Review

Criterion-Related Validity of Field-Based Fitness Tests in Adults: A Systematic Review

Jose Castro-Piñero [1,2], Nuria Marin-Jimenez [1,2,*], Jorge R. Fernandez-Santos [1,2], Fatima Martin-Acosta [1,2], Victor Segura-Jimenez [1,2], Rocio Izquierdo-Gomez [1,2], Jonatan R. Ruiz [3] and Magdalena Cuenca-Garcia [1,2]

1   GALENO Research Group, Department of Physical Education, Faculty of Education Sciences, University of Cádiz, Avenida República Saharaui s/n, Puerto Real, 11519 Cádiz, Spain; jose.castro@uca.es (J.C.-P.); jorgedelrosario.fernandez@uca.es (J.R.F.-S.); fatima.martin@uca.es (F.M.-A.); victor.segura@uca.es (V.S.-J.); rocio.izquierdo@uca.es (R.I.-G.); magdalena.cuenca@uca.es (M.C.-G.)
2   Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), 11009 Cádiz, Spain
3   PROmoting FITness and Health through Physical Activity Research Group (PROFITH), Sport and Health University Research Institute (iMUDS), Department of Physical and Sports Education, School of Sports Science, University of Granada, 18007 Granada, Spain; ruizj@ugr.es
*   Correspondence: nuria.marin@uca.es; Tel.: +34-956-016-253

Abstract: We comprehensively assessed the criterion-related validity of existing field-based fitness tests used to indicate adult health (19–64 years, with no known pathologies). The medical electronic databases MEDLINE (via PubMed) and Web of Science (all databases) were screened for studies published up to July 2020. Each original study's methodological quality was classified as high, low and very low, according to the number of participants, the description of the study population, statistical analysis and systematic reviews which were appraised via the AMSTAR rating scale. Three evidence levels were constructed (strong, moderate and limited evidence) according to the number of studies and the consistency of the findings. We identified 101 original studies (50 of high quality) and five systematic reviews examining the criterion-related validity of field-based fitness tests in adults. Strong evidence indicated that the 20 m shuttle run, 1.5-mile, 12 min run/walk, YMCA step, 2 km walk and 6 min walk test are valid for estimating cardiorespiratory fitness; the handgrip strength test is valid for assessing hand maximal isometric strength; and the Biering–Sørensen test to evaluate the endurance strength of hip and back muscles; however, the sit-and reach test, and its different versions, and the toe-to-touch test are not valid for assessing hamstring and lower back flexibility. We found moderate evidence supporting that the 20 m square shuttle run test is a valid test for estimating cardiorespiratory fitness. Other field-based fitness tests presented limited evidence, mainly due to few studies. We developed an evidence-based proposal of the most valid field-based fitness tests in healthy adults aged 19–64 years old.

Keywords: cardiorespiratory fitness; muscular strength; motor fitness and flexibility; validation; fitness testing; adulthood

1. Introduction

Physical fitness is an integrated measure of all the functions and structures involved in performing physical activity [1]. Nowadays, physical fitness is one surrogate marker of overall adult health (19–64 years), especially cardiorespiratory fitness and muscular strength. Cardiorespiratory fitness is inversely associated with cardiovascular diseases [2], obesity [3], osteoporosis [4] diabetes [5], different cancer types [6,7], and is a predictor of all-cause of mortality [8–12] and cardiovascular disease [10,12–15]. Likewise, in the psychological sphere, high levels of cardiorespiratory fitness are associated with well-being [16,17], improved cognitive function [18] and a reduced risk of Alzheimer's disease [19] and other mental conditions such as anxiety, panic and depression [20]. Muscular strength demonstrates a protective effect against all-cause mortality [21,22]; and is inversely associated

with weight gain and adiposity-related hypertension occurrence and the prevalence and incidence of the metabolic syndrome, [22] and mental health clinical presentations [23,24]. Consequently, physical fitness assessment is a vital tool of prevention and health diagnoses.

Laboratory testing is an objective and accurate method of assessing physical fitness. However, due to the cost of sophisticated instruments, time constraints and the need for qualified technicians, laboratory testing is limited to sport clubs, schools, population-based studies, and offices or clinical settings. However, field-based fitness testing can offer useful and practical alternatives as screening tools, since they are relatively safe and time-efficient, involve minimal equipment and low cost, and can be easily administered to multiple people simultaneously.

The validity of field-based fitness tests needs to be considered when deciding which test to use [25]. Criterion-related validity refers to the extent to which a field-based test of a physical fitness component correlates with the criterion measure (i.e., the gold standard) [26]. Since the early interest in physical fitness testing in the 1950–1960s, many field-based fitness tests have been proposed [27]. It would be desirable to summarise the criterion-related validity of the existing field-based fitness tests in adults. There have been attempts to summarize the criterion-related validity of a certain test [28,29] or several tests with a common characteristic [30–32]; however, no attempts have been made to summarise the criterion-related validity of all the existing field-based fitness tests in adults.

Therefore, the aim of the present systematic review was to comprehensively study the criterion-related validity of the existing field-based fitness tests used in adults. The findings of this review will provide an evidence-based proposal for most valid field-based fitness tests for healthy adults, aged 19–64 years old.

2. Materials and Methods

The review was registered in PROSPERO (registration number: CRD42019118482) and the applied methodology followed the guidelines drawn in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [33].

2.1. Literature Search

The search was performed in the MEDLINE (via Pubmed) and Web of Science electronic databases from inception until July 2020. We screened studies conducted for criterion-related validity in adults, where one or more field-based fitness tests were carried out. Thus, the keywords selected were based on terms related to "criterion-related validity", "adults" and "field-based fitness test". The search syntax was adapted to the indexing terms of each database (see Supplementary Material 1). Searching was restricted to articles published in humans and the English or Spanish languages.

2.2. Eligibility Criteria

The inclusion criteria for this systematic review were the (1) age criterion: adults (19–64 years old). During this review, we faced the problem that some studies sampled adults and older adults, or adults or adolescents together. In these cases, we observed whether these studies performed stratified analyses by age groups, isolating the adult population from the rest; if so, the study was included and information concerning the adult population reported. In contrast, when the authors analysed the whole sample together, we only included the study if the age of the sample was predominantly within our study age range; (2) participants: the study population was based on a generally healthy population, who did not present any injury, physical and/or mental disabilities, irrespective of body mass index (BMI), diabetes or other cardiovascular risks (i.e., hypertension, hypercholesterolemia, lipid profiles, glucose levels, insulin sensitivity); and (3) study design: original studies or systematic reviews/meta-analysis. The original studies that were selected for the analysis of their criterion-related validity but which were also included in the selected systematic reviews were excluded; (4) language criterion: articles were only published in English or Spanish; (5) topic criterion: studies examining the criterion-related validity of

the field-based fitness test. Studies examining the relationship between field-based fitness tests were excluded. Likewise, studies that analysed the criterion-related validity of tests designed for exclusive use in sports or clinical settings were not included.

Two authors (J.C.P. and J.R.F.S.) independently assessed the titles and abstracts of the articles retrieved by the search strategy for eligibility. Then, the full texts of the selected articles were acquired, and the same two researchers independently screened them to determine whether to include the article based on the inclusion criteria. When no consensus was reached between both researchers, a third research (N.M.J.) made the final decision with regard to inclusion. Reasons for the exclusion of identified articles were recorded.

2.3. Data Extraction

Two researchers (N.M.J. and F.M.A.) independently extracted the following information from each eligible original study according to the standardized form: (1) the author's name; (2) participants (sex and number); (3) age of participants; (4) filed-based test; (5) criterion measure (gold standard); (6) statistical methods; (7) main outcome; and (8) conclusions.

The same researchers independently extracted the following information from the systematic reviews: (1) author´s name, date and years covered by the review; (2) type of review and number of included studies; (3) age of participants; (4) filed-based test; (5) criterion measure (gold standard); (6) main outcome; and (7) conclusions.

Disagreements in the extracted data were discussed between studies until a consensus was reached.

2.4. Criteria for Risk of Bias Assessment

Due to the heterogeneity of statistical methods employed by the original studies selected, the high number of tests included, and the limited number of studies per test, a meta-analysis was not conducted. An assessment of risk of bias in selected original studies and systematic reviews was made for each eligible study by two studies (N.M.J. and F.M.A.) independently. Discrepancies were solved in a consensus meeting. Inter-rater agreement for the risk of bias between researchers was calculated by the percentage agreement (96% (Kappa = 0.962) before consensus, and 100% agreement after consensus meeting).

The assessing risk of bias criteria in original studies were determined according to quality assessment list employed by Castro-Piñero et al. [27], which include the three following criteria: (1) the adequate number of participants; (2) an adequate description of the study population; and (3) adequate statistical analysis (see Supplementary Table S1). Each criterion was rated from 0 to 2, being 2 the best score. For all studies, a total score was calculated by counting up the number of positive items (a total score between 0 and 6). Studies were categorized as very low quality (0–2), low quality (3–4) and high quality (5–6).

The methodological quality of each systematic review was appraised using the 'Assessment of Multiple Systematic Reviews' (AMSTAR) rating scale [34]. AMSTAR contains 11-items to assess the methodological aspects of reviews with items scored as 1 if the answer was "Yes", and 0 if the answer was "No", "Cannot Answer" or "Not Applicable" (see Supplementary Table S2). The total score ranged from 0 to 11. The item on conflict of interest requires that the systematic review and all primary studies be assessed. We modified this item to only assess the review itself as Biddle et al. [35] proposed, given that PRISMA does not require a conflict-of-interest assessment for each primary study. The final quality rates were computed by tertiles, where the first tertile ranged from 0 to 3 points (low quality); the second tertile from 4 to 7 points (medium quality); and the third tertile from 8 to 11 points (high quality).

2.5. Levels of Evidence

Three evidence levels [27] were constructed: (1) strong evidence: consistent findings in three or more high-quality studies; (2) moderate evidence: consistent findings in two high-quality studies; and (3) limited evidence: consistent findings in multiple low-quality studies, inconsistent results found in multiple high-quality studies, or results based on one single study. The degree of criterion-related validity of the field-based fitness test will be discussed for those tests on which we found strong or moderate evidence that the test is (or not) valid. The results of low- or very low-quality studies can be seen in the Supplementary Material 2.

3. Results

The literature search yielded 9202 and 27 additional records were identified through other sources (see the PRISMA flowchart in Figure 1). After the removal of duplicate references (1805 studies), and the screening of titles and abstracts (7233 studies), we excluded 9038 studies. A total of ■191 full-text studies were assessed for eligibility, and 85 studies (six systematic reviews) were excluded due to reasons indicated in Figure 1.
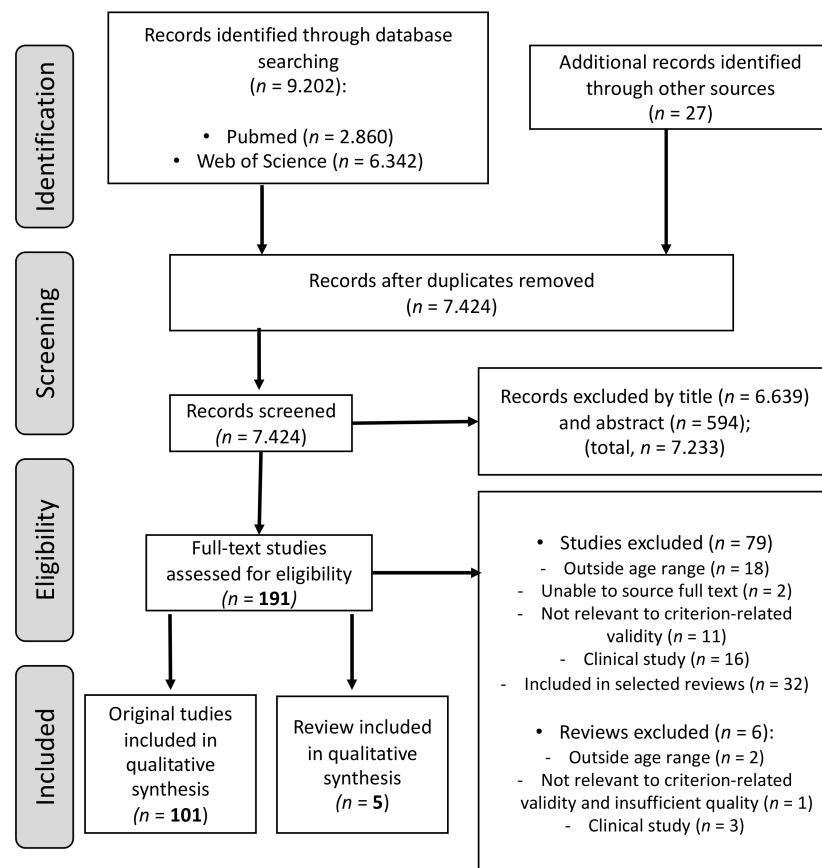

Figure 1. Flow chart of retrieved and selected articles.

Finally, a total of 101 original studies (see Supplementary Table S3) addressed the criterion-related validity of field-based fitness tests in adults aged 19–64 years. The sample size involved 10,632 participants (see Supplementary Table S4). Eighty-six and seventy-eight original studies reported female (*n* = 5539) and male (*n* = 4722) sample proportions, respectively; however, in 7 seven studies, sex was not specified.

A total of four meta-analyses [28–31] and one systematic review [32] were included in the present systematic review (see Supplementary Table S2). The sample size involved 9985 participants with ages ranging from 19 to 64 years (see Supplementary Table S5).

3.1. Quality Assessment

Of the 101 original studies included in the present systematic review, 11 and 40 studies were classified as very low (a total score less than 2) and low quality (a total score of 3 and 4), respectively (see Supplementary Table S3). A total of 50 original studies were classified as high-quality (a total score higher than 4). Of these 40, nine and one analysed the criterion-related validity of cardiorespiratory fitness, muscular strength and flexibility field-based fitness tests in adults, respectively. No study of those classified as high quality analysed the criterion-related validity of motor fitness (i.e., speed, agility, balance and coordination).

Two meta-analyses [28,30] and one systematic review [32] were ranked as high quality (all eight points), and two meta-analyses [29,31] were ranked as medium quality (both seven points) (see Supplementary Table S2). Three of them assessed the criterion-related validity of field-based cardiorespiratory fitness tests: the 20 m shuttle run test [28]; distance and time-based run/walk tests [30]; and the step tests [32]—whilst and two of them studied the criterion-related validity of the sit-and-reach [31] and toe-to-touch tests [29].

* References of high-quality studies are presented in Supplementary Material 3.

3.2. Criterion-Related Validity

Table 1 shows a summary of the different levels of evidence found for the criterion-related validity of cardiorespiratory fitness tests.

Table 1. Levels of evidence of cardiorespiratory fitness tests.

Field-Based Fitness Test	Strong	Moderate	Limited
Shuttle run tests			
20 m shuttle run	●		
20 m square shuttle		●	
Incremental shuttle walk			■
Distance and time-based run/walk test			
1.5-mile run/walk	●		
12 min run/walk	●		
5000 m run/walk	◑		
3 miles run/walk	◑		
2 miles run/walk	◑		
3.000 m run/walk	◑		
1000 m run/walk	◑		
600 m run/walk	◑		
600 yd run/walk	◑		
$\frac{1}{2}$-mile run/walk	◑		
$\frac{1}{4}$-mile run/walk	◑		
9 min run/walk			●
2 km walk	●		
6 min walk	●		
1-mile walk			■
$\frac{1}{4}$-mile walk			●
3 min walk			●
Treadmill jogging			■
Mankato submaximal exercise			●
Modified Astrand–Ryhming			●
University Montreal			●
Ruffier			○

Table 1. *Cont.*

Field-Based Fitness Test	Strong	Moderate	Limited
Step tests			
YMCA step	●		
Chester step			■
Modified Harvard step			●
6 min single 15 cm-step			●
Modified Canadian aerobic fitness step			●
Tecumseh step			●
Astrand–Ryhming step			●
Danish step			○
Queen's College step			■
2 min step			○

● Indicates high validity; ○ moderate validity; ◖ low/null validity; ■ inconclusive validity.

3.2.1. Cardiorespiratory Fitness

Distance and Time-Based Run/Walk Tests

Seventeen high-quality studies examined the criterion-related validity of the distance run/walk or walk tests (see Supplementary Table S4). Four and two studies showed that the 2 km walk [36–39] and 1.5-mile run/walk [40,41] tests, respectively, were valid for assessing cardiorespiratory fitness (r = 0.80–0.93, all $p < 0.05$). Four studies [42–45] observed that the 1-mile walk test was an accurate test for estimating VO_{2max} (r = 0.81–0.88, all $p < 0.05$), while another two studies [46,47] showed that it was not a valid test (r = 0.69, 13.3% E, $p < 0.05$; mean differences range from 2.360 to 9.131 mL/kg/min, all $p < 0.001$, respectively). The treadmill jogging test reported contradictory results: one study [48] found it to have high validity for assessing cardiorespiratory fitness (r = 0.84, both $p < 0.001$); whereas another study [41] revealed that it was not a valid test (r = 0.50, $p < 0.05$).

Five high-quality studies investigated the criterion-related validity of the time-based run/walk or walk tests (see Supplementary Table S4). These studies showed that the 3 min walk, [49] 6 min walk, [50–52] and the 12 min run/walk [41] tests were valid for assessing cardiorespiratory fitness (r = 0.70–0.95, all $p < 0.05$). Additionally, one original high-quality study reported that the University Montreal test [53] was valid for estimating cardiorespiratory fitness (r = 0.71, $p < 0.001$; mean difference = 0.025 ± 7.445 mL/kg/min., $p > 0.05$).

A meta-analysis [30] consisting of 102 studies on adults determined that the criterion-related validity of the distance run/walk field tests for estimating cardiorespiratory fitness ranged from low to high, with the 1.5-mile (r_p = 0.80; 95% CI: 0.72–0.80) and 12 min run/walk tests (r_p = 0.79; 95% CI: 0.71–0.87) being the best predictors (see Supplementary Table S5).

Twenty-Metre Shuttle Run Test

Nine high-quality studies analysed the criterion-related validity of the 20 m shuttle run test [41,54–58] or modifications of it [55,57,59–61] (see Supplementary Table S4). Four studies [41,55–57] reported that the 20 m shuttle run was a valid test for assessing cardiorespiratory fitness (r = 0.82–0.94, all $p < 0.05$). However, one study [58] concluded that this test was not valid for assessing cardiorespiratory fitness (mean differences range from −0.54 ± 6.23 to −2.94 ± 6.55 mL/kg/min, all $p < 0.01$). Two studies [59,60] proved that the incremental shuttle walk test was not valid (r = 0.72, 19% E, both $p < 0.001$), while one study [61] found that this test was valid for assessing cardiorespiratory fitness (mean difference = 0.14 ± 9.27mL/kg/min, $p > 0.05$). Moreover, two studies [55,57] reported that the 20 m square shuttle run test was valid (r = 0.95, both $p < 0.001$).

A meta-analysis [28] which included 24 studies in adults found that the 20 m shuttle run test had a moderate-to-high criterion-related validity for estimating VO_{2max} (r_p = 0.79–0.94; 95% CI: 0.56–1.00) (see Supplementary Table S5).

Step Tests

Eleven high-quality studies analysed the criterion-related validity of the step tests (see Supplementary Table S4). Four studies observed that the Danish step [62], the Queen's College step [63], and the 2 min step [64] tests were not valid for estimating VO_{2max} (r = 0.034–0.72, all p < 0.05). However, another eight studies proved the validity of the modified Canadian aerobic fitness [65], 6 min single 15 cm step [66], YMCA step [67–71], Tecumseh step [70] and modified Harvard step [72] tests (r = 0.80–0.91, all p < 0.05).

A systematic review [32] comprised of 11 studies on adults investigated the criterion-related validity of the step tests (see Supplementary Table S5). Validity measures were varied, and a broad range of correlation coefficients were reported across the 11 studies (r = 0.469–0.95; all p < 0.005) with conflicting results in most of the step test protocols. The study concluded that the Chester step test was the best predictor for assessing cardiorespiratory fitness.

3.2.2. Muscular Strength

Table 2 shows a summary of the different levels of evidence found for the criterion-related validity of muscular strength, flexibility and motor fitness tests.

Table 2. Levels of evidence of muscular strength, flexibility and motor fitness tests.

Field-Based Fitness Test	Strong	Moderate	Limited
Maximal isometric strength			
Handgrip strength (TKK)	●		
Handgrip strength (Jamar)		◑	
Handgrip strength (DynEx)			▪
Hip and back endurance strength			
Biering–Sørensen	●		
Abdominal endurance strength			
Prone bridging			●
Original/modifications curl-up			○
Lower body endurance strength			
Sit-to-stand			▪
Lower body explosive strength			
Sargent jump			●
Upper body endurance strength			
Original/modification flexed-arm hang			▪
Baumgartner modified pull-up			▪
Standard push-up			▪
Hand-release push-up			▪
Bent-knee push-up			▪
Revised push-up			▪
Lower back flexibility			
Original/modifications sit-and-reach	○		
Hamstring flexibility			
Original/modifications sit-and-reach	◑		
Toe-touch	◑		
Agility			
Ten-step			◑
Balance			
Romberg test			▪

● Indicates high validity; ○ moderate validity; ◑ low/null validity; ▪ inconclusive validity.

Maximal Isometric Strength

Four high-quality studies assessed the criterion-related validity of hand maximal isometric strength, using the handgrip strength tests (see Supplementary Table S4). Three high-quality studies reported that the TKK dynamometer [73–75] was valid (mean difference range -0.20, $p > 0.05$ to 2.02 kg $p < 0.001$) (r = 0.98, $p < 0.001$). However, three studies showed inconclusive results about the validity of the DynEx dynamometer [73,75,76], and two studies observed that the Jamar dynamometer [73,76] was less accurate than the TKK and DynEx dynamometer for estimating hand maximal isometric strength.

Endurance Strength

Four high-quality studies assessed the criterion-related validity of trunk endurance strength (see Supplementary Table S4). Two studies [77,78] suggested that the Biering–Sørensen (r = 0.84–98, $p < 0.01$) test was valid, whereas another study [79] reported acceptable validity (r = 0.60–0.71, $p < 0.05$). One study showed that the prone bridging test [80] was valid for assessing trunk endurance strength (no mean difference, $p > 0.05$).

Explosive Strength

Only one high-quality study assessed the criterion-related validity of explosive strength (see Supplementary Table S4). This study concluded that the Sargent test [81] was not valid (mean difference: 4.4 ± 5.1, $p < 0.001$) for estimating lower body explosive strength.

3.2.3. Flexibility

Only one study [82] that examined the criterion-related validity of flexibility tests was classified as high quality (see Supplementary Table S4). They found that the sit-and-reach was not a valid test (r = 0.44–0.48, $p < 0.05$).

A meta-analysis [31] which included 28 studies on adults (see Supplementary Table S5) found that the sit-and-reach test and its different versions, had moderate validity for estimating hamstring extensibility (r_p ranged from 0.49; 95% CI: 0.29–0.68 to 0.68; 95% CI: 0.55–0.80), but a low validity for estimating lumbar extensibility (r_p ranged from 0.16; 95% CI: -0.10–0.41 to 0.35; 95% CI: 0.15–0.54). Moreover, another meta-analysis [29] carried out on adults (of six studies) reported that the toe-touch test had moderate validity for assessing hamstring extensibility (r_p = 0.66; 95% CI: 0.56–1.00).

3.2.4. Motor Fitness

No study investigating the criterion-related validity of motor fitness tests was classified as high quality (see Supplementary Table S3).

4. Discussion

The present systematic review comprehensively studied the criterion-related validity of the existing field-based fitness tests used in adults. The findings of this review provide an evidence-based proposal for most valid field-based fitness tests for adult population.

4.1. Cardiorespiratory Fitness

The gold standard to assess VO_{2max} is the Douglas bag method, although there is agreement that the respiratory gas analyser is a valid method of assessing oxygen uptake [83]. All high-quality studies measured VO_{2max} or peak oxygen consumption when performing a submaximal/maximal treadmill or cycle test, except Manttari et al. [52], who directly measured VO_{2max} when performing the 6 min walk test.

4.1.1. Distance and Time-Based Run/Walk Tests

The run/walk field tests are probably the most widely used tests [27,84], however, until recently, there was no consensus regarding the most appropriate distance or time to use for these tests [85]. Mayorga et al. [30] performed a meta-analysis which examined the criterion-related validity of the 5000 m, 3 mile, 2 mile, 3000 m, 1.5-mile, 1-mile,

1000 m, $\frac{1}{2}$-mile, 600 m, 600 yd, $\frac{1}{4}$-mile, 15 min, 12 min, 9 min, and 6 min run/walk tests. They found that the criterion-related validity of the run/walk tests, only considering the performance score, ranged from low to high, with the 1.5-mile and the 12 min run/walk tests being the most appropriate tests for estimating cardiorespiratory fitness in adults aged 19–64 years. Sex, age or VO_{2max} level did not affect criterion-related validity, whereas when multiple predictors (i.e., performance score, sex, age or body mass) were considered, the criterion-related validity values were higher. In this sense, two high-quality original studies reinforced these results, and showed that the 12 min [41] and the 1.5-mile [40,41] run/walk tests were fairly accurate for estimating cardiorespiratory fitness in adults aged 18–26 years (r = 0.87–0.93, $p < 0.05$).

Overall, the run/walk tests are not user-friendly tests, due to the difficulty of developing an appropriate pace, which may affect the test outcome (some participants start too fast, so they are unable to maintain their speed throughout the test; others start too slow, so when they wish to increase their speed the test is already finished). These problems are more likely to occur in longer distance tests. Other factors affecting the test outcome include the individual's willingness to endure the discomfort of strenuous exercise, a short attention span, poor motivation, and limited interest in a monotonous task [86–88].

The 2 km and 6 min walk tests are probably the most widely used walk tests in adults [39,51]. Both tests require submaximal effort, thus avoiding the problem of enduring the discomfort of strenuous exercise. In addition, it allows to evaluate those people with a low level of physical fitness or is unable to run. Three high-quality studies [36,37,39] observed that Oja's equation derived from the 2 km walk test has high validity (r = 0.80–0.87, all $p < 0.05$) in untrained and/or overweight/obese adults aged 20–64 years. One high-quality study reported that the 2 km walk test [38] is a reasonably valid field test for estimating the cardiorespiratory fitness of moderately active adults aged 35–45 years, but not in adults with very high maximal aerobic power.

Many studies developed prediction equations for the 6 min test based on spirometry [89]. However, only three high-quality studies [50–52] analysed the criterion-related validity of the 6 min test based on VO_{2max} in adults. They showed a moderate-to-high validity (r = 0.70–0.93, all $p < 0.001$) in obese and healthy adults aged 18–64 years. Burr et al. [90] suggested that, on its own, the 6 min walk test can be useful to discriminate between broad categories of high, moderate and low fitness, but that this approach may be associated with a degree of error, especially in the high fitness group.

According to these findings, the 2 km and 6 min walk tests are valid for use in adults aged 19–64 years with low or moderate fitness levels, but not in adults with a high fitness level.

Regarding the 1 mile walk test, conflicting results were found, especially when examining the accuracy of the Kline's [42] and Dolgener's [46] equations in adults aged 19–64 years.

4.1.2. Twenty-Metre Shuttle Run Test

The 20 m shuttle run test was developed by Leger at al. [91] to solve the pace issue of the run/walk tests. The test consists of 1 min stages of continuous running at an increasing speed. Recently, a meta-analysis [28] showed that the performance score of the 20 m shuttle run test had a moderate-to-high criterion-related validity for estimating VO_{2max} ($r_p = 0.66$–0.84) in youth and adults aged 18–64 years, higher than when other variables (i.e., sex, age or body mass) were accounted for ($r_p = 0.78$–0.95). This study also reported that Leger's protocol had a greater average criterion-related validity coefficient ($r_p = 0.84$; 95% CI: 0.80–0.89) than Eurofit, QUB and Dong-HO protocols; and Leger's protocol was statistically higher for adults ($r_p = 0.94$, 0.87–1.00) than for children ($r_p = 0.78$; 95% CI: 0.72–0.85). These values are higher than those reported for the 1500 m and 12 min run/walk tests [30]. Moreover, the meta-analysis showed that sex did not seem to affect the criterion-related validity values.

On the other hand, Cooper et al. [54] showed that Brewer's protocol and equation were not valid for assessing active young people aged 18–26 years (mean difference = 1.8 ± 6.3 mL/kg/min; $p = 0.004$). In line with these findings, Kim et al. [58] observed that Leger's protocol and equation were more accurate than Brewer's protocol and equation (mean difference -0.54 mL/kg/min; %CV: 1.39 vs. mean difference -2.944 mL/kg/min; %CV: 8.87) in Korean adults, especially in women. Nonetheless, the authors suggested the need to develop new equations for Korean adults.

It is important to note that the 20 m square shuttle run test [55,57] was proposed as an alternative to the 20 m shuttle run test to reduce the test's turning angle from 180 to 90. This test was the best predictor of VO_{2max} than the 20 m shuttle run test in young male adults aged 18–25 years.

4.1.3. Step Tests

Step tests are a safe, simple, inexpensive and practical method of assessing cardiorespiratory fitness under submaximal conditions, which require minimum space [32]; they are also a great alternative to laboratory tests in clinical settings. There are a wide variety of step test protocols which differ in terms of stepping frequency, test duration and number of test stages. Bennett et al. [32] analysed the criterion-related validity of different step tests (the Chester step test, a personalised step test, the STEP tool step test, the Queen's College step test, the Skubic and Hodgkins step test, a height-adjusted, rate-specific, single-state step test, the Astrand–Ryhming step test, and a modified YMCA 3 min step test) in adults aged 18–64 years. The validity of these tests ranged from moderate to high, and they suggested that the Chester step test was the most valid step test to evaluate cardiorespiratory fitness in adults. However, this systematic review only included two studies with contradictory results, similarly to the Queen's College step test.

Analysing the 12 high-quality studies that examined the criterion-related validity of the step tests in adults aged 19–64 years, we can conclude that the YMCA step test [67,71] seemed to be the most appropriate step test to estimate VO_{2max} in adults aged 19–64 years. However, it is important to note that there is no single equation, since the result of the equation depends on the sample used. Santo and Golding [92] even altered the protocol by adjusting the step height to the individual participant's height in order to increase the accuracy of this test.

4.1.4. Levels of Evidence

Strong evidence indicated that (a) the 20 m shuttle run test using Leger's equation, the 2 km walk using Oja's equation, the 6 min and the YMCA step tests are valid for estimating cardiorespiratory fitness; and (b) the criterion-related validity of the distance and time-based run/walk tests range from low to high, with the 1.5-mile and 12 min run/walk tests being the best predictors. Moderate evidence indicated that the 20 m square shuttle run test is valid for estimating cardiorespiratory fitness. Due to the inconsistent results found in high-quality studies, limited evidence was found for the validity of the 1-mile walk, treadmill jogging, incremental shuttle walking, Chester, and Queen's College step tests. Due to the low number of high-quality studies, limited evidence indicated that (a) the 3 min walk, the $\frac{1}{4}$-mile walk, Mankato submaximal, modified Astrand–Ryhming, University Montreal, modified Canadian aerobic fitness step, 6 min single 15 cm step, Tecumseh step, modified Harvard step and Astrand–Ryhming Step tests are valid for estimating cardiorespiratory fitness; and (b) the YMCA cycle, Ruffier, Danish step, and 2 min step tests are not valid for estimating cardiorespiratory fitness. Due to the consistent results found in multiple low-quality studies, limited evidence supported using the 6 min step test for estimating cardiorespiratory fitness.

4.2. Muscular Strength

The specificity of the type of muscular work performed and the use of different energy systems are both major challenges for establishing a gold standard method for maximal,

endurance and explosive muscular strength tests [93]. One repetition maximum (1RM) and repetitions to a certain percentage of 1RM (i.e., 50% of 1RM or 70% of 1RM) [27], isokinetic dynamometer strength [94–96], and electromyography [78,80] were used as gold standards.

4.2.1. Maximal Isometric Strength

The TKK dynamometer [73–75] seemed the most appropriate test to assess maximal isometric strength in adults. All the studies used the "known weights" as the criterion reference.

Several studies examined whether the elbow position (extended or flexed at 90 degrees) affected the hand maximal isometric strength score in children [75], adolescents [97] and young adults [98]. They observed that performing the handgrip strength test with the elbow extended seems the most appropriate protocol to evaluate hand maximal isometric strength in these populations—which is in accordance with the protocol recommended by the American Center for Disease Control and Prevention [99].

Ruiz et al. [100] also investigated whether the position (grip span) on the standard grip dynamometer determined the hand maximal isometric strength in adults. They found that when measuring hand maximal isometric strength in women, hand size must be taken into consideration, providing the mathematical equation ($y = x/5 + 1.5$ cm) to adapt optimal grip span (y) to hand size (x). In adult men, optimal grip span could be set at a fixed value (5.5 cm) and is not influenced by hand size.

Importantly, just like the step test, the handgrip strength test can be very useful in clinical settings because it requires minimal equipment and space, is time-efficient and easy to administer.

4.2.2. Endurance Strength

The Biering–Sørensen test, a trunk holding test in an antigravity prone position, is commonly used to measure the back and hip muscle endurance strength, which is associated with lower back pain [101]. Mannion et al. [77] and Coorevits et al. [78] showed that the test endurance time was highly associated with isometric/endurance hip and back musculature strength ($r = 0.84–98$, $p < 0.01$). On the other hand, Kankaanpää et al. [79], found that this association was moderate ($r = 0.60–0.71$, $p < 0.05$). However, when BMI ($r = -0.49–0.51$, $p < 0.001$) in women and age ($r = 0.25–0.29$, $p < 0.05$) in men were accounted for in the prediction model, the explained variance increased considerably. Thus, the Biering–Sørensen test might be considered as valid for measuring back muscle endurance strength.

Assessing abdominal muscle functionality is clinically relevant since it is considered to be related to lower back pain [102,103]. The curl-up test, or its different versions, was the field test originally used to assess this capacity. In the present review, no original studies evaluating the criterion-related validity of this test were classified as high quality. An alternative of the curl-up test could be the prone bridging test, an isometric holding test in prone position which is currently being used to supposedly measure abdominal endurance strength. The prone bridging test time is inversely associated with lower back pain [104,105]. In relation to the validity of this test, De Blaiser et al. [80] found a higher activation of the abdominal core musculature during the test than for the back and hip musculature, showing a high association between test time and abdominal endurance strength. Future high-quality studies are necessary to clarify the validity of this test.

It should be noted that no study that analysed the criterion-related validity of lower and upper body endurance strength tests were classified as high quality.

4.2.3. Explosive Strength

The standing long jump is proposed in health-related fitness test batteries in preschool children [106], as well as children and adolescents [107] to assess lower body explosive strength, given its criterion-related and predictive validity. However, to our knowledge, the criterion-related validity of this test has not been studied. Bui et al. observed that the Sargent jump test [81] is not appropriate to evaluate lower body explosive strength,

because its overestimates the height of a vertical jump and its accuracy is reduced as the jump height increases (mean difference: 4.4 ± 5.1, $p < 0.001$). Due to the close relationship that lower body maximal/explosive strength has on adult health [22,23], more high-quality studies are required to analyse the criterion-related validity of these tests in future research.

4.2.4. Levels of Evidence

Strong evidence indicated that (a) the handgrip strength test with the elbow extended and with the grip span adapted to the hand size and sex (using the TKK dynamometer) is a valid test for assessing hand maximal isometric strength; and (b) the Biering–Sørensen test offers a valid test for assessing endurance strength of hip and back muscles. Moderate evidence indicated that handgrip strength (Jamar) has acceptable validity for assessing hand maximal isometric strength. Due to (a) the low number of high-quality studies, limited evidence (only one study) was found supporting the use of prone bridging for assessing abdominal endurance strength and the Sargent jump test for assessing lower body explosive strength; (b) the inconsistent results found in multiple high-quality studies, limited evidence was found for the validity of using handgrip strength (DynEx) for assessing hand maximal isometric strength; and (c) the consistent results found in multiple low or very low-quality studies, the curl-up test, or its different versions, are not valid for assessing abdominal endurance strength.

4.3. Flexibility

Radiography seems to be the best criterion measurement of flexibility, but goniometry is also used as a criterion measure [108,109].

Goniometers are relatively easy to obtain; nevertheless, their use requires a certain technical qualification since it is a sensitive method, and thus it is not feasible for use in all settings [110]. Traditionally, the sit-and-reach test, originally designed by Wells and Dillon [111], and its different versions, are included in the fitness test batteries for measuring hamstring and lower back flexibility, which are probably the most widely used measures of flexibility [27].

Mayorga et al. [31] performed a meta-analysis to analyse the criterion-related validity of the sit-and-reach and its different versions (modified sit-and-reach, back-saver sit-and-reach, modified back-saver sit-and-reach, V sit-and-reach, modification V sit-and-reach, unilateral sit-and-reach and chair sit-and-reach). These tests showed moderate validity for estimating hamstring extensibility, but low validity for estimating lumbar extensibility. They also found that the classic sit-and-reach test had the highest criterion-related validity coefficient in both hamstring and lumbar extensibility, compared to the other test, which does not seem to justify the use of the classic protocol modifications in order to solve the problems attributed to itself (i.e., the length proportion between the upper and lower limbs or the position of the head and ankles).

The toe-touch test is another field-based test for measuring hamstring flexibility, in which the individuals were assessed standing instead of sitting on the floor [112]. Although this test is easy to administer and can be an alternative to the sit-and-reach test, when the participant has problems being measured sitting, it is not proposed for any filed-based fitness test battery. A meta-analysis [29] analysed the criterion-related validity of the toe-touch test for measuring hamstring flexibility, reporting similar validity coefficients to those of the classic sit-and-reach.

It is interesting to highlight that Nuzzo [113] has recently suggested that flexibility should be invalidated as a major component of fitness, due to its lack of predictive and concurrent validity in terms of meaningful health and performance outcomes.

Levels of Evidence

Strong evidence indicated that (a) the sit-and-reach test and its modified versions have moderate validity for estimating hamstring extensibility, but low validity for estimating

lumbar extensibility; and (b) the toe-to-touch test has moderate validity for estimating hamstring extensibility.

4.4. Motor Fitness

The validity of motor fitness tests is the least studied in adults. None of the three studies that analysed the criterion-related validity in motor fitness tests were classified as high quality. Given that the motor fitness tests (i.e., gait/walking speed, balance, timed up and go) are associated with all-cause mortality [114–116], falls and fractures [117], disability in activities of daily living [118] and depression [119], it would be useful to know their criterion-related validity.

Levels of Evidence

Due to the consistent results found in multiple low-quality studies, we found limited evidence that the ten-step test had moderate validity in assessing agility.

5. Conclusions

The systematic review emphasized important major points regarding the criterion-related validity of adult field-based fitness tests (Figure 2):
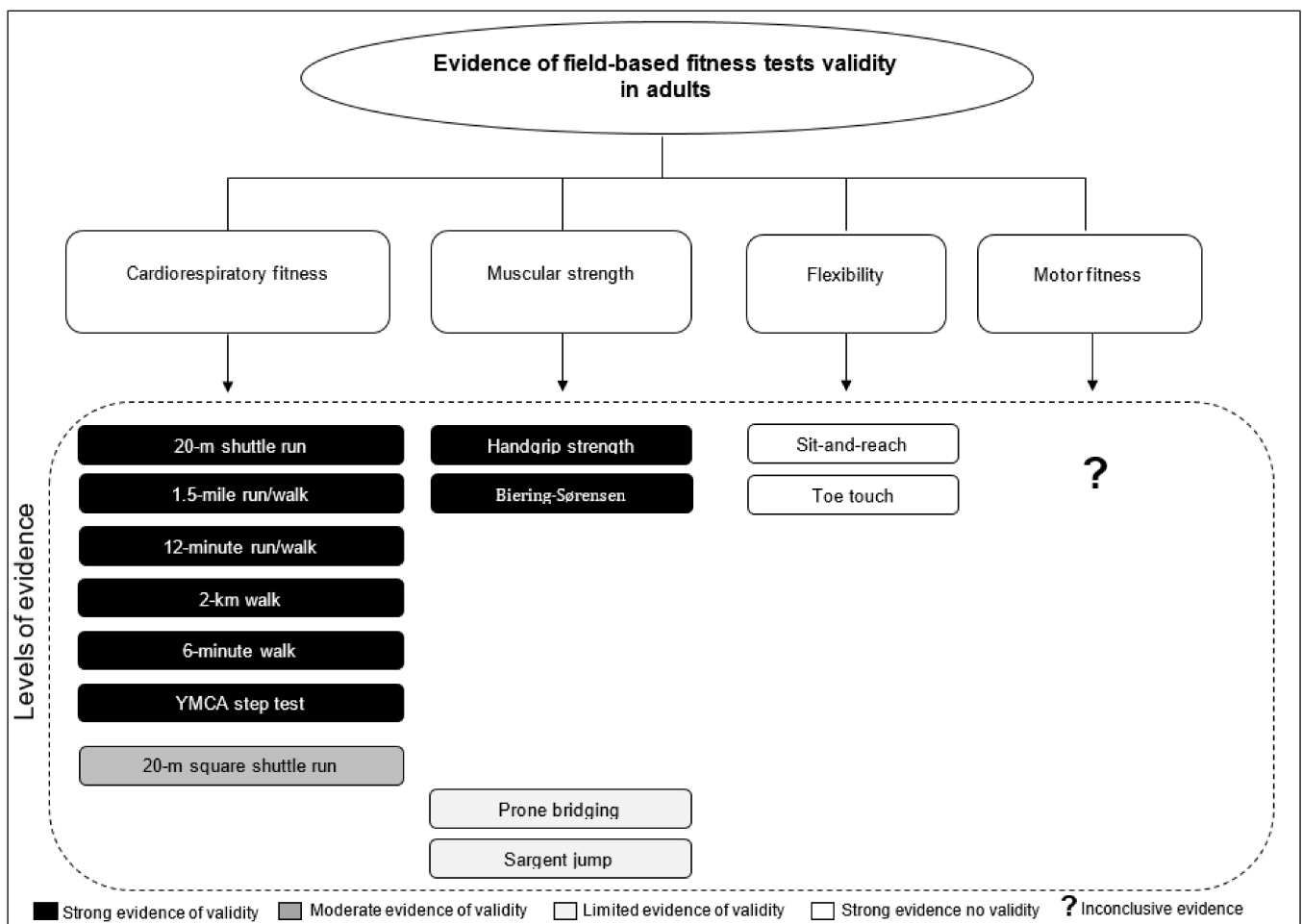

Figure 2. Major points regarding criterion-related validity of adult field-based fitness tests.

Cardiorespiratory fitness: the 20 m shuttle run tests best assessed cardiorespiratory fitness using Leger's equation. Alternatively, the 1.5-mile, 12 min run/walk and YMCA step tests were other cardiorespiratory testing options. When low-level cardiorespiratory

fitness existed, or if running was possible, the 2 km, then Oja's equation or 6 min walk tests were appropriate alternatives.

Muscular strength: strong evidence indicated that (a) the handgrip strength test, with the elbow extended and with the grip span adapted to the individual's hand size (using the TKK dynamometer), offers a valid means to assess hand maximal isometric strength; and (b) the Biering–Sørensen test estimated the endurance strength of hip and back muscles. Limited evidence (only one study) supported the prone bridging and Sargent jump tests as abdominal endurance strength and lower body explosive strength surrogate markers, respectively.

Flexibility: strong evidence supported the sit-and-reach test and its different versions, and that the toe-to-touch tests is not valid for assessing hamstring and lower back flexibility.

Motor fitness: limited evidence about the criterion-related validity of motor fitness existed.

When there are problems of space and time, as in clinical settings, the YMCA step and the handgrip strength tests are good alternatives for assessing cardiorespiratory fitness and isometric muscular strength, respectively.

Author Contributions: J.C.-P. and M.C.-G. conceived the study idea. J.C.-P. led the writing of the review and carried out methodological aspects with N.M.-J., F.M.-A. and J.R.F.-S., F.M.-A., V.S.-J., R.I.-G. and J.R.R. contributed writing—review and editing the final manuscript. All authors discussed the results and contributed to the final manuscript, and agreed with the order of presentation of the authors. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Castillo Garzon, M.J.; Ortega Porcel, F.B.; Ruiz Ruiz, J. Improvement of physical fitness as anti-aging intervention. *Med. Clin.* **2005**, *124*, 146–155.
2. LaMonte, M.J.; Barlow, C.E.; Jurca, R.; Kampert, J.B.; Church, T.S.; Blair, S.N. Cardiorespiratory fitness is inversely associated with the incidence of metabolic syndrome: A prospective study of men and women. *Circulation* **2005**, *112*, 505–512. [CrossRef]
3. Fung, M.D.; Canning, K.L.; Mirdamadi, P.; Ardern, C.I.; Kuk, J.L. Lifestyle and weight predictors of a healthy overweight profile over a 20-year follow-up. *Obesity* **2015**, *23*, 1320–1325. [CrossRef] [PubMed]
4. Howe, T.E.; Rochester, L.; Neil, F.; Skelton, D.A.; Ballinger, C. Exercise for improving balance in older people. *Cochrane Database Syst. Rev.* **2011**, *11*, Cd004963. [CrossRef] [PubMed]
5. Balducci, S.; Cardelli, P.; Pugliese, L.; D'Errico, V.; Haxhi, J.; Alessi, E.; Iacobini, C.; Menini, S.; Bollanti, L.; Conti, F.G.; et al. Volume-dependent effect of supervised exercise training on fatty liver and visceral adiposity index in subjects with type 2 diabetes The Italian Diabetes Exercise Study (IDES). *Diabetes Res. Clin. Pract.* **2015**, *109*, 355–363. [CrossRef] [PubMed]
6. Pletnikoff, P.P.; Laukkanen, J.A.; Tuomainen, T.P.; Kauhanen, J.; Rauramaa, R.; Ronkainen, K.; Kurl, S. Cardiorespiratory fitness, C-reactive protein and lung cancer risk: A prospective population-based cohort study. *Eur. J. Cancer* **2015**, *51*, 1365–1370. [CrossRef] [PubMed]
7. Sui, X.; Lee, D.C.; Matthews, C.E.; Adams, S.A.; Hebert, J.R.; Church, T.S.; Lee, C.D.; Blair, S.N. Influence of cardiorespiratory fitness on lung cancer mortality. *Med. Sci. Sports Exerc.* **2010**, *42*, 872–878. [CrossRef]
8. Blair, S.N.; Kohl, H.W., 3rd; Paffenbarger, R.S., Jr.; Clark, D.G.; Cooper, K.H.; Gibbons, L.W. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* **1989**, *262*, 2395–2401. [CrossRef]
9. Farrell, S.W.; Fitzgerald, S.J.; McAuley, P.A.; Barlow, C.E. Cardiorespiratory fitness, adiposity, and all-cause mortality in women. *Med. Sci. Sports Exerc.* **2010**, *42*, 2006–2012. [CrossRef]

10. Barry, V.W.; Baruth, M.; Beets, M.W.; Durstine, J.L.; Liu, J.; Blair, S.N. Fitness vs. fatness on all-cause mortality: A meta-analysis. *Prog. Cardiovasc. Dis.* **2014**, *56*, 382–390. [CrossRef]

11. Ortega, F.B.; Lavie, C.J.; Blair, S.N. Obesity and Cardiovascular Disease. *Circ. Res.* **2016**, *118*, 1752–1770. [CrossRef]

12. McAuley, P.A.; Beavers, K.M. Contribution of cardiorespiratory fitness to the obesity paradox. *Prog. Cardiovasc. Dis.* **2014**, *56*, 434–440. [CrossRef]

13. Lavie, C.J.; McAuley, P.A.; Church, T.S.; Milani, R.V.; Blair, S.N. Obesity and cardiovascular diseases: Implications regarding fitness, fatness, and severity in the obesity paradox. *J. Am. Coll. Cardiol.* **2014**, *63*, 1345–1354. [CrossRef]

14. Lavie, C.J.; Ozemek, C.; Carbone, S.; Katzmarzyk, P.T.; Blair, S.N. Sedentary Behavior, Exercise, and Cardiovascular Health. *Circ. Res.* **2019**, *124*, 799–815. [CrossRef]

15. Barry, V.W.; Caputo, J.L.; Kang, M. The Joint Association of Fitness and Fatness on Cardiovascular Disease Mortality: A Meta-Analysis. *Prog. Cardiovasc. Dis.* **2018**, *61*, 136–141. [CrossRef]

16. Oktay, A.A.; Lavie, C.J.; Kokkinos, P.F.; Parto, P.; Pandey, A.; Ventura, H.O. The Interaction of Cardiorespiratory Fitness With Obesity and the Obesity Paradox in Cardiovascular Disease. *Prog. Cardiovasc. Dis.* **2017**, *60*, 30–44. [CrossRef] [PubMed]

17. Delextrat, A.A.; Warner, S.; Graham, S.; Neupert, E. An 8-Week Exercise Intervention Based on Zumba Improves Aerobic Fitness and Psychological Well-Being in Healthy Women. *J. Phys. Act. Health* **2016**, *13*, 131–139. [CrossRef] [PubMed]

18. Ortega, F.B.; Lee, D.C.; Sui, X.; Kubzansky, L.D.; Ruiz, J.R.; Baruth, M.; Castillo, M.J.; Blair, S.N. Psychological well-being, cardiorespiratory fitness, and long-term survival. *Am. J. Prev. Med.* **2010**, *39*, 440–448. [CrossRef] [PubMed]

19. Boots, E.A.; Schultz, S.A.; Oh, J.M.; Larson, J.; Edwards, D.; Cook, D.; Koscik, R.L.; Dowling, M.N.; Gallagher, C.L.; Carlsson, C.M.; et al. Cardiorespiratory fitness is associated with brain structure, cognition, and mood in a middle-aged cohort at risk for Alzheimer's disease. *Brain Imaging Behav.* **2015**, *9*, 639–649. [CrossRef] [PubMed]

20. Willis, B.L.; Gao, A.; Leonard, D.; Defina, L.F.; Berry, J.D. Midlife fitness and the development of chronic conditions in later life. *Arch. Intern. Med.* **2012**, *172*, 1333–1340. [CrossRef] [PubMed]

21. Zhu, N.; Jacobs, D.R., Jr.; Schreiner, P.J.; Launer, L.J.; Whitmer, R.A.; Sidney, S.; Demerath, E.; Thomas, W.; Bouchard, C.; He, K.; et al. Cardiorespiratory fitness and brain volume and white matter integrity: The CARDIA Study. *Neurology* **2015**, *84*, 2347–2353. [CrossRef] [PubMed]

22. Garcia-Hermoso, A.; Cavero-Redondo, I.; Ramirez-Velez, R.; Ruiz, J.R.; Ortega, F.B.; Lee, D.C.; Martinez-Vizcaino, V. Muscular Strength as a Predictor of All-Cause Mortality in an Apparently Healthy Population: A Systematic Review and Meta-Analysis of Data From Approximately 2 Million Men and Women. *Arch. Phys. Med. Rehabil.* **2018**, *99*, 2100–2113.e5. [CrossRef]

23. Garcia-Hermoso, A.; Ramirez-Velez, R.; Peterson, M.D.; Lobelo, F.; Cavero-Redondo, I.; Correa-Bautista, J.E.; Martinez-Vizcaino, V. Handgrip and knee extension strength as predictors of cancer mortality: A systematic review and meta-analysis. *Scand. J. Med. Sci. Sports* **2018**, *28*, 1852–1858. [CrossRef] [PubMed]

24. Kettunen, O.; Kyrolainen, H.; Santtila, M.; Vasankari, T. Physical fitness and volume of leisure time physical activity relate with low stress and high mental resources in young men. *J. Sports Med. Phys. Fit.* **2014**, *54*, 545–551.

25. Currell, K.; Jeukendrup, A.E. Validity, reliability and sensitivity of measures of sporting performance. *Sports Med.* **2008**, *38*, 297–316. [CrossRef] [PubMed]

26. Docherty, D. Field tests and test batteries. In *Measurement in Pediatric Exercise Science*; Docherty, D., Ed.; Human Kinetics: Champaign, IL, USA, 1996; pp. 285–334.

27. Castro-Pinero, J.; Artero, E.G.; Espana-Romero, V.; Ortega, F.B.; Sjostrom, M.; Suni, J.; Ruiz, J.R. Criterion-related validity of field-based fitness tests in youth: A systematic review. *Br. J. Sports Med.* **2009**, *44*, 934–943. [CrossRef]

28. Mayorga-Vega, D.; Aguilar-Soto, P.; Viciana, J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. *J. Sports Sci. Med.* **2015**, *14*, 536–547.

29. Mayorga-Vega, D.; Viciana, J.; Cocca, A.; Merino-Marban, R. Criterion-related validity of toe-touch test for estimating hamstring extensibility: A metaanalysis. *J. Hum. Sport Exerc.* **2014**, *9*, 188–200. [CrossRef]

30. Mayorga-Vega, D.; Bocanegra-Parrilla, R.; Ornelas, M.; Viciana, J. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: A systematic review and Meta-analysis. *PLoS ONE* **2016**, *11*, e0151671. [CrossRef]

31. Mayorga-Vega, D.; Merino-Marban, R.; Viciana, J. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *J. Sports Sci. Med.* **2014**, *13*, 1–14. [PubMed]

32. Bennett, H.; Parfitt, G.; Davison, K.; Eston, R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. *Sports Med.* **2016**, *46*, 737–750. [CrossRef]

33. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef]

34. Shea, B.J.; Hamel, C.; Wells, G.A.; Bouter, L.M.; Kristjansson, E.; Grimshaw, J.; Henry, D.A.; Boers, M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J. Clin. Epidemiol.* **2009**, *62*, 1013–1020. [CrossRef] [PubMed]

35. Biddle, S.J.; García Bengoechea, E.; Wiesner, G. Sedentary behaviour and adiposity in youth: A systematic review of reviews and analysis of causality. *Int. J. Behav. Nutr. Phys. Act.* **2017**, *14*, 43. [CrossRef] [PubMed]

36. Oja, P.; Laukkanen, R.; Pasanen, M.; Tyry, T.; Vuori, I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *Int. J. Sports Med.* **1991**, *12*, 356–362. [CrossRef]

37. Laukkanen, R.; Oja, P.; Pasanen, M.; Vuori, I. Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults. *Int. J. Obes. Relat. Metab. Disord.* **1992**, *16*, 263–268.

38. Laukkanen, R.M.T.; Oja, P.; Pasanen, M.E.; Vuori, I.M. Criterion validitv of a two-kilometer walking test for predicting the maximal oxygen uptake of moderately to highly active middle-aged adults. *Scand. J. Med. Sci. Sports* **1993**, *3*, 267–272. [CrossRef]

39. Laukkanen, R.M.T.; Kukkonen-Harjula, T.K.; Oja, P.; Pasanen, M.E.; Vuori, I.M. Prediction of change in maximal aerobic power by the 2-km walk test after walking training in middle-aged adults. *Int. J. Sports Med.* **2000**, *21*, 113–116. [CrossRef]

40. Larsen, G.E.; George, J.D.; Alexander, J.L.; Fellingham, G.W.; Aldana, S.G.; Parcell, A.C. Prediction of maximum oxygen consumption from walking, jogging, or running. *Res. Q. Exerc. Sport* **2002**, *73*, 66–72. [CrossRef]

41. McNaughton, L.; Hall, P.; Cooley, D. Validation of several methods of estimating maximal oxygen uptake in young men. *Percept. Mot. Ski.* **1998**, *87*, 575–584. [CrossRef]

42. Kline, C.; Porcari, J.P.; Hintermeister, R.; Freedson, P.S.; Ward, A.; McCarron, R.F.; Ross, J.; Rippe, J. Estimation of from a one-mile track walk, gender, age and body weight. *Med. Sports Exerc.* **1987**, *19*, 253–259.

43. George, J.D.; Fellingham, G.W.; Fisher, A.G. A modified version of the Rockport Fitness Walking Test for college men and women. *Res. Q. Exerc. Sport* **1998**, *69*, 205–209. [CrossRef]

44. Lunt, H.; Roiz De Sa, D.; Roiz De Sa, J.; Allsopp, A. Validation of one-mile walk equations for the estimation of aerobic fitness in British military personnel under the age of 40 years. *Mil. Med.* **2013**, *178*, 753–759. [CrossRef]

45. Greenhalgh, H.A.; George, J.D.; Hager, R.L. Cross-validation of a quarter-mile walk test using two VO2 max regression models. *Meas. Phys. Educ. Exerc. Sci.* **2001**, *5*, 139–151. [CrossRef]

46. Dolgener, F.A.; Hensley, L.D.; Marsh, J.J.; Fjelstul, J.K. Validation of the Rockport Fitness Walking Test in college males and females. *Res. Q. Exerc. Sport* **1994**, *65*, 152–158. [CrossRef]

47. Seneli, R.M.; Ebersole, K.T.; O'Connor, K.M.; Snyder, A.C. Estimated VO2max from the Rockport Walk Test on a Nonmotorized Curved Treadmill. *J. Strength Cond. Res.* **2013**, *27*, 3495–3505. [CrossRef] [PubMed]

48. George, J.D.; Vehrs, P.R.; Allsen, P.E.; Fellingham, G.W.; Fisher, A.G. Development of a submaximal treadmill jogging test for fit college-aged individuals. *Med. Sci. Sports Exerc.* **1993**, *25*, 643–647. [CrossRef]

49. Cao, Z.-B.; Miyatake, N.; Aoyama, T.; Higuchi, M.; Tabata, I. Prediction of maximal oxygen uptake from a 3-minute walk based on gender, age, and body composition. *J. Phys. Act. Health* **2013**, *10*, 280–287. [CrossRef]

50. Di Thommazo-Luporini, L.; Pinheiro Carvalho, L.; Luporini, R.; Trimer, R.; Falasco Pantoni, C.B.; Catai, A.M.; Arena, R.; Borghi-Silva, A. The six-minute step test as a predictor of cardiorespiratory fitness in obese women. *Eur. J. Phys. Rehabil. Med.* **2015**, *51*, 793–802. [PubMed]

51. Di Thommazo-Luporini, L.; Carvalho, L.P.; Luporini, R.L.; Trimer, R.; Falasco Pantoni, C.B.; Martinez, A.F.; Catai, A.M.; Arena, R.; Borghi-Silva, A. Are cardiovascular and metabolic responses to field walking tests interchangeable and obesity-dependent? *Disabil. Rehabil.* **2016**, *38*, 1820–1829. [CrossRef]

52. Manttari, A.; Suni, J.; Sievanen, H.; Husu, P.; Vaha-Ypya, H.; Valkeinen, H.; Tokola, K.; Vasankari, T. Six-minute walk test: A tool for predicting maximal aerobic power (VO2 max) in healthy adults. *Clin. Physiol. Funct. Imaging* **2018**. [CrossRef] [PubMed]

53. Bonet, J.B.; Magalhaes, J.; Viscor, G.; Pages, T.; Javierre, C.F.; Torrella, J.R. A field tool for the aerobic power evaluation of middle-aged female recreational runners. *Women Health* **2020**, *60*, 839–848. [CrossRef] [PubMed]

54. Cooper, S.M.; Baker, J.S.; Tong, R.J.; Roberts, E.; Hanford, M. The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *Br. J. Sports Med.* **2005**, *39*, e19. [CrossRef] [PubMed]

55. Flouris, A.D.; Koutedakis, Y.; Nevill, A.; Metsios, G.S.; Tsiotra, G.; Parasiris, Y. Enhancing specificity in proxy-design for the assessment of bioenergetics. *J. Sci. Med. Sport* **2004**, *7*, 197–204. [CrossRef]

56. Metsios, G.S.; Flouris, A.D.; Koutedakis, Y.; Nevill, A. Criterion-related validity and test-retest reliability of the 20m square shuttle test. *J. Sci. Med. Sport* **2008**, *11*, 214–217. [CrossRef]

57. Flouris, A.D.; Metsios, G.S.; Koutedakis, Y. Enhancing the efficacy of the 20 m multistage shuttle run test. *Br. J. Sports Med.* **2005**, *39*, 166–170. [CrossRef]

58. Kim, J.; Jung, S.H.; Cho, H.C. Validity and Reliability of Shuttle-Run Test in Korean Adults. *Int. J. Sports Med.* **2011**, *32*, 580–585. [CrossRef]

59. Jurio-Iriarte, B.; Gorostegi-Anduaga, I.; Rodrigo Aispuru, G.; Perez-Asenjo, J.; Brubaker, P.H.; Maldonado-Martin, S. Association between Modified Shuttle Walk Test and cardiorespiratory fitness in overweight/obese adults with primary hypertension: EXERDIET-HTA study. *J. Am. Soc. Hypertens.* **2017**, *11*, 186–195. [CrossRef] [PubMed]

60. Jurio-Iriarte, B.; Brubaker, P.H.; Gorostegi-Anduaga, I.; Corres, P.; Martinez Aguirre-Betolaza, A.; Maldonado-Martin, S. Validity of the modified shuttle walk test to assess cardiorespiratory fitness after exercise intervention in overweight/obese adults with primary hypertension. *Clin. Exp. Hypertens.* **2018**, *41*, 336–341. [CrossRef] [PubMed]

61. Lima, L.P.; Leite, H.R.; Matos, M.A.; Neves, C.D.C.; Lage, V.; Silva, G.P.D.; Lopes, G.S.; Chaves, M.G.A.; Santos, J.N.V.; Camargos, A.C.R.; et al. Cardiorespiratory fitness assessment and prediction of peak oxygen consumption by Incremental Shuttle Walking Test in healthy women. *PLoS ONE* **2019**, *14*, e0211327. [CrossRef]

62. Aadahl, M.; Zacho, M.; Linneberg, A.; Thuesen, B.H.; Jorgensen, T. Comparison of the Danish step test and the watt-max test for estimation of maximal oxygen uptake: The Health2008 study. *Eur. J. Prev. Cardiol.* **2013**, *20*, 1088–1094. [CrossRef]

63. Kumar, S.K.; Khare, P.; Jaryal, A.K.; Talwar, A. Validity of heart rate based nomogram fors estimation of maximum oxygen uptake in Indian population. *Indian J. Physiol. Pharmacol.* **2012**, *56*, 279–283. [PubMed]

64. Ricci, P.A.; Cabiddu, R.; Jürgensen, S.P.; André, L.D.; Oliveira, C.R.; Di Thommazo-Luporini, L.; Ortega, F.P.; Borghi-Silva, A. Validation of the two-minute step test in obese with comorbibities and morbidly obese patients. *Braz. J. Med Biol. Res.* **2019**, *52*, e8402. [CrossRef] [PubMed]

65. Weller, I.M.; Thomas, S.G.; Cox, M.H.; Corey, P.N. A study to validate the Canadian Aerobic Fitness Test. *Can. J. Public Health* **1992**, *83*, 120–124. [CrossRef]

66. Carvalho, L.P.; Di Thommazo-Luporini, L.; Aubertin-Leheudre, M.; Bonjorno Junior, J.C.; de Oliveira, C.R.; Luporini, R.L.; Mendes, R.G.; Lopes Zangrando, K.T.; Trimer, R.; Arena, R.; et al. Prediction of cardiorespiratory fitness by the six-minute step test and its association with muscle strength and power in sedentary obese and lean young women: A cross-sectional study. *PLoS ONE* **2015**, *10*, e0145960. [CrossRef] [PubMed]

67. Teren, A.; Zachariae, S.; Beutner, F.; Ubrich, R.; Sandri, M.; Engel, C.; Loeffler, M.; Gielen, S. Incremental value of veterans specific activity questionnaire and the ymca-step test for the assessment of cardiorespiratory fitness in population-based studies. *Eur. J. Prev. Cardiol.* **2016**, *23*, 1221–1227. [CrossRef] [PubMed]

68. Beutner, F.; Ubrich, R.; Zachariae, S.; Engel, C.; Sandri, M.; Teren, A.; Gielen, S. Validation of a brief step-test protocol for estimation of peak oxygen uptake. *Eur. J. Prev. Cardiol.* **2015**, *22*, 503–512. [CrossRef] [PubMed]

69. Lee, O.; Lee, S.; Kang, M.; Mun, J.; Chung, J. Prediction of maximal oxygen consumption using the Young Men's Christian Association-step test in Korean adults. *Eur. J. Appl. Physiol.* **2019**, *119*, 1245–1252. [CrossRef] [PubMed]

70. Kieu, N.T.V.; Jung, S.J.; Shin, S.W.; Jung, H.W.; Jung, E.S.; Won, Y.H.; Kim, Y.G.; Chae, S.W. The validity of the YMCA 3-minute step test for estimating maximal oxygen uptake in healthy Korean and Vietnamese adults. *J. Lifestyle Med.* **2020**, *10*, 21–29. [CrossRef]

71. Hong, S.H.; Yang, H.I.; Kim, D.I.; Gonzales, T.I.; Brage, S.; Jeon, J.Y. Validation of submaximal step tests and the 6-min walk test for predicting maximal oxygen consumption in young and healthy participants. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4858. [CrossRef]

72. Hansen, D.; Jacobs, N.; Thijs, H.; Dendale, P.; Claes, N. Validation of a single-stage fixed-rate step test for the prediction of maximal oxygen uptake in healthy adults. *Clin. Physiol. Funct. Imaging* **2016**, *36*, 401–406. [CrossRef] [PubMed]

73. Espana-Romero, V.; Artero, E.G.; Santaliestra-Pasias, A.M.; Gutierrez, A.; Castillo, M.J.; Ruiz, J.R. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *J. Hand Surg.* **2008**, *33*, 378–384. [CrossRef] [PubMed]

74. Cadenas-Sanchez, C.; Sanchez-Delgado, G.; Martinez-Tellez, B.; Mora-Gonzalez, J.; Löf, M.; España-Romero, V.; Ruiz, J.R.; Ortega, F.B. Reliability and validity of different models of TKK hand dynamometers. *Am. J. Occup. Ther.* **2016**, *70*, 7004300010. [CrossRef] [PubMed]

75. Kolimechkov, S.; Castro-Piñero, J.; Petrov, A.; Alexandrova, A. The effect of elbow position on the handgrip strength test in children: Validity and reliability of TKK 5101 and DynX dynamometers. *Pedagog. Phys Cult Sports* **2020**, *24*, 240–247. [CrossRef]

76. Shechtman, O.; Gestewitz, L.; Kimble, C. Reliability and validity of the DynEx dynamometer. *J. Hand Ther.* **2005**, *18*, 339–347. [CrossRef] [PubMed]

77. Mannion, A.F.; Dolan, P. Electromyographic median frequency changes during isometric contraction of the back extensors to fatigue. *Spine* **1994**, *19*, 1223–1229. [CrossRef]

78. Coorevits, P.; Danneels, L.; Cambier, D.; Ramon, H.; Vanderstraeten, G. Assessment of the validity of the Biering-Sørensen test for measuring back muscle fatigue based on EMG median frequency characteristics of back and hip muscles. *J. Electromyogr. Kinesiol.* **2008**, *18*, 997–1005. [CrossRef]

79. Kankaanpää, M.; Laaksonen, D.; Taimela, S.; Kokko, S.M.; Airaksinen, O.; Hänninen, O. Age, sex, and body mass index as determinants of back and hip extensor fatigue in the isometric Sørensen back endurance test. *Arch. Phys. Med. Rehabil.* **1998**, *79*, 1069–1075. [CrossRef]

80. De Blaiser, C.; De Ridder, R.; Willems, T.; Danneels, L.; Roosen, P. Reliability and validity of trunk flexor and trunk extensor strength measurements using handheld dynamometry in a healthy athletic population. *Phys. Ther. Sport* **2018**, *34*, 180–186. [CrossRef]

81. Bui, H.T.; Farinas, M.-I.; Fortin, A.-M.; Comtois, A.-S.; Leone, M. Comparison and analysis of three different methods to evaluate vertical jump height. *Clin. Physiol. Funct. Imaging* **2015**, *35*, 203–209. [CrossRef]

82. Kawano, M.M.; Ambar, G.; Oliveira, B.I.R.; Boer, M.C.; Cardoso, A.P.R.G.; Cardoso, J.R. Influence of the gastrocnemius muscle on the sit-and-reach test assessed by angular kinematic analysis. *Braz. J. Phys. Ther.* **2010**, *14*, 10–15. [CrossRef] [PubMed]

83. Bassett, D.R., Jr.; Howley, E.T.; Thompson, D.L.; King, G.A.; Strath, S.J.; McLaughlin, J.E.; Parr, B.B. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. *J. Appl. Physiol.* **2001**, *91*, 218–224. [CrossRef] [PubMed]

84. Meredith, M.D.; Welk, G.J. *Fitnessgram & Activitygram Test Administration Manual*, 4th ed.; Human Kinetics: Champaign, IL, USA, 2010.

85. Ruiz, J.R.; Ortega, F.B.; Castro-Piñero, J. Validity and reliability of the 1/4 mile run-walk test in physically active children and adolescents. *Nutr. Hosp.* **2014**, *31*, 875–882.

86. Krahenbuhl, G.S.; Pangrazi, R.P.; Burkett, L.N.; Schneider, M.J.; Petersen, G. Field estimation of VO2 max in children eight years of age. *Med. Sci. Sports* **1977**, *9*, 37–40. [CrossRef] [PubMed]

87. McCormack, W.P.; Cureton, K.J.; Bullock, T.A.; Weyand, P.G. Metabolic determinants of 1-mile run/walk performance in children. *Med. Sci. Sports Exerc.* **1991**, *23*, 611–617. [CrossRef]

88. Shephard, R.J. Tests of maximum oxygen intake. A critical review. *Sports Med.* **1984**, *1*, 99–124. [CrossRef]

89. Duncan, M.J.; Mota, J.; Carvalho, J.; Nevill, A.M. An Evaluation of Prediction Equations for the 6 Minute Walk Test in Healthy European Adults Aged 50–85 Years. *PLoS ONE* **2015**, *10*, e0139629.

90. Burr, J.F.; Bredin, S.S.; Faktor, M.D.; Warburton, D.E. The 6-minute walk test as a predictor of objectively measured aerobic fitness in healthy working-aged adults. *Physician Sportsmed.* **2011**, *39*, 133–139. [CrossRef]

91. Leger, L.A.; Mercier, D.; Gadoury, C.; Lambert, J. The multistage 20 m shuttle run test for aerobic fitness. *J. Sports Sci.* **1988**, *6*, 93–101. [CrossRef]

92. Santo, A.; Golding, L.A. Predicting maximum oxygen uptake from a modified 3-minute step test. *Res. Q. Exer. Sport* **2003**, *74*, 110–115. [CrossRef]

93. Mayhew, J.L.; Ball, T.E.; Ward, T.E.; Hart, C.L.; Arnold, M.D. Relationships of structural dimensions to bench press strength in college males. *J. Sports Med. Phys. Fit.* **1991**, *31*, 135–141.

94. Stark, T.; Walker, B.; Phillips, J.K.; Fejer, R.; Beck, R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: A systematic review. *PM R.* **2011**, *3*, 472–479. [CrossRef] [PubMed]

95. Paul, D.J.; Nassis, G.P. Testing strength and power in soccer players: The application of conventional and traditional methods of assessment. *J. Strength Cond. Res.* **2015**, *29*, 1748–1758. [CrossRef] [PubMed]

96. De Ste Croix, M.; Deighan, M.; Armstrong, N. Assessment and interpretation of isokinetic muscle strength during growth and maturation. *Sports Med.* **2003**, *33*, 727–743. [CrossRef] [PubMed]

97. España-Romero, V.; Ortega, F.B.; Vicente-Rodríguez, G.; Artero, E.G.; Rey, J.P.; Ruiz, J.R. Elbow position affects handgrip strength in adolescents: Validity and reliability of Jamar, DynEx, and TKK dynamometers. *J. Strength Cond. Res.* **2010**, *24*, 272–277. [CrossRef] [PubMed]

98. Balogun, J.A.; Akomolafe, C.T.; Amusa, L.O. Grip strength: Effects of testing posture and elbow position. *Arch. Phys. Med. Rehabil.* **1991**, *72*, 280–283. [PubMed]

99. *NHANES, Muscle Strength Procedures Manual*; National Health and Nutrition Examination Survey (NHANES); CDC: Druid Hills, GA, USA, 2013.

100. Ruiz-Ruiz, J.; Mesa, J.L.; Gutiérrez, A.; Castillo, M.J. Hand size influences optimal grip span in women but not in men. *J. Hand Surg.* **2002**, *27*, 897–901. [CrossRef] [PubMed]

101. Kim, W.J.; Kim, K.J.; Song, D.G.; Lee, J.S.; Park, K.Y.; Lee, J.W.; Chang, S.H.; Choy, W.S. Sarcopenia and Back Muscle Degeneration as Risk Factors for Back Pain: A Comparative Study. *Asian Spine J.* **2020**, *14*, 364–372. [CrossRef]

102. Abdelraouf, O.R.; Abdel-Aziem, A.A. The relationship between core endurance and back dysfunction in collegiate male athletes with and without nonspecific low back pain. *Int. J. Sports Phys. Ther.* **2016**, *11*, 337–344.

103. Ozcan Kahraman, B.; Salik Sengul, Y.; Kahraman, T.; Kalemci, O. Developing a Reliable Core Stability Assessment Battery for Patients with Nonspecific Low Back Pain. *Spine* **2016**, *41*, E844–E850. [CrossRef]

104. Arab, A.M.; Salavati, M.; Ebrahimi, I.; Ebrahim Mousavi, M. Sensitivity, specificity and predictive value of the clinical trunk muscle endurance tests in low back pain. *Clin. Rehabil.* **2007**, *21*, 640–647. [CrossRef]

105. del Pozo-Cruz, B.; Mocholi, M.H.; del Pozo-Cruz, J.; Parraca, J.A.; Adsuar, J.C.; Gusi, N. Reliability and validity of lumbar and abdominal trunk muscle endurance tests in office workers with nonspecific subacute low back pain. *J. Back Musculoskelet. Rehabil.* **2014**, *27*, 399–408. [CrossRef] [PubMed]

106. Ortega, F.B.; Cadenas-Sanchez, C.; Sanchez-Delgado, G.; Mora-Gonzalez, J.; Martinez-Tellez, B.; Artero, E.G.; Castro-Pinero, J.; Labayen, I.; Chillon, P.; Lof, M.; et al. Systematic review and proposal of a field-based physical fitness-test battery in preschool children: The PREFIT battery. *Sports Med.* **2015**, *45*, 533–555. [CrossRef]

107. Ruiz, J.R.; Castro-Piñero, J.; Espana-Romero, V.; Artero, E.G.; Ortega, F.B.; Cuenca, M.M.; Jimenez-Pavon, D.; Chillon, P.; Girela-Rejon, M.J.; Mora, J.; et al. Field-based fitness assessment in young people: The ALPHA health-related fitness test battery for children and adolescents. *Br. J. Sports Med.* **2011**, *45*, 518–524. [CrossRef]

108. Leighton, J.R. An instrument and technic for the measurement of range of joint motion. *Arch. Phys. Med. Rehabil.* **1955**, *36*, 571–578.

109. Kanbur, N.O.; Duzgun, I.; Derman, O.; Baltaci, G. Do sexual maturation stages affect flexibility in adolescent boys aged 14 years? *J. Sports Med. Phys. Fit.* **2005**, *45*, 53–57.

110. Castro-Pinero, J.; Chillon, P.; Ortega, F.B.; Montesinos, J.L.; Sjostrom, M.; Ruiz, J.R. Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *Int. J. Sports Med.* **2009**, *30*, 658–662. [CrossRef] [PubMed]

111. Wells, K.F.; Dillon, E.K. The sit-and-reach. A test of back and leg flexibility. *Res. Q. Exerc. Sport* **1952**, *23*, 115–118. [CrossRef]

112. Kraus, H.; Hirschland, R. Minimum muscular fitness of the school children. *Res. Q.* **1954**, *25*, 178–188. [CrossRef]

113. Nuzzo, J.L. The Case for Retiring Flexibility as a Major Component of Physical Fitness. *Sports Med.* **2020**, *50*, 853–870. [CrossRef] [PubMed]

114. Elbaz, A.; Sabia, S.; Brunner, E.; Shipley, M.; Marmot, M.; Kivimaki, M.; Singh-Manoux, A. Association of walking speed in late midlife with mortality: Results from the Whitehall II cohort study. *Age* **2013**, *35*, 943–952. [CrossRef] [PubMed]

115. Niiranen, T.J.; Enserro, D.M.; Larson, M.G.; Vasan, R.S. Multisystem Trajectories Over the Adult Life Course and Relations to Cardiovascular Disease and Death. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2019**, *74*, 1778–1785. [CrossRef] [PubMed]

116. Cooper, R.; Strand, B.H.; Hardy, R.; Patel, K.V.; Kuh, D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. *BMJ* **2014**, *348*, g2219. [CrossRef]

117. Nitz, J.C.; Stock, L.; Khan, A. Health-related predictors of falls and fractures in women over 40. *Osteoporos. Int.* **2013**, *24*, 613–621. [CrossRef] [PubMed]
118. Wang, D.X.M.; Yao, J.; Zirek, Y.; Reijnierse, E.M.; Maier, A.B. Muscle mass, strength, and physical performance predicting activities of daily living: A meta-analysis. *J. Cachexia Sarcopenia Muscle* **2020**, *11*, 3–25. [CrossRef] [PubMed]
119. Briggs, R.; Carey, D.; Claffey, P.; McNicholas, T.; Donoghue, O.; Kennelly, S.P.; Kenny, R.A. Do Differences in Spatiotemporal Gait Parameters Predict the Risk of Developing Depression in Later Life? *J. Am. Geriatr. Soc.* **2019**, *67*, 1050–1056. [CrossRef] [PubMed]