Rese

3FUFTU 3FQFBUBCJMJUZ PG .PUPS BOE .VTDVM
)FBMUI .POJUPSJOH PG "EVMU 1PQVMBUJPOT

Jaana Helena Suni[1]*, Marjo Birgitta Rinne[1] and Jonatan R Ruiz[2]

[1]Urho Kaleva Kekkonen Institute for Health Promotion Research (UKK Institute), Finland
[2]Department of Physical Education and Sport, School of Sport Sciences, University of Granada, Spain

Abstract

Background and purpose: [redacted]

Methods: [redacted]

Keywords: Physical tness; Reliability; Surveillance; Adult populations; Musculoskeletal functioning; Health

Abbreviation: CV: Coe cient of Variation

Introduction

Adequate physical activity (PA) and tness are considered as one of the key factors in current public health promotion. Physical tness also re ects the e ects of regular physical activity. It has been recommended that assessment and monitoring of PA and tness could be part of a public health strategy [1,2] to deliver interventions to communities likely to increase population PA and tness levels. Furthermore, data collected from populations should provide evidence based information for health policy planning [1,2].

ere is strong evidence that low performance levels in several factors of physical tness are risk factors for various health problems including the major non-communicable diseases [3-5] musculoskeletal disability related to mobility limitations, [6-8] and increasing evidence for low back pain (LBP) [9-11]. Field-based methods of tness that show meaningful relationship with health and physical functional ability are needed for promotion of PA and tness for health. At best these assessments of health-related tness [12] can be used to monitor the level of tness in di erent populations, and to identify those with increased health risks due to inadequate levels of tness [8,13].

In order to apply tness tests to large populations the test methods need to be safe, economic and easy to administer under conditions available in ordinary communities [12]. Furthermore, their measurement error (reliability) needs to be established in relation to the measurement purposes [14]. Regarding population based tness measurements, the testers categorize subjects into di erent levels of performance (i.e. into tness classes), make comparisons between individuals and groups, and monitor tness changes over time [14].

Retest repeatability concerns the consistency of the observed values when the measurement is repeated in same environment, tester and participants. Within-subject variation is the most important type of repeatability measures: the smaller the within-subject variation (i.e. typical measurement error) the better precision of single measurements and better observation of changes [15,16]. In order to correctly categorize individuals into tness classes, the typical (standard) measurement error of the test needs to be smaller than the average range of applied tness classes [17]. Correct categorization is a critical issue in targeting interventions to low- t population groups and individuals with increased risk of diseases or disability, and in epidemiological follow-up studies estimating the predictive e ect of tness level on future morbidity and mortality.

Systematic change in the mean, a non-random change in the measurement value between two test sessions, is an important issue when volunteers perform a series of test trials as part of a monitoring program [15,16]. A typical example of a systematic change in physical tness testing results is a learning e ect (bias) i.e. the participants perform better in the second test session than the rst, because they bene t from the experience of the rst test session. To indicate true improvement or deterioration, the change in person's tness level over time needs to be bigger than the systematic change in the mean. e true change is a critical issue also in epidemiological studies that estimate the predictive e ects of tness changes on future morbidity and mortality [18,19].

e purpose of this study was to evaluate the retest repeatability of a set of health-related tness tests which have potential to be used in public health monitoring to predict changes in future health and musculoskeletal functioning.

Materials and Methods

Subjects

Volunteers from two departments of the University of Tampere, Finnish Railway Company Ltd, and a private gear manufacturing enterprise were recruited by sending an invitation letter with detailed information on the study by email to persons in supervising positions, who then further delivered the message. Altogether 51 subjects participated in the study, 25 women and 26 men between ages 22 to 64. e mean age of the men was 42.5 (SD 13.4) and of the women 39.4 (SD 9.8). Background characteristics of the participants are given in Table 1.

	MEN n=26	WOMEN n=25	ALL n=51
Civil status			
[redacted]			
[redacted]			
Education			
[redacted]			
[redacted]			
[redacted]			
Smoking			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
Physical strain at work*			
[redacted]			
[redacted]			
[redacted]			
Perceived health†			
[redacted]			
[redacted]			
[redacted]			
Musculoskeletal problems			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
Overall level of physical activity‡			
[redacted]			
[redacted]			
[redacted]			
6HOI UHSRUWHG ¿WQHVV OHYHO FRPSDUHG WR RZQ DJH JURXS			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			

[redacted]

Table 1: [redacted]

Procedures

e subjects were asked to contact the health promotion research institute by phone or email to make appointments for two measurement sessions approximately a week apart. e mean number of days between the two measurement sessions was 7.0 (SD 2.1, range from 2 to 15), 82% (n=40) managed to have a retest within 5-9 days.

A pre-testing health screening was conducted according to the safety model of the Health-Related Fitness Test Battery for Middle-aged Adults [20]. It included a modi ed version of the Physical Activity Readiness Questionnaire (MPAR-Q) [21], and questions on perceived health status, overall level of physical activity [22] (including weekly frequency and intensity), and type of exercise. One experienced tness tester was responsible for conducting all the tness measurements.

Assessment of health-related tness

Nine tness tests were individually performed by each subject in a pre-set sequence with standard methods [17,23,24]: three motor tests measuring balance and agility, and six musculoskeletal tests measuring shoulder-neck exibility, muscular strength of upper and lower extremities, and endurance of the trunk muscles. e tness items with description of their purpose, exclusion criteria, test performance and instructions, and scoring are presented in Appendix.

Statistical analyses

e mean with standard deviation (SD) or the median, and the minimum and maximum of test-retest results in women and men are presented as descriptive statistics. e estimates of repeatability for interval scale measurements were calculated as suggested by Hopkins [15]: (1) typical (standard) error of measurement (s), indicating within-subject variation, was calculated as the standard deviation of test-retests di erence (SD_{di}) divided by the square root of two ($s=SD_{di}/\sqrt{2}$). To compare the repeatability of di erent tests the typical error was presented also as the relative measurement error i.e. (2) coe cient of variation (CV): typical error divided by the mean of two tests ($CV=s/mean(test_i+retest_i)/2$). e (3) systematic chance in the mean with 95% con dence interval was calculated using paired samples t-test in SPSS 17.0 for Windows so ware (SPSS Inc., Chicago, IL). e mean change between the two measurement sessions was considered statistically signi cant if the 95% con dence interval (CI) did not include value zero. e (4) percent change in the mean performance between the rst and second test session was also calculated. In addition, (5) the Bland and Altman plots (data not shown) were screened for heteroscedasticity [14]. For ordinal scale measurements (6) weighted Kappa coe cient was used to estimate the reproducibility between two test sessions.

e rationale for the criteria on adequacy of repeatability in the present study

Currently there are no consensus statements or standards for acceptable measurement precision for monitoring physical tness in population level. Other risk factors of non-communicable diseases such as blood pressure, cholesterol and length of waist circumference have been systematically monitored among large populations in several

Citation: [illegible]

Citation: [illegible]

controlled in the present study (see Appendix). Low grip strength has been associated with a greater likelihood of premature mortality and the development of disability in middle-aged and elderly populations [36,37], and it is measured to detect sarcopenia [34].

e intra-individual variation in test-retest sessions of jump-and-reach test (i.e. vertical jump) was adequate (CV 6%) with no systematic bias and very low change in the mean (1%). e nding is line with a former study, which reported standard error of 3.0 cm [17]. According to the review by Hopkins et al. [16] the CV of tape measures of vertical jump height among athletes varied between 3.8-4.8%. Athletes are likely to be measured regularly however our results with more "novice" jumpers are well in line with these former results. In the present study protocol, a practice trial was performed before the two test trials in both test sessions. is is important, while the CV between subsequent trials during same measurement session [16] is much smaller between second and third trial than (0.2%) than the rst and second trial (1.2%). Vertical jump requires ability to activate fast type of motor units in a

rA((p)12(r)13(o)11(t)6(o)(es)19 41 s)5(u)4(bse in)]TJalperfootnd a1(in)]Tlder1b(l5(d a)9(d.(an)19(t s)5717(m)4h]TJ7(p)-5(lm)4(emidd)(t)pTbse f))5(6

d arlp [16]ch eld(s)]TJ0 Tw T*([3va)19(ts)ermeyTld6lder(o)16u13aw6srjl1(in)]Ts, bsme9()9(o)e ot 26 suregsdm p [16r [16r p

Citation: