac{n_1 + n_2}{n_1 n_2} S_{\rm pooled}^2$ where
			$S_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_1 - 1)S_2^2}{n_1 + n_2 - 2}}$
TE	Log-transformed	$TE = \ln s + \frac{1}{2(n-1)}$	$V_{TE} = 1/2(n-1)$
ICC	Fisher's z	$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right)$	$V_{icc} = 1/(n - 3/2)$
r	Fisher's z	$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right)$	$V_r = 1/(n-3)$

 Table 4
 Effect estimates and their sampling variances

ICC intraclass correlation coefficient, *In* natural logarithm (log-transformed), *M*₁ group mean test 1, *M*₂ group mean test 2, *MD* mean difference, *n*₁ and *n*₂ sample size, *r* correlation, *TE* typical error of measurement, s *TE*, *S*₁ group standard deviation test 1, *S*₂ group standard deviation test 2, *V*_{TE} typical error of measurement, s *TE*, *S*₁ group standard deviation test 1, *S*₂ group standard deviation test 2, *V*_{TE} typical error of measurement sampling variance, *V*_{icc} intraclass correlation sampling variance; *V*, correlation coefficient sampling variance; *z* Fisher's *z*. ICC sampling variance was adjusted following previous recommendations (48)

similar consecutive SMFT and provided the results from each matched test-retest. Two studies examining convergent validity [7, 45] implemented within-group repeated measures of the same tests at different time-points across the season (Table 2 and 3 provide the sources of multiple effect estimates within studies).

Given the hierarchical structure in our datasets (multiple effect estimates are nested within groups), as well as the likelihood of statistical dependency, we employed a more recently developed approach using multilevel mixed-effects meta-analysis and robust variance estimation with adjustments for small-samples [55]. This approach allows exploration of the variance present across multiple levels and, hence, within- and betweengroup variance [56], and provides a robust method for meta-analysis while accounting for the dependency among effect estimates derived from common samples [57]. In such cases, replacing sampling variance with the entire 'V matrix', indicating the variance-covariance matrix of the estimates, can further account for the correlation between effect estimates [55, 58]. As it was not possible to attain the correlation between effect estimates drawn from the same participants in most of the included studies, our analyses were conducted using an assumed constant correlation of $\rho = 0.6$ [55]. In Additional file 4: Table S3 we report sensitivity analyses, whereby a range of correlation values were used to evaluate the influence of the changes in the within-group covariance on the pooled estimates and its variance components. Collectively, these analyses showed identical pooled estimates and nearly similar variance components.

Heterogeneity and Modifying Effects

To describe the extent of heterogeneity, we calculated restricted maximum likelihood estimates of the withinand between-group variances (SD; tau $[\tau]$) [59], as well as the I^2 statistic in each level [60] which implies the percentages of variance which are due to study heterogeneity rather than sampling error [60]. Log-likelihoodratio tests were also computed to determine whether the within- and between-group variances were significant [61]. We then conducted regression analyses to examine possible source of heterogeneity, including parameters related to SMFT (e.g., exercise intensity) or athlete's (e.g., mean age) characteristics. Modifying effects with at least 6 independent samples and overall 10 effect estimates in each level [62] were added separately to the models. These were analysed independently, rather than in combination, due to a wide range of sample sizes for different modifying effects that resulted in a relatively low number of estimates for the requirement of meta-regression [62].

Categorical effects were evaluated as the differences between levels and consisted of athlete's age category (youth, senior), HRex collection method (reliability: fixed, mean range, mean overall; convergent validity: fixed, mean range) and competition level (elite, sub-elite; for convergent validity only). Continuous effects included athlete's mean age (years), SMFT exercise intensity (expressed as HRex) and SMFT duration (minutes), and were explored as the changes associated with pre-defined clinically relevant values: (1) for mean age, we explored the effect of 3-year change in age; (2) SMFT exercise intensity effect was evaluated as the changes associated with 5-point % change in HRex; and (3) the effect of SMFT duration was investigated in reference to 2-min change in duration. Finally, model strength was assessed as the percentage of variance explained by the modifying effect with pseudo- R^2 statistic (i.e., intercept-only versus full-model) [41]. When variances in both levels were identified, we scrutinised R^2 in the within (R^2) and between (R_3^2) group variance separately [56].

Inferences

Uncertainty in meta-analysis and regression estimates was expressed using 95% compatibility (confidence) intervals (CI), representing ranges of values compatible with our models and assumptions [63]. We also derived 95% prediction intervals (PI), which convey the likely range of the true measurement properties in similar future studies. Similar to others [64], we sought to avoid dichotomizing the presence or absence of an effect using traditional null hypothesis testing. Instead, we considered the entire range of 95% CI, relating mostly to the point estimate and uncertainty of predicted values. Effects were then discussed in terms of their compatibility (or coverage) with practically significant or practically equivalent values. Qualitative inferences were made for standardised effects only (i.e., ICC and r) using thresholds of > 0.99, extremely high; 0.90-0.99, very high; 0.75-0.90, high; 0.50–0.75, moderate; 0.20–0.50, low; < 0.20, very low for ICC [65] and 0.10 small; 0.30 moderate; 0.50 large; 0.70 very large; and 0.90 extremely large for r [66]. Thresholds for the smallest meaningful correlation (i.e., 0.1 and 0.2 for r and ICC, respectively) were used to declare modifying effects as substantial [44]. We elected to not make formal decisive interpretation on the substantiality of MD and TE because these are presented in units of the outcome measure (i.e., HRex), and we are as yet unaware of the true, minimum practically important difference for team sport athletes during SMFT.

Study Qualitative Assessment

Qualitative assessment of individual studies was conducted using the Risk of Bias Assessment Tool for Nonrandomised Studies (RoBANS) [67, 68]. The RoBANS comprises six-dimension criteria which were adapted to answer questions that may influence the reliability or convergent validity results reported in the included studies. Answer categories of 'low', 'unclear' and 'high' risk of bias were assigned in each domain. To promote consistency in how domains were evaluated, three reviewers (TS, RL, SJM) created a decision rule for each criterion. Assessments were then performed by one reviewer (TS), with two other authors (RL, SJM) checking for accuracy. A summary detailing the quality assessment criteria in each domain is provided as Additional file 5: Table S1.

Small-Study Effects and Influence Analysis

Small-study effects and asymmetry were visually inspected using funnel plots [69]. To confirm our visual impression, the Begg's rank correlation [70] and Egger's regression (by fitting the square root of the sampling variance as a moderator) [69] tests were employed. Further, to assess the influence of each effect size on the summary effect and heterogeneity, Cook's distance analysis [71] and Baujat plots [72] were obtained. A standard rule of thumb is that Cook distance values greater than three times the mean were used [73], and potential outliers were excluded from the dataset to examine if the change in the summary estimates turned substantial.

Results

Selected Reliability Studies and Characteristics

Table 2 summarises the characteristics of included reliability studies. The overall dataset included 29 samples derived from 21 studies. The MD dataset consisted of 69 effect sizes nested within 25 samples (mean = 2.8 [effect sizes per sample], median = 2, range of 1–24), derived from 18 studies. The TE dataset included 52 effect sizes nested within 25 samples (mean = 2.1, median = 1, range of 1–12), from 18 studies. The ICC dataset included 54 effect sizes nested within 27 samples (mean = 2.0, median = 1, range of 1–12), from 20 studies. The total number of team sport athlete inclusions was 406 in MD, 411 in TE and 480 in ICC datasets. The studies involved male players only, and the majority originated from

 Table 5
 Meta-analysis of measurement properties

soccer (76%), followed by rugby codes (10%), Australian football (7%), ice hockey and handball (one study). Elite players were included in 76% of the studies, with a further 16% being mixed and 8% being sub-elite. The proportion of youth and senior athletes was 55% and 45%, respectively. SMFT categories were distributed as follows: intermittent incremental (59%), continuous fixed (24%) and intermittent variable (17%).

Selected Convergent Validity Studies and Characteristics

Table 3 summarises the characteristics of our included convergent validity studies. The overall dataset included 73 effect sizes nested within 29 samples (mean=2.5, median = 2, range of 1-10), derived from 20 studies, with a total number of 1055 team sport athlete inclusions. Most studies (90%) involved male players, and the majority originated from soccer (83%), ice hockey and rugby codes (6.5% each) and Australian football (one study). Elite players were included in 61% of the studies, with a further 21% being sub-elite and 18% being mixed. Unlike reliability, 62% and 38% of studies were conducted on senior and youth athletes, respectively. Most (73%) studies were conducted during the in-season phase, and the distribution of SMFT categories was similar: intermittent incremental modality (76%), continuous fixed (17%) and intermittent variable (7%).

Overall Meta-Analyses

Table 5 provides a summary of the reliability and convergent validity meta-analyses, with forest plots of the weighted points estimates, 95% CI and PI presented in Additional file 6: Fig. S1–4. The pooled estimates derived from the reliability analyses indicated overall good absolute and high relative reliability. The pooled estimate derived from r dataset indicated an inverse, large relationship between SMFT HRex and endurance performance measures.

Heterogeneity

We found no evidence for heterogeneity in the MD metaanalysis (τ_2 =0.09 [95% CI 0.00 to 0.49], τ_3 =0.00 [95%

Measurement property	Statistic	Number of	f	Pooled Effe	ct	
		Studies	Estimates	Estimate	95%Cl (lower to upper)	95%PI (lower to upper)
Reliability	MD	25	69	0.5	0.1 to 0.9	0.1 to 0.9
	TE	25	52	1.6	1.4 to 1.9	0.9 to 3.0
	ICC	27	54	0.88	0.84 to 0.91	0.59 to 0.97
Convergent validity	r	29	73	-0.58	– 0.62 to — 0.54	- 0.77 to - 0.31

CI confidence intervals, ICC intraclass correlation coefficient, MD mean difference, PI prediction intervals, TE typical error of measurement, r correlation coefficient

CI 0.00 to 0.72], $I_2^2 = 0.5\%$ [95% CI 0.0% to 14.2%] and $I_3^2 = 0\%$ [95% CI 0.0% to 26.4%]) and therefore did not employ meta-regression analysis. In contrast, notable heterogeneity was observed in all other reliability estimates: TE ($\tau_2 = 0.23$ [95% CI 0.15 to 0.32], $\tau_3 = 0.18$ [95% CI 0.00 to 0.34], $I_2^2 = 43.4\%$ [95% CI 42.8% to 76.8%], $I_3^2 = 28.4\%$ [95% CI 0.0% to 78.3%]), and ICC ($\tau_2 = 0.26$ [95% CI 0.18 to 0.36], $\tau_3 = 0.20$ [95% CI 0.00 to 0.40], $I_2^2 = 39.4\%$ [95% CI 35.0% to 68.3%], $I_3^2 = 24.4\%$ [95% CI 0.0% to 72.5%]). A greater source of within- versus between-group variance was evident in both TE and ICC. Heterogeneity in *r* was present in the within-group level only ($\tau_2 = 0.16$ [95% CI 0.11 to 0.21], $\tau_3 = 0.00$ [95% CI 0.0% to 42.6%]).

Meta-Regression Analysis

Figures 3, 4 and 5 illustrate meta-regression bubble plots of the modifying effects, with different colours indicative of SMFT category. The full meta-regression results are presented in Table S2 in Additional file 4. The TE regression coefficients ranged between $\ln_{diff} = -0.27$ to 0.06, which were approximately equal to differences of 0.01 to 0.45 when comparing between TE predicted values (Fig. 3, panel A–E). The ICC regression coefficients ranged between $z_{diff} = -0.10$ to 0.16 and resulted in approximate differences of 0.00 to 0.03 between ICC predicted values (Fig. 4, panel A–E). The *r* regression coefficients ranged between $z_{diff} = -0.08$ to 0.10, indicating approximate correlation differences of -0.07 to -0.02 between *r* predicted values (Fig. 5, panel A–F). Collectively, none of these modifying effects appeared to have a meaningful influence on the reliability or convergent validity estimates. There was, however, a considerable reduction in the overall heterogeneity owing to the inclusion of some of the variables.

Study Qualitative Assessment

Risk of bias assessment is demonstrated in Tables 2 and 3, including the sum of each risk category (low, unclear or high) in individual articles. A graphical overview of the



Fig. 3 Mixed-effects meta-regression of the changes in Intraclass Typical Error of measurement (TE) while controlling for the effects of mean age (A); SMFT exercise intensity (HRex) (B); SMFT duration (C); age category (senior vs. youth) (D); and HRex collection method (fixed vs. mean range vs. mean overall) (E). Data points represent individual effect estimates included in our meta-analysis, and the size of the data point is proportional to their weighting. Green, red, blue and purple points represent studies using continuous fixed, intermittent incremental, intermittent variable small-sided games, and intermittent variable passing drills category, respectively. Solid lines represent the estimate of the modifying effect. Dashed and dotted lines represent 95% confidence and prediction limits, respectively. *HRex* exercise heart rate, *y* years, *min* minutes



Fig. 4 Mixed-effects meta-regression of the changes in Intraclass Correlation Coefficient (ICC) while controlling for the effects of mean age (A); SMFT exercise intensity (HRex) (B); SMFT duration (C); age category (senior vs. youth) (D); and HRex collection method (fixed vs. mean range vs. mean overall) (E). Data points represent individual effect estimates included in our meta-analysis, and the size of the data point is proportional to their weighting. Green, red, blue and purple points represent studies using continuous fixed, intermittent incremental, intermittent variable small-sided games, and intermittent variable passing drills category, respectively. Solid lines represent the estimate of the modifying effect. Dashed and dotted lines represent 95% confidence and prediction limits, respectively. *HRex* exercise heart rate, *y* years, *min* minutes

overall risk in each domain from all studies is illustrated in Fig. 6.

Small-Study Effects

No indications for small-study effects were observed in the funnel plots (Fig. 7). Begg's rank correlation and Egger's regression tests revealed no significant asymmetries for all estimates. Overall, the exclusion of potential outliers in all datasets did not have a practically meaningful influence on the results obtained in the original models (Additional file 7).

Discussion

In this review, we sought to meta-analyse the measurement properties of SMFT HRex, while additionally examining the modifying effects of athlete and protocol characteristics. Reliability and convergent validity data were obtained from 1218 and 3032 individual observations, drawn from 522 and 1055 team sport athlete inclusions and nested within 29 and 29 independent samples, respectively. We found that the overall TE of SMFT HRex was 1.6% points, with 95% CI ranging from 1.4 to 1.9% points. The width of the 95% PI suggested that future studies may expect to observe TE ranges between 0.9 to 3.0% points. The ICC overall estimate and 95% CI (0.88 [95% CI 0.84 to 0.91]) indicated high to very high relative reliability, and 95% PI ranged from 0.59 to 0.97, implying moderate to very high magnitudes in future studies. We also observed a large inverse relationship (r = -0.58) between SMFT HRex and endurance test results, suggesting its validity as a marker of maximal endurance performance when comparing between athletes. Correlation magnitudes remained consistent across the 95% CI (-0.54 to -0.62, large), whereas the interpretation of 95% PI suggested that future studies are likely to observe moderate to very large magnitudes (-0.31 to -0.77). Meta-regression analyses suggested that variables related to athlete or protocol characteristics do not have a meaningful effect on these measurement properties.

Our results support previous conclusions [24-26, 34, 74] suggesting no effect of age on the magnitude of SMFT HRex reliability and convergent validity. Although a trend for slightly improved ICC, TE and *r* was observed



Fig. 5 Mixed-effects meta-regression of the changes in Correlation Coefficient (*r*) while controlling for the effects of mean age (**A**); SMFT exercise intensity (HRex) (**B**); SMFT duration (**C**); age category (senior vs. youth) (**D**); HRex collection method (fixed vs. mean range) (**E**); and level (elite vs. sub-elite) (**F**). Data points represent individual effect estimates included in our meta-analysis, and the size of the data point is proportional to their weighting. Green, red and blue points represent studies using continuous fixed, intermittent incremental and intermittent variable modality, respectively. Solid lines represent the estimate of the modifying effect. Dashed and dotted lines represent 95% confidence and prediction limits, respectively. *HRex* exercise heart rate, y years, *min* minutes

in older players (Fig. 3–5, panel A and D), the differences were not substantial (Table S2 in Additional file 4). Overall, these results imply that the high-degree of SMFT HRex reliability and convergent validity can be confidently applied among a wide range of age groups, at least bounded by the ages we examined. In this respect, when comparisons were made directly between age extremities of the data (i.e., 10 vs 30 years), there did appear to be meaningful improvements in measurement properties. However, this comparison (e.g., 20 years difference) is likely not relevant in practice, as it spans across a team sport career.

We attempted to enhance the understanding of athlete's performance status by categorising the studies in reference to their competition level (elite versus subelite), although this was only possible for the convergent validity dataset. While the relationship between SMFT HRex and established endurance tests appeared slightly larger in athletes who compete in the highest level of their sport compared to athletes who compete in subdivisions (Fig. 5, panel F), the difference was negligible (sub-elite to elite: r = -0.56 versus -0.62, $z_{diff} = -0.08$ [95% CI -0.26 to 0.09]). Additionally, while this effect could not be explored in the reliability datasets, the results attained in the individual reliability studies did not indicate any potential trend [25, 28]. These observations would further suggest that athlete characteristics do not appear to meaningfully affect SMFT HRex measurement properties.

We examined the modifying effects of a variety of SMFT characteristics, including the SMFT duration, intensity, and HRex collection method. Collectively, all these modifying effects were trivial or had no clearly meaningful effect. Nonetheless, these findings are practically useful to guide SMFT programming factors in research and practice. Besides being non-exhaustive, another advantage of SMFT is their short duration. The datasets in our meta-analysis included SMFT ranging between 2 and 12 min. Several previous reviews dealing with monitoring HRex [4, 9], together with individual studies investigating the effect of SMFT duration in our meta-analysis [25, 37, 46, 54], have suggested a duration of 3–4 min considered the minimum for HR to attain steady-state during submaximal exercise [4, 75]. In

А



Fig. 6 Distribution of Risk of Bias for reliability (A) and convergent validity (B) included studies

accordance with this physiological rationale, we found that a 4-min SMFT is adequate for obtaining comparable reliability and convergent validity results, with no further meaningful improvements beyond.

Although part of our included studies [25, 46], taken together with previous research in endurance athletes [76, 77], have highlighted an improved absolute reliability (i.e., reduced TE) at higher SMFT intensities, one of the key findings of the current meta-analysis was that SMFT exercise intensity (expressed as HRex) had no effect upon absolute and relative reliability estimates (Figs. 3 and 4, panel B). Indeed, we found somewhat improved estimates of absolute reliability (TE) in lower versus higher intensities, but these differences were clearly trivial $(ln_{diff} = 0.06)$ [95% CI -0.02 to 0.15] for every 5-point % increase in exercise intensity). For example, computing the predicted TE estimates at 80% and 85% HR_{max} indicated 1.48 (95% CI 1.24 to 1.76) and 1.57 (95% CI 1.37 to 1.80) % points, respectively. Likewise, the predicted ICC values for the same intensities were 0.91 (95% CI 0.88 to 0.94) and 0.89 (95% CI 0.86 to 0.92), respectively. Therefore, while marginal differences are evident, we deem them to have no practical substantiality and might therefore be considered trivial.

While higher SMFT HRex values are hypothetically expected to strengthen the relationships with fitness test performance outcomes, because they are collected at intensities closer to exercise cessation, studies examining this effect have reported disparate findings. For example, some of the included studies reported progressive improvements in the relationships with higher SMFT HRex values [35, 74, 78], while others showed the opposite [46, 79] or no differences (comparable correlation magnitudes) [6]. Our meta-regression analysis demonstrated that this effect was mostly compatible with no meaningful change in the magnitude of $r (z_{\text{diff}} = -0.05)$ [95% CI - 0.16 to 0.05] for every 5-point % increase in HRex; Fig. 5, panel B). To illustrate, predicted estimates for SMFT intensities at 80% and 85% $\mathrm{HR}_{\mathrm{max}}$ are -0.52(95% CI - 0.64 to - 0.38) and -0.56 (95% CI - 0.62 to-0.49), respectively, with only the lower CI bounds indicating a possibly small effect. These findings have practical importance and suggest that practitioners do not necessarily require SMFT that elicit higher intensities for adequate HRex reliability and convergent validity. This may be of particular relevance considering SMFT are often administered as part of the warm-up of a training



Fig. 7 Funnel plots for small-study effects with confidence levels of 90% (white), 95% (dark grey) and 99% (light grey): **A** Intraclass Correlation Coefficient (ICC); **B**: Mean Difference (MD); **C**: Typical Error of measurement (TE); and (**D**): Correlation Coefficient (*r*). Egger's regression test: **A** F = 0.00, p = 0.95; **B** F = 0.03, p = 0.87; **C** F = 0.14, p = 0.73; and **D** F = 0.01, p = 0.91. Begg's rank correlation test: **A** Kendall's tau = 0.11, p = 0.25; **B** Kendall's tau = -0.05, p = 0.54; **C** Kendall's tau = -0.10, p = 0.29; and **D** Kendall's tau = -0.05, p = 0.57

[3, 6, 34, 53], in which an exposure to high intensities maybe considered impractical.

The studies analysed herein used three main collection methods: fixed time point(s) during the test, the mean HR over particular time frames before test cessation (e.g., last 30, 60 s), or mean HR during the overall test (particularly during intermittent variable protocols such as passing drills or small-sided games (SSG); refer to Tables 2 and 3). While using the mean HR during the overall test in intermittent variable (e.g., SSG) SMFT seems a reasonable approach, the rationale for using a particular method and/or time span in other SMFT protocols was not provided in the included studies. While we observed trivial differences between methods (Figs. 3, 4, 5, panel E), mean HR over a particular time-frame before the SMFT cessation maybe deemed preferable [4, 9] since it minimises a potential measurement noise such as HR spikes owing to signal error. In view of the above, it should be noted that in some of the studies [28, 45, 78, 79] which were categorised as 'fixed' collection method, HRex was recorded in 2-5 s intervals by default. Moreover, while the collection method used could, in theory, alter HRex values when different intermittent SMFT are administered (see discussion on SMFT category in the following paragraphs), future research may be necessary to elucidate whether different collection approaches should be selected based upon the SMFT category and subsequently the protocol used.

While the present review could not examine the effect of SMFT category, given the small number of independent studies adopting protocols other than those classified as an intermittent incremental, a theoretical discussion is warranted. A fundamental assumption inherent to the use of HRex as a key monitoring measure in SMFT is based on its ability to provide a valid marker of withinathlete relative exercise intensity due to its linear relationship with oxygen uptake at a wide spectrum of submaximal intensities [4, 9]. From a physiological perspective, an essential component of this assumption is that the exercise regimen should be continuous (intends to elicit steady-state) [4, 80]. Interestingly, most of our included studies administered SMFT protocols characterised as intermittent (incremental), including shorter versions of the most commonly used intermittent shuttle fitness tests such as Yo-YoIR1 and 30-15_{IFT}. We assume that these particular SMFT were chosen since most of the studies investigated both reliability and convergent validity and thereby match the characteristics of the submaximal and maximal tests or administer them simultaneously. Furthermore, due to the absence of a criterion-standard of evidence-based SMFT protocol(s) [8], it is perhaps more logical to administer shorter versions of established field tests. However, considering the nature of such assessments, which include intermittent running bouts separated by rest periods, the final HRex result is not entirely derived from steady-state exercise and may be influenced by factors related to the reactivation of the cardiac parasympathetic system such as heart rate recovery following each exercise bout. In addition, the multiple CODs required during the test may add a neuromuscular component which may influence the cardiovascular response [75].

To illustrate this concept, one of the most investigated SMFT is the 6-min Yo-YoIR1 [8]. This test requires repeated 2×20 m incremental running bouts (shuttles with 180° CODs), interceded by 10 s rest periods [2]. At 6-min of the test, the players are required to run at a mean velocity of 14.5 km h^{-1} (~10 s for each running bout and work/rest ratio of 1:1). Assuming that HRex is collected in the last 30 s of the test, the final value includes the mean HRex during 20 s of activity and 10 s of recovery. Another example is the $30-15_{IET}$, which includes a fixed work/rest ratio of 2:1, and a longer absolute recovery period between running bouts (15 s) [21]. Assuming a similar volume (6 min) and collection method (mean HRex during the last 30 s), the final HRex value would be derived from 15 s of activity and recovery, respectively. Albeit speculative, this can lead to a less precise and reproducible HRex values considering that heart rate recovery is associated with an inferior degree of reliability [8], and indeed, studies using continuous protocols generally observed better reliability outcomes. Conversely, it could also be argued that intermittent incremental SMFT are preferable given their greater ecological validity to team sports [81] which subsequently can enhance SMFT HRex convergent validity. A primary purpose of future research should therefore include a more explicit examination of continuous versus intermittent SMFT protocols and their link to SMFT HRex measurement properties.

Another aspect of SMFT categories is the use of intermittent variable protocols (SSG or passing drill formats). Despite the relatively large number of the total reliability estimates, only five independent studies met inclusion criteria (four SSG [27, 33, 82, 83] and one passing drill [27]). Collectively, these studies suggested further exploration of such SMFT as they showed comparable absolute and relative reliability values with the values observed in other SMFT generic categories. Nonetheless, considering the range of contextual factors (e.g., technical and tactical elements, exercise structure) that influence locomotor variability [27, 81, 82] and the increased neuromuscular demands [75], caution is necessary in their interpretation. With regards to convergent validity, only two independent studies administering SSG met the inclusion criteria. Interestingly, while one study [82] among senior professional soccer players using 5v5 SSG observed a significant large relationship (r = -0.56) between HRex and the distance covered in Yo-YoIR1, a different study [32] using 3v3 SSG in a similar cohort observed no association (r = -0.22 to -0.25) with the final velocity achieved in the $30-15_{IFT}$. Therefore, before using SSG as a standardised SMFT protocol, more research addressing its validity is warranted. Further, to appropriately quantify the influence of locomotor variables upon HRex in the longer-term, we recommend the use of linear regression techniques and avoiding ratios (i.e., SMFT HRex divided by an external intensity parameter) [84].

There are several limitations to our meta-analysis that warrant acknowledgment. First, although we adopted strict inclusion criteria, all of the included studies were observational and conducted with less rigorous methodological design and reporting standards (e.g., STROBE). Second, while it is well known that the main outcome measure (HRex) is influenced by a large number of confounding variables [8] (e.g., environmental conditions, training loads, circadian effect, menstrual cycle stage), only some of the included studies (8 of 30 overall studies; 27%) provided an adequate amount of information regarding how all these confounders were controlled prior and during the test(s) (refer to Fig. 6 'confounding variables' and in Additional file 5: Table S1 for a detailed description). This information is important as such factors may have influenced outcomes of the studies included in our analyses, in particular the reliability findings. For this reason, future studies should report these data descriptively and if needed, account for the factors in the analysis (for example, adjusted HRex based on 'heat index' [85]). Third, during the screening process, we identified more potential studies for inclusion. However, since their effect sizes could not be appropriately estimated or were not provided by personal contact, we were not able to include them. With this in mind, we decided to employ a considerably careful evaluation of missing data and selective reporting in the included studies (Fig. 6 and Table S1 in Additional file 5, dimension 5 and 6). As discussed earlier, we also employed different methods to estimate data related to SMFT and/or athlete's characteristics, all of which could have introduced some errors. Finally, our unbalanced datasets with regard to sex and sport limit the generalization of the findings to female athletes and athletes competing in a particular team sport.

Practical Implications

We propose several practical implications for researchers and partitioners wishing to use SMFT HRex with a view to determining athlete's physiological state.

Protocol Selection

Exercise Regimen

While it is as yet unclear whether SMFT protocol category influences different measurement properties of HRex, we recommend the use of continuous protocols until new evidence is available to suggest otherwise. Continuous fixed protocols are likely the easiest to implement into team sport settings, with pitch markings/poles used to demarcate distance and audio cues to direct the running speed (e.g., dictated by a whistle, metronome, pacing tool). When shuttle courses are used, it is suggested to implement longer distances (where applicable) to reduce the number of total CODs and consequently neuromuscular loading.

For example, on a FIFA standard football (soccer) pitch (105 m in length and 68 m in width), performing a 4 min continuous shuttle of pitch width in 20 s repetitions would result in a running speed of 12.68 km h^{-1} and a total number of 12 CODs, assuming the time taken to complete a 180° COD is 0.7 s [86]. Here, it may be pragmatic to marginally adjust the running speed so that the time taken to complete the shuttle is an integer. Another option is to administer a 'digital figure of eight' course (e.g., 15 s for 50 m line portions at a mean velocity of 12 km h^{-1}), with smaller groups starting at course landmarks. In this approach athletes are exposed to lower neuromuscular loading due to lower COD angles (90° left and right).

Exercise Intensity

The current results demonstrated that exercise intensity does not appear to meaningfully affect HRex validity or reliability. Therefore, we recommend the prescription of fixed intensities that serve the training purposes (e.g., integrated into the warm-up and at what stage, start versus end), and must be easily repeated over time (i.e., individual constraints and resources). Running speeds between 10 and 14 km h⁻¹ depending on age, level and sport appear reasonable (likely to elicit HRex between ~75 and 85% HR_{max} in all team members).

Exercise Volume

Protocol durations of 3–4 min have the strongest theoretical and empirical rationale (refer to earlier section discusses SMFT duration in "Discussion" section).

HRex Collection Method

During continuous protocols, practitioners are advised to use the mean HR over the last 30–60 s of the test, as athletes have most likely reached a physiological (primarily HR) steady-state and this time window potentially minimises measurement noise related to the wearable device. When using intermittent variable protocols, the mean HRex throughout the overall test is more logical to retain, since exercise intensity is sporadically variable by definition. Regardless of the method used, it is important to visually inspect the HR trace for outliers prior to analysis.

Conceptual Interpretation of the Data

SMFT HRex can provide a valid marker of endurance performance when comparing between athletes in the same testing bout. That is, an athlete with a lower SMFT HRex might typically be expected to perform better on a standard maximal test on that day versus an athlete with a higher SMFT HRex (and vice versa). This does not necessarily extrapolate to within-athlete comparisons, however, which are usually the focus of training monitoring. Future studies should therefore examine the validity of SMFT HRex to track intra-athlete changes in aerobic capacity.

Statistical Interpretation of the Data

Under standardised training setting and environmental conditions, SMFT HRex TE can be considered between 1-2% points. The TE can be used to establish thresholds for changes that are beyond the measurement error, which may be useful when interpreting the data to inform decision making regarding an athlete's physiological state. For example, using our TE pooled estimate of 1.6% points and a z-distribution, confidence limits at 80%, 90%, 95% and 99% for an individual change in SMFT HRex are $\pm 2.9, \pm 3.7, \pm 4.4$ and $\pm 5.8\%$ points, respectively. These limits are synonymous with the minimum detectable change (MDC), which is the smallest change that can be detected beyond measurement error (which includes normal biological variation in SMFT HRex) [87]. Note that this is not the minimum practically important difference (MPID) or smallest worthwhile change and should not be used as such. The MPID requires separate discussion and can be combined with the MDC to establish both 'true' and meaningful changes in SMFT HRex [88].

Conclusion

Our meta-analysis is the first to provide quantitative syntheses of SMFT HRex reliability and convergent validity. Results demonstrate good absolute and high



Fig. 8 Practical implications for designing and monitoring SMFT HRex in team sports. temp. temperature

relative reliability, as well as a large convergent validity, which are not affected by athlete or protocol-related characteristics. Practitioners can use these findings to inform SMFT protocol selection ("Practical Implications" section and Fig. 8), interpret data and identify true physiological effects in individual athletes. Future research should focus on SMFT HRex sensitivity to physiological state (within-athlete changes in aerobic capacity), as well as the methodological elements related to SMFT characteristics, HRex collection and analysis approaches that may influence key measurement properties.

Abbreviations

SMFT	Submaximal fitness tests
HRex	Exercise heart rate
TE	Typical error of measurement
ICC	Intraclass correlation coefficient
MD	Mean difference
r	Pearson's correlation coefficient
Yo-YoIR	Yo-Yo intermittent recovery test
30-15 _{IFT}	30-15 Intermittent fitness test
PRISMA	Preferred Reporting Items for Systematic reviews and
	Meta-Analyses
HR _{max}	Heart rate maximum
CV	Coefficient of variation
SD	Standard deviation
CI	Confidence intervals
PI	Prediction intervals
Robans	Risk of bias assessment tool for non-randomised studies
COD	Change of direction
MDC	Minimum detectable change
MPID	Minimum practically important difference

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40798-023-00564-w.

Additional file 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist.

Additional file 2. Methodological overview of the searching strategy, screening process and protocol registration.

Additional file 3. Typical error simulation summary and results.

Additional file 4. Heterogeneity and meta-regression results from metaanalysis of measurement properties.

Additional file 5. A summary detailing the quality assessment criteria of the included studies.

Additional file 6. Meta-analysis forest plots of the weighted points estimates.

Additional file 7. Sensitivity and influence analyses from metaanalysis of measurement properties.

Acknowledgements

The authors would like to thank the authors who provided their data for the studies included in our systematic review and meta-analysis. We also thank the reviewers for their constructive feedback and contribution.

Author contributions

TS, SJM and RL conceptualised the idea and developed the project. TS and RL developed the searching strategy. TS SJM and RL constructed the inclusionexclusion criteria. TS firstly extracted the data, then checked and approved by SJM. Discrepancies in the extracted data were compared and resolved through a discussion (TS, SJM and RL). TS and SJM performed the data analysis. TS wrote the first manuscript and all other versions. SJM, RL, MB, TJS, SB and DN contributed substantially to the conceptual direction, content and provided feedback to the manuscript. All authors read and approved the final manuscript. TS coordinated the submission and revision process. All authors read and approved the final manuscript.

Funding

No sources of funding were used to assist in the preparation of this article.

Availability of data and materials

Supplementary materials for this paper are also available on the Open Science Framework at https://osf.io/mgnt9/.

Declarations

Ethical approval and consent to participate Not applicable.

Consent for publication

Not applicable.

Competing interest

Tzlil Shushan, Ric Lovell, Martin Buchheit, Tannath J. Scott, Steve Barret, Dean Norris and Shaun J. McLaren declare that they have no competing interests.

Author details

¹School of Health Sciences, Western Sydney University, Sydney, NSW, Australia.
²Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, NSW, Australia. ³HIIT Science, Revelstoke, BC, Canada. ⁴Laboratory of Sport, Expertise and Performance, French National Institute of Sport (INSEP), 7370 Paris, France. ⁵Kitman Labs, Performance Research Intelligence Initiative, Dublin, Ireland. ⁶Institute for Health and Sport, Victoria University, Melbourne, VIC, Australia. ⁷Netball Australia, Victoria, Australia. ⁸Carnegie Applied Rugby Research (CARR) Centre, Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, UK. ⁹Department of Sport Science Innovation, Playermaker, London, UK. ¹⁰Newcastle Falcons Rugby Club, Newcastle Upon Tyne, UK. ¹¹Institute of Sport, Manchester Metropolitan University, Manchester, UK.

Received: 16 May 2022 Accepted: 8 March 2023 Published online: 24 March 2023

References

- Akenhead R, Nassis GP. Training load and player monitoring in high-level football: current practice and perceptions. Int J Sports Physiol Perform. 2016;11:587–93. https://doi.org/10.1123/ijspp.2015-0331.
- Bangsbo J, laia FM, Krustrup P. The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. Sports Med. 2008;38:37–51. https://doi.org/10.2165/00007256-20083 8010-00004.
- Scott TJ, McLaren SJ, Caia J, et al. The reliability and usefulness of an individualised submaximal shuttle run test in elite rugby league players. Sci Med Football. 2018;2:184–90. https://doi.org/10.1080/24733938.2018. 1448937.
- Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol. 2014. https://doi.org/10.3389/fphys.2014. 00073.
- Jeffries AC, Marcora SM, Coutts AJ, et al. Development of a revised conceptual framework of physical training for use in research and practice. Sports Med. 2021. https://doi.org/10.1007/s40279-021-01551-5.
- Hulin BT, Gabbett TJ, Johnston RD, et al. Sub-maximal heart rate is associated with changes in high-intensity intermittent running ability in professional rugby league players. Sci Med Football. 2019;3:50–6. https:// doi.org/10.1080/24733938.2018.1475748.
- Fanchini M, Castagna C, Coutts AJ, et al. Are the Yo-Yo intermittent recovery test levels 1 and 2 both useful? Reliability, responsiveness and interchangeability in young soccer players. J Sports Sci. 2014;32:1950–7. https://doi.org/10.1080/02640414.2014.969295.
- 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;52:2605–26. https://doi.org/10. 1007/s40279-022-01712-0.
- 9. Schneider C, Hanakam F, Wiewelhove T, et al. Heart rate monitoring in team sports—a conceptual framework for contextualizing heart rate

measures for training and recovery prescription. Front Physiol. 2018;9:639. https://doi.org/10.3389/fphys.2018.00639.

- Bosquet L, Merkari S, Arvisais D, et al. Is heart rate a convenient tool to monitor over-reaching? A systematic review of the literature. Br J Sports Med. 2008;42:709–14. https://doi.org/10.1136/bjsm.2007.042200.
- Buchheit M, Voss SC, Nybo L, et al. Physiological and performance adaptations to an in-season soccer camp in the heat: associations with heart rate and heart rate variability: monitoring responses to training in the heat. Scand J Med Sci Sports. 2011;21:e477–85. https://doi.org/10.1111/j. 1600-0838.2011.01378.x.
- 12. Buchheit M, Simpson BM, Garvican-Lewis LA, et al. Wellness, fatigue and physical performance acclimatisation to a 2-week soccer camp at 3600 m (ISA3600). Br J Sports Med. 2013;47:i100–6. https://doi.org/10.1136/bjsports-2013-092749.
- Brink MS, Visscher C, Schmikli SL, et al. Is an elevated submaximal heart rate associated with psychomotor slowness in young elite soccer players? Eur J Sport Sci. 2013;13:207–14. https://doi.org/10.1080/17461391.2011. 630101.
- 14. Robertson S, Kremer P, Aisbett B, et al. Consensus on measurement properties and feasibility of performance tests for the exercise and sport sciences: a Delphi study. Sports Med Open. 2017;3:2.
- Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30:1–15. https://doi.org/10.2165/00007256-20003 0010-00001.
- Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26:217–38. https://doi.org/10.2165/00007256-199826040-00002.
- 17. Matheson GJ. We need to talk about reliability: making better use of testretest studies for study design and interpretation. PeerJ. 2019;7:e6918.
- Krabbe PFM. The measurement of health and health status: concepts, methods and applications from a multidisciplinary perspective. 1st ed. Amsterdam: Elsevier; 2017.
- Thorpe RT, Strudwick AJ, Buchheit M, et al. Tracking morning fatigue status across in-season training weeks in elite soccer players. Int J Sports Physiol Perform. 2016;11:947–52. https://doi.org/10.1123/ijspp.2015-0490.
- Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:160. https://doi.org/10.1136/bmj.n160.
- 21. Buchheit M. The 30–15 intermittent fitness test: 10 year review. Myorobie J. 2010;1:278.
- Lemmink KAPM, Visscher C, Lambert MI, et al. The interval shuttle run test for intermittent sport players: evaluation of reliability. J Strength Cond Res. 2004;18:821–7. https://doi.org/10.1519/13993.1.
- 23. Buchheit M, Millet GP, Parisy A, et al. Supramaximal training and postexercise parasympathetic reactivation in adolescents. Med Sci Sports Exerc. 2008;40:362–71. https://doi.org/10.1249/mss.0b013e31815aa2ee.
- 24. Deprez D, Fransen J, Lenoir M, et al. The Yo-Yo intermittent recovery test level 1 is reliable in young high-level soccer players. Biol Sport. 2015;32:65–70. https://doi.org/10.5604/20831862.1127284.
- Deprez D, Coutts AJ, Lenoir M, et al. Reliability and validity of the Yo-Yo intermittent recovery test level 1 in young soccer players. J Sports Sci. 2014;32:903–10. https://doi.org/10.1080/02640414.2013.876088.
- Hulse M, Morris J, Hawkins R, et al. A field-test battery for elite, young soccer players. Int J Sports Med. 2012;34:302–11. https://doi.org/10.1055/s-0032-1312603.
- Iacono AD, Beato M, Unnithan V. Comparative effects of game profilebased training and small-sided games on physical performance of elite young soccer players. J Strength Cond Res. 2019. https://doi.org/10.1519/ jsc.000000000003225.
- Ingebrigtsen J, Brochmann M, Castagna C, et al. Relationships between field performance tests in high-level soccer players. J Strength Cond Res. 2014;28:942–9. https://doi.org/10.1519/jsc.0b013e3182a1f861.
- Owen C, Jones P, Comfort P. The reliability of the submaximal version of the Yo-Yo intermittent recovery test in elite youth soccer. J Trainol. 2017;6:31–4.
- Pereira LA, Abad CCC, Leiva DF, et al. Relationship between resting heart rate variability and intermittent endurance performance in novice soccer players. Res Q Exerc Sport. 2019;90:355–61. https://doi.org/10.1080/02701 367.2019.1601666.

- Ryan S, Pacecca E, Tebble J, et al. Measurement characteristics of athlete monitoring tools in professional Australian football. Int J Sports Physiol Perform. 2019. https://doi.org/10.1123/ijspp.2019-0060.
- Younesi S, Rabbani A, Clemente FM, et al. Relationships between aerobic performance, hemoglobin levels, and training load during small-sided games: a study in professional soccer players. Front Physiol. 2021;12:649870. https://doi.org/10.3389/fphys.2021.649870.
- Younesi S, Rabbani A, Manuel Clemente F, et al. Session-to-session variations of internal load during different small-sided games: a study in professional soccer players. Res Sports Med. 2021. https://doi.org/10. 1080/15438627.2021.1888103.
- Buchheit M, Mendez-Villanueva A, Quod MJ, et al. Determinants of the variability of heart rate measures during a competitive period in young soccer players. Eur J Appl Physiol. 2010;109:869–78. https://doi.org/10. 1007/s00421-010-1422-x.
- Bradley PS, Mohr M, Bendiksen M, et al. Sub-maximal and maximal Yo-Yo intermittent endurance test level 2: heart rate response, reproducibility and application to elite soccer. Eur J Appl Physiol. 2011;111:969–78. https://doi.org/10.1007/s00421-010-1721-2.
- Scott TJ, McLaren SJ, Lovell R, et al. The reliability, validity and sensitivity of an individualised sub-maximal fitness test in elite rugby league athletes. J Sports Sci. 2022;9:1–13. https://doi.org/10.1080/02640414.2021.2021047.
- Buchheit M, Lefebvre B, Laursen PB, et al. Reliability, usefulness, and validity of the 30–15 intermittent ice test in young elite ice hockey players. J Strength Cond Res. 2011;25:1457–64. https://doi.org/10. 1519/jsc.0b013e3181d686b7.
- Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol. 2001;37:153–6. https://doi.org/10.1016/ s0735-1097(00)01054-8.
- Vigh-Larsen JF, Beck JH, Daasbjerg A, et al. Fitness characteristics of elite and subelite male ice hockey players: a cross-sectional study. J Strength Cond Res. 2019;33:2352–60. https://doi.org/10.1519/jsc.00000 00000003285.
- Francini L, Rampinini E, Bosio A, et al. Association between match activity, endurance levels and maturity in youth football players. Int J Sports Med. 2019;40:576–84. https://doi.org/10.1055/a-0938-5431.
- Borenstein M, editor. Introduction to meta-analysis. Chichester: Wiley; 2009.
- Hopkins WG. Spreadsheets for analysis of validity and reliability. Sportscience. 2015;19:36–45.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
- Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Erlbaum Associates: Hillsdale; 1988.
- Mohr M, Krustrup P. Yo-Yo intermittent recovery test performances within an entire football league during a full season. J Sports Sci. 2014;32:315–27. https://doi.org/10.1080/02640414.2013.824598.
- Veugelers KR, Naughton GA, Duncan CS, et al. Validity and Reliability of a submaximal intermittent running test in elite Australian football players. J Strength Cond Res. 2016;30:3347–53. https://doi.org/10.1519/jsc. 000000000001441.
- Lignell E, Fransson D, Krustrup P, et al. Analysis of high-intensity skating in top-class ice hockey match-play in relation to training status and muscle damage. J Strength Cond Res. 2018;32:1303–10. https://doi. org/10.1519/jsc.000000000001999.
- Fisher R. Statistical methods for research workers. 12th ed. Edinburgh: Oliver and Boyd; 1954.
- Nakagawa S, Poulin R, Mengersen K, et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. Methods Ecol Evol. 2015;6:143–52. https://doi.org/10.1111/2041-210X.12309.
- Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft. 2010. https://doi.org/10.18637/jss.v036.i03.
- Pustejovsky J. clubSandwich: cluster-robust (Sandwich) variance estimators with small-sample corrections. R package version 0.2.3. 2017. Available from: https://github.com/jepusto/clubSandwich.
- RStudio: Integrated development for R. RStudio, Inc., Boston, MA. Available from: https://www.rstudio.com.
- Rabbani A, Kargarfard M, Twist C. Reliability and validity of a submaximal warm-up test for monitoring training status in professional soccer

players. J Strength Cond Res. 2018;32:326–33. https://doi.org/10.1519/ jsc.00000000002335.

- Doncaster G, Scott M, Iga J, et al. Reliability of heart rate responses both during and following a 6 min Yo-Yo IR1 test in highly trained youth soccer players. Sci Med Football. 2019;3:14–20. https://doi.org/ 10.1080/24733938.2018.1476775.
- Pustejovsky JE, Tipton E. Meta-analysis with robust variance estimation: expanding the range of working models. Prev Sci. 2021;7:1–14. https:// doi.org/10.1007/s11121-021-01246-3.
- Cheung MWL. Modeling dependent effect sizes with three-level metaanalyses: a structural equation modeling approach. Psychol Methods. 2014;19:211–29. https://doi.org/10.1037/a0032968.
- Hedges LV, Tipton E, Johnson MC. Robust variance estimation in metaregression with dependent effect size estimates. Res Synth Method. 2010;1:39–65. https://doi.org/10.1002/jrsm.5.
- Cheung MWL. A guide to conducting a meta-analysis with non-independent effect sizes. Neuropsychol Rev. 2019;29:387–96. https://doi. org/10.1007/s11065-019-09415-6.
- Higgins JPT. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol. 2008;37:1158– 60. https://doi.org/10.1093/ije/dyn204.
- Borenstein M, Higgins JPT, Hedges LV, et al. Basics of meta-analysis: l2 is not an absolute measure of heterogeneity. Res Synth Methods. 2017;8:5–18. https://doi.org/10.1002/jrsm.1230.
- Assink M, Wibbelink CJM. Fitting three-level meta-analytic models in R: a step-by-step tutorial. TQMP. 2016;12:154–74. https://doi.org/10.20982/ tqmp.12.3.p154.
- 62. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane handbook for systematic reviews of interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from: www.training.cochrane.org/handbook.
- Greenland S. Valid P-values behave exactly as they should: some misleading criticisms of P-values and their resolution with S-values. Am Stat. 2019;73:106–14. https://doi.org/10.1080/00031305.2018.1529625.
- Halperin I, Malleron T, Har-Nir I, et al. Accuracy in predicting repetitions to task failure in resistance exercise: a scoping review and exploratory meta-analysis. Sports Med. 2022;52:377–90. https://doi.org/10.1007/ s40279-021-01559-x.
- Malcata RM, Vandenbogaerde TJ, Hopkins WG. Using athletes' world rankings to assess countries' performance. Int J Sports Physiol Perform. 2014;9:133–8. https://doi.org/10.1123/ijspp.2013-0014.
- Hopkins WG, Marshall SW, Batterham AM, et al. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 2009;41:3–13. https://doi.org/10.1249/mss.0b013e31818cb278.
- 67. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. J Clin Epidemiol. 2013;66:408–14. https://doi.org/10.1016/j.jclin epi.2012.09.016.
- Park J, Lee Y, Seo H, Jang B, Son H, Kim S, Shin S, Hahn S. Risk of bias assessment tool for non-randomized studies (RoBANS): development and validation of a new instrument. In: Abstracts of the 19th cochrane colloquium; 2011 19–22 Oct; Madrid, Spain. Wiley; 2011.
- Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34. https://doi. org/10.1136/bmj.315.7109.629.
- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50:1088–101.
- Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for metaanalysis. Res Synth Method. 2010;1:112–25. https://doi.org/10.1002/jrsm. 11.
- Baujat B, Mahé C, Pignon J-P, et al. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med. 2002;21:2641–52. https://doi.org/10.1002/sim.1221.
- Algur SP, Biradar JG. Cooks distance and mahanabolis distance outlier detection methods to identify review spam. Ijecs. 2017;6:21638–49.
- Bradley PS, Bendiksen M, Dellal A, et al. The application of the Yo-Yo intermittent endurance level 2 test to elite female soccer populations: Yo-Yo IE2 testing in female soccer players. Scand J Med Sci Sports. 2014;24:43– 54. https://doi.org/10.1111/j.1600-0838.2012.01483.x.

- 75. Laursen P, Buchheit M, editors. Science and application of high-intensity interval training. Champaign: Human Kinetics; 2019.
- Lamberts RP, Lemmink KAPM, Durandt JJ, et al. Variation in heart rate during submaximal exercise: implications for monitoring training. J Strength Cond Res. 2004;18:641–5.
- Lamberts RP, Lambert MI. Day-to-day variation in heart rate at different levels of submaximal exertion: implications for monitoring training. J Strength Cond Res. 2009;23:1005–10. https://doi.org/10.1519/jsc.0b013 e3181a2dcdc.
- Araújo M, Baumgart C, Freiwald J, et al. Contrasts in intermittent endurance performance and heart rate response between female and male soccer players of different playing levels. Biol Sport. 2019;36:323–31. https://doi.org/10.5114/biolsport.2019.88755.
- Ingebrigtsen J, Bendiksen M, Randers MB, et al. Yo-Yo IR2 testing of elite and sub-elite soccer players: performance, heart rate response and correlations to other interval tests. J Sports Sci. 2012;30:1337–45. https://doi. org/10.1080/02640414.2012.711484.
- Bock AV, Vancaulaert C, Dill DB, et al. Studies in muscular activity: IV. The 'steady state' and the respiratory quotient during work. J Physiol. 1928;66:162–74. https://doi.org/10.1113/jphysiol.1928.sp002515.
- Lacome M, Simpson B, Broad N, et al. Monitoring players' readiness using predicted heart-rate responses to soccer drills. Int J Sports Physiol Perform. 2018;13:1273–80. https://doi.org/10.1123/ijspp.2018-0026.
- Owen AL, Newton M, Shovlin A, et al. The use of small-sided games as an aerobic fitness assessment supplement within elite level professional soccer. J Hum Kinet. 2020;71:243–53. https://doi.org/10.2478/ hukin-2019-0086.
- Hulka K, Weisser R, Belka J, et al. Stability of internal response and external load during 4-a-side football game in an indoor environment. Acta Gymnica. 2015;45:21–5. https://doi.org/10.5507/ag.2015.003.
- Wiseman RM. On the use and misuse of ratios in strategic management research. Research methodology in strategy and management, vol. 5. Emerald Group Publishing Limited; 2009. p. 75–110.
- Lacome M, Simpson BM, Buchheit M. Monitoring training status with player-tracking technology: still on the road to Rome. Aspetar Sports Med J. 2018;7:54–63.
- Buchheit M. The 30–15 intermittent fitness test: accuracy for individualizing interval training of young intermittent sport players. J Strength Cond Res. 2008;22:365–74. https://doi.org/10.1519/jsc.0b013e3181635b2e.
- Buchheit M, Simpson BM, Lacome M. Monitoring cardiorespiratory fitness in professional soccer players: is it worth the prick? Int J Sports Physiol Perform. 2020;15:1437–41. https://doi.org/10.1123/ijspp.2019-0911.
- de Vet HCW, Terwee CB. The minimal detectable change should not replace the minimal important difference. J Clin Epidemiol. 2010;63:804– 5. https://doi.org/10.1016/j.jclinepi.2009.12.015.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com