Health-Related Physical Fitness Measures: Reference Values and Reference Equations for Use in Clinical Practice

Anne Therese Tveter, PT, MSc,a Hanne Dagfinrud, PT, PhD,a,b Tuva Moseng, PT, MSc,a Inger Holm, PT, PhDa,c

From the aInstitute of Health and Society, Medical Faculty, University of Oslo, Oslo;bNational Resource Center for Rehabilitation in Rheumatology, Diakonhjemmet Hospital, Oslo; and cDivision of Surgery and Clinical Neuroscience, Orthopaedic Department, Section of Research, Oslo University Hospital, Oslo, Norway.

Abstract

Objective: To provide reference values and reference equations for frequently used clinical field tests of health-related physical fitness for use in clinical practice.

Design: Cross-sectional design.

Setting: General community.

Participants: Convenience sample of volunteers (N ≥ 370) between 18 and 90 years of age were recruited from a wide range of settings (ie, work sites, schools, community centers for older adults) and different geographic locations (ie, urban, suburban, rural) in southeastern Norway.

Interventions: Not applicable.

Main Outcome Measures: The participants conducted 5 clinical field tests (6-minute walk test, stair test, 30-second sit-to-stand test, handgrip test, fingertip-to-floor test).

Results: The results of the field tests showed that performance remained unchanged until approximately 50 years of age; after that, performance deteriorated with increasing age. Grip strength (79%), meters walked in 6 minutes (60%), and seconds used on the stair test (59%) could be well predicted by age, sex, height, and weight in participants ≥50 years of age, whereas the performance on all tests was less well predicted in participants <50 years of age.

Conclusions: The reference values and reference equations provided in this study may increase the applicability and interpretability of the 6-minute walk test, stair test, 30-second sit-to-stand test, handgrip test, and fingertip-to-floor test in clinical practice.
specific than the more sophisticated laboratory-based tests, they are shown to be applicable for measuring cardiorespiratory endurance, muscle strength, and flexibility.

To improve the interpretability and clinical usefulness of clinical field tests, relevant reference values and reference equations are needed. Reference values and reference equations are previously established for tests (eg, 6-minute walk test [6MWT], handgrip test, 30-second sit-to-stand test [30sSTS]), but these values were derived from studies based on multiple trials and presented only for some specific age groups. For use in clinical practice, reference values and reference equations for men and women in all age groups are needed. The aim of this study was, therefore, to provide age- and sex-specific reference values for health-related physical fitness measures in the general population.

Methods

This study is part of a large-scale research program (FYSIOPRIM, a research program on physiotherapy in primary care) focusing on several methodologic and clinical aspects of physical therapy in primary health care. The research program is developed and led by a group of experienced researchers and clinicians. One of the aims of FYSIOPRIM is to establish a core set of physical fitness measures for use in clinical practice. The predefined criteria for inclusion of fitness measures in the core set were that they had to be applicable in a busy clinical practice (ie, easy to perform, time efficient and require a minimum amount of equipment), making field tests the most clinical feasible tools to use. The selection of relevant field tests was based on a thorough literature search followed by discussions and an informal consensus process in the research group.

A convenience sample of volunteers was included in sex and age groups with 10 year increments (a total of 14 groups). Power calculations were based on the mean of the 6MWT of the first 80 participants in the ongoing data collection (range, 437–714m) with a group SD of approximately 83m. The significance level was set to .05, and the required power was set to at least .80. The sample size was estimated to 20 to 25 participants per group. The recruitment period lasted from June 2011 to August 2012.

To ensure a representative sample, participants were included from a wide range of settings (ie, work sites, schools, community centers for older adults) and different geographic locations (ie, urban, suburban, rural), mainly in the southeast part of Norway. Participants from different work sites were recruited to cover different professions. When approval was given from the general manager or a superior at the site, the employees, users of community centers, students, and so forth were asked to volunteer. In addition, people were also recruited from other settings (network connections) to capture other workplaces and people who were retired but not visiting community centers for older adults (fig 1).

Two physical therapists (A.T.T. and T.M.) tested all the participants. Pilot testing was conducted before the study, and all participants were tested according to a standardized test protocol. To be included, the participants had to be ≥18 years old, understand written and spoken Norwegian language, and live at home. Participants with self-reported serious heart disease or other diseases that restricted participation in moderate physical activity were not included. Participants who were unable to climb stairs were excluded. We considered this the most demanding test and expected the participants to be able to complete the other tests if they could complete the stair test.

All participants answered a set of sociodemographic questions, including age, sex, employment status, occupation, smoking habits, and comorbidities. Body composition was measured by body weight and height and was presented as body mass index (BMI) (kg/m²). To assess physical activity level, the participants answered the International Physical Activity Questionnaire Short Form, consisting of 7 questions on the time spent in vigorous-intensity activities, moderate-intensity activities, walking and sedentary activities. The results were transformed into metabolic equivalent task minutes per week scores and categorized into low, moderate, and high level of participation in physical activity according to the guidelines for the International Physical Activity Questionnaire Short Form, where a moderate to high level of participation is regarded as health-enhancing (more information on the questionnaire is available at www.ipaq.ki.se).

To assess cardiorespiratory endurance, the 6MWT and a stair test were used. The 6MWT is described as a simple and inexpensive walk test and can be used as a predictor of aerobic capacity. Participants were instructed to walk as fast as possible (without running) back and forth between 2 cones on a flat, hard surface for 6 minutes. With no significant difference between walking courses of 15 to 50m, a distance of 15m between the 2 cones was used to be applicable in a clinical outpatient setting. The walking distance was measured in meters. The stair test is described as a submaximal cardiopulmonary exercise test. We used a revised version of the stair test; for practical reasons we used the staircases available at each test location. All participants were instructed to ascend and descend 18 average-sized steps (17±1cm) 3 consecutive times. All the stairs comprised a platform (or repos) in-between the steps, implying that the participants had to take an additional step on level ground before continuing the steps. Participants were instructed to use all steps, they were allowed to run, and for safety reasons they could use the bannister if needed. The results were measured in seconds. Heart rate was recorded after both the 6MWT and stair test using a heart rate monitor. Perceived exertion was measured after the 6MWT with Borg’s rating of perceived exertion, which is a 15-point scale ranging from 6 (very, very light) to 20 (very, very hard).

Muscle strength was assessed with a handgrip test and the 30sSTS. The handgrip test is a simple method of assessing muscle strength in the upper extremities. The grip strength was measured using a hydraulic hand dynamometer with 5 handle positions; the second position was used for all participants. The testing was conducted with the participant seated with the upper arm alongside the trunk and the elbow at 90° of flexion. The dominant hand was tested first, and the mean of 2 trials was used in the analysis of the right and left hand. The 30sSTS is a measure of lower extremity strength. Starting from a seated position with arms folded across the chest, the participants were instructed to complete as many full stands as possible in 30 seconds. For practical reasons, chairs available at the different test locations were used, but all chairs were of standard height (44–45cm). The number of full stands was recorded.

List of abbreviations:
- BMI: body mass index
- FTF: fingertip-to-floor test
- MSC: musculoskeletal condition
- 6MWT: 6-minute walk test
- 30sSTS: 30-second sit-to-stand test
To evaluate flexibility, we used the fingertip-to-floor test (FTF), which has been described as a measure of mobility of the spine, pelvic girdle, and hamstrings. With knees fully extended, the participants were asked to reach as far down toward the floor as possible while standing on a stool. The results were measured in centimeters (negative values refer to an inability to reach the stool, whereas positive values reflect the ability to reach beyond the level of the stool).

Analysis

The data were analyzed using IBM SPSS version 20, and the variables are presented as mean ± SD if normally distributed or median and interquartile range (25th and 75th quartiles) if skewed. The results are presented for the total group and for separate sex and age groups (10y groups from 18 to 29y old, 30 to 39y old, 40 to 49y old, and up to 80 to 90y old). Sex differences were analyzed with independent t test or Mann-Whitney U test. Changes in heart rate and perceived exertion with increasing age groups were analyzed with a 1-way analysis of variance. Kernel plots were applied to show the associations between test scores of the clinical field tests and age and sex groups. Equations for calculating individually adopted reference values based on age (y), height (cm), weight (kg), and sex (woman =0, man =1) were derived from multiple linear regression analyses (backward deletion method). Only statistically significant variables were kept in the final model ($P<.05$). The explained variance of the equation estimates was given as $R^2$ values, whereas the 95% prediction interval for the estimates was derived from the SD of the prediction error.

Ethical considerations

Each participant gave their written consent before participation. Ethical approval was granted by the Regional Ethical Committee in Norway. Because all clinical field tests resemble daily activities, we considered the tests to be of minimal danger to the participants.

Results

In the study, 370 controls between 18 and 90 years of age were included in the study (table 1). Men and women were equal with regard to age, but statistically significant differences were observed for height, weight, and BMI ($P<.001$) (see table 1). Of the participants, 58% reported themselves as healthy with no known diseases. Heart disease, osteoarthritis, and other MSCs...
were most frequently reported, with heart disease and osteoarthritis mainly being reported in older participants. Of the participants, 77% reported to be within the recommended level of health-enhancing physical activity participation (see www.ipaq.ki.se for further information).

Distribution of scores of the clinical field tests is presented for different age and sex groups in table 2. Significant differences were found between the total group of men and women for all field tests (P < 0.001). Heart rate reported after the 6MWT and the stair test decreased with increasing age (P < 0.001) (see table 2). Heart rate was significantly higher for the stair test compared with the 6MWT for all sex and age groups (P < 0.001). Perceived exertion after the 6MWT showed a mean ± standard deviation of 12 ± 2, with no significant age and sex differences.

The distribution of scores of the clinical field tests was visualized in kernel plots (figs 2A–F), showing that performance on all tests, except the FTF, was similar until about 50 years of age, whereas performance deteriorated with increasing age in older age groups. The 30sSTS (see fig 2C) and FTF (see fig 2F) showed the largest variability in the distribution of scores, whereas the handgrip test (see figs 2D and 2E) showed the largest sex differences.

Reference equations for the different field tests are presented in table 3. Based on the distribution of scores, the equations were presented separately for participants <50 and ≥50 years of age. Explained variance of the equations ranged from 79% to 0%, with handgrip being highly explained by age, height, and sex, whereas the FTF remained unexplained by the same variables in participants aged <50 years. A reference equation for this test could, therefore, not be calculated, and only the constant is given. The explained variance was higher in participants aged ≥50 years for all the clinical field tests (see table 3).

Discussion

The reference values for health-related physical fitness in the general population showed that performance on the clinical field tests (except the FTF) were similar in the age groups <50 years, after which the performance deteriorated with increasing age. For this reason, more precise estimates of performance based on easily obtained characteristics (age, sex, height, weight) could be provided for participants ≥50 years of age.

Except for a few studies,24,28,46 the reference values for performance tests have mostly been presented as the mean value of multiple trials. However, because patients in busy outpatient physical therapy clinics most often are tested only once, and most field tests are shown to have a learning effect,47 the values calculated in our study were based on unrepeated trials, making them comparable with the results from clinical practice. Further, the reference equations provided for the different clinical field tests may facilitate the prediction of the individual patient’s fitness based on their age, sex, and height. The use of the equations can be exemplified by the prediction of grip strength (right hand) for a 62-year-old, 167-cm-tall woman: the predicted grip strength is 29.6kg. The 95% prediction interval for this estimate is ±11.4kg, indicating that grip strength between 18 and 41kg should be considered normative for this person. These equations will form the basis for an easily accessible web-based application providing age- and sex-specific reference values for use in clinical practice.

To improve the clinical applicability, the reference values were calculated for sex-specific, 10-year age spans. Although comparison with values presented in previous studies is limited because of different methods and population groups, similar values are found for the 6MWT5,4 and the handgrip test,3,36,49 supporting the validity of the age- and sex-specific values provided in this study.

To account for the curvilinear relation between performance and age,50 reflecting the natural aging of biologic function and physical performance,59 we chose to present reference equations derived separately for participants ≥50 and <50 years of age based on the kernel plots (see figs 2A–E). The reference equations presented for the 6MWT, stair test, and handgrip test showed that performance could be well predicted by easily obtained participant characteristics (age, height, weight, sex), especially in participants ≥50 years of age. In contrast, only some or no variance of the 30sSTS and FTF could be explained by age, sex, weight, and height. The large variability, low explained variance, and wide prediction intervals indicate that the predictions of lower extremity strength and flexibility by these methods are uncertain in participants <50 years of age.

The large number of participants is a strength of this study. In large, normally distributed samples, some standardized residuals in a multiple regression will be outside ±3, and actions toward these are not necessary.50 Based on this, 10 participants with residual values between 3.1 and 3.9 were kept in the analyses because they did not influence the results of the regression equations. On the other hand, 2 participants were excluded from all analyses of the stair test because they constituted standardized residual values of 5.8 (showing stair test performance of 141.2s) and 8.4 (showing stair test performance of 198.2s), thereby making the distribution of the residuals skewed.

Study limitations

A limitation of this study was the use of a convenience sample. To account for this, much effort was put into recruiting a representative sample. The participants were recruited from different settings and geographic locations and covered several economic activities (agriculture; manufacturing; construction; retail trade; transport and storage; accommodation and food service activities;
<table>
<thead>
<tr>
<th>Age Group</th>
<th>Heart Rate 6MWT (bpm)</th>
<th>ST (s)</th>
<th>Heart Rate ST (bpm)</th>
<th>30sSTS (n)</th>
<th>Handgrip Right Hand (kg)</th>
<th>Handgrip Left Hand (kg)</th>
<th>FTF (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–29y</td>
<td>649 (611–687)</td>
<td>152 (142–162)</td>
<td>33.0 (31.3–34.8)</td>
<td>172 (167–177)</td>
<td>26 (23–29)</td>
<td>32.4 (30.7–34.2)</td>
<td>32.1 (30.2–34.0)</td>
</tr>
<tr>
<td>30–39y</td>
<td>650 (617–683)</td>
<td>154 (145–163)</td>
<td>34.0 (31.9–36.2)</td>
<td>170 (164–176)</td>
<td>26 (24–27)</td>
<td>31.2 (28.8–33.5)</td>
<td>31.7 (29.6–33.9)</td>
</tr>
<tr>
<td>40–49y</td>
<td>664 (639–689)</td>
<td>143 (138–149)</td>
<td>35.0 (33.2–36.7)</td>
<td>164 (160–168)</td>
<td>25 (23–27)</td>
<td>32.9 (31.2–34.7)</td>
<td>33.8 (32.5–35.4)</td>
</tr>
<tr>
<td>50–59y</td>
<td>638 (614–662)</td>
<td>146 (139–154)</td>
<td>38.8 (36.6–41.1)</td>
<td>160 (154–166)</td>
<td>24 (22–26)</td>
<td>30.3 (29.2–31.5)</td>
<td>30.4 (28.9–31.9)</td>
</tr>
<tr>
<td>60–69y</td>
<td>573 (545–600)</td>
<td>137 (131–143)</td>
<td>46.2 (42.1–50.2)</td>
<td>149 (145–154)</td>
<td>21 (18–23)</td>
<td>26.9 (25.2–28.6)</td>
<td>27.7 (25.9–29.5)</td>
</tr>
<tr>
<td>70–79y</td>
<td>510 (488–531)</td>
<td>128 (123–134)</td>
<td>57.0 (53.5–60.4)</td>
<td>145 (139–152)</td>
<td>17 (16–19)</td>
<td>24.3 (23.0–25.5)</td>
<td>24.8 (23.5–26.2)</td>
</tr>
<tr>
<td>80–90y</td>
<td>438 (399–476)</td>
<td>120 (112–128)</td>
<td>74.6 (64.8–84.5)</td>
<td>133 (125–142)</td>
<td>14 (13–16)</td>
<td>21.2 (19.0–23.4)</td>
<td>20.3 (18.3–22.2)</td>
</tr>
<tr>
<td>Total</td>
<td>648 (633–663)</td>
<td>136 (133–140)</td>
<td>33.9 (32.3–35.2)</td>
<td>151 (148–154)</td>
<td>24 (23–25)</td>
<td>46.8 (45.4–48.2)</td>
<td>47.5 (46.2–48.8)</td>
</tr>
</tbody>
</table>

NOTE. Values are presented as mean (95% CI) or as otherwise indicated.
Abbreviations: bpm, beats per minute; CI, confidence interval; ST, stair test.
* One participant excluded from the sample.
† Presented as median (95% CI).
Fig 2  Distribution of scores is shown for the different age groups for the (A) 6MWT, (B) stair test, (C) 30sSTS, handgrip test for the (D) right and (E) left hands, and (F) FTF.
information and communication activities; professional, scientific, and technical activities; administrative and support service activities; education, human health, and social work activities; art, entertainment, and recreation activities) (for more information see http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-015/EN/KS-RA-07-015-EN.PDF) and, therefore, covered professions with different degrees of education and physical demands. The sample was comparable with regard to height, weight, activity level, and smoking habits with a large population-based study in 1 of the 19 counties in Norway and with the prevalence of MSCs in another Norwegian population study. The sample was also comparable with a Danish population-based study with regard to demographic variables (eg, height, BMI). The proportion of participants with BMI >30 was lower in the present study compared with a Norwegian population study, indicating that the reference values derived from the present study are mainly representative for people with BMI <30, which comprises 75% to 80% of the general population in Norway.

The use of different equipment (staircases and chairs) at the different locations might have increased the variability of the results. However, generally, no differences were found between the different settings in the age and sex groups; therefore, the results may be generalized to settings where a standard height of stairs and chairs are used.

Conclusions
The age- and sex-specific reference values and reference equations for the 6MWT, stair test, 30sSTS, handgrip test, and FTF provided in this study may improve the interpretability of patients’ health-related physical fitness and the applicability of fitness measures in clinical practice.

Suppliers
a. Polar FT4 heart rate monitor; Polar Electro Oy, Professorintie 5, 90440 Kempele, Finland.
b. Baseline hydraulic hand dynamometer; Fabrication Enterprises Inc, PO Box 1500, White Plains, NY 10602.
c. IBM Corporation, New Orchard Rd, Armonk, NY 10504-1722.

Keywords
Methods; Physical therapists; Reference values; Rehabilitation

Corresponding Author
Anne Therese Tvester, PT, MSc, Institute of Health and Society, Medical Faculty, University of Oslo, PO Box 1089 Blindern, NO-0318 Oslo, Norway. E-mail address: a.t.tveter@medisin.uio.no.

References
1. Ryan CG, Grant PM, Dall PM, Gray H, Newton M, Granat MH. Individuals with chronic low back pain have a lower level, and an altered pattern, of physical activity compared with matched controls: an observational study. Aust J Physiother 2009;55:53-8.
2. Hodselmans AP, Dijkstra PU, Geertzen JH, van der Schans CP. Nonspecific chronic low back pain patients are deconditioned and have an increased body fat percentage. Int J Rehabil Res 2010;33:268-70.

11. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken) 2011;63(Suppl 1):S350-70.


