

European normative values for physical fitness in children and adolescents aged 9–17 years: results from 2 779 165 Eurofit performances representing 30 countries

Grant R Tomkinson,^{1,2} Kevin D Carver,¹ Frazer Atkinson,¹ Nathan D Daniell,² Lucy K Lewis,^{2,3} John S Fitzgerald,¹ Justin J Lang,⁴ Francisco B Ortega^{5,6}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bjsports-2017-098253>).

¹Department of Kinesiology and Public Health Education, University of North Dakota, Grand Forks, North Dakota, USA

²Alliance for Research in Exercise, Nutrition and Activity (ARENA), School of Health Sciences & Sansom Institute for Health Research, University of South Australia, Adelaide, Australia

³Discipline of Physiotherapy, Flinders University, Adelaide, Australia

⁴Healthy Active Living and Obesity (HALO) Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

⁵The PROFITH Research Group, Department of Physical Education and Sports, Faculty of Sports Sciences, University of Granada, Granada, Spain

⁶Department of Biosciences and Nutrition, Karolinska Institute, Huddinge, Sweden

Correspondence to

Dr Grant R Tomkinson, Department of Kinesiology and Public Health Education, University of North Dakota, Grand Forks, ND 58202, USA; grant.tomkinson@und.edu

Accepted 30 October 2017
Published Online First
30 November 2017

ABSTRACT

Objective To develop sex-specific and age-specific normative values for the nine Eurofit tests in European children and adolescents aged 9–17 years.

Methods A systematic review was undertaken to identify papers that explicitly reported descriptive results for at least one of nine Eurofit tests (measuring balance, muscular strength, muscular endurance, muscular power, flexibility, speed, speed-agility and cardiorespiratory fitness (CRF)) on children and adolescents. Data were included on apparently healthy (free from known disease/injury) children and adolescents aged 9–17 years. Following harmonisation for methodological variation where appropriate, pseudodata were generated using Monte Carlo simulation, with population-weighted sex-specific and age-specific normative centiles generated using the Lambda Mu Sigma (LMS) method. Sex-specific and age-specific differences were expressed as standardised differences in means, with the percentage of children and adolescents with healthy CRF estimated at the sex-age level.

Results Norms were displayed as tabulated centiles and as smoothed centile curves for the nine Eurofit tests. The final dataset included 2 779 165 results on children and adolescents from 30 European countries, extracted from 98 studies. On average, 78% of boys (95% CI 72% to 85%) and 83% of girls (95% CI 71% to 96%) met the standards for healthy CRF, with the percentage meeting the standards decreasing with age. Boys performed substantially (standardised differences >0.2) better than girls on muscular strength, muscular power, muscular endurance, speed-agility and CRF tests, but worse on the flexibility test. Physical fitness generally improved at a faster rate in boys than in girls, especially during the teenage years.

Conclusion This study provides the largest and most geographically representative sex-specific and age-specific European normative values for children and adolescents, which have utility for health and fitness screening, profiling, monitoring and surveillance.

BACKGROUND

Physical fitness is a good summative measure of the body's ability to perform physical activity and exercise, and it also provides an important summative indicator of health.¹ In adults, cardiorespiratory fitness (CRF) and musculoskeletal fitness (MSF) are strongly associated with mortality and cancer, independent of obesity and

physical activity levels.^{2–5} Several studies have shown considerably stronger inverse relationships between CRF and mortality than between physical activity and mortality,^{6–7} indicating that changes in CRF may be more important to monitor in response to intervention (eg, exercise training). In children and adolescents, favourable associations have been reported linking CRF and MSF to cardiometabolic disease risk, adiposity, mental health and cognition as well as MSF to bone health.^{1–10} Direct evidence has also emerged indicating that low CRF and MSF in adolescence are significantly associated with all-cause mortality later in life.^{11–13} In addition to the health implications, physical fitness is an important determinant of success for many popular youth sports and athletic events (eg, hockey, basketball, football (soccer), running, swimming, rugby).¹⁴

Since its inception in 1988, the Eurofit has become the most popular test battery used to assess the physical fitness of European children and adolescents and the effectiveness of national physical education curricula.^{15–16} The Eurofit comprises numerous health-related and skill-related fitness tests, including: (1) flamingo balance (balance), plate tapping (upper body speed), sit-and-reach (extent flexibility), standing broad jump (lower body muscular power), handgrip strength (upper body muscular strength), sit-ups (abdominal muscular endurance), bent arm hang (upper body muscular endurance), 10×5 m agility shuttle run (running speed-agility) and the 20 m shuttle run (CRF) (see online supplement 1); (2) anthropometric tests measuring height, mass and skinfold (various sites) and (3) age-identification and sex-identification data.¹⁷ The Eurofit has excellent field-based utility because it is cheap and simple to administer, is practical in the school and club settings, requires minimal equipment and personnel and is appropriate for mass testing.¹⁶ The Eurofit tests demonstrate very good test-retest reliability and good criterion validity for tests where appropriate criterion measures have been identified (eg, the 20 m shuttle run, standing broad jump, handgrip strength),^{18–21} suggesting that it is a good test battery to measure physical fitness in youth. Criterion-referenced standards have also been developed for some Eurofit tests (eg, CRF) to help identify children and adolescents with apparently healthy cardiometabolic



To cite: Tomkinson GR, Carver KD, Atkinson F, et al. *Br J Sports Med* 2018;**52**:1445–1456.

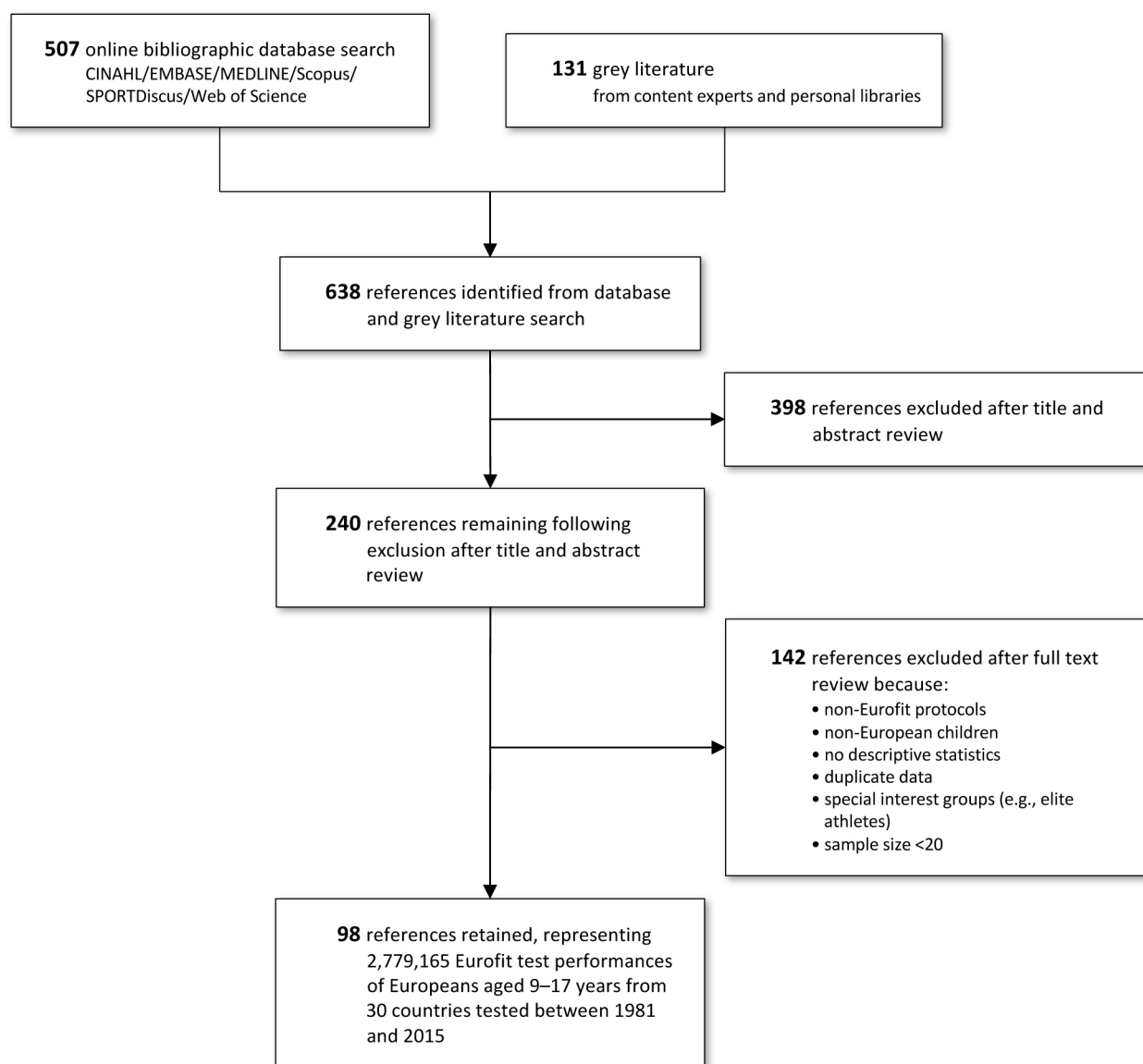


Figure 1 PRISMA flow chart outlining the flow of studies through the review.

profiles.^{22,23} Several of the Eurofit tests have been supported by European experts from the ALPHA (Assessing Levels of Physical Activity) project²⁰ and by North American experts from the IOM (Institute of Medicine) report,²⁴ both of which provide strong and consistent guidelines about fitness testing in children and adolescents.

In order to extend the utility of the Eurofit as a surveillance instrument, there is a clear need for European normative-referenced standards to help interpret test scores, which are currently only available at the local, state/provincial or national level.^{25–29} Previously, Tomkinson *et al*¹⁶ used a method to match and compare Eurofit data in children and adolescents by standardising differences in test protocols and performance metrics. These data helped describe the geographical variability in the Eurofit performance of 1.2 million European children and adolescents aged 7–18 years from 23 countries,¹⁶ and could be updated to provide European norms. Thus, the primary aim of this study was to develop sex-specific and age-specific normative values for physical fitness in European children and adolescents using the Eurofit, which implies a 10-year update to the previous

Tomkinson *et al* review.¹⁶ The secondary aim was to estimate the sex-related differences in Eurofit test performance as well as the percentage of European children and adolescents meeting the new international criterion-referenced standards for healthy CRF.²³

METHODS

Data sources

A systematic review of the scientific literature was prospectively registered (PROSPERO 2013:CRD42013003646) and completed to locate studies that reported descriptive Eurofit data on European children and adolescents aged 9–17 years (see online supplement 2). This review was undertaken according to the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) guidelines for systematic reviews.³⁰ Studies were identified from January 1988 up until December 2016 using the following bibliographic databases: CINAHL, EMBASE, MEDLINE, Scopus, SPORTDiscus and Web of Science. This search strategy was developed by the author group

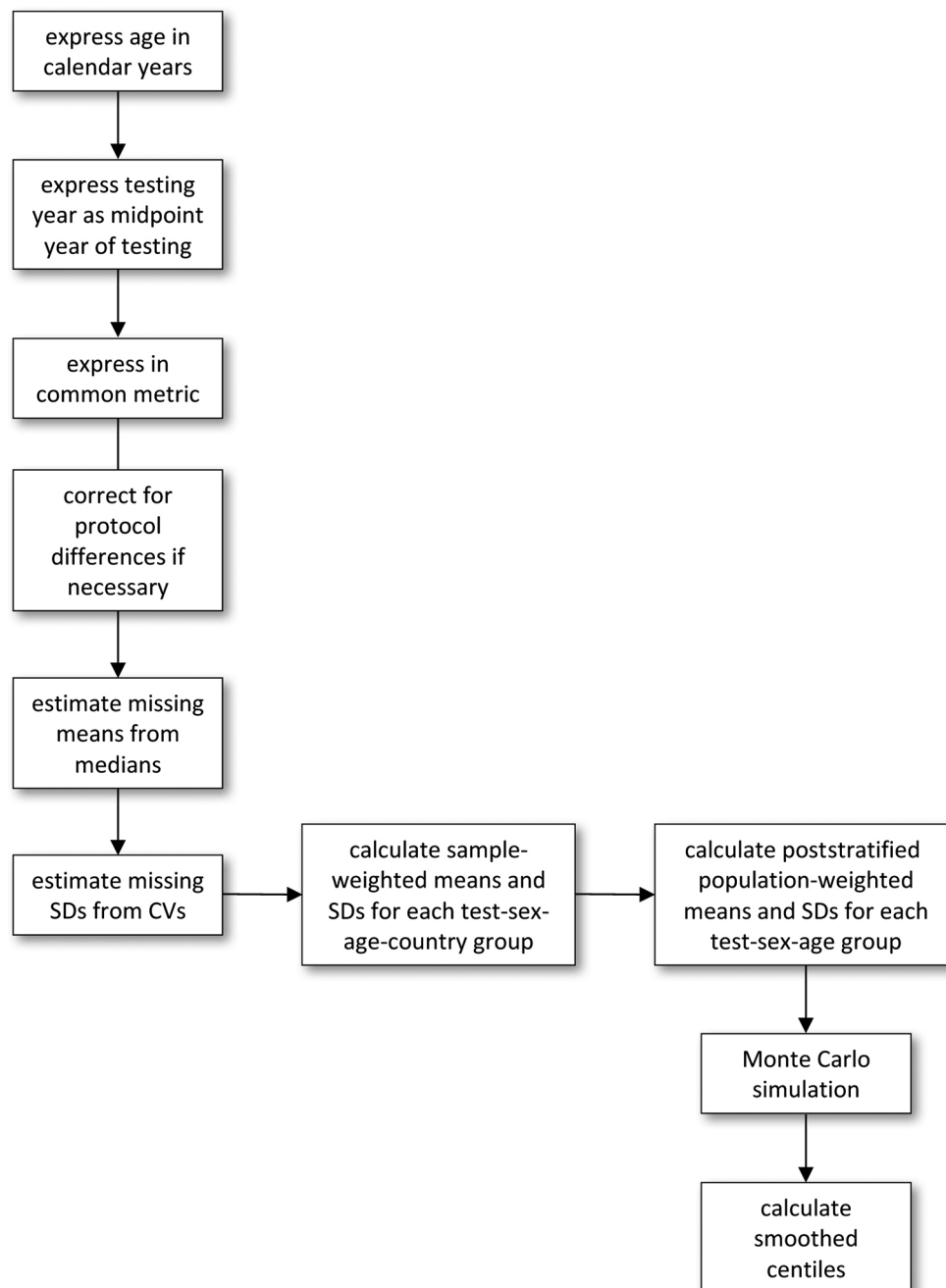


Figure 2 Flow chart showing the methodological procedure used in this study. Results from studies were first expressed in a common metric and corrected for protocol differences. Following the estimation of missing means and SDs if necessary, poststratified population-weighted means and SDs were estimated for each test-sex-age group, with pseudodata and smoothed centiles subsequently generated. CV, coefficient of variation.

in conjunction with a trained academic librarian. The search strategy included the term: Eurofit; with child*, OR adolescen*, OR youth, OR boy*, OR girl*, OR teen*, OR paediatric*, OR pediatric*, as search term modifiers. All studies were extracted as text files, imported into RefWorks (ProQuest, Ann Arbor, Michigan, USA) and assigned a unique reference identification number. Duplicate studies were first removed using RefWorks with the remaining duplicates removed manually. Two independent reviewers screened all titles and abstracts for eligibility, with full-text copies obtained for all studies meeting initial screening criteria according to at least one reviewer. These two independent reviewers then examined all full-text articles and discrepancies

were resolved by discussion and consensus. A third reviewer examined an article when the two reviewers were unable to reach consensus, with consensus reached for all included articles. Email contact with the corresponding authors of studies occurred when necessary, in order to provide clarification, to avoid 'double counting' previously reported data and/or to request additional descriptive or raw data. The reference lists of all included studies were manually reviewed by two reviewers to identify new studies. Reviewers contacted content experts to obtain grey literature. In addition, the personal libraries of the authors were examined for relevant studies not identified through the search strategy.

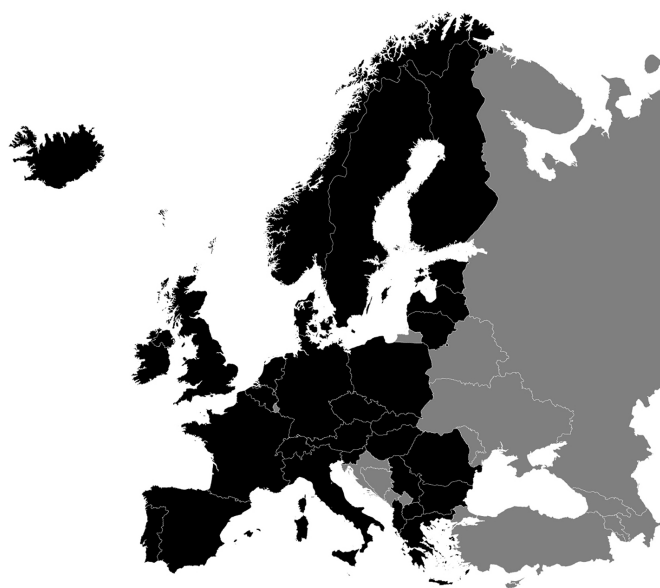


Figure 3 European map indicating the 30 countries (filled in black) for which Eurofit data on children and adolescents aged 9–17 years were available.

Inclusion/exclusion criteria

Studies were included if they explicitly reported descriptive Eurofit data at the test-sex-age-country-year level. Study participants must have been apparently healthy (free from known disease or injury) European children and adolescents aged 9–17 years who were tested from 1981 onwards—the inception year of the provisional Eurofit test battery. Studies were excluded if they reported descriptive Eurofit data on: (1) test-sex-age-country-year groups for which the sample size was less than 20 (because the means and SDs for smaller samples were too labile); (2) duplicate data published in another included study or (3) on only special interest groups that were atypical of their source population (eg, elite athletes, physically or mentally impaired children). [Figure 1](#) shows a PRISMA flow chart of the included studies.

Data treatment and statistical analysis

All descriptive data were extracted into Excel (Microsoft Office 2010, USA) using a standardised data extraction table. The following descriptive data were extracted by one author and checked for accuracy by another: authors, country of testing, year of testing, sex, age, Eurofit test (including data on the name of test, measurement units, sample size, mean, SD and median), sampling method and the sampling base. Mean data were examined for anomalies by running range checks and examining sex-specific and age-specific scatter plots, with means ± 2 SEs of the mean away from the respective sex-age-test level mean identified and checked for transcription errors. Only data on children and adolescents aged 9–17 years were retained for further analysis.

The general procedure used to generate the sex-specific and age-specific normative centiles from extracted data is described elsewhere³¹ and summarised in [figure 2](#). Age was reported as age at last birthday (70% or 69/98 studies), a span of years (6% or 6/98 studies) or as mean and SD years (24% or 23/98 studies). Testing year was recorded as the midpoint year of testing (47% or 46/98 studies), a span of testing years (38% or 37/98 studies) or not reported at all (15% or 15/98 studies). Age and testing

year were therefore expressed as age at last birthday and the midpoint year of testing, respectively.³¹

To combine data from different studies, all Eurofit data were standardised to a common metric and protocol. Measurement units reported in the Eurofit handbook¹⁷ were used as the test-specific common metrics and for the presentation of normative centiles. All 20 m shuttle run data were standardised to Léger's 1-min protocol,³² which starts at a speed of 8.5 km/hour and increases by 0.5 km/hour each minute and the speed at the last completed stage using the procedures described elsewhere.^{31 33} The accuracy of the 20 m shuttle run data standardisation procedure is excellent.³³

As part of the modelling procedure used to generate sex-specific and age-specific norms, means and SDs were required at the study-test-sex-age-country-year level. If no mean was available (1% or 1/98 studies), then mean values were estimated from the reported median values. This was done by first locating all studies reporting both median and mean values at the study-test-sex-age-country-year level and second, by determining the best-fitting and most parsimonious linear or curvilinear (second-order and third-order polynomials) regression models between median (predictor variable) and mean (response variable) values. Furthermore, 4% (4/98) of studies did not report SD values. Missing SD values were estimated by first locating all studies reporting both means and SDs at the study-test-sex-age-country-year level; second, by calculating the corresponding coefficient of variation (CV) values and third, by calculating the sample-weighted mean CVs for boys and girls separately.

Sample-weighted means and SDs (the latter calculated from sample-weighted mean CVs) were then calculated at the test-sex-age-country level. While these data represent the best available Eurofit data, in order to best generate European representative sex-specific and age-specific normative centiles and to correct for systematic bias associated with oversampling and undersampling, means and SDs were corrected using a poststratification population-weighting procedure.³⁴ This procedure ensures that our norms were standardised to underlying country-sex-age demographics. Thus, population estimates standardised to the mean testing year of 2000 were extracted from the United Nations World Population Prospects report.³⁵ Monte Carlo simulation was then used to create pseudodata using the detailed methods described elsewhere.³⁶ This simulation procedure attempts to 'recreate' the unavailable raw data by using a random number generator to produce data points based on population-weighted means and SDs at the sex-age level. Monte Carlo simulation assumes that the distributions are approximately normal, which was not true of all available raw Eurofit data. The simulation procedure described by Tomkinson *et al*³⁶ however allowed for the recreation of both normal and non-normal pseudodata, with Eurofit data considered to be either normal or non-normal following the assessment of normality by the d'Agostino-Pearson K^2 test³⁷ using available raw data of the same test. Pseudo-datasets were repeatedly generated until the calculated mean differed from the reported mean by $<0.5\%$, and the calculated SD differed from the reported SD by $<2.5\%$. These pseudo-datasets were then used to generate sex-specific and age-specific normative centiles in LMSchartmaker Pro (V2.43, The Institute of Child Health, London, UK), which analyses data using the Lambda Mu Sigma (LMS) method.³⁸ The LMS method fits smooth centile curves to reference data by summarising the changing distribution of three sex-specific and age-specific curves representing the skewness (L; expressed as a Box-Cox power), the median (M) and the CV (S). Using penalised likelihood, the curves can be fitted as cubic splines using non-linear regression, and the extent

Table 1 Flamingo balance (n/60 s) centiles by age and sex based on 123 655 test performances of children and adolescents aged 9–17 years representing 19 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	3691	24	21	18	15	13	12	10	9	7	5	4
10	5140	25	22	18	16	14	12	10	8	7	5	3
11	6409	26	22	18	16	14	12	10	8	7	4	3
12	8313	26	23	18	16	14	12	10	8	7	4	3
13	8750	26	23	18	16	14	12	10	8	6	4	3
14	9466	25	21	18	15	13	11	10	8	6	4	3
15	7605	21	18	15	13	11	10	9	7	6	4	3
16	6665	21	18	15	13	11	10	8	7	6	4	3
17	5940	21	18	15	13	11	10	8	7	6	4	3
Girls												
9	3654	23	20	17	14	13	11	10	8	7	5	3
10	4935	23	20	17	15	13	11	10	8	7	5	3
11	6247	24	20	17	15	13	11	10	8	7	5	3
12	8271	24	21	17	15	13	11	10	8	7	5	3
13	8958	23	20	17	15	13	11	10	8	7	5	3
14	9279	23	20	16	14	13	11	10	8	7	5	3
15	7956	21	18	15	13	12	10	9	8	6	4	3
16	6644	19	17	14	12	11	9	8	7	6	4	3
17	5732	18	16	13	12	10	9	8	7	5	4	3

Note: the ages shown represent age at last birthday (eg, 9=9.00–9.99).

of smoothing required can be expressed in terms of smoothing parameters or equivalent df.³⁹

The percentage of children and adolescents with healthy CRF (ie, healthy cardiometabolic profiles) was estimated using the new international criterion-referenced standards of 42 and 35 mL/kg/min for boys and girls, respectively.²³ Sex-specific differences in mean Eurofit performance were expressed as standardised differences. Positive differences indicated that Eurofit performances for boys were better than those for girls. Standardised

differences of 0.2, 0.5 and 0.8 were used as thresholds for small, moderate and large effect sizes (ES), respectively.⁴⁰

RESULTS

The final dataset included 2 779 165 Eurofit test performances of European children and adolescents aged 9–17 years (6458 study-sex-age-country-year groups extracted from 98 studies), representing 30 countries (figure 3). These 30 countries

Table 2 Plate tapping (s) centiles by age and sex based on 148 093 test performances of children and adolescents aged 9–17 years representing 19 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	7543	24.05	22.04	20.00	18.74	17.78	16.96	16.21	15.48	14.70	13.73	13.02
10	9090	21.55	19.90	18.19	17.13	16.31	15.61	14.97	14.33	13.65	12.80	12.17
11	8198	19.48	18.11	16.68	15.77	15.07	14.46	13.90	13.35	12.75	12.00	11.44
12	9799	17.91	16.74	15.51	14.72	14.10	13.57	13.07	12.58	12.05	11.37	10.87
13	9104	16.44	15.44	14.37	13.69	13.15	12.68	12.25	11.81	11.34	10.74	10.28
14	9964	15.12	14.26	13.34	12.74	12.27	11.86	11.48	11.09	10.67	10.13	9.72
15	7797	14.00	13.25	12.45	11.92	11.51	11.14	10.80	10.45	10.07	9.59	9.22
16	7217	13.38	12.70	11.95	11.46	11.08	10.74	10.42	10.10	9.74	9.29	8.94
17	6157	13.11	12.45	11.73	11.26	10.89	10.56	10.25	9.94	9.59	9.15	8.82
Girls												
9	7121	25.25	22.05	19.29	17.77	16.70	15.83	15.06	14.34	13.60	12.72	12.09
10	8904	22.35	19.95	17.77	16.54	15.64	14.90	14.25	13.62	12.97	12.19	11.63
11	8561	19.93	18.11	16.38	15.38	14.63	14.01	13.45	12.91	12.35	11.66	11.16
12	10 089	18.41	16.96	15.53	14.68	14.04	13.50	13.01	12.53	12.03	11.41	10.95
13	9031	16.92	15.76	14.60	13.89	13.35	12.88	12.46	12.05	11.60	11.05	10.64
14	9476	15.51	14.58	13.63	13.03	12.57	12.18	11.81	11.45	11.06	10.58	10.21
15	7690	14.95	14.12	13.25	12.70	12.28	11.91	11.57	11.24	10.87	10.41	10.07
16	6790	14.58	13.80	12.99	12.48	12.07	11.73	11.41	11.08	10.74	10.30	9.97
17	5562	14.54	13.77	12.96	12.45	12.05	11.71	11.39	11.07	10.72	10.28	9.95

Table 3 Sit-and-reach (cm) centiles by age and sex based on 464 807 test performances of children and adolescents aged 9–17 years representing 27 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	34 495	6.0	8.1	10.7	12.7	14.4	16.0	17.6	19.4	21.4	24.3	26.8
10	35 532	6.0	8.1	10.8	12.7	14.4	16.1	17.7	19.4	21.5	24.5	26.9
11	35 413	6.0	8.1	10.8	12.7	14.4	16.1	17.7	19.4	21.5	24.5	26.9
12	29 962	6.0	8.2	10.8	12.8	14.5	16.1	17.8	19.6	21.7	24.6	27.1
13	26 840	6.1	8.3	11.1	13.1	14.8	16.5	18.2	20.0	22.2	25.2	27.7
14	25 302	6.7	9.1	12.1	14.3	16.2	18.0	19.9	21.9	24.2	27.5	30.3
15	21 644	7.7	10.3	13.7	16.1	18.3	20.3	22.4	24.6	27.2	30.9	34.0
16	16 285	8.4	11.1	14.6	17.1	19.3	21.4	23.6	25.9	28.6	32.4	35.6
17	9 696	9.1	11.9	15.5	18.1	20.4	22.6	24.8	27.2	30.0	33.9	37.2
Girls												
9	33 008	7.9	10.2	13.1	15.2	16.9	18.6	20.3	22.1	24.2	27.2	29.6
10	34 803	8.5	10.8	13.7	15.7	17.5	19.2	20.9	22.7	24.8	27.7	30.1
11	35 250	9.4	11.7	14.5	16.6	18.4	20.1	21.7	23.5	25.6	28.6	31.0
12	29 835	10.6	12.9	15.8	17.9	19.7	21.4	23.1	24.9	27.1	30.0	32.5
13	26 090	11.9	14.4	17.3	19.5	21.3	23.1	24.8	26.7	28.9	31.9	34.4
14	24 563	13.1	15.6	18.6	20.8	22.7	24.5	26.3	28.2	30.4	33.5	36.1
15	20 540	13.9	16.4	19.5	21.7	23.6	25.4	27.2	29.1	31.3	34.4	37.0
16	16 197	14.4	16.9	20.0	22.2	24.1	25.9	27.6	29.5	31.8	34.9	37.5
17	9 352	14.7	17.2	20.3	22.5	24.4	26.1	27.9	29.8	32.1	35.2	37.8

Note: a score of 15 cm corresponds to the participant reaching their toes.

represented approximately 65% of Europe's population and 49% of Europe's land area and included 25 high-income and five upper-middle-income countries. Online supplement 3 provides a summary of the 98 included studies.

Tables 1–9 provide normative values as tabulated centiles from 5% to 95% for all nine Eurofit tests. Smoothed centile curves are presented in figure 4 with additional 20 m shuttle run norms (speed at last completed stage, number of laps and relative $\dot{V}O_{2peak}$) presented in online supplement 4.

On average, 78% of boys (95% CI 72% to 85%) and 83% of girls (95% CI 71% to 96%) had healthy CRF, with the percentage of those with healthy CRF decreasing by about 3% (boys) and 7% (girls) per year from the age of 9 years onwards (figure 5). There was considerable variability in healthy CRF levels among different European countries, which increased with age (see online supplement 5). When dividing Europe into two segments at the 45th parallel north,^{41 42} a gradient existed where Northern-Central European countries had a higher percentage

Table 4 Standing broad jump (cm) centiles by age and sex based on 464 900 test performances of children and adolescents aged 9–17 years representing 29 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	35 148	100.5	107.9	116.8	123.2	128.7	133.8	138.9	144.3	150.7	159.5	166.8
10	36 069	107.6	115.3	124.6	131.3	137.0	142.4	147.7	153.4	160.1	169.3	176.9
11	35 618	115.4	123.5	133.3	140.3	146.3	151.9	157.5	163.5	170.5	180.2	188.2
12	30 631	122.5	131.0	141.2	148.5	154.8	160.7	166.5	172.8	180.1	190.3	198.6
13	24 760	129.7	138.5	149.3	157.0	163.6	169.7	175.9	182.5	190.2	200.9	209.7
14	24 061	138.7	148.1	159.6	167.8	174.8	181.4	188.0	195.0	203.2	214.6	223.9
15	20 334	147.8	157.8	169.8	178.5	186.0	192.9	199.8	207.2	215.9	227.9	237.8
16	18 967	154.2	164.5	176.9	185.9	193.6	200.8	207.9	215.6	224.6	237.0	247.2
17	12 108	158.3	168.9	181.6	190.7	198.5	205.8	213.1	221.0	230.1	242.7	253.2
Girls												
9	34 339	91.2	98.4	107.1	113.4	118.9	123.9	129.0	134.5	140.8	149.7	157.1
10	35 339	98.5	105.9	114.9	121.4	127.0	132.3	137.5	143.2	149.8	159.0	166.6
11	34 992	105.6	113.3	122.6	129.4	135.2	140.6	146.0	151.9	158.7	168.2	176.1
12	29 974	111.1	119.0	128.6	135.6	141.6	147.1	152.7	158.7	165.8	175.6	183.7
13	23 749	113.9	121.9	131.6	138.7	144.8	150.4	156.1	162.2	169.3	179.3	187.5
14	22 416	115.6	123.7	133.6	140.7	146.8	152.5	158.3	164.4	171.6	181.7	190.0
15	16 394	116.8	124.9	134.8	142.0	148.1	153.9	159.6	165.8	173.1	183.1	191.5
16	18 459	117.5	125.6	135.5	142.7	148.8	154.6	160.4	166.6	173.8	183.9	192.2
17	11 542	119.0	127.2	137.2	144.4	150.6	156.4	162.3	168.5	175.8	186.0	194.4

Table 5 Handgrip strength (kg) centiles by age and sex based on 203 295 test performances of children and adolescents aged 9–17 years representing 24 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	10 180	8.6	10.1	11.9	13.2	14.3	15.3	16.4	17.5	18.8	20.6	22.1
10	11 965	9.5	11.1	13.0	14.5	15.7	16.8	18.0	19.2	20.6	22.6	24.2
11	11 358	10.8	12.6	14.8	16.4	17.7	19.0	20.3	21.6	23.2	25.4	27.2
12	13 107	13.1	15.2	17.7	19.6	21.2	22.6	24.1	25.7	27.6	30.1	32.3
13	13 070	16.9	19.4	22.5	24.7	26.6	28.4	30.2	32.1	34.3	37.4	39.9
14	13 843	21.6	24.5	27.9	30.4	32.6	34.6	36.6	38.7	41.2	44.7	47.6
15	10 944	25.9	28.9	32.5	35.2	37.4	39.5	41.6	43.9	46.5	50.1	53.2
16	10 062	29.1	32.1	35.8	38.5	40.7	42.9	45.0	47.2	49.9	53.6	56.7
17	8157	31.3	34.3	38.0	40.6	42.9	45.0	47.1	49.4	52.1	55.7	58.8
Girls												
9	9690	7.2	8.7	10.4	11.6	12.6	13.6	14.6	15.6	16.8	18.5	19.9
10	11 804	8.0	9.6	11.5	12.9	14.1	15.2	16.3	17.5	18.8	20.7	22.3
11	11 582	9.4	11.2	13.4	14.9	16.3	17.5	18.8	20.1	21.7	23.9	25.6
12	13 331	12.0	13.9	16.2	17.9	19.3	20.6	21.9	23.3	25.0	27.3	29.1
13	13 182	16.1	18.0	20.3	21.9	23.3	24.6	25.9	27.3	29.0	31.2	33.1
14	13 168	18.5	20.4	22.7	24.3	25.7	27.1	28.4	29.8	31.4	33.7	35.6
15	10 586	19.1	21.1	23.5	25.2	26.7	28.0	29.4	30.8	32.5	34.9	36.8
16	9672	19.3	21.2	23.6	25.4	26.9	28.2	29.6	31.1	32.8	35.2	37.2
17	7594	19.4	21.4	23.8	25.5	27.0	28.4	29.8	31.3	33.0	35.5	37.4

of children and adolescents with healthy CRF than Southern European countries (average difference in means (range): 7% (0% to 27%) at the sex-age level).

On average, boys performed substantially better than girls at each age group on muscular strength (ES: large), muscular power (ES: large), muscular endurance (ES: moderate to large), speed-agility (ES: moderate) and CRF (ES: large) tests, with the magnitude of the sex-specific differences increasing with age and accelerating from about 12 years (figure 6). Boys also developed at a faster rate than girls on these tests, especially

during the teenage years. Conversely, girls performed substantially better at each age group on the flexibility test (ES: moderate), with boys and girls developing with age at similar rates. There were negligible sex-specific differences overall on the balance and upper body speed tests, although boys developed at a faster rate than girls on the upper body speed test.

DISCUSSION

This study systematically analysed 2 779 165 Eurofit performances of children and adolescents aged 9–17 years to generate

Table 6 Sit-ups (n/30 s) centiles by age and sex based on 481 032 performances of children and adolescents aged 9–17 years representing 23 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	31 757	9	11	13	15	16	17	18	20	21	23	25
10	33 748	11	13	15	17	18	19	20	22	23	25	27
11	35 559	13	14	16	18	19	20	22	23	24	26	28
12	29 338	14	15	17	19	20	21	22	24	25	27	29
13	30 805	14	16	18	20	21	22	23	24	26	28	29
14	29 024	15	17	19	20	22	23	24	25	27	29	30
15	22 541	17	18	20	22	23	24	25	26	28	30	31
16	18 751	18	19	21	22	24	25	26	27	29	30	32
17	12 059	18	20	22	23	24	25	27	28	29	31	33
Girls												
9	31 091	9	11	13	14	15	17	18	19	21	23	25
10	33 131	10	12	14	16	17	18	19	20	22	24	26
11	34 525	11	13	15	16	17	19	20	21	22	24	26
12	31 415	12	13	15	17	18	19	20	21	23	24	26
13	29 168	12	14	15	17	18	19	20	21	23	24	26
14	27 377	12	14	16	17	18	19	20	21	23	25	26
15	21 072	13	14	16	17	19	20	21	22	23	25	26
16	18 365	13	15	16	18	19	20	21	22	23	25	27
17	11 306	13	15	17	18	19	20	21	22	24	25	27

Table 7 Bent-arm hang (s) centiles by age and sex based on 189 673 test performances of children and adolescents aged 9–17 years representing 23 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	8282	1.48	2.13	3.29	4.49	5.85	7.48	9.55	12.38	16.74	25.36	35.62
10	9584	1.56	2.25	3.48	4.76	6.20	7.92	10.10	13.08	17.65	26.62	37.23
11	11079	1.63	2.35	3.66	5.00	6.51	8.32	10.60	13.71	18.46	27.73	38.62
12	11899	1.71	2.48	3.87	5.29	6.89	8.79	11.19	14.44	19.39	28.99	40.19
13	12321	1.90	2.77	4.33	5.92	7.70	9.81	12.44	15.99	21.34	31.57	43.30
14	12550	2.50	3.67	5.72	7.78	10.05	12.70	15.96	20.26	26.61	38.39	51.45
15	10576	3.73	5.40	8.26	11.05	14.04	17.43	21.50	26.72	34.18	47.44	61.48
16	9165	5.19	7.39	10.98	14.36	17.87	21.75	26.28	31.94	39.77	53.13	66.71
17	7425	6.48	9.03	13.07	16.74	20.45	24.46	29.04	34.64	42.19	54.66	66.92
Girls												
9	7681	0.98	1.43	2.24	3.08	4.02	5.14	6.55	8.46	11.36	16.94	23.40
10	9287	0.97	1.42	2.24	3.08	4.03	5.15	6.57	8.50	11.42	17.06	23.60
11	10942	0.96	1.42	2.23	3.08	4.03	5.16	6.59	8.53	11.48	17.18	23.79
12	13198	0.96	1.41	2.23	3.08	4.03	5.17	6.60	8.54	11.50	17.22	23.86
13	13613	0.96	1.41	2.23	3.08	4.03	5.18	6.62	8.58	11.56	17.33	24.04
14	13322	0.94	1.40	2.22	3.09	4.06	5.23	6.72	8.73	11.82	17.83	24.86
15	11324	0.92	1.38	2.23	3.11	4.13	5.35	6.91	9.05	12.34	18.80	26.41
16	9639	0.91	1.38	2.27	3.21	4.30	5.63	7.33	9.68	13.33	20.57	29.19
17	7786	0.93	1.43	2.40	3.45	4.67	6.16	8.11	10.82	15.07	23.61	33.92

the largest and most geographically representative sex-specific and age-specific European normative values for physical fitness. These norms add to existing norms across a range of other cardiometabolic risk factors, including adiposity (eg, body mass index^{43 44} and waist circumference,^{45–49} blood pressure,^{50 51} cholesterol,⁵¹ triglycerides⁵¹ and glucose).⁵¹ More importantly, they expand the normative data bank for health-related fitness, building on existing norms studies such as the recently published international CRF norms⁵¹ and other European health-related fitness norms.^{52 53}

Despite these norms not being linked to a health outcome, they nonetheless have utility for health and fitness screening, profiling, monitoring and surveillance by identifying the centile rank of children and adolescents in comparison with their peers. For instance, several authors^{31 52 54} have suggested using a normative quintile-based framework to classify the fitness levels of children and adolescents, where those below the 20th centile are classified as ‘very low/poor’; 20–40th centiles as ‘low/poor’; 40–60th centiles as ‘moderate’; 60–80th centiles as ‘high/good’ and those above the 80th centile as ‘very high/

Table 8 10×5 m agility shuttle run (s) centiles by age and sex based on 258 618 test performances of children and adolescents aged 9–17 years representing 19 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	15409	29.26	27.58	25.79	24.64	23.73	22.94	22.20	21.46	20.66	19.64	18.87
10	16773	28.00	26.54	24.98	23.96	23.15	22.44	21.78	21.11	20.38	19.44	18.73
11	17925	26.77	25.53	24.16	23.27	22.55	21.92	21.33	20.73	20.07	19.22	18.57
12	16152	25.68	24.59	23.39	22.60	21.96	21.40	20.86	20.32	19.72	18.94	18.35
13	18549	24.77	23.79	22.70	21.98	21.40	20.88	20.39	19.88	19.33	18.61	18.05
14	16914	24.10	23.18	22.15	21.47	20.92	20.43	19.96	19.48	18.95	18.27	17.73
15	12649	23.61	22.72	21.73	21.06	20.53	20.05	19.60	19.13	18.62	17.95	17.43
16	11783	23.22	22.35	21.37	20.72	20.20	19.73	19.28	18.83	18.32	17.67	17.16
17	6423	22.89	22.03	21.07	20.43	19.91	19.45	19.01	18.56	18.06	17.42	16.91
Girls												
9	16273	30.96	28.96	26.93	25.67	24.70	23.88	23.12	22.37	21.57	20.57	19.83
10	15703	28.87	27.35	25.76	24.74	23.95	23.27	22.63	21.99	21.30	20.43	19.78
11	15063	27.11	25.92	24.64	23.81	23.15	22.58	22.04	21.50	20.90	20.14	19.57
12	18344	26.36	25.29	24.13	23.37	22.77	22.24	21.74	21.24	20.68	19.97	19.43
13	16678	26.06	25.03	23.90	23.16	22.58	22.06	21.58	21.08	20.54	19.85	19.32
14	15589	25.98	24.95	23.83	23.09	22.51	22.00	21.51	21.03	20.49	19.79	19.27
15	11479	25.97	24.94	23.82	23.09	22.51	22.00	21.51	21.02	20.48	19.79	19.26
16	11018	25.95	24.92	23.81	23.07	22.49	21.98	21.50	21.01	20.47	19.78	19.25
17	5895	25.93	24.90	23.79	23.06	22.48	21.96	21.48	20.99	20.46	19.77	19.24

Table 9 20 m shuttle run (min/stages) centiles by age and sex based on 445 092 test performances of children and adolescents aged 9–17 years representing 24 countries

Age (years)	n	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅
Boys												
9	36 079	1.27	1.96	2.80	3.41	3.93	4.43	4.92	5.45	6.08	6.95	7.68
10	36 935	1.53	2.25	3.13	3.77	4.31	4.83	5.34	5.90	6.55	7.46	8.22
11	30 786	1.79	2.53	3.45	4.11	4.68	5.22	5.75	6.33	7.01	7.96	8.75
12	26 552	2.04	2.82	3.77	4.46	5.06	5.61	6.18	6.78	7.49	8.47	9.30
13	29 467	2.31	3.12	4.11	4.82	5.44	6.02	6.60	7.23	7.97	8.99	9.85
14	28 262	2.71	3.55	4.57	5.31	5.95	6.55	7.15	7.80	8.56	9.62	10.51
15	23 754	3.08	3.92	4.95	5.70	6.34	6.95	7.56	8.21	8.98	10.05	10.94
16	13 417	3.35	4.19	5.22	5.96	6.61	7.21	7.81	8.47	9.23	10.30	11.19
17	11 326	3.80	4.64	5.67	6.42	7.06	7.66	8.26	8.91	9.67	10.74	11.63
Girls												
9	35 027	0.87	1.41	2.08	2.56	2.98	3.38	3.77	4.20	4.70	5.40	5.98
10	36 270	1.03	1.60	2.29	2.79	3.22	3.63	4.04	4.48	5.00	5.72	6.33
11	30 751	1.31	1.91	2.64	3.18	3.64	4.07	4.51	4.98	5.53	6.30	6.94
12	26 119	1.27	1.89	2.66	3.21	3.69	4.14	4.60	5.08	5.66	6.46	7.13
13	20 066	1.25	1.87	2.64	3.20	3.68	4.13	4.58	5.07	5.65	6.46	7.13
14	19 557	1.24	1.87	2.64	3.20	3.68	4.13	4.58	5.07	5.65	6.46	7.13
15	15 682	1.24	1.87	2.63	3.19	3.67	4.13	4.58	5.07	5.65	6.46	7.13
16	13 317	1.21	1.84	2.61	3.17	3.66	4.11	4.57	5.06	5.64	6.45	7.13
17	11 725	1.20	1.83	2.60	3.17	3.65	4.11	4.56	5.06	5.64	6.45	7.13

Note: 20 m shuttle run centiles are available for other metrics in online supplement 4.

good'. Single test measures can be qualitatively interpreted using these quintile-based thresholds and longitudinal changes tracked against centile bands to identify expected, better than expected or worse than expected developmental changes. In addition, long-term intervention studies are required to determine whether changes in fitness in response to exercise training are over and above expected developmental changes illustrated by our age-related reference values. While individual fitness test scores can be benchmarked and tracked, a composite or overall fitness score could also be generated as an aggregate score summarising centiles across all fitness components or across multiple components or subdomains of interest (eg, a composite score for health-related fitness should aggregate centiles for CRF, MSF and flexibility). This scoring structure, similar to that used in the Canadian Assessment of Physical Literacy,^{55 56} could help identify the fitness components/subdomains in need of attention in order to provide appropriate feedback and advice to children about how to best improve their overall physical fitness. In this context, the lowest quintile has extensively been used as a threshold for defining low fitness or unfit youth.⁵⁷ In prospective cohort studies, this group has been shown to have a disproportionately higher risk for future diseases.⁵⁸ Even more stringent cut-points (eg, 10th centile) have been proposed for individuals who should be checked for the existence of other risk factors or developmental problems. In a cohort study conducted in more than 1 million Swedish male adolescents, it was observed that those in the lowest decile of muscular strength had significantly higher risk of all-cause mortality, cardiovascular disease mortality and suicide mortality, supporting the notion that this should be considered a group at risk.¹²

To date, research examining criterion-referenced standards in children and adolescents has focused on CRF,^{22 23 59} with new international standards recently published for healthy CRF recently published.²³ While not the first study to estimate the percentage of European children and adolescents with apparently healthy CRF,⁵² this study provides the most current and

best available estimate using the new international criterion-referenced standards. This study is consistent with previous studies showing a latitudinal gradient, where children and adolescents from Northern-Central Europe typically have better CRF than their peers from Southern Europe.^{16 41 42} This study also identified considerable variability in healthy CRF levels among different European countries. Variability in CRF was previously identified as a strong unfavourable correlate of country-specific income inequality (operationalised as the Gini index); meaning, countries with a large population spread of income tend to have poor CRF levels.⁴² The observed age gradient in healthy CRF levels may reflect that children are generally healthier than adolescents or it may be an artefact of the new international standards being age-independent. Unfortunately, criterion-referenced standards for fitness components other than CRF do not currently exist. In addition, CRF criterion-referenced standards do not exist for outcomes other than cardiometabolic health (ie, poor bone health, mental health, cognitive health and so on), which is a limitation and represents an area for future research.

This study systematically identified and quantified the sex-specific differences in Eurofit performance, showing that boys outperformed girls on CRF, MSF and speed-agility tests and experienced larger age-specific changes, while girls outperformed boys on the flexibility test. While the underlying causes of the sex-specific differences are clear for some fitness components (eg, differences in MSF are largely explained by physical differences such as differences in body size/composition), they are less clear for others (eg, differences in CRF may be explained by physiological differences such as differences in mechanical efficiency and/or the fractional utilisation of oxygen).^{21 60 61} It is, nonetheless, beyond the scope of this paper to discuss these mechanistic causes. However, there is a need for longitudinal cohort studies to better understand what mechanisms drive sex-specific and age-specific differences in physical fitness throughout childhood and adolescence.

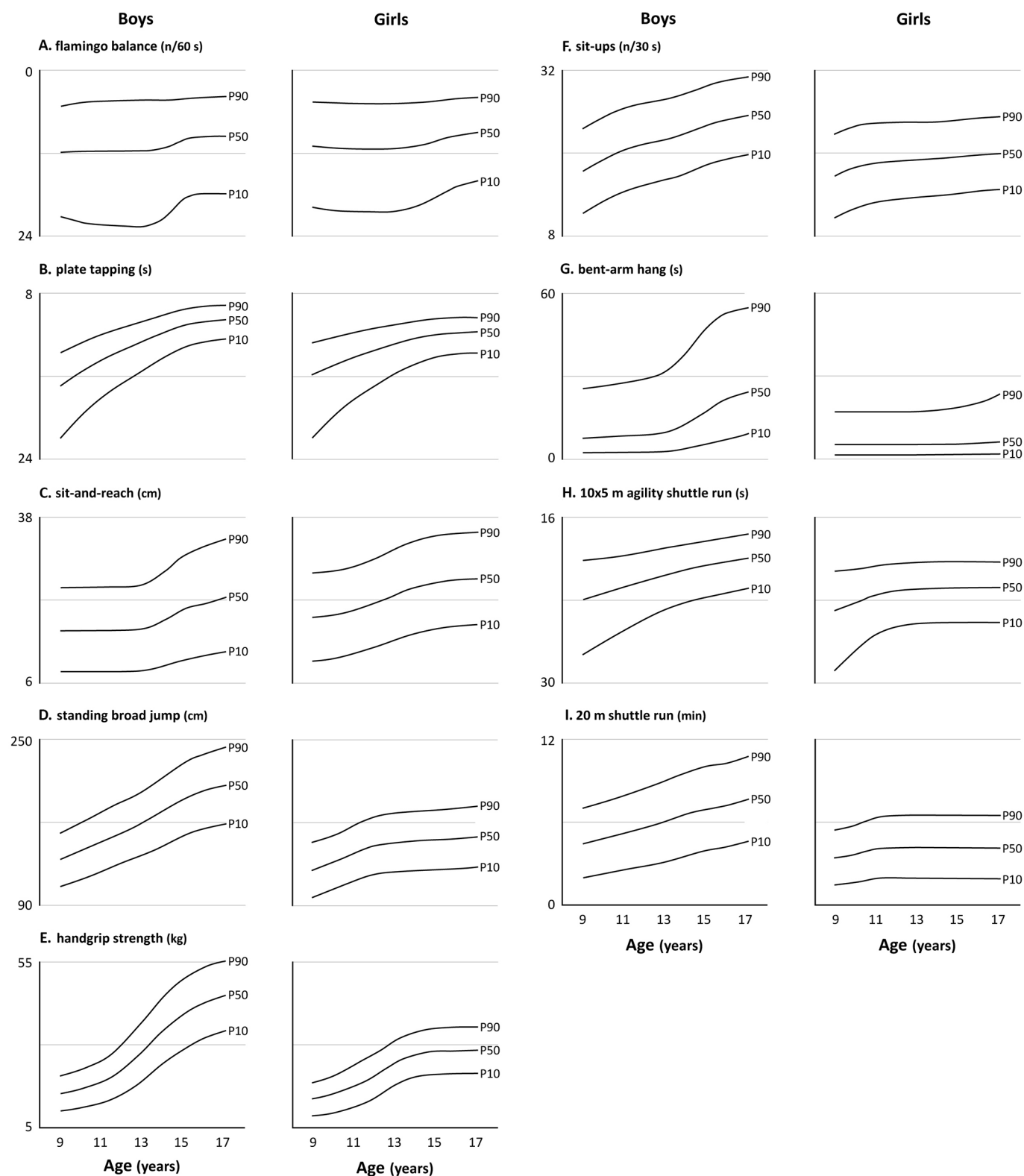


Figure 4 Smoothed centile curves (P_{10} , P_{50} and P_{90}) for (A) flamingo balance (n/60 s), (B) plate tapping (s), (C) sit-and-reach (cm), (D) standing broad jump (cm), (E) handgrip strength (kg), (F) sit-ups (n/30 s), (G) bent-arm hang (s), (H) 10×5 m agility shuttle run (s) and (I) 20 m shuttle run (min).

Strengths and limitations

This study summarised cross-sectional Eurofit data from 98 studies to generate probably Europe's largest physical fitness database for children and adolescents. Although not the first comprehensive review of children's Eurofit performance, it does provide an update to a previous review¹⁶ by: (1) extending the

data coverage from 2001 to 2015 through a rigorous systematic review process, (2) producing sex-specific and age-specific European normative values and (3) estimating the percentage of European children and adolescents with healthy CRF.

Despite the strengths of this study, it is not without limitations. First, we pooled data from studies that used different

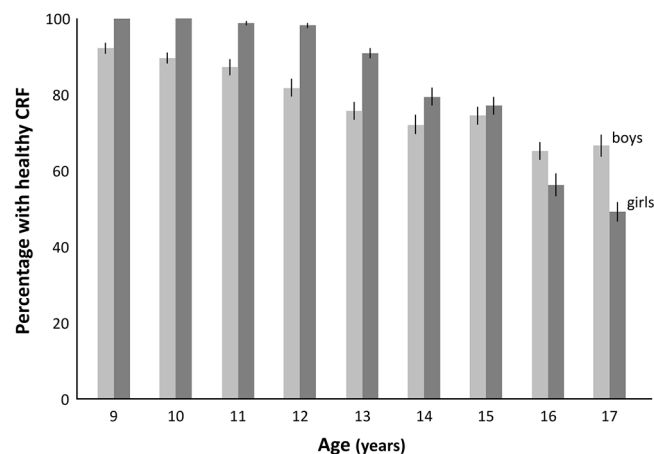


Figure 5 Percentage of European children and adolescents aged 9–17 years meeting the new international criterion-referenced standards of 42 mL/kg/min (boys, light grey bars) and 35 mL/kg/min (girls, dark grey bars) for healthy CRF. The thin black vertical lines show the 95% CIs. CRF, cardiorespiratory fitness.

sampling methods (probability and non-probability sampling) and sampling frames (national-level, state/provincial-level and community-level), which raises the issue of representativeness. However, we used the best available data and a poststratification population weighted approach to control for oversampling and undersampling across studies and countries. Second, differences in testing conditions (eg, climate, altitude, practice and testing surfaces) and measurement errors (eg, methodological drift and diurnal variation) might have occurred, although the large number of included data points should have minimised these issues. Third, the vigorous nature of the Eurofit may have resulted in difficulties in testing, or exclusion of, individuals with a lower level of physical function. The absence of data from these populations may have inflated our norms within the lower centile range. Fourth, our sex-specific and age-specific norms and differences in Eurofit performance are also limited by the potential for unmeasured confounding. For example, biological maturation, which was rarely reported in the included studies and was therefore not included in our analysis, confounds sex-specific and age-specific differences in physical fitness.⁶² Large-scale longitudinal studies focused on the influence of maturation on physical fitness are needed. Finally, Eurofit data were also collected at different times in the period between 1981 and 2015 and given evidence of temporal changes in some (but not all) fitness components in European children,^{21 28 63–69} it is possible that our norms represent a different health-related picture than what would actually be observed today. However, without the availability of temporal trends data for all included countries, temporal corrections of our norms are not possible.

Recommendations

Given the widespread use of the Eurofit and other test batteries such as the ALPHA, there is a need for consistent reporting of results across studies to assist future data pooling and the update of normative values. In addition to recommending that the Eurofit be routinely administered (in part or in whole) in schools to improve national and regional surveillance of health and fitness, we also make the following recommendations:

1. An online multilingual operations and procedures manual, including instructional videos, should be made available (eg, the ALPHA project manual, <http://profith.ugr.es/>

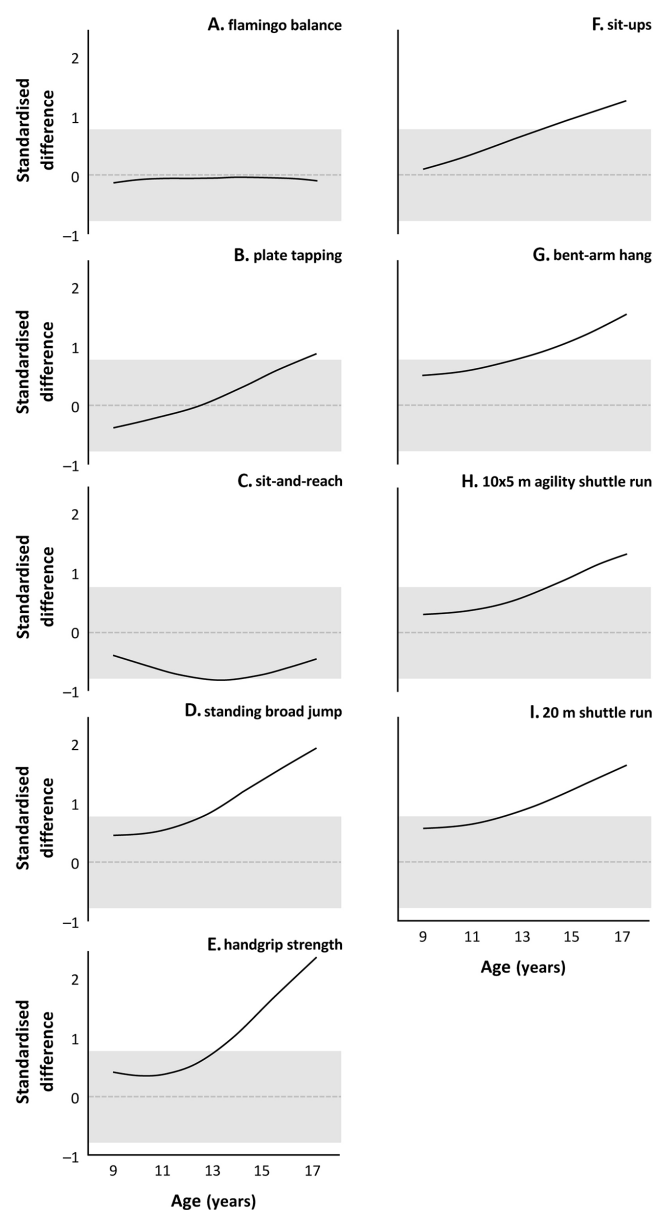


Figure 6 Standardised sex-specific differences in mean Eurofit performance for European children and adolescents aged 9–17 years. The limits of the grey zone represent the threshold for a large standardised difference (ie, 0.8 or –0.8). Positive differences indicated that Eurofit performances for boys were better than those for girls.

alpha-children). Researchers should make de-identified raw data available through an online data repository^{42 70} in order to help improve surveillance efforts across the region. For example, scheduled for official release in 2018 is a free website (<http://www.activehealthykids.org/kids-fit-guide/>) that will compute a report comparing individual 20 m shuttle run performances to national, regional and international normative values and criterion-referenced standards, providing researchers with valuable analytical support.

2. Care should be taken to minimise and report factors that may impact fitness test performance (eg, climate, temperature, humidity, altitude, clothing, ground surfaces/conditions, pre-test instructions and test familiarisation). Studies should be conducted to assess the effect of these factors on fitness test performance.

What are the new findings?

- This study presents the largest and most geographically representative sex-specific and age-specific European normative values for physical fitness in children and adolescents.
- This study estimated that 78% (95% CI 72% to 85%) of boys and 83% (95% CI 71% to 96%) of girls met the new international criterion-referenced standards of 42 and 35 mL/kg/min respectively for healthy cardiorespiratory fitness (CRF), with the percentage meeting the standards decreasing with age.
- This study showed that boys performed better than girls on muscular strength, muscular power, muscular endurance, speed-agility and CRF tests, but worse on the flexibility test. Boys' fitness also generally improved at a faster rate than girls' fitness, especially during the teenage years.

How might it impact on clinical practice in the future?

- Sex-specific and age-specific European normative values for physical fitness in children and adolescents are important for health and fitness screening, profiling, monitoring and surveillance.

- Best practice should include that: (1) test protocols be followed and test results be reported as per the operations and procedures manual; (2) biological age (sexual maturation) be measured (if appropriate) in addition to chronological age; (3) descriptive statistics (sample sizes, means and SDs) be reported in 1 year age and sex groups based on age at last birthday and (4) the year(s) of testing be reported.

CONCLUSION

Physical fitness is an important indicator of good health, and the Eurofit is probably the most popular way to measure physical fitness throughout Europe. This study pooled 2 779 165 Eurofit performances, representing children and adolescents from 30 European countries. This large summary analysed the best available Eurofit data to: (1) provide the largest and most geographically representative sex-specific and age-specific European normative values for physical fitness in children and adolescents and (2) estimate the percentage of children and adolescents with healthy CRF according to the new international criterion-referenced standards. These data have utility for both health and sport promotion given that they help to identify children and adolescents with: (1) very low/poor fitness in order to set appropriate fitness goals, monitor longitudinal changes and promote positive health-related fitness behaviours (eg, physical activity and exercise promotion) and (2) very high/good fitness in the hope of recruiting them into sporting or athletic development programmes.

Acknowledgements We would like to thank the authors of the included studies for generously clarifying details of their studies and/or for providing raw data.

Contributors GRT developed the systematic review research question and objectives. GRT, ND and LL created the search strategy and provided guidance on review methodology. KC, FA, LL and ND screened and extracted the data. GRT and KC led the data analysis, data synthesis and writing of the manuscript. All authors contributed to interpretation of the results, edited, reviewed and approved the final manuscript.

Funding A College Research Council Summer Research Professorship from the College of Education and Human Development at the University of North Dakota supported this project. FBO research activity is by the Spanish Ministry of Economy and Competitiveness—MINECO (RYC-2011-09011, DEP2016-79512-R); from the University of Granada, Plan Propio de Investigación 2016, Excellence actions: Units of Excellence; Unit of Excellence on Exercise and Health (UCEES); from the EXERNET Research Network on Exercise and Health in Special Populations (DEP 2005-00046/ACTI) and from the SAMID III network, RETICS, funded by the PN I+D+I 2017-2021 (Spain), ISCIII-Sub-Directorate General for Research Assessment and Promotion, the European Regional Development Fund (ERDF) (Ref. RD16/002).

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

© Article author(s) or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Ortega FB, Ruiz JR, Castillo MJ, *et al.* Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes* 2008;32:1–11.
- Lee DC, Sui X, Ortega FB, *et al.* Comparisons of leisure-time physical activity and cardiorespiratory fitness as predictors of all-cause mortality in men and women. *Br J Sports Med* 2011;45:504–10.
- Kodama S, Saito K, Tanaka S, *et al.* Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA* 2009;301:2024–35.
- Katzmarzyk PT, Craig CL. Musculoskeletal fitness and risk of mortality. *Med Sci Sports Exerc* 2002;34:740–4.
- Sawada SS, Lee IM, Naito H, *et al.* Cardiorespiratory fitness, body mass index, and cancer mortality: a cohort study of Japanese men. *BMC Public Health* 2014;14:1012.
- Erikssen G, Liestøl K, Bjørnholt J, *et al.* Changes in physical fitness and changes in mortality. *Lancet* 1998;352:759–62.
- Slattery ML, Jacobs DR. Physical fitness and cardiovascular disease mortality. The US railroad study. *Am J Epidemiol* 1988;127:571–80.
- Smith JJ, Eather N, Morgan PJ, *et al.* The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis. *Sports Med* 2014;44:1209–23.
- Ruiz JR, Castro-Piñero J, Artero EG, *et al.* Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med* 2009;43:909–23.
- Ortega FB, Labayen I, Ruiz JR, *et al.* Improvements in fitness reduce the risk of becoming overweight across puberty. *Med Sci Sports Exerc* 2011;43:1–7.
- Höglström G, Nordström AN. Aerobic fitness in late adolescence and the risk of early death: a prospective cohort study of 1.3 million Swedish men. *Int J Epi* 2016;45:1159–68.
- Ortega FB, Silventoinen K, Tynelius P, *et al.* Muscular strength in male adolescents and premature death: cohort study of one million participants. *BMJ* 2012;345:e7279.
- Sato M, Kodama S, Sugawara A, *et al.* Physical fitness during adolescence and adult mortality. *Epidemiology* 2009;20:463–4.
- Armstrong N, Tomkinson G, Ekelund U. Aerobic fitness and its relationship to sport, exercise training and habitual physical activity during youth. *Br J Sports Med* 2011;45:849–58.
- Kemper HCG, Van Mechelen W. Physical fitness testing of children: a European perspective. *Pediatr Exerc Sci* 1996;8:201–14.
- Tomkinson GR, Olds TS, Borms J. Who are the Eurofittest? *Med Sport Sci* 2007;50:104–28.
- Council of Europe. *Eurofit: handbook for the Eurofit tests of physical fitness*. Rome: Council of Europe, 1988.
- Artero EG, España-Romero V, Castro-Piñero J, *et al.* Reliability of field-based fitness tests in youth. *Int J Sports Med* 2011;32:159–69.
- Castro-Piñero J, Artero EG, España-Romero V, *et al.* Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med* 2010;44:934–43.
- Ruiz JR, Castro-Piñero J, España-Romero V, *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;45:518–24.
- Tomkinson G, Olds T. Field tests of fitness. In: Armstrong N, Van Mechelen W, eds. *Paediatric exercise science and medicine*. 2 ed. United Kingdom: Oxford, 2008:109–28.
- Lang JJ, Tremblay MS, Ortega FB, *et al.* Review of criterion-referenced standards for cardiorespiratory fitness: what percentage of 1 142 026 international children and youth are apparently healthy? *Br J Sports Med* 2017 doi: 10.1136/bjsports-2016-096955. [Epub ahead of print 02 Mar 2017].
- Ruiz JR, Caverio-Redondo I, Ortega FB, *et al.* Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents; what level of fitness should raise a red flag? A systematic review and meta-analysis. *Br J Sports Med* 2016;50:1451–8.
- IOM (Institute of Medicine). *Fitness measures and health outcomes in youth*. Washington, DC: The National Academies Press, 2012.

- 25 Cauderay M, Narring F, Michaud P-A. A cross-sectional survey assessing physical fitness of 9- to 19-year-old girls and boys in Switzerland. *Pediatr Exerc Sci* 2000;12:398–412.
- 26 Haugen T, Høigaard R, Seiler S. Normative data of BMI and physical fitness in a Norwegian sample of early adolescents. *Scand J Public Health* 2014;42:67–73.
- 27 Jürimäe T, Volbekiene V. Eurofit test results in Estonian and Lithuanian 11 to 17-year-old children: a comparative study. *European Journal of Physical Education* 1998;3:178–84.
- 28 Lefèvre J, Bouckaert J, Duquet W, et al. De barometer van de fysieke fitheid van de Vlaamse jeugd: de resultaten. *Sport* 1998;4:16–22.
- 29 Tambalis KD, Panagiotakos DB, Psarra G, et al. Physical fitness normative values for 6-18-year-old Greek boys and girls, using the empirical distribution and the lambda, mu, and sigma statistical method. *Eur J Sport Sci* 2016;16:736–46.
- 30 Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- 31 Tomkinson GR, Lang JJ, Tremblay MS, et al. International normative 20 m shuttle run values from 1 142 026 children and youth representing 50 countries. *Br J Sports Med* 2017;51:1545–54.
- 32 Léger LA, Mercier D, Gadoury C, et al. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988;6:93–101.
- 33 Tomkinson GR, Léger LA, Olds TS, et al. Secular trends in the performance of children and adolescents (1980-2000): an analysis of 55 studies of the 20m shuttle run test in 11 countries. *Sports Med* 2003;33:285–300.
- 34 Levy PS, Lemeshow S. Stratification random sampling: further issues. In: Levy PS, Lemeshow S, eds. *Sampling of populations: methods and application*. Hoboken, NJ: John Wiley & Sons, Inc, 2008:143–88.
- 35 United Nations, Department of Economic and Social Affairs, Population Division. *World population prospects: the 2015 revision, key findings and advance tables working paper*, 2015. No. ESA/P/WP.241.
- 36 Tomkinson GR, Hamlin MJ, Olds TS. Secular trends in anaerobic test performance in Australasian children and adolescents. *Pediatr Exerc Sci* 2006;18:314–28.
- 37 d'Agostino RB, Pearson ES. Tests of departure from normality: empirical results for the distribution of b2 and $\sqrt{b1}$. *Biometrika* 1973;60:613–22.
- 38 Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
- 39 Pan H, Cole T. *User's guide to LMSChartmaker*. UK: Medical Research Council, 2010:1–42.
- 40 Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. New Jersey: Lawrence Erlbaum, 1988.
- 41 Ortega FB, Ruiz JR, Labayen I, et al. Health inequalities in urban adolescents: role of physical activity, diet, and genetics. *Pediatrics* 2014;133:e884–e895.
- 42 Lang JJ, Tremblay MS, Léger L, et al. International variability in 20 m shuttle run performance in children and youth: who are the fittest from a 50-country comparison? A systematic literature review with pooling of aggregate results. *Br J Sports Med* 2016 doi: 10.1136/bjsports-2016-096224. [Epub ahead of print 20 Sept 2016].
- 43 Cole TJ, Bellizzi MC, Flegal KM, et al. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;320:1240–3.
- 44 Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. *Arch Dis Child* 1995;73:25–9.
- 45 Eisenmann JC. Waist circumference percentiles for 7- to 15-year-old Australian children. *Acta Paediatr* 2005;94:1182–5.
- 46 Fernández JR, Redden DT, Pietrobello A, et al. Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J Pediatr* 2004;145:439–44.
- 47 McCarthy HD, Jarrett KV, Crawley HF. The development of waist circumference percentiles in British children aged 5.0–16.9 y. *Eur J Clin Nutr* 2001;55:902–7.
- 48 Shields M. Overweight and obesity among children and youth. *Health Rep* 2006;17:27–42.
- 49 Tremblay MS, Shields M, Laviolette M, et al. Fitness of Canadian children and youth: results from the 2007–2009. *Health Rep* 2009;21:1–14.
- 50 Jackson LV, Thalange NK, Cole TJ. Blood pressure centiles for Great Britain. *Arch Dis Child* 2007;92:298–303.
- 51 Jolliffe CJ, Janssen I. Development of age-specific adolescent metabolic syndrome criteria that are linked to the Adult Treatment Panel III and International Diabetes Federation criteria. *J Am Coll Cardiol* 2007;49:891–8.
- 52 Ortega FB, Artero EG, Ruiz JR, et al. Physical fitness levels among European adolescents: the HELENA study. *Br J Sports Med* 2011;45:20–9.
- 53 De Miguel-Etayo P, Gracia-Marco L, Ortega FB, et al. Physical fitness reference standards in European children: the IDEFICS study. *Int J Obes* 2014;38 Suppl 2:S57–S66.
- 54 Catley MJ, Tomkinson GR. Normative health-related fitness values for children: analysis of 85347 test results on 9–17-year-old Australians since 1985. *Br J Sports Med* 2013;47:98–108.
- 55 Longmuir PE, Boyer C, Lloyd M, et al. The Canadian assessment of physical literacy: methods for children in grades 4 to 6 (8 to 12 years). *BMC Public Health* 2015;15:767.
- 56 Francis CE, Longmuir PE, Boyer C, et al. The Canadian assessment of physical literacy: development of a model of children's capacity for a healthy, active lifestyle through a delphi process. *J Phys Act Health* 2016;13:214–22.
- 57 Ortega FB, Ruiz JR, Labayen I, et al. The Fat but Fit paradox: what we know and don't know about it. *Br J Sports Med* 2017.
- 58 Ortega FB, Lavie CJ, Blair SN. Obesity and cardiovascular disease. *Circ Res* 2016;118:1752–70.
- 59 Tomkinson G. Aerobic fitness thresholds for cardio metabolic health in children and adolescents. *Br J Sports Med* 2011;45:686–7.
- 60 Rowland TW. Evolution of maximal oxygen uptake in children. *Med Sport Sci* 2007;50:200–9.
- 61 Armstrong N, Welsman JR. Aerobic fitness: what are we measuring? *Med Sport Sci* 2007;50:5–25.
- 62 Baxter-Jones ADG. Growth and maturation. In: Armstrong N, Van Mechelen W, eds. *Children's sport and exercise medicine*. 3 ed. Oxford, United Kingdom: Oxford University Press, 2017:13–24.
- 63 Costa AM, Costa MJ, Reis AA, et al. Tendências seculares dos níveis antropométricos e de aptidão física em crianças portuguesas. *Acta Med Port* 2017;30:108–14.
- 64 Ekblom O, Oddsson K, Ekblom B. Health-related fitness in Swedish adolescents between 1987 and 2001. *Acta Paediatr* 2004;93:681–6.
- 65 Jürimäe T, Volbekiene V, Jürimäe J, et al. Changes in Eurofit test performance of Estonian and Lithuanian children and adolescents (1992–2002). *Med Sport Sci* 2007;50:129–42.
- 66 Mahmoud O, Mészáros J, Szabo T. Secular trend and motor performance scores in Hungarian schoolboys. *Kinesiology* 2002;2:127–33.
- 67 Matton L, Duvinéaud N, Wijndaele K, et al. Secular trends in anthropometric characteristics, physical fitness, physical activity, and biological maturation in Flemish adolescents between 1969 and 2005. *Am J Hum Biol* 2007;19:345–57.
- 68 Moliner-Urdiales D, Ruiz JR, Ortega FB, et al. Secular trends in health-related physical fitness in Spanish adolescents: the AVENA and HELENA studies. *J Sci Med Sport* 2010;13:584–8.
- 69 Sjølie A, Mønness E. Truncus endurance, hip and ankle mobility and aerobic fitness in 15-year-old Norwegian adolescents in 1968 and 1997. *Scand J Med Sci Sports* 2007;17:488–96.
- 70 Lang JJ, Tomkinson GR, Janssen I, et al. Making a case for cardiorespiratory fitness surveillance among children and youth. *Exerc Sport Sci Rev*. In Press.