



## Criterion-related validity of field-based fitness tests in youth: a systematic review

J Castro-Piñero, E G Artero, V España-Romero, et al.

*Br J Sports Med* 2010 44: 934-943 originally published online April 12, 2009  
doi: 10.1136/bjasm.2009.058321

---

Updated information and services can be found at:

<http://bjsm.bmj.com/content/44/13/934.full.html>

---

### Data Supplement

*These include:*

"Web Only Data"

<http://bjsm.bmj.com/content/suppl/2010/09/14/bjasm.2009.058321.DC1.html>

### References

This article cites 79 articles, 14 of which can be accessed free at:

<http://bjsm.bmj.com/content/44/13/934.full.html#ref-list-1>

Article cited in:

<http://bjsm.bmj.com/content/44/13/934.full.html#related-urls>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://journals.bmj.com/cgi/ep>

# Criterion-related validity of field-based fitness tests in youth: a systematic review

J Castro-Piñero,<sup>1</sup> E G Artero,<sup>2</sup> V España-Romero,<sup>2,3</sup> F B Ortega,<sup>2,3</sup> M Sjöström,<sup>3</sup> J Suni,<sup>4</sup> J R Ruiz<sup>3</sup>

► Additional data are published online only. To view this file please visit the journal online (<http://bjsm.bmj.com>).

<sup>1</sup>Department of Physical Education, School of Education, University of Cadiz, Puerto Real, Spain

<sup>2</sup>Department of Physiology, School of Medicine, University of Granada, Granada, Spain

<sup>3</sup>Department of Biosciences and Nutrition at NOVUM, Unit for Preventive Nutrition, Karolinska Institutet, Huddinge, Sweden 4UKK Institute for Health Promotion Research, Tampere, Finland

## Correspondence to

J R Ruiz, Department of Biosciences and Nutrition, Unit for Preventive Nutrition, NOVUM, 14157, Huddinge, Sweden; [ruizj@ugr.es](mailto:ruizj@ugr.es)

Accepted 2 April 2009

Published Online First

12 April 2009

## ABSTRACT

The objective of this systematic review was to comprehensively study the criterion-related validity of the existing field-based fitness tests used in children and adolescents. The studies were scored according to the number of subjects, description of the study population and statistical analysis. Each study was classified as high, low and very low quality. Three levels of evidence were constructed: strong evidence, when consistent findings were observed in three or more high quality studies; moderate evidence, when consistent findings were observed in two high quality studies; and limited evidence when consistency of findings and/or the number of studies did not achieve the criteria for moderate. The results of 73 studies (50 of high quality) addressing the criterion-related validity of field-based fitness tests in children and adolescents indicate the following: that there is strong evidence indicating that the 20 m shuttle run test is a valid test to estimate cardiorespiratory fitness, that the hand-grip strength test is a valid measure of musculoskeletal fitness, that skin fold thickness and body mass index are good estimates of body composition, and that waist circumference is a valid measure to estimate central body fat. Moderate evidence was found that the 1-mile run/walk test is a valid test to estimate cardiorespiratory fitness. A large number of other field-based fitness tests presented limited evidence, mainly due to a limited number of studies (one for each test). The results of the present systematic review should be interpreted with caution due to the substantial lack of consistency in reporting and designing the existing validity studies.

Physical fitness refers to the full range of physical qualities (ie, cardiorespiratory fitness, muscular strength, agility, coordination and flexibility).<sup>1</sup> It can be understood as an integrated measurement of all functions (skeletal-muscular, cardiorespiratory, haematocirculatory, psychoneurological and endocrine/metabolic) and structures involved in the performance of physical activity and/or physical exercise.<sup>2</sup> Physical fitness, especially cardiorespiratory fitness and muscular strength, is considered an important marker of health in adults,<sup>3,4</sup> as well as in young people.<sup>5,6</sup>

Physical fitness can be objectively and accurately measured through laboratory tests. However, due to their high cost, necessity for sophisticated instruments and qualified technicians, and time constraints, their use is limited in school settings and in population-based studies. Field-based tests provide a reasonable alternative since they are time-efficient, low in cost and equipment requirements and can be easily administered to a large number of people simultaneously.

Field-based fitness assessment depends on the prediction techniques, and thus is prone to error. In order for a test or a fitness test battery to be considered “good”, it should measure what it is supposed to measure (ie, validity).<sup>7</sup> Criterion-related validity refers to the extent to which a field test of a fitness component correlates with the criterion measure (ie, the gold standard).<sup>8</sup> In deciding whether or not to use a test, the user should be satisfied that the test has established validity. In the 1990s, Safran<sup>9</sup> summarised the criterion-related validity of several fitness tests; yet, despite the growing interest in this area, no other attempt has been made to summarise the criterion-related validity of the existing field-based fitness tests in youth.

During the last two decades a great deal of attention has been devoted to the fitness of children and adolescents. As a result, numerous field-based fitness test batteries have been developed to assess fitness in this population (table 1).

The objective of the present systematic review was to comprehensively study the validity of the existing field-based fitness tests used in children and adolescents. To better understand whether or not a field-based test has established validity will help physical educators, exercise scientists, health agencies and private organisations dealing with sport, fitness and health to decide which field test should be used to assess physical fitness.

## METHODS

The present systematic review is produced as a part of the ALPHA (for “instruments for Assessing Levels of PPhysical Activity and fitness”) study.<sup>25</sup> The ALPHA study aims to provide a set of instruments for assessing levels of physical activity as well as health-related physical fitness in a comparable way within the European Union.

## Procedures

The electronic databases MEDLINE, SCOPUS and SPORTS DISCUS were screened for criterion-related validity studies in children and adolescents where one or more field-based fitness test were carried out. All the fitness tests from the most commonly used fitness test batteries in youth were included (table 1).

The keywords used (in various combinations) were: criterion validity, validity, validation, cross-validation, estimation, prediction, physical fitness, fitness, aerobic capacity, cardiorespiratory fitness, maximum oxygen consumption, strength, flexibility, motor, endurance, speed, agility, balance, body

**Table 1** Existing field-based physical fitness test batteries for children and adolescents

Acronym	Society/organisation	Country/region	Age, years
EUROFIT <sup>10</sup>	Council of Europe Committee for the Development of Sport	Europe	6–18
FITNESSGRAM <sup>11</sup>	The Cooper Institute	USA	5–17
PCHF <sup>12</sup>	President's Challenge: Health Fitness. The President's Council on Physical Fitness and Sports/American Association for Health, Physical Education, and Recreation (AAHPER)	USA	6–17
PCPF <sup>13</sup>	President's Challenge: Physical Fitness. The President's Council on Physical Fitness and Sports/American Association for Health, Physical Education, and Recreation (AAHPER)	USA	6–17
AAUTB <sup>14</sup>	Amateur Athletic Union Test Battery. Chrysler Foundation/Amateur Athletic Union	USA	6–17
YMCA YFT <sup>15</sup>	YMCA Youth Fitness Test	USA	6–17
NYFPF <sup>16</sup>	National Youth Physical Program. The United States Marines Youth Foundation	USA	5–17
HRFT <sup>17</sup>	Health-Related Fitness Test, American Association for Health, Physical Education, and Recreation (AAHPER)	USA	5–18
Physical Best <sup>18</sup>	American Association for Health, Physical Education, and Recreation (AAHPER)	USA	5–18
IPFT <sup>19</sup>	International Physical Fitness Test (United States Sports Academic/General Organization of Youth and Sport of Bahrain)	USA	9–19
CAHPER-FPT II <sup>20</sup>	Fitness Performance Test II. Canadian Association for Health, Physical Education and Recreation (CAHPER)	Canada	7–69
CPAFLA <sup>21</sup>	The Canadian Physical Activity, Fitness & Lifestyle Approach (Canadian Society for Exercise Physiology)	Canada	15–69
NFTP-PRC <sup>22</sup>	National Fitness Test Program in the Popular Republic China (China's National Sport and Physical Education Committee)	China	9–19+
NZFT <sup>23</sup>	New Zealand Fitness Test. Russell/Department of Education	New Zealand	6–12
AFEA <sup>24</sup>	Australian Fitness Education Award. The Australian Council for Health, Education and Recreation, ACHER	Australia	9–19

YMCA, Young Men's Christian Association.

composition, anthropometry, Body Mass Index (BMI), skin folds and waist circumference. The specific names of the tests were also included. Tables 2 to 5 summarise the field-based fitness tests used to assess cardiorespiratory fitness, musculoskeletal fitness, motor fitness and body composition, respectively.

The computer-based searches were limited to papers published from January 1990 to December 2008, full reports published in English or Spanish, in humans, and all children (0–18 years). An additional search using adolescents (13–18 years) was also performed. Additional studies were identified from reference lists.

The results of the most recent reviews were summarised first, and then the studies potentially relevant for the selected topics were screened for retrieval. Finally, a snowball search was performed, in which reference lists of the selected articles were checked for titles including validity of physical fitness.

### Quality assessment of the study

The quality of the selected studies was scored using a quality assessment list. The list included three items based on number of study subjects, description of the study population and statistical methods; see table 6. The items were rated from 0 to 2, 2 being the best score. For all studies, a total quality score was calculated by counting up the number of positive items (a total score between 0 and 6). Studies were defined as high quality if they had a total score of 5 or higher. A total score of 3 or 4 was defined as low quality and a score lower than 3 was defined as very low quality. Two reviewers (JCP and JRR) evaluated the quality of the studies, separately. A consensus meeting was arranged to sort out differences between both reviewers. The articles were not blinded for authors because the reviewers who performed the quality assessment were familiar with the literature.

### Levels of evidence

Three levels of evidence were constructed: (1) strong evidence: consistent findings in three or more high quality studies; (2) moderate evidence: consistent findings in two high quality studies; (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 have found strong or moderate evidence that the test is (or not) valid.

### Data extraction

We extracted information on fitness quality, population characteristics, fitness test, gold standard, statistical methods, main outcome and conclusions from studies defined as high quality. We regarded results with a  $p \leq 0.05$  as statistically significant.

## RESULTS

### Quality assessment

The literature search identified 73 studies addressing the criterion-related validity of field-based fitness tests in children and adolescents (see supplementary material, table 1). Of these, 23 studies were of low quality and were not included in this manuscript. There were no studies with a score  $\leq 2$ , that is, with a very low quality. A total of 31 high quality studies had the highest score (score = 6). The overall agreement between the two reviewers was 90% ( $\kappa = 0.813$ ). Disagreement was solved in a consensus meeting.

### Levels of evidence

#### Cardiorespiratory fitness

The 20 m shuttle run test (20mSRT) was investigated in eight studies,<sup>26–33</sup> and the 1-mile run/walk test was investigated in three studies.<sup>34–36</sup> (see supplementary material, table 2).

**Table 2** Field-based fitness tests used to assess cardiorespiratory fitness

Fitness quality	Fitness test	Fitness test battery (country)														
		EUROFIT (EU)	FITNESSGRAM (USA)	PCHF (USA)	PCPF (USA)	AAUTB (USA)	YMCAYFT (USA)	NYPFP (USA)	HRFT (USA)	Physical Best (USA)	IPFT (USA)	CAHPER-FPT II (CAN)	CPAFLA (CAN)	NFTP-PRC (CHIN)	NZFT (NZ)	AFEA (AUS)
Aerobic capacity	20 m shuttle run	•	•													•
	50 × 8 m shuttle run												•	•		
	4/3/2 min 25 m shuttle run															
	Hoosier endurance shuttle run			•												
	1.5-mile run/walk test								•							
	1-mile run/walk test								•							
	1000 m run									•						
	1/2-mile run/walk test										•					
	1/4-mile run/walk test											•				
	Cooper test															
	9-min running															•
	1-min jump rope															•

Organisational abbreviations given in table 1.

EU, Europe; CAN, Canada; CHIN, China; NZ, New Zealand; AUS, Australia.

The Douglas bag method is considered the gold standard to assess maximal oxygen consumption ( $\text{VO}_{2\text{max}}$ ),<sup>37</sup> yet there is agreement on that respiratory gas analyser is a valid method to assess oxygen uptake.<sup>37</sup> All the studies measured  $\text{VO}_{2\text{max}}$  or peak oxygen consumption ( $\text{VO}_{2\text{peak}}$ ) when performing a maximal treadmill test to measure, except Ruiz *et al*<sup>26 27</sup> that measured  $\text{VO}_{2\text{max}}$  when performing the 20mSRT.

### 20mSRT

Several studies<sup>26 28 30–32</sup> attempted to develop an equation to estimate  $\text{VO}_{2\text{max}}$ . McVeigh *et al*<sup>28</sup> showed that the estimation of  $\text{VO}_{2\text{peak}}$  from the 20mSRT might be improved by including skin fold thickness measurements in the regression model, particularly for girls ( $R^2 = 0.85$ , standard error of estimate (SEE): 2.4 ml/kg/min for girls and  $R^2 = 0.68$ , SEE: 3.23 ml/kg/min for boys), which concurs with others.<sup>30 31</sup> In contrast, Mahar *et al*<sup>32</sup> showed that a model including sex, number of laps completed and body weight or BMI was not accurate to estimate  $\text{VO}_{2\text{peak}}$  ( $R^2 = 0.65$ , SEE: 6.35 ml/kg/min) in boys and girls aged 12–14 years. More recently, we have developed a new equation to estimate  $\text{VO}_{2\text{max}}$  from 20mSRT performance (stage), sex, age, weight and height in adolescents aged 13–19 years using a more advance mathematical model, that is, artificial neural network modelling ( $R^2 = 0.92$ , percentage error: 7.30%, SEE: 2.84 ml/kg/min).<sup>26</sup>

Several studies have cross-validated the mentioned equations.<sup>26 27 29 33</sup> Pitteti *et al*<sup>29</sup> cross-validated the Leger and the Fernhall equations, and found significant but modest relationships between both regression equations and  $\text{VO}_{2\text{peak}}$  ( $r = 0.57$ ,  $p < 0.01$ ;  $r = 0.66$ ,  $p < 0.01$ , respectively). More recently, Ruiz *et al*<sup>27</sup> assessed the validity of five different equations (ie, the Ruiz *et al*,<sup>26</sup> Léger *et al*,<sup>38</sup> Barnett *et al* (a),<sup>30</sup> Barnett *et al* (b)<sup>30</sup> and Matsuzaka *et al*<sup>31</sup> equations) for estimating  $\text{VO}_{2\text{max}}$  from the 20mSRT test in a relatively small sample of 48 Portuguese adolescents 13–19 years of age. They reported that equations to estimate  $\text{VO}_{2\text{max}}$  from the 20mSRT should not be used at an individual level, and suggested that the equations reported by Barnett (b)<sup>30</sup> and Ruiz<sup>26</sup> seems to be the most accurate to estimate  $\text{VO}_{2\text{max}}$  in adolescents.

### Distance run/walk tests

The most commonly used equation to estimate  $\text{VO}_{2\text{peak}}$  from the 1-mile run/walk test is the Cureton equation,<sup>34</sup> which was selected for the FITNESSGRAM battery to estimate  $\text{VO}_{2\text{peak}}$ .<sup>11</sup> We have examined the criterion-related validity of Cureton equation in 66 endurance trained children and adolescents aged 8–17 years.<sup>35</sup> We observed that there was a significant mean difference between measured and estimated  $\text{VO}_{2\text{peak}}$  (10.01 ml/kg/min, 95% confidence interval (CI) 9.2 to 11.8,  $p < 0.001$ ). The findings did not materially change when the analyses were performed by sex, age groups and weight status, which suggests that this equation is not accurate for estimating  $\text{VO}_{2\text{peak}}$  in endurance trained children.<sup>35</sup> Buono *et al*<sup>36</sup> also developed an equation to estimate  $\text{VO}_{2\text{peak}}$  from the 1-mile run/walk time, and reported a SEE of 4.3 ml/kg/min ( $R^2 = 0.84$ ).

The nature of this test deserves several comments. The 1-mile run/walk test is not a friendly test, especially in young people. One of its major problems is the participant's capacity to develop an appropriate pace. Participants may either start too fast so that they are not able to keep up the speed all through the test, or they may start too slow so that when they want to increase speed, the test is already finished. To ameliorate this problem, several versions were developed such as the 1-mile

**Table 3** Field-based fitness tests used to assess musculoskeletal fitness

Fitness quality	Fitness test	Fitness test battery (country)														
		EUROFIT (EU)	FITNESSGRAM (USA)	PCHF (USA)	PCPF (USA)	AAUTB (USA)	YMCAYFT (USA)	NYPPF (USA)	HRFT (USA)	Physical Best (USA)	IPFT (USA)	CAHPER-FPT II (CAN)	CPAFLA (CAN)	NFTP-PRC (CHIN)	NZFT (NZ)	AFEA (AUS)
Maximal isometric strength	Hand-grip strength	•														
	Trunk lift	•														
Endurance strength: Upper body	Bent arm hang	•	•	•	•	•										
	Pull-ups	•	•	•	•	•										
	Modified pull-ups	•	•	•	•	•										
	Parallel-bars dips	•	•	•	•	•										
	Push-ups	•	•	•	•	•										
	Modified push-ups	•	•	•	•	•										
Trunk	Isometric push-ups	•	•	•	•	•										
	Sit-ups	•	•	•	•	•										
	Curl-ups	•	•	•	•	•										
Lower body	Partial curl-ups	•	•	•	•	•										
	Phantom chair	•	•	•	•	•										
Explosive strength: Upper body	Basketball throw															
	Medicine ball throw															
	Sand ball throw															
	(0.25 kg)															
	Shot put (1, 2, 3, 4 or 5 kg)															
	Back throw (8 pounds)															
Lower body	Standing broad jump	•														
	Vertical jump	•														
Flexibility	Sit and reach	•	•	•	•	•										
	V sit and reach	•	•	•	•	•										
	Back saver sit and reach	•	•	•	•	•										
	Shoulder stretch	•	•	•	•	•										

Organisational abbreviations given in table 1.

EU, Europe; CAN, Canada; CHIN, China; NZ, New Zealand; AUS, Australia.

**Table 4** Field-based fitness tests used to assess motor fitness

Fitness test battery (country)																
Fitness quality	Fitness test	EUROFIT (EU)	FITNESSGRAM (USA)	PCHF (USA)	PCPF (USA)	AAUTB (USA)	YMCAYFT (USA)	NYPPF (USA)	HRFT (USA)	Physical Best (USA)	IPFT (USA)	CAHPER-FPT II (CAN)	CPAFLA (CAN)	NFTP-PRC (CHIN)	NZFT (NZ)	AFEA (AUS)
Speed and agility	Shuttle run (10×5 m)	•														
	Shuttle run with sponges (10×4 m)		•			•					•					
	10 sec 25 m shuttle run															•
Speed	100 m dash					•										
	50 m dash										•					•
Speed of limbs	Plate tapping	•									•					
	Fleming balance															•

Organisational abbreviations given in table 1.

EU, Europe; CAN, Canada; CHIN, China; NZ, New Zealand; AUS, Australia.

**Table 5** Field-based fitness tests used to assess body composition

Fitness test battery (country)																	
Fitness quality	Fitness test	EUROFIT (EU)	FITNESSGRAM (USA)	PCHF (USA)	PCPF (USA)	AAUTB (USA)	YMCAYFT (USA)	NYPPF (USA)	HRFT (USA)	Physical Best (USA)	IPFT (USA)	CAHPER-FPT II (CAN)	CPAFLA (CAN)	NFTP-PRC (CHIN)	NZFT (NZ)	AFEA (AUS)	
Anthropometric indices	Height and weight	•	•	•			•		•		•						
	Waist circumference												•				
	Hip circumference												•				
	Waist-to-hip ratio												•				
	Waist-to-height ratio												•				
	Body mass index		•	•									•			•	
	Skin fold thickness	•	•				•						•			•	

Organisational abbreviations given in table 1.

EU, Europe; CAN, Canada; CHIN, China; NZ, New Zealand; AUS, Australia.



**Table 6** Quality assessment criteria for criterion-related validity studies

Grading system parameter	Grade	Criterion
Number of study subjects	0	$n \leq 10$
	1	$n = 11-50$
	2	$n \geq 51$
Description of the study population with respect to age, sex, health status, fitness levels, pubertal status, ethnicity, physical activity patterns, body composition, etc.	0	Less items than required for grade 1
	1	At least age, sex, health status and fitness levels
	2	More items than required for grade 1
Statistical analysis included in the study	0	Those not included in 1
	1	Error indexes or regression analysis
	2	$\geq 3$ items or Bland-Altman plot and/or ANOVA for repeated measurements

Rating for total score: high quality = 5-6 low quality = 3-4 very low quality = 0-2

walk,<sup>39</sup> the submaximal 1-mile track jog test (pacing test),<sup>40</sup> and the 1/2-mile run/walk test (table 2).<sup>41</sup> We assessed the criterion-related validity of the 1/2-mile run/walk test in children aged 6-17 years, and also examined the criterion-related validity of the Fernhall equation in a subgroup of children aged 10-17 years.<sup>41</sup> We computed a regression equation that was assessed through several error measures and the Bland-Altman method. We found that the 1/2-mile run/walk time, sex and BMI were significantly associated with  $VO_{2peak}$ . There was no systematic bias in the validation group nor in the cross-validation group ( $p > 0.1$ ), and the root mean squared error (RMSE) and the percentage error were 6.5 ml/kg/min and 13.9%, respectively. The newly developed equation had a lower RMSE and percentage error than the Fernhall equation in the subgroup of children aged 10-17 years (7.2 vs 17.7 ml/kg/min and 16.0% vs 50.4%, respectively,  $p < 0.001$ ).

In conclusion, there is strong evidence indicating that the 20mSRT is a suitable test to estimate cardiorespiratory fitness. From the developed equations, it seems that the Barnett (b) and Ruiz equations yielded the most promising results to estimate  $VO_{2max}$ . There is moderate evidence in the case of the 1-mile run/walk test, and limited evidence that the 1-mile walk test, the submaximal 1-mile track jog test (pacing test) and the 1/2-run/walk test are valid tests to estimate  $VO_{2max}$  (or  $VO_{2peak}$ ). The Cureton equation seems the best equation to predict  $VO_{2peak}$  from the 1-mile run/walk test, but the fitness levels of the individuals may affect its validity.

### Musculoskeletal fitness

Two studies examined the criterion-related validity of the hand-grip strength test (maximal isometric strength),<sup>42, 43</sup> one study examined the criterion-related validity of the bent arm hang, push-ups, pull-ups and modified pull-ups tests<sup>44</sup> (upper body endurance strength) and one examined the criterion-related validity of the standing broad jump and vertical test.<sup>45</sup> Two studies assessed the criterion-related validity of flexibility: one analysed the back saver sit and reach test,<sup>45</sup> and the other one analysed the trunk lift test<sup>46</sup> (see supplementary material, table 2).

There is no established gold standard for most of the musculoskeletal fitness tests, which makes it difficult to determine the criterion-related validity of these tests. 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 muscular strength and endurance strength tests.<sup>47</sup> One repetition maximum (1RM) and repetitions to a certain percentage of 1RM (ie, 50% of 1RM or 70% of 1RM) have been used as gold standards.<sup>43, 44</sup> Concerning flexibility, radiography seems to be the best criterion measurement, but goniometry has also been used as a criterion measure.<sup>45, 48-50</sup>

### Maximal isometric strength

We have studied the criterion-related validity of the hand-grip strength test using Jamar, DynEx and TKK dynamometers in adolescents aged 12-16 years.<sup>42</sup> We used known weights (ranging from 20 to 70 kg) as the criterion measure. We observed a negative systematic bias (underestimation) for the Jamar and DynEx dynamometers (-1.92 and -1.43 kg, respectively,  $p < 0.05$ ), whereas a marginal positive overestimation was observed for the TKK dynamometer (0.49 kg,  $p < 0.05$ ). These results concur with those reported in studies performed in adults.<sup>51-59</sup> We also examined whether the elbow position (extended or flexed at 90 degrees) affects the hand-grip strength in adolescents. We observed that performing the hand-grip strength test with the elbow extended seems the most appropriate protocol to evaluate maximal hand-grip strength in adolescents when using the TKK dynamometer. We have also conducted a series of studies in children<sup>60</sup> and adolescents,<sup>61</sup> to determine if there is an optimal grip span for determining the maximum hand-grip strength, and if the optimal grip span was related to hand size. We found that there was an optimal grip span to which the dynamometer should be adjusted when measuring hand-grip strength in children<sup>60</sup> and adolescents.<sup>61</sup> We provided sex and age specific equations to adjust the grip span of the dynamometer to the hand size of the individual in order to obtain the actual maximal hand-grip strength. Milliken *et al*<sup>43</sup> analysed the association between hand-grip strength (using the TKK dynamometer) and 1RM chest press in children aged 7-12 years. They found that the hand-grip strength test is valid to assess upper body maximal strength.

### Upper body endurance strength

Woods *et al*<sup>44</sup> studied the criterion-related validity of the bent arm hang, push-up, pull-ups and two modified pull-up tests using 1RM and repetitions at 50% of 1RM as criterion reference in children aged 9-11 years. They concluded that these tests are not valid to assess muscular endurance and that body fat percentage was the main determinant of performance. We observed that muscular strength is highly influenced by body weight in children aged 6-17 years,<sup>62</sup> especially in relation to weight bearing tests. We showed that out of 2778, a total of 1037 (85%) of the girls and 889 (60%) of the boys were not able to perform a single repetition in the pull-up test. Likewise, a total of 478 (39%) of the girls and 409 (28%) of the boys were not able to perform for more than 0 s in the bent arm hang test. Collectively, these findings suggest that these tests are not appropriate to measure upper body endurance strength in children and adolescents.

### Lower body explosive strength

Milliken *et al*<sup>45</sup> studied the criterion-related validity of the standing broad jump test and the vertical jump test using 1RM of leg press as criterion measure in children aged 7-12 years. They reported that the standing broad jump and vertical jump test, with BMI, accounted for 44.4% and for 40.8% of the variation in 1RM leg press, respectively.

### Flexibility

Patterson *et al* studied the criterion-related validity of the back saver sit and reach test using goniometry (hamstring flexibility) and the MacRae and Wright method (low back flexibility) as gold standard in children and adolescents aged 11–15 years.<sup>45</sup> The results suggested that this test has a moderate validity to assess hamstring flexibility ( $r = 0.51$  to  $0.72$ ), and a low validity to assess the lower back flexibility ( $r = 0.10$  to  $0.25$ ). Paterson *et al*<sup>46</sup> reported that the correlations of trunk lift scores and the goniometry scores were moderate ( $r = 0.70$  for boys and  $r = 0.68$  for girls) in children of similar ranges of age.

In conclusion, there is strong evidence that the hand-grip strength test with the elbow extended and with the grip span adapted to the individual's hand size (using the TKK dynamometer) is a valid test to assess isometric muscular strength. Due to a limited number of studies, we found limited evidence that: (1) the bent arm hang, push-up, pull-ups and two modified pull-ups tests are not valid to assess muscular endurance and (2) the back saver sit and reach test and the trunk lift test have moderate validity to measure hamstring flexibility and lumbar flexibility.

### Body composition

A total of 22 studies investigated the criterion-related validity of BMI,<sup>63–84</sup> 18 investigated the validity of skin fold thickness,<sup>64–66 72–76 80 82 85–92</sup> and 7 studied the validity of circumferences and/or ratios (ie, waist-to-hip ratio) (see supplementary material, table 2).<sup>63 65 66 82 83 92 93</sup>

Imaging methods (ie, axial CT and MRI),<sup>94</sup> dual energy x ray absorptiometry (DXA),<sup>95</sup> ultrasonography,<sup>96</sup> and air displacement plethysmography<sup>97 98</sup> are considered gold standards for assessing body composition in youth. Bioelectrical impedance analysis (BIA) has also been used a reference method for body fat determination,<sup>99 100</sup> whereas hydrodensitometry is considered the gold standard in adults, but not in children due to the fact that they have difficulties with the breathing manoeuvre involved in determining underwater weight.<sup>85</sup>

### Skin fold thickness

The validity of different skin fold\* equations to estimate percentage body fat, mainly the Slaughter equation, have been extensively analysed by using the Bland–Altman method<sup>72 82 85 87–89</sup> and/or ANOVA for repeated measures.<sup>82 86</sup> Rodríguez *et al*<sup>88</sup> reported that the Slaughter equations using either triceps and subscapular or triceps and calf skin folds had the best agreement in male and female adolescents. Likewise, Buisson *et al*<sup>89</sup> found that the Brook equation is a valid alternative to measure percentage body fat in children aged 7–10 years (mean difference =  $-1.4\%$  for percentage body fat; with limits of agreement of  $\pm 12.2\%$ ). Treuth *et al*<sup>87</sup> reported that the Slaughter equation appears to be valid to estimate percentage body fat in prepubertal multiethnic girls ( $R^2 = 0.69$ ). Gutin *et al*<sup>85</sup> reported a strong correlation between the Slaughter equation and the criterion measure (BIA and DXA), yet the limits of agreement were high (DXA vs skin fold thickness  $-3.65$  to  $9.50$  and BIA vs skin fold thickness  $-10.81$  to  $9.89$ ). Likewise, Campanozzi *et al*<sup>91</sup> reported similar results in obese children and adolescents when they compared the Brook equation, BIA and DXA. It is noteworthy that discrepancies between methods seem to increase with the degree of obesity, which indicates the presence of heteroscedascity.

Ihmels *et al*<sup>90</sup> showed a good overall agreement between estimates from BIA and sum of triceps and subscapular skin fold

(classification agreement values: 82.8–92.6%). Guida *et al*<sup>64</sup> reported, with BIA vector distribution, that triceps skin fold had moderate validity to assess percentage body fat ( $r = 0.79$ ,  $p < 0.001$ ), and that it was not affected by body size. Goran *et al*<sup>65</sup> developed several equations to estimate intra-abdominal adipose tissue and subcutaneous abdominal adipose tissue. They concluded that intra-abdominal adipose tissue was best predicted by abdominal skin fold, ethnicity and subscapular skin fold ( $R^2 = 0.82$ , SEE:  $9.8 \text{ cm}^2$ ), whereas subcutaneous abdominal adipose tissue was best predicted by waist circumference, subscapular skin fold, height and abdominal skin fold ( $R^2 = 0.92$ , SEE:  $28.8 \text{ cm}^2$ ). However, in obese children and adolescents, using ultrasound as a gold standard method, triceps and subscapular skin fold did not show a good validity ( $r = 0.13$  to  $0.34$ ,  $r = 0.02$  to  $0.37$ , respectively).<sup>66</sup> Differences in criterion methods, statistical methods and techniques to measure skin fold thickness make comparison among studies difficult.

### BMI

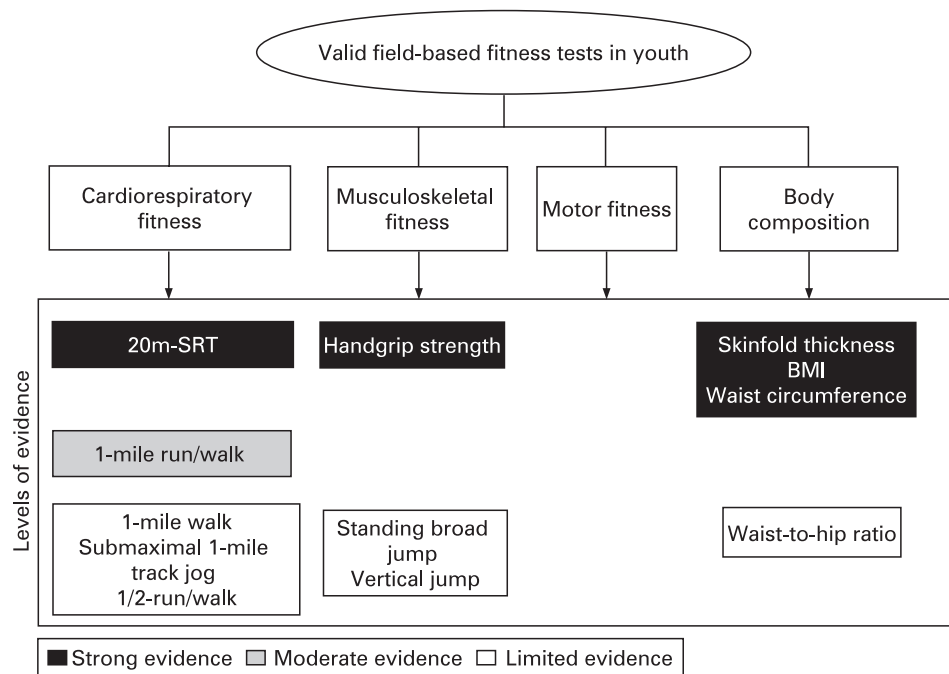
Correlations between BMI and body fat measured by more advanced methods generally exceed 0.50 and are frequently much higher.<sup>67 68</sup> This finding supports, a priori, the validity of BMI, yet several studies highlight that BMI should be used with caution when comparing groups with different demographic characteristics and using DXA as criterion measure.<sup>68–70 77 78 81</sup> Ellis *et al*<sup>69</sup> reported in children aged 3–10 years that the correlations between percentage body fat and BMI were significant for girls ( $R^2 = 0.70$ ,  $p < 0.001$ ) and boys ( $R^2 = 0.34$ ,  $p < 0.001$ ); however, when a linear model was used, the ability of BMI to accurately estimate percentage body fat was poor (SEE:  $4.7\%$  for girls and SEE:  $7.3\%$  for boys). Daniels *et al*<sup>70</sup> studied whether BMI was a representative equivalent measure of body fat independent of age, race, gender, sexual maturation and distribution of fat in children and adolescents aged 7–17 years. They showed that BMI, gender, race, sexual maturation and distribution of fat were all significant independent correlates of the percentage body fat ( $R^2 = 0.77$ ).

In order to detect possible changes in body composition, Guida *et al*<sup>64</sup> analysed the association between BIA vector distribution and BMI in children aged 8 years. They showed a relationship between fat mass using BIA and BMI ( $r = 0.92$ ,  $p < 0.001$ ) and fat free mass using BIA and BMI ( $r = 0.58$ ,  $p < 0.001$ ). However, they showed that BMI was not able to differentiate whether the weight change was due to a variation in body fat or in fat free mass, which concurs with other studies.<sup>66 70</sup> Moreno *et al*<sup>71</sup> tried to improve the International Obesity Task Force BMI cut-off values, in terms of prediction of percentage body fat in adolescents aged 13–17 years. They concluded that BMI cut-off points seem to be useful as an approximate classification of obesity status, but cannot accurately predict a specific individual's percentage body fat. Therefore, they suggested that BMI could be used as a screening test, but that in clinical setting the percentage body fat should be measured by using a more accurate method such as DXA.

Several studies have compared the accuracy of BMI and skin fold thickness to estimate body fat.<sup>72–76</sup> Steinberger *et al*<sup>72</sup> reported that the Slaughter equation and BMI were highly correlated with DXA children and adolescents aged 11–17 years. Similarly, Sarria *et al*<sup>73</sup> developed several equations to estimate body density from underwater weighing in boys aged 7–16 years. They found that the correlations between body density and  $\log_{\Sigma 4}$  skin folds ( $r = -0.781$  to  $0.820$ ) were higher than those with BMI ( $r = -0.586$  to  $-0.798$ ) at all ages. The best estimators of body density were  $\log_{\Sigma 4}$  skin folds or a



**Figure 1** Evidence-based proposal of most valid field-based fitness tests in youth. <sup>1</sup>20mSRT with the equation reported by Barnett (b) and Ruiz. <sup>2</sup>1-mile run/walk with the equation reported by Cureton, but not in trained children. <sup>3</sup>TKK dynamometer with age-specific and sex-specific grip span equations. 20mSRT: 20 m shuttle run test; BMI, Body Mass Index.



combination of BMI and triceps skin fold. Freedman *et al*<sup>74</sup> examined the additional information provided by skin fold thickness on body fatness, beyond that conveyed by BMI for age, among children and adolescents aged 5–18 years. The use of sum of skin folds reduced the overall prediction errors (absolute value of the residuals) for percentage body fat by 20% to 30%, whereas among overweight children, the sum of skin folds reduced the prediction errors for percentage body fat by only 7% to 9%. Likewise, BMI was shown to be a less accurate predictor of degree of fatness than the sum of skin folds in lean female adolescents ( $r = 0.67$ ,  $p < 0.001$  between BMI and percentage of body fat-DXA;  $r = 0.80$ ,  $p < 0.001$  between skin fold sum and percentage of body fat-DXA).<sup>75</sup>

These findings support the idea that skin fold thicknesses are, in general, better predictors of childhood obesity than BMI is,<sup>72–74</sup> although the accuracy of BMI as an indicator of adiposity varies by the degree of body fatness,<sup>71 72 74 80 84</sup> markedly improving at higher levels of BMI.<sup>66 71 72 80 84</sup>

#### Waist circumference (WC) and waist-to-hip (WHR) ratio

Goran *et al*<sup>65</sup> developed several equations to estimate intra-abdominal adipose tissue and subcutaneous abdominal adipose tissue by skin folds, circumferences and DXA in prepubertal children. They found that WHR had low correlation with intra-abdominal adipose tissue ( $r = 0.32$ ) and subcutaneous abdominal adipose tissue ( $r = 0.40$ ), whereas WC was strongly correlated with intra-abdominal adipose tissue ( $r = 0.84$ ) and subcutaneous abdominal adipose tissue ( $r = 0.93$ ). This finding is consistent with the study by Taylor *et al*<sup>93</sup> They showed that WC was valid measure of central adiposity in children and adolescents ( $r = 0.92$ ,  $p < 0.001$  in girls and boys), but not WHR ( $r = -0.40$  and  $r = -0.04$  in girls and boys, respectively). Likewise, Brambilla *et al*<sup>83</sup> using MRI in a greater sample, also support that WC is a good predictor of body fat distribution. WC was the best single predictor of visceral adipose tissue (64.8% of variance) and BMI was the best predictor of subcutaneous adipose tissue (88.9% of variance), while WC explained 80.4%. Finally, one study examined the best anthropometric measure to estimate fat distribution in obese children,

using ultrasound as criterion measure.<sup>66</sup> The results showed that in obese children, BMI provided the best estimate of visceral adipose tissue ( $R^2 = 0.53$ ), while skin fold thickness, WC and WHR showed a lower association ( $r = 0.02$  to  $0.37$ ,  $r = 0.08$  to  $0.42$ ,  $r = 0.08$  to  $0.42$ , respectively). In the control group, visceral adiposity was significantly correlated with BMI, skin fold thickness and WC ( $p < 0.05$ ), but not with WHR ( $p > 0.05$ ).

In conclusion, there is strong evidence that skin fold thickness and BMI are good predictors of body fat. The Slaughter equations are valid to estimate body fat in youth, but not in obese children, where BMI seems to be the best indicator. The

#### What is already known on this topic

- ▶ In the last two decades, much attention has been devoted to the status of physical fitness in youth, and many surveillance systems have been used across the world.
- ▶ Despite the growing interest in this area, no attempt has been made to summarise the criterion-related validity of the existing field-based fitness tests in youth.

#### What this study adds

- ▶ We have formulated an evidence-based proposal of the most valid field-based fitness tests in youth.
- ▶ Our finding are: (i) The 20 m shuttle run test is a valid test to assess cardiorespiratory fitness, (ii) the hand-grip strength test is a valid test to assess upper body muscular strength and the standing broad jump is a valid test to assess lower body muscular strength, (iii) skin fold thickness is a valid measure to estimate body fat in non-obese youth, where body mass index seems to be the best indicator and (iv) waist circumference is a valid measure to estimate central fatness.

accuracy of BMI varies by the degree of body fatness, significantly improving at higher levels of body fat. There is strong evidence that WC is a good measure to estimate central body fat, and there is limited evidence that WHR is a good measure to estimate central body fat.

## DISCUSSION

In summary, the present systematic review highlights a number of important key points about the criterion-related validity of field-based fitness tests in children and adolescents (fig 1):

- ▶ **Cardiorespiratory fitness:** the 20mSRT seems to be the most appropriate test to assess cardiorespiratory fitness. The Barnett (b) and Ruiz equations yielded the most promising results to estimate  $\text{VO}_{2\text{max}}$ . An alternative to the 20mSRT might be the 1-mile run/walk test, using the Cureton equation, but not in trained children.
- ▶ **Musculoskeletal fitness:** there is strong evidence that the hand-grip strength test with the elbow extended and with the grip span adapted to the individual's hand size (using the TKK dynamometer) is a valid test to assess isometric muscular strength. Due to a reduced number of studies, there is limited evidence that the standing broad jump and vertical jump tests are valid to assess explosive strength.
- ▶ **Motor fitness:** there is no study, ranked as high quality, assessing the criterion-related validity of motor fitness.
- ▶ **Body composition:** the Slaughter equations, based on skin fold thickness are valid to estimate body fat in youth, but not in obese children, where BMI seems to be the best indicator. The accuracy of BMI varies by the degree of body fatness, significantly improving at higher levels of body fat. There is strong evidence that WC is a valid measure to estimate central fatness.

**Acknowledgements** We would like to thank Professor Willem van Mechelen, Professor Pekka Oja, Professor Han CG Kemper, Professor Kari Ba and Professor Jorge Mota for their valuable contributions to the conception and strategy of this review.

**Funding** This work was supported by the European Union within the framework of the Public Health Programme (ALPHA project, Ref: 2006120), the Swedish Council for Working Life and Social Research, the Spanish Ministry of Education (EX-2007-1124; AP-2004-2745; and AP2005-4358) and the Spanish Ministry of Education and Science-FEDER funds (Acciones Complementarias DEP2007-29933-E).

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

1. **Ruiz JR**, Ortega FB, Gutierrez A, *et al*. Health-related fitness assessment in childhood and adolescence: a European approach based on the AVENA, EYHS and HELENA studies. *J Public Health* 2006;**14**:269–77.
2. **Castillo Garzon MJ**, Ortega Porcel FB, Ruiz Ruiz J. Improvement of physical fitness as anti-aging intervention [in Spanish]. *Med Clin (Barc)* 2005;**124**:146–55.
3. **Blair SN**, Kohl HW 3rd, Paffenbarger RS Jr, *et al*. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 1989;**262**:2395–401.
4. **Ruiz JR**, Sui X, Lobelo F, *et al*. Association between muscular strength and mortality in men: prospective cohort study. *BMJ* 2008;**337**:a439.
5. **Ortega FB**, Ruiz JR, Castillo MJ, *et al*. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes (Lond)* 2008;**32**:1–11.
6. **Ruiz JR**, Castro-Pinero J, Artero EG, *et al*. Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med*. Published Online First: 21 January 2009. doi:10.1136/bjsm.2008.056499
7. **Safrit MJ**. *Introduction to measurement in physical education and exercise science*. St Louis, Missouri, USA: Times Mirror/Mosby College, 1986.
8. **Docherty D**. Field tests and test batteries. In: Docherty D, ed. *Measurement in pediatric exercise science*. Champaign, Illinois, USA: Human Kinetics, 1996: 285–334.
9. **Safrit MJ**. The validity and reliability of fitness tests for children: a review. *Pediatr Exerc Sci* 1990;**2**:9–28.
10. **Council of Europe Committee for the Development of Sport**. *Eurofit. Handbook for the EUROFIT tests of physical fitness*. Rome, Italy: Edigraf Editoriale Grafica, 1988.
11. **Cooper Institute for Aerobics Research**. *The prudential fitnessgram: test administration manual*, 3rd edn. Champaign, Illinois, USA: Human Kinetics, 2004.
12. **The President's Council on Physical Fitness and Sport**. The President's Challenge: The Physical Fitness Test. [http://www.presidentschallenge.org/educators/program\\_details/physical\\_fitness\\_test.aspx](http://www.presidentschallenge.org/educators/program_details/physical_fitness_test.aspx) (accessed 4 October 2007).
13. **The President's Council on Physical Fitness and Sport**. The President's Challenge: The Physical Fitness Test. [http://www.presidentschallenge.org/educators/program\\_details/physical\\_fitness\\_test.aspx](http://www.presidentschallenge.org/educators/program_details/physical_fitness_test.aspx) (accessed 4 October 2007).
14. **Amateur Athletic Union**. Physical fitness test manual. Bloomington, In: Franks, DB, ed. *YMCA Youth Fitness Test manual*. Champaign, Illinois, USA: Human Kinetics, 1989.
15. **Franks DB**. *YMCA Youth Fitness Test manual*. Champaign, Illinois, USA: Human Kinetics, 1989.
16. **National Youth Physical Program**. The United States Marines Youth Foundation. <http://marineyouthfoundation.org/YPFBook.pdf> (accessed 11 October 2007).
17. **American Alliance for Health, Physical Education, Recreation, and Dance (AAHPERD)**. *Health-related fitness test*. Reston, Virginia, USA: AAHPERD, 1980.
18. **American Alliance for Health, Physical Education, Recreation, and Dance (AAHPERD)**. *Physical best: the American Alliance physical fitness education and assessment program*. Reston, Virginia, USA: AAHPERD, 1980.
19. **United States Sports Academy**. United States Sports Academy in cooperation with the General Organization of Youth and Sport (State of Bahrain). International physical fitness test. <http://www.thesportjournal.org/VOL1NO2/fittest.HTM> (accessed 13 October 2007).
20. **Canadian Association for Health, Physical Education and Recreation (CAHPER)**. *CAHPER fitness performance II: test manual*. Ottawa, Canada: CAHPER, 1980.
21. **Canadian Society for Exercise Physiology (CSEP)**. *The Canadian Physical Activity, Fitness & Lifestyle Approach (CPAFLA): CSEP-health & fitness program's health-related appraisal and counseling strategy*, 3rd edn. Ottawa, Canada: CSEP, 2003.
22. **China's National Sports and Physical Education Committee**. *The national fitness testing methods*. Beijing, China: National Sports and Physical Education Committee, 1990.
23. **Russell DG**, Isaac A, Wilson PG. *New Zealand fitness test handbook*. Wellington, New Zealand: Department of Education, 1989.
24. **Australian Council for Health Physical Education and Recreation (ACHPER)**. *Handbook for the Australian Fitness Education Award manual*. Hindmarsh, Australia: ACHPER, 1996.
25. **Meusel D**, Ruiz JR, Ortega FB, *et al*. Assessing levels of physical activity in the European population – the ALPHA project. *Selección* 2007;**16**:9–12.
26. **Ruiz JR**, Ramirez-Lechuga J, Ortega FB, *et al*. Artificial neural network-based equation for estimating  $\text{VO}_{2\text{max}}$  from the 20 m shuttle run test in adolescents. *Artif Intell Med* 2008;**44**:233–45.
27. **Ruiz JR**, Silva G, Nordin O, *et al*. Criterion related validity of 20 m shuttle run test in adolescents aged 13–19 years. *J Sports Sci*. In press.
28. **McVeigh SK**, Payne AC, Scott S. The reliability and validity of the 20-meter shuttle test as a predictor of peak oxygen uptake in Edinburgh school children, age 13 to 14 years. *Pediatr Exerc Sci* 1995;**7**:69–79.
29. **Pitetti KH**, Fernhall B, Figoni S. Comparing two regression formulas that predict  $\text{VO}_{2\text{peak}}$  using the 20-m shuttle run for children and adolescents. *Pediatr Exerc Sci* 2002;**14**:125–34.
30. **Barnett A**, Chan LYS, Bruce IC. A preliminary study of the 20-m multistage shuttle run as a predictor of peak  $\text{VO}_2$  in Hong Kong Chinese students. *Pediatr Exerc Sci* 1993;**5**:42–50.
31. **Matsuzaka A**, Takahashi Y, Yamazoe M, *et al*. Validity of the multistage 20-m shuttle-run test for Japanese children, adolescents, and adults. *Pediatr Exerc Sci* 2004;**16**:113–25.
32. **Mahar MT**, Welk GJ, Rowe DA, *et al*. Development and validation of a regression model to estimate  $\text{VO}_{2\text{peak}}$  from PACER 20-m shuttle run performance. *J Phys Act Health* 2006;**3**:S34–46.
33. **Suminski RR**, Ryan ND, Poston CS, *et al*. Measuring aerobic fitness of Hispanic youth 10 to 12 years of age. *Int J Sports Med* 2004;**25**:61–7.
34. **Cureton KJ**, Sloniger MA, O'Bannon JP, *et al*. A generalized equation for prediction of  $\text{VO}_{2\text{peak}}$  from 1-mile run/walk performance. *Med Sci Sports Exerc* 1995;**27**:445–51.
35. **Castro-Piñero J**, Mora J, Gonzalez-Montesinos JL, *et al*. Criterion-related validity of the one-mile run/walk test in children aged 8–17 years. *J Sports Sci* 2009;**27**:405–13.
36. **Buono MJ**, Roby JJ, Micale FG, *et al*. Validity and reliability of predicting maximum oxygen uptake via field tests in children and adolescents. *Pediatr Exerc Sci* 1991;**3**:250–5.
37. **Bassett DR Jr**, Howley ET, Thompson DL, *et al*. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. *J Appl Physiol* 2001;**91**:218–24.
38. **Leger LA**, Mercier D, Gadoury C, *et al*. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988;**6**:93–101.
39. **McSwegin PJ**, Plowman SA, Wolff GM, *et al*. The validity of a one mile walk test for high school age individuals. *Meas Phys Educ Exerc Sci* 1998;**2**:47–63.
40. **Hunt BR**, George JD, Vehrs PR, *et al*. Validity of a submaximal 1-mile track jog test in predicting  $\text{VO}_{2\text{max}}$  in fit teenagers. *Pediatr Exerc Sci* 2000;**12**:80–90.
41. **Castro-Piñero J**, Ortega FB, Mora J, *et al*. Criterion related validity of 1/2 mile run-walk test for estimating  $\text{VO}_{2\text{peak}}$  in children aged 6–17 years. *Int J Sports Med* 2009;**30**:366–71.
42. **España-Romero V**, Ortega FB, Vicente-Rodríguez G, *et al*. Elbow position affects handgrip strength in adolescents: validity and reliability of Jamar, DynEx and TKK dynamometers. *J Strength Cond Res*. In press.
43. **Milliken LA**, Faigenbaum AD, Loud RL, *et al*. Correlates of upper and lower body muscular strength in children. *J Strength Cond Res* 2008;**22**:1339–46.

44. **Woods JA**, Pate RR, Burgess ML. Correlates to performance on field tests of muscular strength. *Pediatr Exerc Sci* 1992;**4**:302–11.
45. **Patterson P**, Wiksten DL, Ray L, *et al*. The validity and reliability of the back saver sit-and-reach test in middle school girls and boys. *Res Q Exerc Sport* 1996;**67**:448–51.
46. **Patterson P**, Rethwisch N, Wiksten D. Reliability of the trunk lift in high school boys and girls. *Meas Phys Educ Exerc Sci* 1997;**1**:145–51.
47. **Mayhew JL**, Ball TE, Ward TE, *et al*. Relationships of structural dimensions to bench press strength in college males. *J Sports Med Phys Fitness* 1991;**31**:135–41.
48. **Leighton JR**. An instrument and technic for the measurement of range of joint motion. *Arch Phys Med Rehabil* 1955;**36**:571–8.
49. **Jackson A**, Langford NJ. The criterion-related validity of the sit and reach test: replication and extension of previous findings. *Res Q Exerc Sport* 1989;**60**:384–7.
50. **Kanbur NO**, Duzgun I, Derman O, *et al*. Do sexual maturation stages affect flexibility in adolescent boys aged 14 years? *J Sports Med Phys Fitness* 2005;**45**:53–7.
51. **Shechtman O**, MacKinnon L, Locklear C. Using the BTE Primus to measure grip and wrist flexion strength in physically active wheelchair users: an exploratory study. *Am J Occup Ther* 2001;**55**:393–400.
52. **Shechtman O**, Gestewitz L, Kimble C. Reliability and validity of the DynEx dynamometer. *J Hand Ther* 2005;**18**:339–47.
53. **Bellace JV**, Healy D, Besser MP, *et al*. Validity of the Dexter Evaluation System's Jamar dynamometer attachment for assessment of hand grip strength in a normal population. *J Hand Ther* 2000;**13**:46–51.
54. **Mathiowetz V**, Vizenor L, Melander D. Comparison of baseline instruments to the Jamar dynamometer and the B&L engineering pinch gauge. *Occup Ther J Res* 2000;**20**:147–62.
55. **Mathiowetz V**. Comparison of Rolyan and Jamar dynamometers for measuring grip strength. *Occup Ther Int* 2002;**9**:201–9.
56. **Hamilton GF**, McDonald C, Chenier TC. Measurement of grip strength: validity and reliability of the sphygmomanometer and jamar grip dynamometer. *J Orthop Sports Phys Ther* 1992;**16**:215–19.
57. **Stephens JL**, Pratt N, Michlovitz S. The reliability and validity of the Tekdyne hand dynamometer: Part II. *J Hand Ther* 1996;**9**:18–26.
58. **Beaton DE**, O'Driscoll SW, Richards RR. Grip strength testing using the BTE work simulator and the Jamar dynamometer: a comparative study. Baltimore Therapeutic Equipment. *J Hand Surg Am* 1995;**20**:293–8.
59. **Lusardi M**, Bohannon R. Hand grip strength: comparability of measurements obtained with a Jamar dynamometer and a modified sphygmomanometer. *J Hand Ther* 1991;**4**:117–22.
60. **Espana-Romero V**, Artero EG, Santaliestra-Pasias AM, *et al*. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *J Hand Surg [Am]* 2008;**33**:378–84.
61. **Ruiz JR**, Espana-Romero V, Ortega FB, *et al*. Hand span influences optimal grip span in male and female teenagers. *J Hand Surg [Am]* 2006;**31**:1367–72.
62. **Castro-Piñero J**, González-Montesinos JL, Mora J, *et al*. Percentile values for muscular strength field tests in children aged 6 to 17 years; influence of weight status. *J Strength Cond Res*. In press.
63. **Rodríguez PN**, Bermudez EF, Rodríguez GS, *et al*. Body composition by simple anthropometry, bioimpedance and DXA in preschool children: interrelationships among methods. *Arch Argent Pediatr* 2008;**106**:102–9.
64. **Guida B**, Pietrobelli A, Trio R, *et al*. Body mass index and bioelectrical vector distribution in 8-year-old children. *Nutr Metab Cardiovasc Dis* 2008;**18**:133–41.
65. **Goran MI**, Gower BA, Treuth M, *et al*. Prediction of intra-abdominal and subcutaneous abdominal adipose tissue in healthy pre-pubertal children. *Int J Obes Relat Metab Disord* 1998;**22**:549–58.
66. **Semiz S**, Ozgoren E, Sabir N. Comparison of ultrasonographic and anthropometric methods to assess body fat in childhood obesity. *Int J Obes (Lond)* 2007;**31**:53–8.
67. **Goulding A**, Gold E, Cannan R, *et al*. DEXA supports the use of BMI as a measure of fatness in young girls. *Int J Obes Relat Metab Disord* 1996;**20**:1014–21.
68. **Pietrobelli A**, Faith MS, Allison DB, *et al*. Body mass index as a measure of adiposity among children and adolescents: a validation study. *J Pediatr* 1998;**132**:204–10.
69. **Ellis KJ**, Abrams SA, Wong WW. Monitoring childhood obesity: assessment of the weight/height index. *Am J Epidemiol* 1999;**150**:939–46.
70. **Daniels SR**, Khoury PR, Morrison JA. The utility of body mass index as a measure of body fatness in children and adolescents: differences by race and gender. *Pediatrics* 1997;**99**:804–7.
71. **Moreno LA**, Blay MG, Rodriguez G, *et al*. Screening performances of the International Obesity Task Force body mass index cut-off values in adolescents. *J Am Coll Nutr* 2006;**25**:403–8.
72. **Steinberger J**, Jacobs DR, Rautz S, *et al*. Comparison of body fatness measurements by BMI and skinfolds vs dual energy x-ray absorptiometry and their relation to cardiovascular risk factors in adolescents. *Int J Obes (Lond)* 2005;**29**:1346–52.
73. **Sarria A**, Garcia-Llop LA, Moreno LA, *et al*. Skinfold thickness measurements are better predictors of body fat percentage than body mass index in male Spanish children and adolescents. *Eur J Clin Nutr* 1998;**52**:573–6.
74. **Freedman DS**, Wang J, Ogden CL, *et al*. The prediction of body fatness by BMI and skinfold thicknesses among children and adolescents. *Ann Hum Biol* 2007;**34**:183–94.
75. **Eliakim A**, Ish-Shalom S, Giladi A, *et al*. Assessment of body composition in ballet dancers: correlation among anthropometric measurements, bio-electrical impedance analysis, and dual-energy x-ray absorptiometry. *Int J Sports Med* 2000;**21**:598–601.
76. **Cassady SL**, Nielsen DH, Janz KF, *et al*. Validity of near infrared body composition analysis in children and adolescents. *Med Sci Sports Exerc* 1993;**25**:1185–91.
77. **Lindsay RS**, Hanson RL, Roumain J, *et al*. Body mass index as a measure of adiposity in children and adolescents: relationship to adiposity by dual energy x-ray absorptiometry and to cardiovascular risk factors. *J Clin Endocrinol Metab* 2001;**86**:4061–7.
78. **Mei Z**, Grummer-Strawn LM, Pietrobelli A, *et al*. Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am J Clin Nutr* 2002;**75**:978–85.
79. **Deurenberg P**, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. *Br J Nutr* 1991;**65**:105–14.
80. **Bray GA**, DeLany JP, Harsha DW, *et al*. Evaluation of body fat in fatter and leaner 10-y-old African American and white children: the Baton Rouge Children's Study. *Am J Clin Nutr* 2001;**73**:687–702.
81. **Taylor RW**, Jones IE, Williams SM, *et al*. Body fat percentages measured by dual-energy x-ray absorptiometry corresponding to recently recommended body mass index cutoffs for overweight and obesity in children and adolescents aged 3–18 y. *Am J Clin Nutr* 2002;**76**:1416–21.
82. **Eisenmann JC**, Heelan KA, Welk GJ. Assessing body composition among 3- to 8-year-old children: anthropometry, BIA, and DXA. *Obes Res* 2004;**12**:1633–40.
83. **Brambilla P**, Bedogni G, Moreno LA, *et al*. Crossvalidation of anthropometry against magnetic resonance imaging for the assessment of visceral and subcutaneous adipose tissue in children. *Int J Obes (Lond)* 2006;**30**:23–30.
84. **Freedman DS**, Wang J, Maynard LM, *et al*. Relation of BMI to fat and fat-free mass among children and adolescents. *Int J Obes (Lond)* 2005;**29**:1–8.
85. **Gutin B**, Litaker M, Islam S, *et al*. Body-composition measurement in 9–11-yr-old children by dual-energy x-ray absorptiometry, skinfold-thickness measurements, and bioimpedance analysis. *Am J Clin Nutr* 1996;**63**:287–92.
86. **Janz KF**, Nielsen DH, Cassady SL, *et al*. Cross-validation of the Slaughter skinfold equations for children and adolescents. *Med Sci Sports Exerc* 1993;**25**:1070–6.
87. **Treuth MS**, Butte NF, Wong WW, *et al*. Body composition in prepubertal girls: comparison of six methods. *Int J Obes Relat Metab Disord* 2001;**25**:1352–9.
88. **Rodríguez G**, Moreno LA, Blay MG, *et al*. Body fat measurement in adolescents: comparison of skinfold thickness equations with dual-energy x-ray absorptiometry. *Eur J Clin Nutr* 2005;**59**:1158–66.
89. **Buisson AM**, Ittenbach RF, Stallings VA, *et al*. Methodological agreement between two-compartment body-composition methods in children. *Am J Hum Biol* 2006;**18**:470–80.
90. **Ihmels M**, Welk GJ, McClain JJ, *et al*. The reliability and convergent validity of field tests of body composition in young adolescents. *JPAH* 2006;**3**:S67–77.
91. **Campanozzi A**, Dabbas M, Ruiz JC, *et al*. Evaluation of lean body mass in obese children. *Eur J Pediatr* 2008;**167**:533–40.
92. **Paineau D**, Chiheb S, Banu I, *et al*. Comparison of field methods to estimate fat mass in children. *Ann Hum Biol* 2008;**35**:185–97.
93. **Taylor RW**, Jones IE, Williams SM, *et al*. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy x-ray absorptiometry, in children aged 3–19 y. *Am J Clin Nutr* 2000;**72**:490–5.
94. **Pietrobelli A**, Tato L. Body composition measurements: from the past to the future. *Acta Paediatr Suppl* 2005;**94**:8–13.
95. **Sopher AB**, Thornton JC, Wang J, *et al*. Measurement of percentage of body fat in 411 children and adolescents: a comparison of dual-energy x-ray absorptiometry with a four-compartment model. *Pediatrics* 2004;**113**:1285–90.
96. **Ferrozzi F**, Zuccoli G, Tognini G, *et al*. An assessment of abdominal fatty tissue distribution in obese children. A comparison between echography and computed tomography [in Spanish]. *Radiol Med (Torino)* 1999;**98**:490–4.
97. **Fields DA**, Hull HR, Chelene AJ, *et al*. Child-specific thoracic gas volume prediction equations for air-displacement plethysmography. *Obes Res* 2004;**12**:1797–804.
98. **Urlando A**, Dempster P, Aitkens S. A new air displacement plethysmograph for the measurement of body composition in infants. *Pediatr Res* 2003;**53**:486–92.
99. **Pietrobelli A**, Andreoli A, Cervelli V, *et al*. Predicting fat-free mass in children using bioimpedance analysis. *Acta Diabetol* 2003;**40**(Suppl 1):S212–15.
100. **Pietrobelli A**, Rubiano F, St-Onge MP, *et al*. New bioimpedance analysis system: improved phenotyping with whole-body analysis. *Eur J Clin Nutr* 2004;**58**:1479–84.