Title: Feasibility and Reliability of the Alpha Health-Related Fitness Test Battery in Adults with Intellectual Disabilities.

Running title: Feasibility and reliability of fitness tests

Authors: Cabeza-Ruiz, Ruth[1], Sánchez-López, Antonio Manuel[1], Trigo, María Eva [2]; Gómez-Píriz, Pedro Tomás[1]

1. *Department of Human Movement and Sport Performance, Faculty of Education, University of Seville. Spain*
2. *Faculty of Psychology, University of Seville. Spain*

Correspondence: R. Cabeza-Ruiz, Department of Human Movement and Sport Performance, Faculty of Education, University of Seville. C/ Pirotecnia s/n, 41013, Office 4.110. Seville, Spain (e-mail: ruthcr@us.es) Tlf: +34 639409868

Funding

The present study did not receive any funding. The tests were carried out with the support of the department of Human Movement and Sport Performance at the University of Seville and the Paz y Bien Association (Seville, Spain), which volunteered its facilities and materials for the research.

Acknowledgments

The research team would like to thank all the personnel at the centres of the Paz y Bien Association for their collaboration to ensure the correct running of the study. Likewise, they must thank the participants´ families for their permission to have their relatives take part in the study and their cooperation at all times.

Conflict of interest

No conflicts of interest have been declared

Abstract

Background

People with intellectual disabilities (ID) seem to have a lower physical fitness (PF) than their peers without disabilities which coincides with reduced autonomy, life expectancy and quality of life. To assess PF in these individuals, it is necessary to use appropriate tools that permit the assessment of their physical capacities taking into account their specific characteristics. The aim of this work is to study the feasibility and reliability of the Alpha-Fit test battery for adults in a group of men and women with mild to moderate ID.

Method

Forty-one adults with ID of both sexes, ranging in age from 20 to 60 years old, participated in the study. To identify the feasibility and reliability of the Alpha-Fit test battery for adults, two complete assessments were done for each one of the tests included in the battery. The assessments were performed for a period of no more than two weeks (test-retest). An intraclass correlation coefficient (ICC) was used to determinate test-retest reliability, and a mixed ANOVA factorial was used for each of the dependent variables. Bland-Altman plots were also used to assess consistency between the two measurements. Feasibility was calculated as the percentage of people who were able to perform the tests correctly (not feasible <50%, fairly feasible 50%-75%, and feasible >75%).In order to determine other psychometric properties, minimal detectable change (MDC) and standard error of measurement (SEM) were also calculated.

Results

Of the 10 tests in the Alpha-Fit test battery, 8 were shown to be feasible. High reliability was obtained (> .90) for the variables related to body composition. In the hand-grip test, reliability was high in the men's group, but low in the women's group. Good reliability results were also found (.80-.89) in the dynamic sit-up test for women, but not for men. Fair reliability (.70-.79) was found in jump-and-reach and neck-shoulder mobility tests. The variables in the 2km walk and agility tests indicated poor reliability. Al tests showed SEM values related to high variability. However, Bland-Altman plots showed results related to lack of consistency.

Conclusions

The feasibility and reliability calculations, as well as the SEM values, confirm that not all the tests of the Alpha-Fit Test Battery for Adults are suitable for the assessment of PF in adults with ID, probably because of the complexity of the motor fitness tests. The authors emphasise the necessity of making adaptations to the protocols used, or of using other tests more appropriate to the characteristics of people with mild to moderate ID.

Keywords: adults, disabilities, physical-fitness, feasibility, reliability.

Introduction

The American Association on Intellectual and Developmental Disabilities (AAIDD) defines intellectual disability (ID) as significant limitations in development and cognitive capacities, such as a reduced capacity to manage adaptive behaviour. Cognitive capacity is related to logical thinking, memory and solving problems. Limitations in adaptive behaviour affect basic survival skills, such as those necessary for the development and maintenance of social relationships, language and communication, time and money management or the use of social resources (Schalock et al., 2010). However, these characteristics, though broadly representative of individuals with ID, are not found in every one of them. This means that not everyone with ID presents all the characteristics associated with their population group, so population heterogeneity is presented as a common group characteristic (Walsh et al., 2017).

People with ID seem to have a lower physical fitness (PF) than their typically developing peers (Blomqvist et al., 2013; Carmeli et al., 2002; Hsieh et al., 2014; Lahtinen et al., 2007). One of the main causes of this reduced PF may be sedentariness, the prevalence of which has been studied in younger (Sundahl *et al.* 2016) and in elderly adults with ID (Hilgenkamp, et al., 2012a). The correlates of sedentary behaviour in adults with ID remains unclear (Oppewal et al., 2018), but what is known is that people with ID maintain more sedentary lifestyles throughout their whole lifespans (Chow *et al.* 2018) and this seems to be related to multimorbidity (Tyrer et al., 2018). The importance of evaluating PF in individuals with ID comes from the necessity of knowing how their state of health-related fitness in order to give them the support necessary for the recovery, improvement or maintenance of their health situation. In the SOPHIE study (Walsh et al., 2017), researchers found that those individuals with ID who took part in Special Olympics competitions had higher fitness levels and more positive health profiles than those who did not.

Appropriate tools are necessary to assess PF in people with ID in order to find reliable and feasible results. Walsh et al. (2017) evaluated the PF of 400 people with ID using the 6-minute walk test (6mwt), body mass index (BMI), waist circumference and accelerometry, but any psychometric properties of the tests were not included. Hilgenkamp et al. (2012b) evaluated the PF of 36 adults with ID, all older than 50, by means of a 7-test battery (walking speed, grip strength, box and block test, 10-m incremental shuttle walking test, the modified sit-and-reach test, response time test, and the 30-s chair stand) for which the subsequently calculated reliability was ICC > .60 and all of the tests had moderate to excellent feasibility (> 50%).

Other studies have evaluated the reliability of batteries designed for people without additional disabilities in the population with Down syndrome, and have yielded good results. Tejero-González et al. (2013) assessed 17 adolescents with Down syndrome to determine the reliability of the Alpha-Fit health-related fitness measures for children and adolescents, and found values of ICC >.64 in conducted tests. Despite the previous studies and the high reliability of the Alpha-Fit test battery for adults in population without disabilities (Suni et al., 1996; Suni et al., 1998; Suni et al., 2009), it is interesting to know the suitability of using the aforementioned tool for the assessment of adults with ID for the purpose of evaluating whether the diversity of this collective influences the consistency of the results. The aim of the present study is to determine the feasibility and reliability of the Alpha-Fit Test Battery for Adults (onwards Alpha-Fit) in women and men with ID.

Methods

Study design

An observational study was undertaken as part of a project between the University of Seville (Spain) and the association for assistance for people with intellectual disabilities Paz y Bien (Spain). To determine the reliability of the Alpha-Fit, two complete assessments were done with each of the tests included in the battery for a maximum period of no more than 2 weeks (test-retest).

Participants

Forty-one individuals with ID of both sexes, and ranging in age from 20 to 60 years old, took part in the study. The selection criteria were: (i) a diagnosis of mild (IQ=79-55) to moderate (IQ=40-54) ID by the relevant government authority; (ii) institutionalisation in a daytime care centre; and (iii) a lack of regular participation in sports activities. Participants who were excluded from the data analysis were those whose ID was related to chromosomic syndromes (such as Down syndrome) and those who were not able to move around independently. Information from the psychological departments of the participating centres was used to identify the ID level of participants.

Procedures and instruments

Before the study began, the families of the participants were informed of its objectives and procedures. For the volunteers to be able to take part, their legal guardians had to sign a participation consent form. In the same way, the participants were required to have medical authorization reflecting their ability to perform physical activities without any risk to their health. The study was approved by the Biomedical Ethics Research Committee of Andalusia (Spain) and followed the Helsinki guidelines for ethical behaviour set out by the World Medical Association (2013).

For the best assessment of the participants, they were divided into groups of 10. Only one group was evaluated per day. All tests were carried out by the same group of evaluators in the same time zone (9.30-11.30 a.m.) at the centres where the participants resided. This was done to make sure that the volunteers did not find themselves in unfamiliar surroundings. The assessment team was composed of researchers with extensive experience in working with people with ID. In addition, the evaluators were trained by the principal investigator on how to run the tests (how to do pre-demonstrations and how to give verbal information). In order to avoid distractions and increase the motivation of the participants, standardised protocols and instructions were eschewed (Wouters et al., 2017).

The Alpha-Fit included the following physical fitness evaluation tests: (i) Body composition: BMI and waist circumference; (ii) Motor fitness: balance (one-leg stand) and agility (figure-of-eight run); (iii) Musculoskeletal fitness: hand grip (upper-body strength), jump-and-reach (lower-body strength), modified push-up and dynamic sit-up (trunk strength), and shoulder-neck mobility (functional mobility); and (iv) cardiorespiratory fitness (2-km walk test). The test battery should be conducted in standardised order: body composition, motor tests, musculoskeletal tests and cardiorespiratory test. The procedures for each of the tests are specified on the ALPHA project webpage, as well as in manuals and videos that can be also found at www.ukkinstituutti.fi/filebank/500-ALPHA_FIT_Testers_Manual.pdf (Suni et al., 2009).

In order to respect the procedures of the Alpha-Fit, no significant adaptations were made to the protocols of each test. However, the following adjustments were introduced: (i) Basic language was used when transmitting the instructions to the participants, so that they understood well what they had to do in each test; (ii) the evaluator gave a demonstration of the test before the participants began the assessment; (iii) participants were verbally encouraged to increase/maintain their motivation for each of the tests; (iv) no warm-up or stretching exercises were allowed before testing; and (v) no trials were conducted prior to the execution of the tests, as is indicated in the Alpha-Fit tester manual.

In addition to these indications, in order to facilitate the execution of the tests and data collection, the one-leg stand test was filmed to accurately determine the length of time that participants were able to balance on one leg. To make sure that participants did the figure-of-eight correctly in the figure-of-eight run test, it was necessary to place visual signals on the floor indicating the direction of the route. Finally, in the 2-km walk test, the assessors designed routes in places close to the occupational centres. The researchers were stationed every 50-100 m along these routes, which were previously measured using a GPS, in order to encourage the participants and avoid them becoming disoriented, stopping or starting to run.

Statistical analysis

Feasibility completion rates were calculated in percentages, i.e.: not feasible < 50%, fairly feasible 50 - 75% and feasible > 75% (Wouters et al., 2017).

To analyse the possible differences between test and retest measurements, a mixed ANOVA was developed for each independent variable, with the test-retest measurements as within-subject factors, sex as a between-groups factor, and weight and age as covariates (see Table 3). Thus, it was possible to test for the absence of interaction effects with the test-retest factor, to ensure the validity of conclusions about test-retest over different levels of sex, age, and weight. The generalised eta-squared was used as effect-size index (Bakeman, 2005).

A post-hoc power analysis was performed using G*Power (Faul et al., 2007) for default values of medium effect size, $r = .50$ between both measurements, and $\alpha = .05$. For our sample of 41 participants in a sex * measurement factorial ANOVA, the power ranged between $1 – \beta = .44$ for the between-groups variable and $1 - \beta = .88$ for the within-subjects variable. Power would have been lower considering the covariables age and weight, but the software did not allow for this calculation.

The test-retest reliability of the Alpha-Fit was determined by the intraclass correlation coefficient (ICC). A two-way mixed model with absolute agreement and with 95% confidence intervals was applied. The ICC values were interpreted as follows: values of .90-.99 reflected high reliability, .80-.89 good reliability, .70-.79 fair reliability and scores equal or under .69 poor reliability (Shrout & Fleiss, 1979). A weighted Kappa coefficient (K_w) was calculated for ordinal scale measures (Haley & Osberg, 1989). Kappa results < .40 showed poor inter-rater values, results between .40 - .70 showed fair to good values, and results > .70 reflected excellent reliability (Landis & Koch, 1977).

Consistency between the measurements (test-retest) was verified using Bland-Altman plots. This graphical method allows the observation of the level of agreement between two measures and the presence or absence of heteroscedasticity and outliers.

Standard error of measurement (SEM) is one indicator of absolute reliability and is calculated to determine the degree to which repeated measurements vary in the subjects. In this study, the SEM was presented following the instructions of Atkinson and Nevill (1998). In addition, the minimal detectable change (MDC) was calculated. This is the minimum change in a subject´s score that guarantees the change is not a consequence of measurement error (Wouters et al., 2017).

No calculations could be conducted with the modified push-up, since only two people could perform the test correctly.

Results

Participants

Results related to the sample can be observed on the flowchart (Figure 1). Seventy-seven adults with ID were contacted. Of these, 11 did not meet the inclusion criteria, 10 did not want to take part in the study and 10 were excluded for other reasons related to their reception/understanding of the information. Finally, 46 subjects were included in the study group, five of whom did not do the retest and were considered disoriented. The abandonment rate of this study is 10.9%. Final participation data for each one of the tests for the group of women and men are shown in Tables 1 and 2. The final sample consisted of 41 subjects; of whom 15 were women (36.9 ± 11.7 y.o.) and 26 were men (38.9 ± 12.5 y.o.).

Feasibility

Out of the 10 tests that make up the Alpha-Fit, eight were shown to be feasible (>93.3% Table 1 and >92.3% Table 2). However, the 2-km walk test had a feasibility value of 53.33% in the group of women (fair). Only 4 participants (2 women and 2 men) performed the correct technique for the modified push-up test, even though the evaluators made several demonstrations to each of the volunteers and allowed the participants to practise more than once (the *Alpha-Fit tester manual* indicates just one practise). Test feasibility scores, as well as the rest of the variables, can be seen in Table 1 (women) and Table 2 (men).

Reliability

Table 3 shows the results for each dependant variable. The results referring to the interactions between test-retest have been excluded, as have those relating to sex, age and weight variables. In no case were the latter statistically significant; they always obtained effect size rates of < .01.

Tables 1 and 2 show the relative results of the coefficients ICC, SEM and MDC in the group of women and the group of men, respectively.

High reliability results (> .90) were achieved in the variables referred to as body composition in the women's and men's groups, which also had a high reliability in the hand-grip test. Good reliability results (.80 - .89) were also achieved in shoulder-neck and dynamic sit-up in the women's group. Fair reliability (.70 - .79) was found for jump and reach in the women's and men's groups, as well as for one-leg stand in the women's group. Hand-grip test, 2-km walk test, and agility showed poor reliability in women. The same could be seen in the remaining tests for the men's group.

Bland-Altman plots revealed some outliers for hand grip and figure of eight run. However, systematic biases were observed for one leg stand, dynamic sit up, jump and reach and 2-km walk test (men and women –see appendix section to see Bland Altman plots-).

The analysis of the SEM and MDC results obtained for body composition show small values with respect to the measurements of the rest of the variables. However, it must be taken into account that since they are not measurements of motor function, these variables are more stable in an interval of two weeks. The rest of the variables have high SEM and MDC values.

Discussion

The main finding of the present study is that the tests included in the Alpha-Fit Test Battery for Adults generally present a high level of feasibility, fairly reliability, and high SEM values when applied to adults with ID.

Body composition

The groups of both men and women showed high feasibility and reliability in body composition variables. In both groups, the confidence intervals showed little data variability and limited variation in the SEM and MDC (see Tables 1 and 2). However, it is necessary to take into account that these are passive tests. In other words, the subjects did not have to do any tasks in which their mobility was compromised. The results obtained in this study related to body composition are similar to those obtained by Waninge et al. (2009), whose study also showed that for the assessment of body composition in individuals with severe ID, the calculation of the BMI and waist circumference were more reliable than skin fold measurements.

Motor fitness

Although the motor fitness tests had high feasibility, people with ID showed performance difficulties during the execution of the tests. These difficulties have been reflected in a previous study (Cabeza-Ruiz & Castro-Lemus, 2016).

This is the case in the monopodal balance test (one-leg stand), which is used in the evaluation of different collectives, such as postmenopausal women (Rikkonen et al., 2018), stroke patients (Kim et al., 2015), healthy subjects (Schlee et al., 2012) and people with ID (Cuesta-Vargas et al., 2011). In relation to the assessment of body balance in adults with ID, Hilgenkamp et al. (2010) suggested more complex tasks that include movements or turns instead of monopodal balance tests, in which people with ID seem to have serious difficulties. The average time the participants in this study could stay on one foot was less than 20 s. Bland-Altman plot showed that the higher the time spent in the test, the poorer reliability (see Appendix).

In the case of the figure-of-eight run, during the development of the study the evaluators observed the fact that having to carry out the figure-of-eight during the route was too big a challenge for some of the participants. Changes in direction caused uncertainty in the subjects, and execution was sometimes affected. However, Bland-Altman plot showed no bias in the figure of eight run. Tejero-González et al. (2013) found the agility test proposed in the Alpha Health Related Fitness Test Battery for Children and Adolescents (4x10 shuttle run test) reliable (ICC = .92) in 17 subjects with Down syndrome, even though it was more complex than the proposed version of the test for adults. Following a revision led by Hilgenkamp et al. (2010), the authors selected the walking-speed test as the ideal means of evaluating agility in older adults with ID, since it is considered a simple, cheap, objective and reliable instrument.

Musculoskeletal

Feasibility was high for both men's and women's groups in the flexibility/shoulder-neck mobility test (> 90%). However, the tests results obtained and the retest indicate moderate reliability in both groups (Tables 2 and 3). These reliability results are similar to those found by Suni et al. (1996), who claim that a test with a scale of only 3 points presents reproducibility problems. The results presented in this study, in accordance with Rivilis et al. (2011), are not very conclusive, so the data could interfere with other variables such as muscular strength. In the meta-analysis conducted by Shin and Park (2012), the results showed a low size-effect. From our point of view, the measurement protocols should be reconceptualised and adequate instruments should be developed to allow researchers to interpret the results with care.

The values obtained for women in the hand-grip strength are consistent with those obtained for our group in a previous study (Cabeza-Ruiz & Castro-Lemus 2017) and other studies (Cuesta-Vargas et al., 2011; Cuesta-Vargas & Hilgenkamp, 2015; Hilgenkamp et al., 2012; Tejero-González et al., 2013). The authors suggest that the differences in the reliability of the test between men and women (.90 Vs .67, Tables 1 and 2) could be due to the fact that women in this study do not do tasks that involve hand strength in their daily activities. This could have limited their ability to do the test, as they may not have known how to apply force to the dynamometer. Although Hilgenkamp et al. (2013) achieved good feasibility results in the test, after our experience in this study, we suggest that individuals with ID have difficulty applying sustained maximum force. A period of learning is necessary for this test.

In relation to lower limb strength, the results for the jump-and-reach test seem to be reliable in both groups. While a jump is a movement that requires coordination components and adjustments, the results seem to have been consistent during the test-retest period established in this study (2 weeks). However, Bland-Altman plots showed that higher and lower scores presented poor reliability. In other studies, in which the standing broad jump has been used, the results have shown good feasibility in the short term but not in the long term (Wouters et al., 2017).

The dynamic sit-up test did not seem to be reliable in the men's group, which was not the case in the group of women. However, plots showed that medium scores presented poor reliability in both groups, which indicates that only the highest and lowest scores are reliable. Caution should be taken with these results due to the sample size.

In the modified push-up test, the feasibility in both the group of men (7.69 %) and the group of women (13.33 %) was very low. Only two people from each group could complete the test in a satisfactory manner, so the results were not included in the statistical analysis. These limitations could be associated with a difficulty people with ID have memorising the test protocol and performing it correctly without a period of previous practice. In addition, this test requires notable strength in the muscles implicated in the movement (Lehman et al., 2006), which the subjects in this study probably do not possess.

At this point, it is important to highlight that SEM needs to be considered, because a high measurement error leads to a bias in the results, which means that some tests, although highly reliable and/or feasible, are not recommended for use on people with ID. Only 2 of the 10 tests met the admission criterion SEM< SD1/2. (Atkinson & Nevill, 1998). In the women's group, neither the lower-limb strength test (jump-and-reach) nor the manual grip strength meet this criterion. In the case of men, the results showed that 4 of the 10 Alpha-Fit tests have SEM>SD1/2 (one-leg stand, figure-of-eight run, jump-and-reach and dynamic sit-up). This high measurement error makes it advisable to be cautious when assessing these physical fitness indicators in people with ID.

Cardiorespiratory fitness

The cardiorespiratory fitness test is fairly feasible, but not reliable, in the population with ID (ICC < .69). This test could estimate the capacity of the subjects to make a moderate effort for a prolonged period (approximately 20-25 min in our study). However, there is sufficient evidence to indicate lack of effectiveness, excessive demands and a high risk of measurement errors. Because of this, other tests whose feasibility and reliability have been demonstrated in the population with ID (with or without Down syndrome), such as the 6-min walk test (Boer & Moss 2016; Guerra-Balic et al., 2015; Nasuti et al., 2013) and that have been recommended by the American College of Sports Medicine (2014) have been suggested. In the case of the 2-km walk test, demotivation could have been a determining factor, despite the encouragement provided. The duration of the test could also have affected participants and resulted in their having difficulty paying attention for a prolonged period of time. This can be observed in Bland-Altman plots where the higher the time spent in the test, the poorer reliability. The participants got distracted and started showing signs of tiredness (e.g. they stopped many times and said they were tired), although they did not look fatigued. Limitations due to reduced physical fitness may also have been a reason for the cases of abandonment.

Study limitations

Two factors could have interfered in the results of this study. One of them is the consumption of antipsychotics, which were not controlled. An elevated percentage of people with ID normally have these prescribed, and the medications affect their attention, sleep, tiredness and ability to make coordinated movements (O'Dwyer et al., 2018; Ramerman et al., 2018). Obtaining the specific medication data of the participants was not possible due to data protection policies in Spain. However, the caregivers informed that many participants had mental illness and were medicated. The second is the diversity of people with ID in relation to their cognitive abilities. There are different ways of receiving, interpreting and processing information, and that produces great variability in the responses of the participants, which may influence the results of the tests.

Conclusion

Feasibility, reliability, SEM and Bland-Altman plots confirm that not all the tests of the Alpha-Fit Test Battery for Adults must be used for the evaluation of the physical fitness of an adult population with ID. The results obtained show that 8 tests were feasible, the 2-km walk test was fairly feasible (50 - 75%) and the modified push-up test was not feasible. Only body composition tests (in both groups) and hand-grip strength test (in the men's group) had high reliability (.90-.99). Shoulder-neck mobility and dynamic sit-up tests had good reliability (.80-.89) in women. The tests that obtained fair reliability were the jump-and-reach and the one-leg stand in women. The rest of the tests had poor reliability. Blan-Altman plots show that the Alpha-Fit Test Battery for Adults seems to be more reliable in those individuals with ID with low or good fitness levels but not in those with intermediate levels.

There were many difficulties in the execution of the motor fitness test (mainly in the one-leg stand and the figure-of-eight run). Consequently, the following recommendations should be established for future physical fitness evaluation of individuals with ID: (i) turns for the evaluation of agility should be avoided; (ii) different measurement scales for flexibility and mobility should be proposed; (iii) a longer learning period for the correct technique should be allowed (for example, hand-grip strength or dynamic sit-up; and (iv) the use of shorter tests to assess cardiorespiratory fitness, as the 6-minute walk test, could be considered.

The results of the study highlight the necessity of reconsidering assessments or modifying existing protocols for tests for people with ID. In addition, it is advisable not use extensive batteries, considering that the total evaluation time must not be excessive, since it makes the evaluation process very difficult and forces the participants to excessive physical and intellectual demands.

In this respect, two batteries have recently shown good values of feasibility and reliability on adults with ID: the SAMU DIS-FIT battery (Alcántara-Cordero, et al. 2020), to be used with adults with ID and the ID-Fitscan for older adults with ID (Oppewal & Hilgenkamp, 2018).

References

Alcántara-Cordero, F., Gómez-Píriz, P., Sánchez-López, A.M. & Cabeza-Ruiz, R. (2020). Feasibility and reliability of a physical fitness tests battery for adults with intellectual disabilities: The SAMU DIS-FIT battery. *Disability and Health Journal*. doi: 10.1016/j.dhjo.2020.100886 (In press)

American College of Sports Medicine. (2014) *ACSM's guidelines for exercise testing and prescription* (9th ed.). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Atkinson G. & Nevill A. M. (1998) Statistical methods for assessing meassurement error (reliability) in variables relevant to Sports Medicine. *Sport Medicine* **26**, 217-238. doi: 10.2165/00007256-199826040-00002

Bakeman, R. (2005) Recommended effect size statistics for repeated measures designs. *Behavior Research Methods* **37**, 379–384. doi: 10.3758/bf03192707

Blomqvist S., Olsson J., Wallin L., Wester A. & Rehn B. (2013) Adolescents with intellectual disability have reduced postural balance and muscle performance in trunk and lower limbs compared to peers without intellectual disability. *Research in Developmental Disabilities* **34**, 198–206. doi: 10.1016/j.ridd.2012.07.008

Boer P. H. & Moss S. J. (2016) Validity of the 16-metre PACER and six-minute walk test in adults with Down syndrome. *Disability and Rehabilitation* **38**, 2575-83. doi:10.3109/09638288.2015.1137982

Cabeza-Ruiz R. & Castro-Lemus N. (2016) Body composition and body balance in people with intellectual disabilities: A descriptive study. RICYDE: *Revista Internacional de Ciencias del Deporte* **12**, 399-410. doi:10.5232/ricyde

Cabeza-Ruiz R. & Castro-Lemus N. (2017) Hand Grip Strength in Adults with Intellectual Disabilities. *Apunts. Educación Física y Deportes* **129**, 44-50. doi:10.5672/apunts.2014-0983.es.(2017/3).129.03

Carmeli E., Ayalon M., Barchad S., Sheklow S. L. & Reznick A. Z. (2002) Isokinetic leg strength of institutionalized older adults with mental retardation with and without Down's syndrome. *Journal of Strength and Conditioning Research* **16**, 316-320. doi: 10.1519/1533-4287(2002)016<0316:ilsoio>2.0.co;2

Chow B.C., Choi P. H. N. & Huang W. Y. J. (2018). Physical Activity and Physical Fitness of Adults with Intellectual Disabilities in Group Homes in Hong Kong. International *Journal of Environmental Research and Public Health* **15**, 1370. doi:10.3390/ijerph15071370

Cuesta-Vargas A. I., Paz-Lourido B. & Rodríguez A. (2011) Physical fitness profile in adults with intellectual disabilities: Differences between levels of sport practice. *Research in Developmental Disabilities* **32**, 788-794. doi: 10.1016/j.ridd.2010.10.023.

Cuesta-Vargas A. I & Hilgenkamp T. (2015) Reference values of grip strength measured with a Jamar Dynamometer in 1526 adults with intellectual disabilities and compared to adults without intellectual disability. *PLoS ONE* **10**, e0129585. doi: 10.1371/journal.pone.0129585.

Faul F., Erdfelder E., Lang A. & Buchner A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* **39**, 175–191. doi: 10.3758/bf03193146

Guerra-Balic M., Oviedo G. R., Javierre C., Fortuño J., Barnet-López S., Niño O. et al. (2015). Reliability and validity of the 6-min walk test in adults and seniors with intellectual disabilities. *Research in Developmental Disabilities* **47**, 144-153. doi: 10.3390/ijerph15071370

Haley S. M. & Osberg J. S. (1989) Kappa Coefficient Calculation Using Multiple Ratings Per Subject: A Special Communication. *Physical Therapy* **69**, 970-974. doi: 10.1093/ptj/69.11.970

Hilgenkamp T. I. M., van Wijck, R. & Evenhuis H. M. (2010) Physical fitness in older people with ID-Concept and measuring instruments: A review. *Research in Developmental Disabilities* **31**, 1027-1038. doi:10.1016/j.ridd.2010.04.012

Hilgenkamp T. I. M., Reis D., van Wijck R. & Evenhuis H. M. (2012a) Physical activity levels in older adults with intellectual disabilities are extremely low. *Research in Developmental Disabilities* **33**, 477-483. doi: 10.1016/j.ridd.2011.10.011.

Hilgenkamp T. I. M., van Wijck R. & Evenhuis H. M. (2012b) Feasibility and reliability of physical fitness tests in older adults with intellectual disability: A pilot study. *Journal of Intellectual & Developmental Disability* **37**, 158-162. doi: 10.3109/13668250.2012.681773.

Hilgenkamp T. I. M., van Wijck R. & Evenhuis H. M. (2013) Feasibility of eight physical fitness tests in 1,050 older adults with intellectual disability: results of the healthy ageing with intellectual disabilities study*. Intellectual and Developmental Disabilities* **51**, 33-47. doi: 10.1352/1934-9556-51.01.033.

Hsieh K., Rimmer J.H. & Heller T. (2014) Obesity and associated factors in adults with intellectual disability. *Journal of Intellectual Disability Research* **58**, 851-863. doi: 10.1111/jir.12100

Kim E. K., Lee D. K. & Kim Y. M. (2015) Effects of aquatic PNF lower extremity patterns on balance and ADL of stroke patients. *Journal of Physical Therapy Science* **27**, 213-215. doi: 10.1589/jpts.27.213.

Lahtinen U., Rintala P. & Malin A. (2007) Physical Performance of Individuals with Intellectual Disability: A 30-Year Follow-Up. *Adapted Physical Activity Quarterly* **24**, 125-143. DOI: 10.1123/apaq.24.2.125

Landis J.R., & Koch G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174. doi: 10.2307/2529310

Lehman G.J., MacMillan B., MacIntyre I., Chivers M. & Fluter M. (2006) Shoulder muscle EMG activity during push up variations on and off a Swiss ball. *Dynamic Medicine* **5**, 7. doi: 10.1186/1476-5918-5-7

Nasuti G., Stuart-Hill L. & Temple V. A. (2013) The six-minute walk test for adults with intellectual disability: A study of validity and reliability. *Journal of Intellectual and Developmental Disability* **38**, 31–38. doi: 10.3109/13668250.2012.748885

O'Dwyer M., Peklar J., Mulryan N., McCallion P., McCarron M. & Henman M.C. (2018). Prevalence, patterns and factors associated with psychotropic use in older adults with intellectual disabilities in Ireland. *Journal of Intellectual Disability Research* **61**, 969-983. doi: 10.1111/jir.12391.

Oppewal A., Schoufour J. D., Van Der Maarl H. J. K., Evenhuis H. M., Hilgenkamp T. I. M. & Festen D. A. (2018) Causes of mortality in older people with intellectual disability: Results from the HA-ID study. *American Journal on Intellectual and Developmental Disabilities* **123**, 61-71. doi: 10.1352/1944-7558-123.1.61.

Oppewal, A. & Hilgenkamp, T.I.M. (2018) Adding meaning to physical fitness test results in individuals with intellectual disabilities. Disability and Rehabilitation, 1-8. doi: 10.1080/09638288.2018.1527399.

Ramerman L., Hoekstrab P. J. & de Kuijpera G. (2018) Health-related quality of life in people with intellectual disability who use long-term antipsychotic drugs for challenging behaviour. *Research in Developmental Disabilities* **75**, 49-58. doi:10.1016/j.ridd.2018.02.011.

Rikkonen T., Poole K., Sirola J., Sund R., Honkanen R. & Kröger H. (2018) Long-term effects of functional impairment on fracture risk and mortality in postmenopausal women. *Osteoporosis International* **29**, 2111-2120. doi: 10.1007/s00198-018-4588-4.

Rivilis I., Hay J., Cairney J., Panagiota K., Liu J. & Faught B. E. (2011) Physical activity and fitness in children with developmental coordination disorder: A systematic review. *Research in Developmental Disabilities* **32**, 894-910. doi: 10.1016/j.ridd.2011.01.017

Schalock R. L., Borthwick-Duffy S. A., Bradley V., Buntix W. H. E., Coulter M. D., Craig E. M. et al. (2010) *Intellectual disability: Definition, classification, and systems of supports* (11th Edition). Washington, D.C: American Association on Intellectual and Developmental Disabilities.

Schlee G., Reckmann D. & Milani T. L. (2012) Whole body vibration training reduces plantar foot sensitivity but improves balance control of healthy subjects. *Neuroscience Letters* **506**, 70-73. doi:10.1016/j.neulet.2011.10.051

Shin I. S. & Park E. Y. (2012) Meta-analysis of the effect of exercise programs for individuals with intellectual disabilities. *Research in Developmental Disabilities* **33**, 1937–1947. doi: 10.1016/j.ridd.2012.05.019

Shrout P. E. & Fleiss J. L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* **86**, 420-428. DOI: 10.1037//0033-2909.86.2.420

Sundahl L., Zetterberg M., Wester A., Rehn B. & Blomqvist S. (2016) Physical activity levels among adolescent and young adult women and men with and without intellectual disability. *Journal of Applied Research in Intellectual Disabilities* **29**, 93-98. doi: 10.1111/jar.12170.

Suni J. H., Husu P. & Rinne M. (2009) Fitness for health: The ALPHA-FIT test battery for adults aged 18-69. Tester's Manual. Tampare, Finland: European Union DS and the UKK Institute for Health Promotion Research. Retrieved from: http://www.ukkinstituutti.fi/en/alpha

Suni J. H., Mülunpalo S. I., Asikainen T. M., Laukkanen R. T., Oja P., Pasanen M. E. et al. (1998) Safety and Feasibility of a Health-Related Fitness Test Battery for Adults. *Physical Therapy* **78**, 134-148. DOI: 10.1093/ptj/78.2.134

Suni J. H., Oja P., Laukkanen R. T., Mülunpalo S. I., Pasanen M. E., Vuori I. M. et al. (1996) Health-related fitness test battery for adults: aspects of reliability. Archives of Physical Medicine and Rehabilitation **77**, 399-405. doi: 10.1016/s0003-9993(96)90092

Tyrer F., Dunkley A. J., Singh J., Kristunas C., Khunti K., Bhaumik S., et al. (2018). Multimorbidity and lifestyle factors among adults with intellectual disabilities: a cross-sectional analysis of a UK cohort. *Journal of Intellectual Disability Research* **63**, 255-265. https://doi.org/10.1111/jir.12571.

Tejero-González C. M., Martínez-Gómez D., Bayón-Serna J., Izquierdo-Gómez R., Castro-Piñero J. & Veiga O. L. (2013) Reliability of the Alpha health-related fitness test battery in adolescents with Down syndrome. *Journal of Strength and Conditioning Research* **27**, 3221–3224. doi: 10.1519/JSC.0b013e31828bed4e.

Walsh D., Belton S., Meegan S., Bowers K., Corby D. & Staines A. (2017) A comparison of physical activity, physical fitness levels, BMI and blood pressure of adults with intellectual disability, who do and do not take part in Special Olympics Ireland programmes: Results from the SOPHIE study. *Journal of Intellectual Disabilities* ***22***, 154-170. doi: 10.1177/1744629516688773.

Waninge A., van der Weide W., Evenhuis I. J., van Wijck R. & van der Schans C. P. (2009) Feasibility and reliability of body composition measurements in adults with severe intellectual and sensory disabilities. *Journal of Intellectual Disability Research* **53**, 377-388. doi: 10.1111/j.1365-2788.2009.01153.x.

World Medical Association (2013). World Medical Association declaration of Helsinki. Ethical principles for medical research involving human subjects. *Journal of the American Medical Association* **310**, 2191-2194.

Wouters M, Van Der Zanden A. M, Evenhuis H. M., Hilgenkamp T. I. M. (2017) Feasibility and reliability of tests measuring health-related physical fitness in children with moderate to severe levels of intellectual disability. *American Journal of Intellectual and Developmental Disabilities* **122**, 422e438. doi: 10.1352/1944-7558-122.5.422.

Tables

Table 1 Test-retest reliability of the extended ALPHA health-related fitness battery in women with intellectual disabilities

	n	TEST Mean (*SD*)	RE-TEST Mean (*SD*)	ICC (95% CI)	SEM	MDC	F %
			Body composition				
Waist circumference (cm)	15	93.80 (12.47)	93.67 (12.74)	.98 (.94 - .99)	1.78	4.94	100
Body Mass Index (kg/m^2)	15	27.28 (5.95)	27.11 (5.98)	.99 (.98 - .99)	0.60	1.65	100
			Motor fitness				
One leg stand (s)	15	18.67 (20.42)	19.53 (20.44)	.76 (.43 - .91)	10.01	27,74	100
Figure of eight run (s)	14	6.00 (2.88)	6.50 (2.50)	.69 (.23 - .88)	1.21	3,34	93.33
			Musculoeskeletal fitness				
Shoulder-neck (points)	14	4.71 (2.87)	4.92 (2.94)	(K_w) .42 (.12 - .72)			93.33
Hand grip (kg)	14	23.60 (5.93)	23.67 (6.62)	.67 (.23 - .88)	3.60	9.98	93.33
Jump and reach (cm)	14	11.89 (6.56)	12.93 (5.64)	.73 (.33 - .90)	3.17	8.78	93.33
Dynamic sit-up (number)	15	11.39 (4.52)	12.31 (3.88)	.83 (.57 - .94)	1.79	4.95	93.33
Modified push-up (number)	2	-	-	-	-	-	13.33
			Cardiorespiratory fitness				
2km walk test (min)	8	24,65 (4.18)	22.86 (3.03)	.50 (.14 - .87)	2.54	7.05	53.33

SD: standard deviation; ICC: intraclass correlation coefficient; (K_w) Kappa coefficient; CI: interval confidence; SEM: standard error of measurement; MDC: minimal detectable change; F: feasibility.

Table 2 Test-retest reliability of the extended ALPHA health-related fitness battery in men with intellectual disabilities

	n	TEST Mean (*SD*)	RE-TEST Mean (*SD*)	ICC (95% CI)	SEM	MDC	F %
Body composition							
Waist circumference (cm)	26	100.90 (10.56)	101.00 (10.33)	.97 (.95 - .99)	1.81	5.01	100
Body Mass Index (kg/m^2)	26	26.67 (3.63)	26.87 (3.88)	.98 (.97 - .99)	0.53	1.47	100
Motor fitness							
One leg stand (s)	25	14.97 (16.54)	13.99 (17.16)	.56 (.21 - .78)	11.31	31.35	96.15
Figure of eight run (s)	23	8.46 (2.15)	8.35 (1.70)	.67 (.37 - .85)	1.11	3.07	88.46
Musculoeskeletal fitness							
Shoulder-neck mobility (points)	24	3.96 (2.63)	4.00 (2.81)	(K_w) .32 (.08 - .56)			92.30
Hand grip (kg)	24	32.41 (10.48)	29.73 (10.43)	.90 (.78 - .95)	3.31	9.16	92.30
Jump and reach (cm)	24	19.97 (6.87)	20.48 (10.00)	.73 (.46 - .87)	4.38	12.15	92.30
Dynamic sit-up (number)	24	11.27 (4.64)	12.96 (3.68)	.34 (.04 - .64)	3.38	9.36	92.30
Modified push-up (number)	2	-	-	-	-	-	7.69
Cardiorespiratory fitness							
2km walk test (min)	20	22.34 (2.99)	23,24 (4.33)	.67 (.35 - .86)	2.10	5.82	76.92

SD: standard deviation; ICC: intraclass correlation coefficient; (K_w) Kappa coefficient; CI: interval confidence; SEM: standard error of measurement; MDC: minimal detectable change; F: feasibility.

Table 3 Results from the mixed ANOVA for each independent variable

Dependent variable	Factor/Covariate	*F*	*df*	*P*	R_G^2
Body Mass Index	Test-retest	0.18	1,37	.671	< .01
	Age	2.01	1,37	.165	.01
	Weight	119.88 **	1,37	< .001	.64
	Sex	26.76 **	1,37	< .001	.14
Waist circumference	Test-retest	0.00	1,37	.961	<.01
	Age	7.06 *	1,37	.012	.05
	Weight	103.25 **	1,37	< .001	.69
	Sex	0.87	1,37	.356	<.01
One leg stand	Test-retest	0.31	1,36	.579	<.01
	Age	9.27 **	1,36	.004	.16
	Weight	0.30	1,36	.588	<.01
	Sex	1.01	1,36	.322	.02
Figure of eight run	Test-retest	,624	1,33	.435	<.01
	Age	11,129 **	1,33	.002	.20
	Weight	,002	1,33	.963	< .01
	Sex	3,421	1,33	.073	.06
Dynamic sit-up	Test-retest	,130	1,35	.721	<.01
	Age	5,300 *	1,35	.027	.08
	Weight	9,839 **	1,35	.003	.14
	Sex	2,985	1,35	.093	.04
Jump and reach	Test-retest	,667	1,34	.420	< .01
	Age	10,508 **	1,34	.003	.15
	Weight	,985	1,34	.328	.01
	Sex	15,472 **	1,34	< .001	.22
Shoulder-neck	Test-retest	2,638	1,34	.114	<.01
	Age	,081	1,34	.778	<.01
	Weight	1,304	1,34	.262	.03
	Sex	3,556	1,34	.068	.08
Hand grip strength	Test-retest	,950	1,34	.337	< .01
	Age	,246	1,34	.623	< .01
	Weight	,264	1,34	.611	< .01
	Sex	3,598	1,34	.066	.08
2 Km walk test	Test-retest	,144	1,24	.708	< .01
	Age	,206	1,24	.654	< .01
	Weight	4,207	1,24	.051	.11
	Sex	2,198	1,24	.151	.06

* $p < .05$; ** $p < .01$

Figure legends

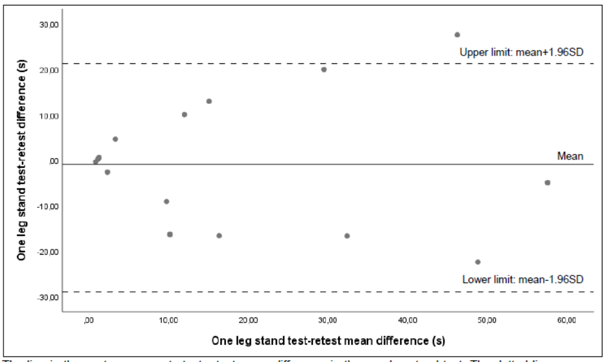
Figure 1 Flowchart inclusion

Appendix. Bland –Altman Plots

Figures 1.A-6.B

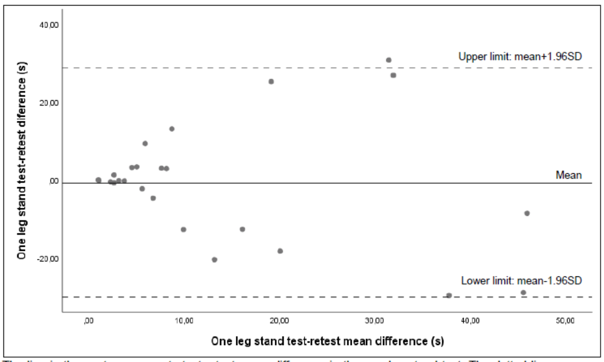
Appendix: Bland-Altman Plots

1.A. One leg stand (women)



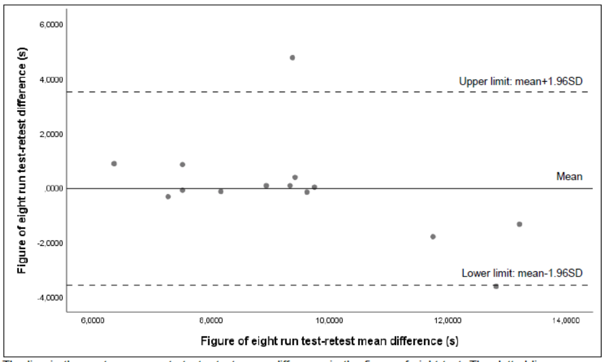
The line in the centre represents test-retest mean difference in the one leg stand test. The dotted lines represent the 95% limit of agreement (upper and lower). Seconds (s). Standard deviation (SD).

1.B. One leg stand (men)



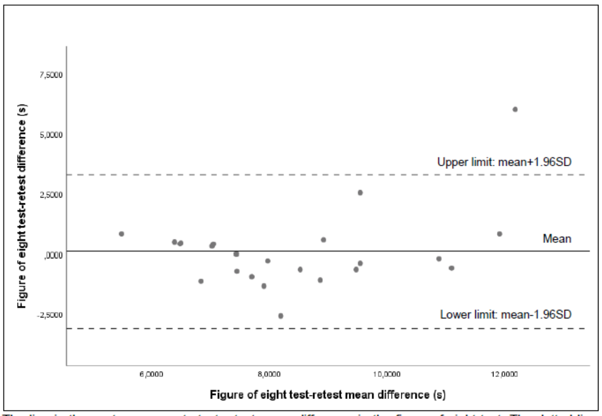
The line in the centre represents test-retest mean difference in the one leg stand test. The dotted lines represent the 95% limit of agreement (upper and lower). Seconds (s). Standard deviation (SD).

2.A. Figure of eight run (women)



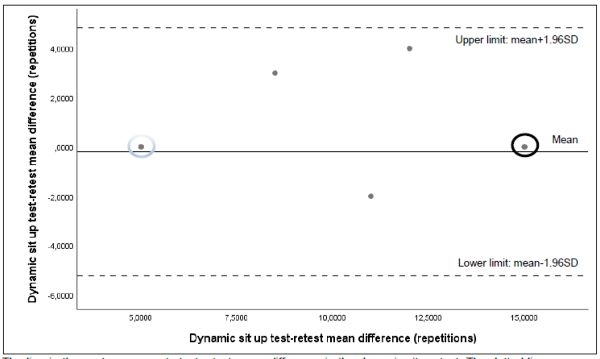
The line in the centre represents test-retest mean difference in the figure of eight test. The dotted lines represent the 95% limit of agreement (upper and lower). Seconds (s). Standard deviation (SD).

2.B. Figure of eight run (men)



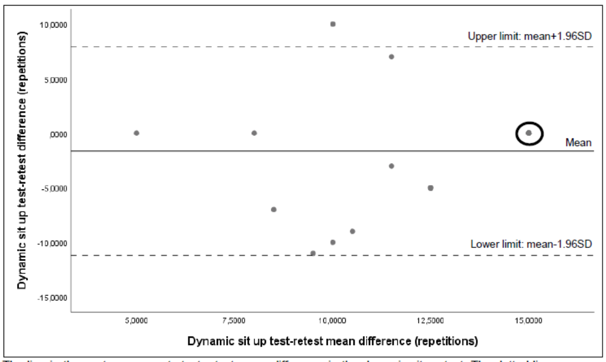
The line in the centre represents test-retest mean difference in the figure of eight test. The dotted lines represent the 95% limit of agreement (upper and lower). Seconds (s). Standard deviation (SD).

3.A. Dynamic sit up (women)



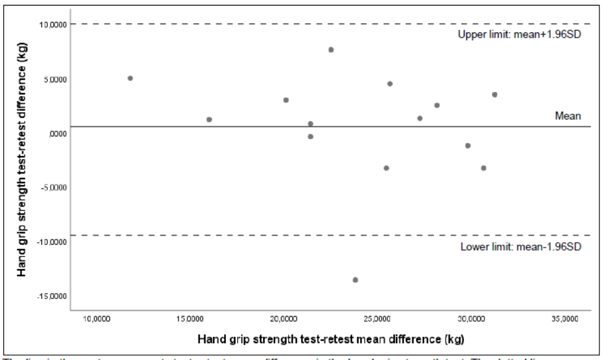
The line in the centre represents test-retest mean difference in the dynamic sit up test. The dotted lines represent the 95% limit of agreement (upper and lower). The grey circled point represents the scores obtained by 3 subjects. The black circled point represents 8 subjects. Standard deviation (SD).

3.B. Dynamic sit up (men)



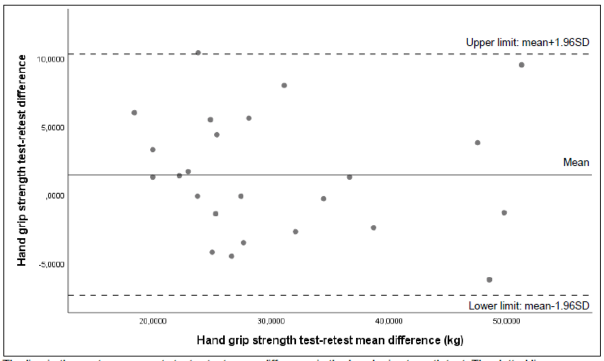
The line in the centre represents test-retest mean difference in the dynamic sit up test. The dotted lines represent the 95% limit of agreement (upper and lower). The highlighted point represents the scores obtained by 13 individuals. Standard deviation (SD).

4.A. Hand grip strength (women)



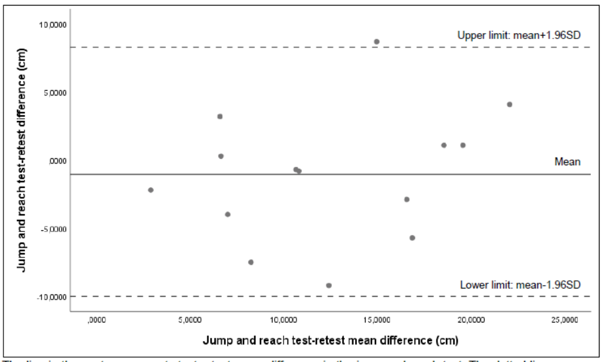
The line in the centre represents test-retest mean difference in the hand grip strength test. The dotted lines represent the 95% limit of agreement (upper and lower). Standard deviation (SD).

4.B. Hand grip strength (men)



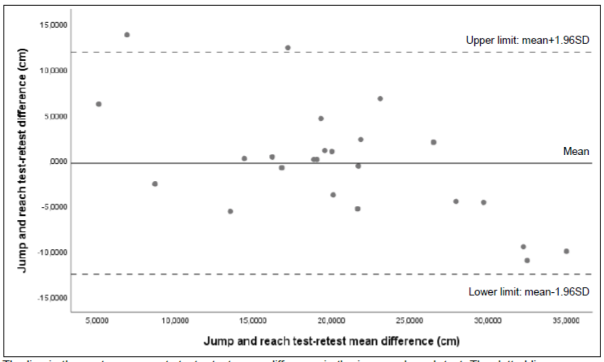
The line in the centre represents test-retest mean difference in the hand grip strength test. The dotted lines represent the 95% limit of agreement (upper and lower). Standard deviation (SD).

5.A. Jump and reach (women)



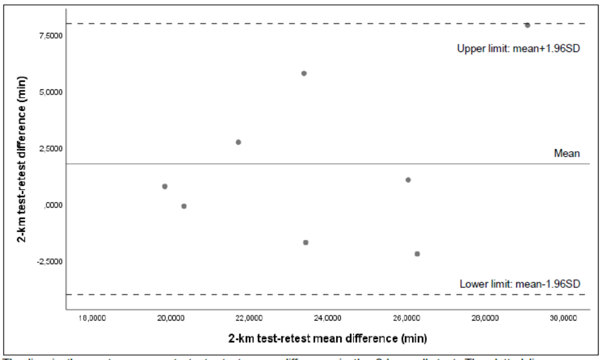
The line in the centre represents test-retest mean difference in the jump and reach test. The dotted lines represent the 95% limit of agreement (upper and lower). Standard deviation (SD).

5.B. Jump and reach (men)



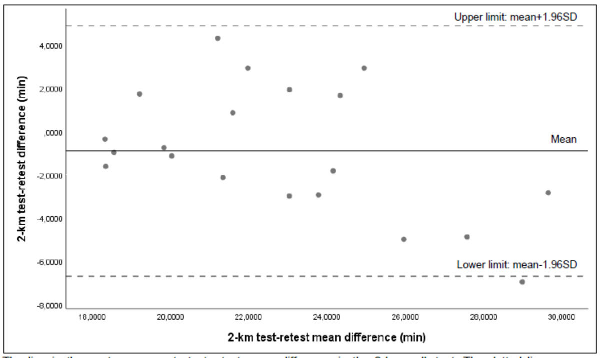
The line in the centre represents test-retest mean difference in the jump and reach test. The dotted lines represent the 95% limit of agreement (upper and lower). Standard deviation (SD).

6.A. 2-km wallk test (women)



The line in the centre represents test-retest mean difference in the 2-km walk test. The dotted lines represent the 95% limit of agreement (upper and lower). Minutes (min). Standard deviation (SD).

6.B. 2-km walk test (men)



The line in the centre represents test-retest mean difference in the 2-km walk test. The dotted lines represent the 95% limit of agreement (upper and lower). Minutes (min). Standard deviation (SD).