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 Tomkinson
*University of North Dakota, grant.tomkinson@und.edu*

Kevin D. Carver

Frazer Atkinson

Nathan D. Daniell

Lucy K. Lewis

*See next page for additional authors*

Follow this and additional works at: [https://commons.und.edu/ehb-fac](https://commons.und.edu/ehb-fac)

Part of the [Health and Physical Education Commons](https://commons.und.edu/ehb-fac)
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, Kevin D Carver, Frazer Atkinson, Nathan D Daniell, Lucy K Lewis, John S Fitzgerald, Justin J Lang, Francisco B Ortega

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. 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. 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), 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 profiles. Several of the Eurofit tests have been supported by European experts from the ALPHA (Assessing Levels of Physical Activity) project and by North American experts from the IOM (Institute of Medicine) report, 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. Previously, Tomkinson et al 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 tested between 1981 and 2005, 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. 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.
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.\textsuperscript{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 in conjunction with a trained academic librarian. The search strategy included the term: Eurofit; with child*, OR adolescent*, 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

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.
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 elsewhere 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. To combine data from different studies, all Eurofit data were standardised to a common metric and protocol. Measurement units reported in the Eurofit handbook 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, 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. The accuracy of the 20 m shuttle run data standardisation procedure is excellent.

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. 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. Monte Carlo simulation was then used to create pseudodata using the detailed methods described elsewhere. 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. 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 K² test 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 (V.2.43, The Institute of Child Health, London, UK), which analyses data using the Lambda Mu Sigma (LMS) method. 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 of smoothing required can be expressed in terms of smoothing parameters or equivalent df.

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. 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.
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 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.
Although boys developed at a faster rate than girls on the upper body strength tests, there were negligible sex 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 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.
This study systematically analysed 2,779,165 Eurofit performances of children and adolescents aged 9–17 years to generate the lastest 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 index43 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
To date, research examining criterion-referenced standards in children and adolescents has focused on CRF, with new international standards recently published for healthy CRF recently published. While not the first study to estimate the percentage of European children and adolescents with apparently healthy CRF, 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. 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. 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). 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.

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 review 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.
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.

Despite the strengths of this study, it is not without limitations. First, we pooled data from studies that used different 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 (e.g., climate, altitude, practice and testing surfaces) and measurement errors (e.g., 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. 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, 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 (e.g., the ALPHA project manual, http://profith.ugr.es/alpha-children). Researchers should make de-identified raw data available through an online data repository 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, pretest instructions and test familiarisation). Studies should be conducted to assess the effect of these factors on fitness test performance.

3. 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.

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.

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.
References


