

VALIDITY AND RELIABILITY OF FIELD TEST TO ASSESS PHYSICAL FITNESS RELATED TO HEALTH IN ADULTS. THE ADULT-FIT PROJECT

**VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR
LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN
ADULTOS. PROYECTO ADULT- FIT**

International Doctoral Thesis / Tesis Doctoral Internacional



Programa oficial de doctorado en Ciencias de la Salud

Facultad de Ciencias de la Educación

Universidad de Cádiz

Nuria Marín Jiménez

Noviembre 2022

INTERNATIONAL DOCTORAL THESIS / TESIS DOCTORAL INTERNACIONAL

PROGRAMA DE DOCTORADO EN CIENCIAS DE LA SALUD

UNIVERSIDAD DE CÁDIZ

**VALIDITY AND RELIABILITY OF FIELD TEST TO ASSESS PHYSICAL
FITNESS RELATED TO HEALTH IN ADULTS. THE ADULT-FIT PROJECT**



VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA
CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS.

PROYECTO ADULT- FIT

NURIA MARÍN JIMÉNEZ

DICIEMBRE 2022



DEPARTAMENTO DE LA EDUCACIÓN FÍSICA, PLÁSTICA Y MUSICAL
FACULTAD DE CIENCIAS DE LA EDUCACIÓN
UNIVERSIDAD DE CÁDIZ

**VALIDITY AND RELIABILITY OF FIELD TEST TO ASSESS PHYSICAL
FITNESS RELATED TO HEALTH IN ADULTS. THE ADULT-FIT PROJECT**

VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA
CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS.
PROYECTO ADULT- FIT

NURIA MARÍN JIMÉNEZ

Directores de Tesis [Thesis Supervisors]

Dr. José Castro Piñero

Catedrático de Universidad

Universidad de Cádiz

España

Dra. M^a Magdalena Cuenca García

Profesora Titular de Universidad

Universidad de Cádiz

España

En Puerto Real, 04 de noviembre de 2022



Dr. D. José Castro Piñero

Catedrático

Departamento de la Educación Física, Plástica y Musical

Facultad de Ciencias de la Educación

Universidad de Cádiz

JOSÉ CASTRO PIÑERO, CATEDRÁTICO DE LA UNIVERSIDAD DE CÁDIZ,
CERTIFICA:

Que la Tesis Doctoral titulada “VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT” que presenta Dña. Nuria Marín Jiménez al superior juicio que del tribunal que designe la Universidad de Cádiz, ha sido realizada bajo mi dirección durante los años 2019-2022, siendo expresión de la capacidad técnica e interpretativa de su autora, en condiciones tan aventajadas que le hacen merecedora del Título de Doctora, siempre y cuando así lo considere el citado Tribunal.

A handwritten signature in blue ink, consisting of a large circle with a vertical line through it and some horizontal strokes at the bottom.

Fdo. José Castro Piñero

En Puerto Real, a 04 de noviembre de 2022



Dra. Dña. M^a Magdalena Cuenca García

Profesora Titular de Universidad

Departamento de la Educación Física, Plástica y Musical

Facultad de Ciencias de la Educación

Universidad de Cádiz

MARIA MAGDALENA CUENCA GARCÍA, PROFESORA TITULAR DE LA UNIVERSIDAD DE CÁDIZ, CERTIFICA:

Que la Tesis Doctoral titulada “VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT” que presenta Dña. Nuria Marín Jiménez al superior juicio que del tribunal que designe la Universidad de Cádiz, ha sido realizada bajo mi dirección durante los años 2019-2022, siendo expresión de la capacidad técnica e interpretativa de su autora, en condiciones tan aventajadas que le hacen merecedora del Título de Doctora, siempre y cuando así lo considere el citado Tribunal.

A handwritten signature in blue ink, appearing to read 'M. Cuenca', with a stylized flourish at the end.

Fdo. M^a Magdalena Cuenca García

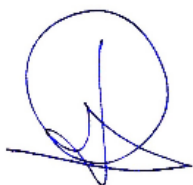
En Puerto Real, a 04 de noviembre de 2022

Tanto la doctoranda, Nuria Marín Jiménez, como sus directores de tesis, D. José Castro Piñero y Dña. M^a Magdalena Cuenca García:

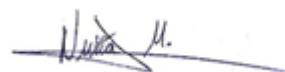
Garantizamos que la presente Tesis Doctoral ha sido redactada bajo la supervisión de los directores de tesis y, hasta donde nuestro conocimiento alcanza, respetando la autoría de los estudios citados, al utilizarse sus resultados o publicaciones.

Directores de tesis

Doctoranda



Fdo. José Castro Piñero



Fdo. Nuria Marín Jiménez



Fdo. M^a Magdalena Cuenca García

En Puerto Real, a 04 de noviembre de 2022

*Un Yo que sigue cambiando es un Yo que continúa viviendo.
Y, por el amor de Dios, no publiques nada antes de los treinta.*

-Virginia Woolf.

TABLE OF CONTENTS

<u>Research projects and funding</u>	17
<u>Publications</u>	21
<u>List of tables and figures</u>	25
<u>Abbreviations</u>	33
<u>ABSTRACT / RESUMEN</u>	37
<u>INTRODUCTION</u>	47
A. GENERAL INTRODUCTION	49
B. CHAPTER-SPECIFIC INTRODUCTION	52
Chapter I	53
Chapter II	53
Chapter III	55
Chapter IV	56
<u>HYPOTHESIS</u>	63
<u>AIMS / OBJETIVOS</u>	67
<u>MATERIAL AND METHODS</u>	73
Measurements	87
Statistical analyses	92
General overview of the studies' methodology	97
<u>RESULTS</u>	103
Chapter I	106
Chapter II	140
Chapter III	317
Chapter IV	377
<u>DISCUSSION</u>	477
Chapter I	480
Chapter II	486
Chapter III	499
Chapter IV	513
<u>LIMITATIONS AND STRENGTHS</u>	523
<u>CONCLUSIONS / CONCLUSIONES</u>	529
<u>FUTURE RESEARCH DIRECTIONS</u>	543
<u>ANNEXES</u>	567
Curriculum Vitae	569
List of studies included and quality indexes	577
Agreement of co-authors	581

RESEARCH PROJECTS AND FUNDING

RESEARCH PROJECTS AND FUNDING

The present International Doctoral Thesis was performed under the framework of the ADULT-FIT project*, which received the following funding:

- Ministry of Economy, Industry and Competitiveness in the 2017, call for R&D Projects of the State Program for Research, Development and Innovation Oriented to the Challenges of the Company; National Plan for Scientific and Technical Research and of Innovation 2017-2020 (DEP2017-88043-R).
- Ministry of Science, Innovation and Universities in the 2018, call for the Procedure for the granting of Grants for the acquisition of Scientific-Technical Equipment corresponding to the State Subprogram for Research Infrastructures and Scientific-Technical Equipment (State Plan R&D&I 2017-2020). Acquisition of Isokinetic equipment for Functional Assessment, Prevention, Rehabilitation and Treatment in Healthy People and People with Pathologies (EQC2018-004475-P).
- Regional Government of Andalusia and University of Cadiz: Research and Knowledge Transfer Fund (PPIT-FPI19).

Moreover, the present International Doctoral Thesis was supported by the following fellowships:

- Programa de fomento e impulso de la investigación y la transferencia en la Universidad de Cádiz 2018-2019. Ayuda para la Realización de Estancias de Investigación en Universidades de Prestigio. UCA Internacional.
Receiving center: Karolinska Institutet, Sweeden. 2021.
- ERASMUS+, movilidad internacional para Personal Docente e Investigador (modalidad KA107).
Receiving center: Tbilisi State University, Georgia. 2022.
- ERASMUS+, movilidad internacional para Personal Docente e Investigador (modalidad KA103).
Receiving center: Tallina Ülikool, Estonia. 2020.

*Visit <https://adultfitstudy.uca.es/> for more information.

PUBLICATIONS

PUBLICATIONS

The present International Doctoral Thesis is composed of six publications. These six publications are presented and grouped into different chapters throughout the Doctoral Thesis document, in the following order:

Chapter I. Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study I*).

- I. **Marín-Jiménez, N.**, Cruz-León, C., Pérez-Bey, A., Conde-Caveda, J., Grao-Cruces, A., Aparicio, V. A., Castro-Piñero, J. & Cuenca-García, M. (2022). Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review. *Journal of Clinical Medicine*, 11(2), 328: Q1.

Chapter II. Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

- II. Castro-Piñero, J., **Marín-Jiménez, N.**, Fernández-Santos, J. R., Martín-Acosta, F., Segura-Jiménez, V., Izquierdo-Gómez, R., R. Ruiz, J. & Cuenca-Garcia, M. (2021). Criterion-Related Validity of Field-Based Fitness Tests in Adults: A Systematic Review. *Journal of clinical medicine*, 10(16), 3743: Q1.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

- III. **Marín-Jiménez, N.**, Cruz-León, C., Sánchez-Oliva, D., Jiménez-Iglesias, J., Caraballo, I., Padilla-Moledo, C., Cadenas-Sánchez, C., Cuenca-García, M. & Castro-Piñero, J. (2022). Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults: A Systematic Review. *Current Obesity Reports, Accepted, In production*: Q1.

Chapter III. Reliability of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study IV*).

- IV. Cuenca-García, M., **Marín-Jiménez, N.**, Pérez-Bey, A., Sánchez-Oliva, D., Camiletti-Moirón, D., Álvarez-Gallardo, I. C., Ortega, F. B. & Castro-Piñero, J. (2022). Reliability of Field-Based Fitness Tests in Adults: A Systematic Review. *Sports Medicine*, 1-20: Q1.

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

- V. **Marín-Jiménez, N.**, Sánchez-Parente, S., Expósito-Carrillo, P., Jiménez-Iglesias, J., C. Álvarez-Gallardo, I., Cuenca-García, M. & Castro-Piñero, J. Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level. *Journal of Science and Medicine in Sport*, Under review: Q1.

Study VI: The standing long jump test.

- VI. **Marín-Jiménez, N.**, Pérez-Bey, A., Cruz-León, C., Conde-Caveda, J., Segura-Jiménez, V., Castro-Piñero, J. & Cuenca-García, M. Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project. *Journal of Strength and Conditioning Research*, Submitted: Q1.

LIST OF TABLES AND FIGURES

LIST OF TABLES

- Table 1.** Inclusion and exclusion criteria of the studies of the present Doctoral Thesis.
- Table 2.** Quality assessment criteria for predictive validity studies (*study I*).
- Table 3.** Quality assessment criteria for physical fitness criterion-related validity studies (*study II*).
- Table 4.** Quality assessment criteria for body composition criterion-related validity studies (*study III*).
- Table 5.** Quality assessment criteria-related reliability studies (*study IV*).
- Table 6.** Summary of the main methodological features of the studies of the present Doctoral Thesis.
- Table 7.** List of included longitudinal studies with quality scores with reference to predictive value of motor fitness and flexibility tests for health outcomes in adults and older adults (*study I*).
- Table 8.** Quality assessment of the systematic reviews included using the AMSTAR rating tool (*study I*).
- Table 9.** Longitudinal validity of motor fitness and flexibility tests for falls and hip fracture risk in adults and older adults (*study I*).
- Table 10.** Longitudinal validity of motor fitness and flexibility tests for cognitive outcomes in adults and older adults (*study I*).
- Table 11.** Longitudinal validity of motor fitness and flexibility tests for depressive symptoms and well-being in adults and older adults (*study I*).
- Table 12.** Longitudinal validity of motor fitness and flexibility tests for mobility limitations and disability risk in adults and older adults (*study I*).
- Table 13.** Longitudinal validity of motor fitness and flexibility tests for CVD risk and mortality in adults and older adults (*study I*).
- Table 14.** Systematic reviews/meta-analysis on predictive validity of motor fitness tests for adverse health outcomes in adults and older adults (*study I*).
- Table 15.** Main characteristics and findings of longitudinal studies investigating the predictive validity of motor fitness and flexibility tests for health outcomes in adults and older adults (*study I*).
- Table 16.** Quality assessment of fitness criterion-related validity studies (*study II*).
- Table 17.** Overview of field-based physical fitness tests criterion-related validity studies in adults (*study II*).

Table 18. Quality assessment of the included meta-analysis/systematic reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) rating tool (*study II*).

Table 19. Overview of field-based physical fitness tests criterion-related validity of included meta-analysis/systematic reviews (*study II*).

Table 20. Levels of evidence of cardiorespiratory fitness tests (*study II*).

Table 21. Levels of evidence of muscular fitness, flexibility and motor fitness tests (*study II*).

Table 22. Quality assessment of field-based body composition validity studies in adults (*study III*).

Table 23. Quality assessment of derived-equation body composition validity studies in adults (*study III*).

Table 24. Overview of field-based body composition estimation studies in adults (*study III*).

Table 25. Levels of evidence of field-based body composition assessment validity studies in adults (*study III*).

Table 26. Durnin and Womersley (1974) generalized equations and, Siri and Brozek converted formulas for body fat calculation in adults (*study III*).

Table 27. Jackson, Pollock and Ward (1980), and Jackson and Pollock (1978) generalized equations and, Siri and Brozek converted formulas for body fat calculation in adults (*study III*).

Table 28. Overview of estimation equation body composition studies in adults (*study III*).

Table 29. Levels of evidence of estimation equations studies in adults (*study III*).

Table 30. Advantages and limitations of the “gold-standard” methods, and valid field-based methods and equations used for the assessment of body composition in adults (*study III*).

Table 31. Quality assessment of fitness reliability studies (*study IV*).

Table 32. Overview of field-based physical fitness tests reliability studies in adults (*study IV*).

Table 33. Levels of evidence of cardiorespiratory fitness tests (*study IV*).

Table 34. Levels of evidence of musculoskeletal and motor fitness tests (*study IV*).

Table 35. Descriptive characteristics of the participants, stratified by sex, age and, physical activity level (*study V*).

Table 36. Bivariate correlation analysis between measured $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) with estimated $\text{VO}_{2\text{max}}$ by 2-km walk test, by 20-meter shuttle run test, and anthropometric variables, distributed by whole sample, sex, age, and physical activity levels (*study V*).

Table 37. Stepwise linear regression model predicting $\text{VO}_{2\text{max}}$ by 2-km walk test and, by 20-meter shuttle run test (*study V*).

Table 38. Test-retest reliability of 2-km walk test and 20-m shuttle run test in the whole sample (*study V*).

Table 39. Test-retest reliability of 2-km walk test and 20-m shuttle run test, distributed by sex, age, and physical activity levels (*study V*).

Table 40. Descriptive characteristics of the participants, stratified by sex, age and physical activity levels (*study VI*).

Table 41. Bivariate correlation analysis between standing long jump field-based test, force platform and isokinetic measurements, and anthropometrics variables (*study VI*).

Table 42. Bivariate correlation analysis between standing long jump field-based test, force platform and isokinetic measurements, and anthropometrics variables, distributed by sex, age groups, and physical activity levels (*study VI*).

Table 43. Stepwise regression models assessing the association of the force platform variables with the standing long jump distance (*study VI*).

Table 44. Test-retest reliability of standing long jump in the whole sample, and in a subsample after a learning period (*study VI*).

Table 45. Test-retest reliability of standing long jump, distributed by sex, age groups, and physical activity levels (*study VI*).

LIST OF FIGURES

Figure 1. Flow-chart of the ADULT-FIT project.

Figure 2. Flow-chart of the study selection process (*study I*).

Figure 3. Number of studies reporting predictive validity or lack of predictive validity of the included motor fitness and flexibility tests for different health outcomes (*study I*).

Figure 4. Cut points of gait speed at usual pace and risk of adverse outcomes found in the selected studies and systematic reviews/meta-analysis (*study I*).

Figure 5. Cut points of timed up&go test and risk of adverse outcomes found in the selected studies and systematic reviews/meta-analysis (*study I*).

Figure 6. Flow chart of retrieved and selected articles (*study II*).

Figure 7. Level of evidence found about the criterion-related validity of field-based physical fitness tests in adults (*study II*).

Figure 8. Flow chart of retrieved and selected articles (*study III*).

Figure 9. Most common methods for body composition assessment (*study III*).

Figure 10. PRISMA flowchart showing the selection of original studies (*study IV*).

Figure 11. Evidence-based proposal of field-based physical fitness test reliability in adults (*study IV*).

Figure 12. Scatterplot of relationship between measured with estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation (estimated from the 2-km walk test) and Leger's equation (estimated from the 20-m shuttle run test); and relationship between measured with total time in the 2-km walk test and final stage in the 20-m shuttle run test (*study V*).

Figure 13. Scatterplot of relationship between measured with estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation (estimated from the 2-km walk test), and relationship between measured with total time in the 2-km walk test; distributed by sex, age and physical activity levels (*study V*).

Figure 14. Scatterplot of relationship between measured with estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Leger's equation (estimated from the 20-m shuttle run test), and relationship between measured with final stage in the 20-m shuttle run test; distributed by sex, age, and physical activity levels (*study V*).

Figure 15. Bland-Altman plot of agreement between measured and estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation (estimated from the 2-km walk test), and Leger's equation (estimated from the 20-m shuttle run test) (*study V*).

Figure 16. Bland-Altman plot of agreement between measured and estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation (estimated from the 2-km walk test) and Leger's equation (estimated from the 20-m shuttle run test), distributed by sex, age, and physical activity levels (*study V*).

Figure 17. Bland-Altman plot of the estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation, and Leger's equation, the 2-km walk test (seconds), and the 20-m shuttle run test (stage) (*study V*).

Figure 18. Bland-Altman plot of the estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation, and Leger's equation, the 2-km walk test (seconds), and the 20-m shuttle run test (final stage), distributed by sex, age, and physical activity levels (*study V*).

Figure 19. Bland-Altman plot of standing long jump test (*study VI*).

Figure 20. Bland-Altman plot of standing long jump distributed by sex, age groups, and physical activity levels (*study VI*).

Figure 21. Bland-Altman plot of standing long jump test after a learning period (*study VI*).

Figure 22. Evidence-based field physical fitness tests in adult population / Test de campo de condición física basados en evidencia científica en población adulta.

ABBREVIATIONS

ABBREVIATIONS

%BF: Body Fat percentage	CV_{ME}: Coefficient of Variation of Method Error
20-m SRT: 20-m shuttle run test	CV_{RSME}: Coefficient of Variation of the Root-Mean-Square Error
³H₂O: Tritium Water	CV_{SEM}: Coefficient of Variation of Standard Error of Measurement
AbC: Abdominal Circumference	D/W: Durnin/Womersley equation
AbC+NC: Equations combining Abdominal Circumference + Neck Circumference	D₂O: Deuterium Oxide
AbC+SKF: Equations combining Abdominal Circumference + Skinfolde	DD: Deltoid Diameter
AC: Arm Circumference	DXA: Dual-energy X-ray Absorptiometry
ACSM: American College of Sport Medicine	E: Error
AD: Arm Diameter	ED: Elbow Diameter
ADP: Air-Displacement Plethysmography	EST: Regression Coefficients Estimates
AkC: Ankle Circumference	FC: Forearm Circumference
AMSTAR: Ameasurement Tool to Assess Systematic Reviews	FFM: Fat Free Mass
ANCOVA: Analysis of Covariance	FRT: Functional Reach Test
ANOVA: Analysis of Variance	GlC: Gluteal Circumference
AUC: Area Under the Curve	GLM: General Linear Models
B: Non-Standardized Regression Coefficient	HC: Head Circumference
BAI: Body Adiposity Index	HC: Hip Circumference
BD: Body Density	HR*: Hazard Ratio
BIA: Bioelectrical Impedance Analysis	HR: Heart rate
BMI: Body mass index	IADL: Disability in Instrumental Activities of Daily Living
C+D+SKF: Equations combining Circumferences + Diameters + Skinfolde	IC: Iliac Circumference
C+SKF: Equations combining Circumferences + Skinfolde	ICC: Intraclass Correlation Coefficient
CC: Calf Circumference	ISAK: International Society for the Advancement of Kinanthropometry
CD: Critical Difference	K⁴⁰: Potassium-40
CE: Constant Error	KC: Knee Circumference
ChC: Chest Circumference	KD: Knee Diameter
ChD: Chest Diameter	LCCC: Lin's Concordance Coefficient
CI: Confidence Intervals	LoA: Limits of Agreement
CPAFLA: Canadian Physical Activity, Fitness and Lifestyle Appraisal	MANOVA: Multiple Analysis Of Variance
CPET: Cardiopulmonary Exercise Test	MD: Mean Difference
CV: Coefficient of Variation	MDC 95% CI: Minimal Detectable Change at 95% of Coefficient of Interval
CVD: Cardiovascular Disease	MDC: Minimal Detectable Change
	ME*: Mean Error
	ME: Method Error

MeSH: Medical Subject Heading	SRD: Smallest Real Difference
MIR: Magnetic Resonance Imaging	S_w: Measurement Error
MSD: Mean Signed Difference	TC: Thigh Circumference
NC: Neck Circumference	TE*: Typical Error
NIR: Near-Infrared Interactance	TE: Total Error
NMF_{slope}: Normalized Median Frequency slope	TEM: Technical Error of Measurement
OR: Odds Ratio	TUG: Timed Up&Go
PAR-Q: Physical Activity Readiness Questionnaire	UWW: Underwater Weighing
PCA: Percentage Close Agreement	VO_{2max}: Maximum Oxygen Uptake
PE: Pure Error	VO_{2peak}: Peak Oxygen Uptake
PF: Physical Fitness	WC: Waist Circumference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis	WC+SKF: Equations combining Waist Circumference + Skinfolds
PROSPERO: International Prospective Register of Systematic Reviews	WC+SKF+BMI: Equations combining Waist Circumference + Skinfolds + Body Mass Index
QoL: Quality of Life	WHR: Waist to Hip Ratio
R: Pearson Correlation Coefficient	WHtR: Waist to Height Ratio
RER: Respiratory Exchange Ratio	WrC: Wrist Circumference
RM: Repetition Maximum	WrD: Wrist Diameter
RMSE: Root Mean Square Error	β: Standardized Regression Coefficient
ROM: Range Of Motion	
RPE: Rated Perceived Exertion	
RR: Relative Risk	
RRSE: Root Relative Square Error	
SD: Standard Deviation	
SD_{diff}: Standard Deviation of the Differences	
SE: Standard Error	
SEE: Standard Error Estimate	
SEM: Standard Error of Measurement	
ShC: Shoulder Circumference	
SKF: Skinfold Thickness	
SKF+BMI: Equations combining Skinfolds + Body Mass Index	
SLJ: Standing Long Jump test	
SPSS: Statistical Package for the Social Sciences	

ABSTRACT
[RESUMEN]

ABSTRACT

Physical fitness is considered a powerful health marker in different populations (i.e., preschool, childhood, adolescence, adults, and older adults). Epidemiological studies have demonstrated an inverse association between cardiorespiratory and muscular fitness, with morbidity and all-cause mortality. Consequently, physical fitness assessment is considered an important prevention and diagnosis tool. Physical fitness can be objectively and accurately measured through laboratory tests. However, due to their high cost, the necessity of sophisticated instruments, qualified technicians, and time constraints, their use is limited in population-based studies. Field-based physical fitness tests provide a reasonable alternative since they are easy to administer, involve minimal equipment, are low in cost, and a large number of participants can be evaluated in a relatively short period of time and simultaneously. Valid, reliable, feasible, safe, and responsiveness are characteristics that need to be assured for any measurement tool. Longitudinal studies have recently contributed to the existing health-related fitness batteries archiving those characteristics in children and adolescents (The ALPHA Study), and preschool children (The PREFIT Study). However, researchers, clinicians, and sport practitioners do not have enough information about which field-based physical fitness tests are more reliable, valid, and informative from the health point of view to be implemented in adults.

Therefore, the **general aim** of this International Doctoral Thesis was: to propose a battery of field-based tests to assess physical fitness related to health in adults based on scientific evidence, valid (i.e. predictive validity and criterion-related validity) and reliable, depending on sex, age, and physical activity level: The ADULT-FIT project [Plan nacional I+D +i 2017-2020, (ref.: DEP2017-88043-R)].

Likewise, the **specific aims** were: (i) to comprehensively analyze the **predictive validity** of the existing motor fitness and flexibility tests in relation to several health outcomes in adults and older adults (*study I*); (ii) to comprehensively study the **criterion-related validity** of the existing field-based physical fitness tests used in adults aged 18-64 years (*study II*), and to identify studies evaluating the **criterion-related validity** of existing field-based methods and equations for body composition estimation used in adults aged 18-64 years (*study III*); (iii) to systematically review studies conducted to examine the **reliability** of field-based physical fitness tests used in adults aged 18-64 years (*study IV*); (iv) to analyze the **criterion-related validity** and the **reliability** of the

2-km walk test and the 20-m SRT for evaluating cardiorespiratory fitness in the adult population, according to sex, age, and physical activity level (*study V*), and to analyze the **criterion-related validity** and **reliability** of the SLJ test to assess the lower-body explosive muscular fitness in adults, according to sex, age, and physical activity level (*study VI*).

The **main results** showed that there is **strong evidence** indicating that: (i) slower gait speed, impaired balance, and worse TUG performance are good indicators of health outcomes in both adults and older adults; and limited evidence about the **predictive validity** of flexibility tests exists (*study I*). (ii) To assess cardiorespiratory fitness, the 20-m shuttle run test (using Leger's equation), 1.5-mile run/walk test, 12-minute run/walk test, 6-minute walk test, YMCA step test, and the 2-km walk test (using Oja's equation) are **valid**; to assess muscular fitness, the handgrip test (using the TKK dynamometer), and the Biering–Sørensen test are **valid**; to assess flexibility, the sit-and-reach test (and its different versions), and that the toe-to-touch tests are **not valid**; limited evidence about the criterion-related validity of motor fitness exists (*study II*). (iii) The field-based methods, waist circumference, body adiposity index, and body mass index are **valid** indicators of body adiposity in adult population; classical equations, such as Durnin/Womersley equation, Jackson/Pollock equation, Jackson, Pollock and Ward equation, and estimation equations implying skinfolds, alone or combined with circumferences, are most **valid** to estimate total body fat mass or body fat percentage in adult population (*study III*). (iv) To assess cardiorespiratory fitness, the 20-m shuttle run test, the step tests and the 6-minute walk test are **reliable**; to assess musculoskeletal fitness, the handgrip test (using a JAMAR dynamometer), back-leg test, Biering–Sørensen (and its modified versions), trunk flexion sustained, 5-reps sit-to-stand, bilateral side and prone bridge, the sit-and-reach (and its modified versions), and the toe-to-touch test are **reliable**; to assess motor fitness, the T-test (and its modified version) is **reliable**, and the single-leg stand test (and its modified versions) is **not reliable** (*study IV*). (v) The 2-km walk test (using) and the 20-m shuttle run test, as well as their corresponding Oja's and Leger's equations, are **valid** and **reliable** for estimating cardiorespiratory fitness in adults aged 18-64 years. However, the 20-m shuttle run test obtained slightly greater criterion-related validity and reliability, regardless of sex, age, and physical activity level (*study V*). (vi) The standing long jump test may be a **valid** tool to assess the lower-body explosive muscular fitness in the adult population, independently of age, and physical

activity levels. The standing long jump test may be used when controlling the possible learning effect to provide **reliability** (*study VI*).

This International Doctoral Thesis provides evidence to propose **a battery of field-based tests to assess physical fitness related to health in adults** based on **valid** and **reliable** scientific evidence, depending on sex, age, and physical activity levels. Future methodological studies are still necessary to establish the validity and/or reliability of those field-based physical fitness tests identified in systematic reviews having insufficient or strong evidence that have not been developed in this International Doctoral Thesis, as well as the feasibility and safety of those field-based physical fitness test identified (in the systematic reviews) as valid and reliable, and the responsiveness of those field-based physical fitness test identified as valid, reliable, feasible and safe, in adult population.

RESUMEN

La **condición física** es considerada un potente marcador de salud en todas las poblaciones (i.e., preescolares, niñez, adolescencia, adultos y adultos mayores). Estudios epidemiológicos han demostrado una asociación inversa entre la capacidad cardiorrespiratoria y fuerza muscular, con morbilidad y mortalidad por todas las causas. Así, la evaluación de la condición física se considera un importante medio de prevención y diagnóstico. Los test de laboratorio son la manera más objetiva y precisa de evaluar la condición física. Sin embargo, éstos suponen un alto coste, debido al uso de equipamientos sofisticados, a la presencia de personal cualificado y limitaciones de tiempo, siendo su uso limitado en estudios poblacionales. Como alternativa, los test de campo son fáciles de administrar y suponen un bajo coste al reducirse la dependencia de equipamiento especializado e incrementarse el número de participantes evaluados en un mismo momento y, en un período relativamente corto de tiempo. Para que un test de campo sea considerado “adecuado” ha de ser válido, fiable, objetivo, además de seguro, viable-aplicable y sensible al cambio de rendimiento. Recientemente se han realizado importantes esfuerzos para disponer de baterías de test que evalúan la condición física relacionada con la salud que cumplan estas características, tanto en niños y adolescentes (Batería ALPHA) como en población infantil (Batería PREFIT). Sin embargo, los investigadores, personal clínico y los usuarios no tienen suficiente información sobre qué test de condición física de campo son más fiables, válidos e informativos desde el punto de vista de la salud para ser implementados en adultos.

Por tanto, el **objetivo general** de la presente Tesis Doctoral Internacional fue: proponer una batería de test de campo para evaluar la condición física relacionada con la salud en adultos basada en la evidencia científica, válida (i.e. validez predictiva y de criterio) y fiable, en función del sexo, la edad y el nivel de actividad física: Proyecto ADULT-FIT [Plan nacional I+D +i 2017-2020, (ref.: DEP2017-88043-R)].

De igual forma, los **objetivos específicos** fueron: (i) analizar la **validez predictiva** de los test de campo existentes de capacidad motora y flexibilidad en relacionados con parámetros de salud en adultos y adultos mayores (*estudio I*); (ii) estudiar la **validez de criterio** de los test de campo de condición física empleados en adultos de 18-64 años (*estudio II*), e identificar los estudios que evalúen la **validez de criterio** de los métodos y ecuaciones de campo para la estimación de la composición corporal en adultos de 18-64 años (*estudio III*); (iii) revisar sistemáticamente los estudios que examinen la **fiabilidad**

de los test de campo de condición física en adultos de 18-64 años (*estudio IV*); (iv) analizar la **validez de criterio** y **fiabilidad** de los test 2-km andando y 20-m ida y vuelta para la evaluación de la capacidad cardiorrespiratoria en la población adulta, según su sexo, edad y nivel de actividad física (*estudio V*), y analizar la **validez de criterio** y **fiabilidad** del salto horizontal a pies juntos para evaluar la fuerza muscular explosiva de la parte inferior del cuerpo en adultos, según su sexo, edad y nivel de actividad física (*estudio VI*).

Los **principales resultados** mostraron que existe **fuerte evidencia** indicando que: (i) la velocidad de marcha más lenta, el equilibrio alterado y el peor rendimiento en el test de levantarse y caminar son buenos indicadores de resultados de salud tanto en adultos como en adultos mayores; y existe evidencia limitada sobre la **validez predictiva** de los test de flexibilidad (*estudio I*). (ii) Para evaluar la capacidad cardiorrespiratoria, el test de 20-m ida y vuelta (usando la ecuación de Leger), los test de 1,5 millas, 12 minutos corriendo/andando, 6-minutos andando, el YMCA step test, y el test de 2-km andando (usando la ecuación de Oja) son **válidos**; para evaluar la fuerza muscular, el test de prensión manual (utilizando el dinamómetro TKK) y el test de Biering-Sørensen son **válidos**; para evaluar la flexibilidad, el sit-and-reach test (y sus diferentes versiones), y el toe-to-touch test **no son válidos**; existe evidencia limitada sobre la validez de criterio de los test de capacidad motora (*estudio II*). (iii) Los métodos de evaluación de campo de la composición corporal, como la circunferencia de cintura, el índice de adiposidad corporal y el índice de masa corporal son los indicadores más **válidos** para evaluar la composición/adiposidad corporal en población adulta; las **ecuaciones clásicas**, como la ecuación de Durnin/Womersley, la ecuación de Jackson/Pollock, la ecuación de Jackson, Pollock y Ward y las ecuaciones que implican pliegues corporales, solas o combinadas con circunferencias, son las más válidas para estimar la grasa corporal total o el porcentaje de grasa corporal en la población adulta (*estudio III*). (iv) Para evaluar la capacidad cardiorrespiratoria, los test de 20-m ida y vuelta, los step test y el test de 6-minutos andando son **fiables**; para evaluar la capacidad musculoesquelética, el test de fuerza de prensión manual (usando un dinamómetro JAMAR), los test back-leg strength, Biering-Sørensen (y sus versiones modificadas), la flexión sostenida del tronco, y 5 repeticiones de sentarse y levantarse, puente lateral y prono, el sit-and-reach test (y sus versiones modificadas) y toe-to-touch test son **fiables**; para evaluar la capacidad motora, el T-test (y su versión modificada) es **fiable**, y el test de equilibrio sobre una sola pierna (y sus versiones modificadas) **no es fiable** (*estudio IV*). (v) Los test de 2-km marcha (usando la

ecuación de Oja) y 20-m ida y vuelta, (usando la ecuación de Leger) son **válidos y fiables** para estimar la capacidad cardiorrespiratoria en adultos de 18 a 64 años. Sin embargo, el test de 20-m ida y vuelta obtuvo una validez y fiabilidad ligeramente mayores, independientemente del sexo, la edad y el nivel de actividad física (*estudio V*). (vi) El test de salto horizontal a pies juntos puede ser una herramienta **válida** para evaluar la fuerza muscular explosiva de la parte inferior del cuerpo en población adulta, independientemente de la edad y nivel de actividad física. El test de salto horizontal a pies juntos es **fiable** cuando se controla el posible efecto de aprendizaje (*estudio VI*).

La presente Tesis Doctoral Internacional proporciona indicios para proponer una **batería de test de campo para evaluar la condición física relacionada con la salud en adultos** basada en la evidencia científica, **válida y fiable**, en función del sexo, edad y el nivel de actividad física. Futuros estudios metodológicos son aún necesarios para establecer la validez y/o fiabilidad de aquellos test de condición física identificados en las revisiones sistemáticas que no han obtenido suficiente o fuerte evidencia que no han podido ser desarrollados en la presente Tesis Doctoral Internacional, así como viabilidad-aplicabilidad y seguridad de aquellos test de condición física de campo identificados (en las revisiones sistemáticas) como válidos y fiables, y la sensibilidad al cambio de rendimiento de aquellos test de condición física de campo identificados como válidos, fiables, viables-aplicables y seguros, en la población adulta.

INTRODUCTION



INTRODUCTION

A. GENERAL INTRODUCTION

Physical fitness is an integrated measure of all the functions and structures involved in performing physical activity¹. Nowadays, physical fitness is considered a powerful marker of health at any age²⁻⁴, especially cardiorespiratory and muscular fitness. Cardiorespiratory fitness is inversely associated with cardiovascular diseases (CVD)⁵, obesity⁶, osteoporosis⁷ diabetes⁸, different types of cancer^{9, 10}, and is a predictor of all-cause of mortality¹¹⁻¹⁵, and CVD^{13, 15-18}. Likewise, in the psychological sphere, high levels of cardiorespiratory fitness are associated with well-being^{19, 20}, improved cognitive function²¹ and a reduced risk of Alzheimer's disease²², and other mental conditions such as anxiety, panic and depression²³. Muscular fitness demonstrates a protective effect against all-cause mortality^{24, 25}; and is inversely associated with weight gain and adiposity-related hypertension occurrence and the prevalence and incidence of the metabolic syndrome²⁵, and mental health clinical presentations^{26, 27}.

Consequently, physical fitness assessment is a vital tool of prevention and health diagnoses.

Health-related physical fitness

Increases in life expectancy in addition to the high proportion of deaths being attributed to unspecified causes and the multi-morbidity mainly common in older adults (≥ 65 years old) are associated with increased rates of healthcare use and cost, resulting in challenges for health policies around the world²⁸. Moreover, disabilities related to pain (such as musculoskeletal pain), depressive disorders, dementia, osteoarthritis, and falls are becoming a great burden in older people²⁸. Impaired musculoskeletal health (including mobility and function limitations) is characterized by a reduced quality of life (QoL), loss of independence²⁹, mental ill-being and mortality³⁰.

As mentioned above, it is well established that cardiorespiratory and muscular fitness are powerful marker of health. Thus, maintaining an adequate physical fitness level is a good indicator of possible adverse health events in different populations including adults and older adults^{29, 31, 32}.

On the other hand, there are some other components of physical fitness that need to be analyzed to clarify whether they have predictive value for diverse health outcomes in adults and older adults. These components of physical fitness are motor fitness and flexibility, which are related to disabilities that can impair movement common in several health complications, which may imply a reduction in QoL²⁹. Therefore, assessment of motor fitness and flexibility could help to identify adults and older adults who may be at risk of suffering from these health outcomes. Through its predictive value, motor fitness and flexibility could be postulated as a health marker in these populations.

Moreover, obesity is a pandemic that affects people from childhood to adulthood, being an important risk factor for CVD, hypertension, diabetes mellitus, cancer, and premature death³³, in addition to entailing elevated healthcare costs³⁴. By 2030, nearly 38% of the world's population will be affected by overweight, and 20% will be affected by obesity³³. Therefore, body composition assessment is an important health outcome in every stage of life. In adult population, maintaining adequate levels of body composition, such as reduced body fat or increased muscle and bone mass, are associated with a greater health prognosis³⁵.

Health-related physical fitness: Definitions

Physical fitness is defined as “*a set of attributes or characteristics that people have or achieve that relates to the ability to perform physical activity*”. More specifically, health-related physical fitness is defined as “*those specific components of physical fitness that have a relationship with good health*”³⁶.

The American College of Sport Medicine (ACSM) compiles the components of health-related physical fitness as following: cardiorespiratory fitness, musculoskeletal fitness (related to muscular fitness and flexibility), motor fitness, and body composition³⁶.

Cardiorespiratory fitness

Cardiorespiratory fitness refers to the ability of the circulatory and the respiratory systems to supply oxygen during sustained physical activity³⁶.

Musculoskeletal fitness

Musculoskeletal fitness relates to the ability of the muscle to exert force³⁷.

Flexibility is that component of physical fitness which refers to the ability to move a joint through its full range of motion (ROM) with ease³⁸.

Motor fitness

Motor fitness is considered as the performance aspect of physical fitness in daily activities which requires quick reactions, speed of movement, agility, coordination and balance³⁹.

Body composition

Body composition refers to the relative amount or percentage of body tissues (bone content, fat mass, muscular mass), that are related to health. The most common health-related measure is the total body fat percentage (%BF)³⁶.

Physical fitness assessment

Laboratory testing is an objective and accurate method of assessing physical fitness, through reference criteria methods or *gold standard*. However, due to the cost of sophisticated instruments, time constraints and the need for qualified technicians, laboratory testing is limited to sport clubs, schools, population-based studies, and offices or clinical settings. Therefore, field-based physical fitness testing can offer useful and practical alternatives as screening tools, since they are relatively safe and time-efficient, involve minimal equipment and low cost, and can be easily administered to multiple people simultaneously^{40, 41}.

A battery test to assess physical fitness related to health must meet a series of fundamental requirements. The tests that compose it must⁴²: (i) be related to the present and future health of the person, as demonstrated by the previously scientific evidence, and (ii) be valid, reliable, feasible, safe and easy to perform, both in context clinical and epidemiological studies.

Criterion-related validity is defined as the degree to which the measurements of a test are related to a reference test or gold standard. Predictive validity determines whether or not a test predicts selected outcomes⁴³.

Reliability refers to the reproducibility of values in repeated trials on the same individual⁴⁴.

Physical fitness tests batteries

Physical fitness tests batteries in children, adolescents and elderly

Different physical fitness batteries have been proposed to assess health-related physical fitness in preschoolers, children, adolescents, and the elderly (i.e., the PREFIT battery⁴⁵, the ALPHA battery⁴⁶, and the Senior Fitness Test battery⁴⁷), based on their validity (i.e., predictive validity and criterion-related validity) and reliability.

Physical fitness tests batteries in adults

Currently, there are numerous test batteries that assess physical fitness in adults, such as the *UKK Health-Related Fitness Test Battery for Adults*⁴⁸, *The Canadian Physical Activity, Fitness and Lifestyle Appraisal (CPAFLA)*⁴⁹, the *AFISAL-INEFC*⁵⁰, and the *Eurofit Fitness Test Battery*⁵¹; however, there is no scientific evidence that they are valid, reliable, safe, applicable and responsiveness to be used in a consensual manner, in addition to not presenting the scalability of the tests. A field-based physical fitness test included in one of these batteries may fulfill all or some of these characteristics, but not the complete battery⁴⁶. In addition, analyzing these batteries we found that different tests evaluate the same component of physical fitness.

To the best of our knowledge, there is no known battery of field-based physical fitness test that meets all these aforementioned characteristics for the adult population.

B. CHAPTER-SPECIFIC INTRODUCTION

The present International Doctoral Thesis is composed of six studies, grouped into four chapters:

Chapter I. Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study I*).

Chapter II. Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

Chapter III. Reliability of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study IV*).

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

Study VI: The standing long jump test.

The specific introduction and basis of each chapter and study are shown below:

Chapter I. Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study I*).

Assessment of motor fitness and flexibility could help to identify adults and older adults who may be at risk of suffering from health outcomes (such as impaired musculoskeletal health, mental illness, QoL or premature death). Through its predictive value, motor fitness and flexibility could be postulated as a health marker in these populations. Nevertheless, little prior scientific literature has systematically addressed this issue^{31, 52-56}, and there are some limitations and gaps that still need to be investigated thoroughly, since they only were based on older adults, gait speed was the predominant motor fitness test evaluated, and other important health outcomes (such as cause-specific mortality, hip fracture, or mental health and well-being) have not been deeply investigated. Furthermore, no previous systematic review has presented an overview of the different wide range of motor fitness and flexibility tests used in standard health practice (i.e., balance assessment, multidimensional measures, or flexibility assessment, in addition to gait speed assessment). Therefore, to comprehensively analyze the **predictive validity** of the existing motor fitness and flexibility tests in relation to several health outcomes in adults and older adults is needed.

Chapter II. Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

Since the early interest in physical fitness testing in the 1950–1960s, many field-based physical fitness tests have been proposed⁵⁷. It would be desirable to summarize the criterion-related validity of the existing field-based physical fitness tests in adults. There have been attempts to summarize the criterion-related validity of a certain test^{58, 59} or several tests with a common characteristic^{60–62}; however, no attempts have been made to summarize the criterion-related validity of all the existing field-based physical fitness tests in adults. Therefore, it is necessary to comprehensively study the **criterion-related validity** of the existing field-based physical fitness tests used in adults.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

Although *in vivo* body composition may be measured by “gold standard” such as air displacement plethysmography (ADP), dual-energy x-ray absorptiometry (DXA), deuterium oxide (D₂O), magnetic resonance imaging (MRI), or underwater weighing (UWW), as a criterion method⁶³, these devices are limited to certain environments (such as clinical and laboratory setting), expensive, require specialized technicians and take much time to be used. Therefore, it is necessary to provide the most valid techniques, as well as affordable devices and ease to use, for body fat assessment/distribution. Simple and accessible anthropometry assessments, also known as field-based methods [(i.e., weight and height, body circumferences, or skinfolds (SKF) thickness] are frequently used to determine body fat distribution not only in epidemiological research⁶³ but also in health-related environments, such as sport or nutrition fields.

Likewise, body mass index (BMI; weight / height squared) is also widely used as it is determinant of unfavorable health consequences^{64, 65}. The advantage of BMI is in its ease of calculation, requiring only the metrics of weight and height. However, BMI cannot distinguish between fat and lean mass, and reliance on measurements of BMI alone is still under discussion^{64, 66}. For instance, athletes may have BMI values similar to those of an individual with obesity⁶⁷. Another limitation is that different cutoffs are established for Caucasian/White, African-American/Black, Hispanic, and the Asian and South Asian populations⁶⁷ since it underestimates the obesity risk in the latter. Finally, BMI seems to be also sex-dependent, showing a higher %BF in females than in males with the same BMI^{68, 69}. Therefore, new indexes, such as body adiposity index [BAI; hip circumference / (height)^{1.5}-18]⁷⁰, have emerged to address these limitations.

Similarly, body circumferences, such as waist circumference (WC), hip circumference (HC), waist-hip circumference ratio (WHR; waist circumference / hip circumference), and waist-height ratio (WHtR; waist circumference / height), are simple methods to indicate abdominal adiposity and are widely associated with adverse health risks, such as cardiovascular disease or mortality in adults⁶⁶. Finally, SKF have the advantage of providing localized information about the thickness of subcutaneous fat tissue. Therefore, different regression equations, based on SKF, were developed to calculate body fat mass and/or %BF⁶³. Moreover, a great amount of estimation equations (based on others anthropometry measurements) have been also proposed. Since the 70s, Durnin and Womersley generalized equations⁷¹, Jackson and Pollock⁷² (in male adults), and Jackson Pollock and Ward⁷³ (in female adults) equations have been widely used to estimate %BF, for their simplicity of calculation. However, a universal equation for predicting body composition cannot simply explain the variation of body types and shapes around the world (i.e., different race-ethnicities and weight statuses). Additionally, it is still under discussion which anthropometric measurements or estimation equations are the most valid to calculate body fat mass and/or BF%⁶³.

Assessing body composition may be difficult in large epidemiological studies with limited access to advanced assessment methods. Therefore, it is of particular interest to practitioners to balance the practicality and feasibility of performing body composition measurements and equations in a non-laboratory setting.

Chapter III. Reliability of Field-Based Physical Fitness Test in Adults: A (*study IV*).

A reliable test is useful for clinical and research purposes as it offers similar results when performed on two or more occasions under the same conditions in the same individual. The reliability of a large number of field-based physical fitness tests has been analyzed in adult populations with different characteristics (i.e., age range, physical fitness level, or health condition). It would be desirable to summarize the reliability determined so far of all existing field-based physical fitness tests and establish their current level of supporting evidence. The assessment of reliability is only a first step in the recommendation process for a complete tool to assess or predict health status through physical fitness⁴⁴. There is a clear need to better inform researchers, clinicians, and health personal trainers of the best physical fitness tests to be used in adults. Accordingly, to

provide an evidence-based proposal of the **reliability** of field-based physical fitness tests is needed.

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

The 2-km walk test and 20-m shuttle run test (20-m SRT) are the most used test to evaluate cardiorespiratory fitness in adults. The 2-km walk test (using Oja's equation) and the 20-m SRT (using Leger's equation), are valid for estimating cardiorespiratory fitness (through VO_{2max} calculation) in adult population (derived from our systematic review, *study II*). Moreover, the 2-km walk test has a high test-retest reliability in adults aged 30-64 years and 20-m SRT has a high test-retest reliability in adults aged 18-64 years (derived from our systematic review, *study IV*).

Nevertheless, most of the studies that analyzed the validity of these two field-based tests to assess cardiorespiratory fitness in adults had a small sample, in which there was no balanced representation of age and sex. In addition, these studies have not taken into account the level of physical activity of the participants; and there are no studies that have evaluated simultaneously the criterion-related validity and reliability of the 2-km walk test and the 20-m SRT in the same sample.

Therefore, to analyze the **criterion-related validity** and the **reliability** of the 2-km walk test and the 20-m SRT for evaluating the maximum oxygen consumption (VO_{2max}) in adult population, according to sex, age, and physical activity levels are needed.

Study VI: The standing long jump test.

Regarding lower-body explosive muscular fitness assessment, the standing long (broad) jump (SLJ) test is the most used, and is proposed in health-related fitness test batteries to assess lower-body explosive muscular fitness in preschoolers⁴⁵, children and adolescents⁴⁶ given its known criterion-related validity and reliability. However, to our knowledge (derived from our systematic review, *study II*), the criterion-related validity of this test has not been explored in adult population⁷⁴. Moreover, a recent systematic reviews (*study IV*) have concluded that the SLJ test has moderate to high reliability, only in adults aged 18-45 years⁷⁵. In addition, the sample of the two included studies were

young adults, and military and civilian adults. Therefore, a broader overview is needed, including the full age range of the adult population (i.e., 18-64 years), as well as whether physical activity levels could affect its reliability.

Due to the close relationship between lower-body muscular fitness and health in adults, it would be advisable to determine the **criterion-related validity** and **reliability** of the SLJ test in this population to use it as a health evaluation tool.

REFERENCES

1. Castillo Garzon MJ, Ortega Porcel FB, Ruiz Ruiz J. [Improvement of physical fitness as anti-aging intervention]. *Med Clin*. 2005;124(4):146-55.
2. Barry VW, Baruth M, Beets MW, Durstine JL, Liu J, Blair SN. Fitness vs. fatness on all-cause mortality: a meta-analysis. *Progress in cardiovascular diseases*. 2014;56(4):382-90.
3. Lin X, Zhang X, Guo J, Roberts CK, McKenzie S, Wu WC, et al. Effects of exercise training on cardiorespiratory fitness and biomarkers of cardiometabolic health: a systematic review and meta-analysis of randomized controlled trials. *Journal of the American heart association*. 2015;4(7):e002014.
4. García-Hermoso A, Cavero-Redondo I, Ramírez-Vélez R, Ruiz JR, Ortega FB, Lee D-C, et al. Muscular strength as a predictor of all-cause mortality in an apparently healthy population: a systematic review and meta-analysis of data from approximately 2 million men and women. *Archives of physical medicine and rehabilitation*. 2018;99(10):2100-13. e5.
5. LaMonte MJ, Barlow CE, Jurca R, Kampert JB, Church TS, Blair SN. Cardiorespiratory fitness is inversely associated with the incidence of metabolic syndrome: a prospective study of men and women. *Circulation*. 2005;112(4):505-12.
6. Fung MD, Canning KL, Mirdamadi P, Ardern CI, Kuk JL. Lifestyle and weight predictors of a healthy overweight profile over a 20-year follow-up. *Obesity* 2015;23(6):1320-5.
7. Howe TE, Rochester L, Neil F, Skelton DA, Ballinger C. Exercise for improving balance in older people. *Cochrane Database Syst Rev*. 2011(11):Cd004963.
8. Balducci S, Cardelli P, Pugliese L, D'Errico V, Haxhi J, Alessi E, et al. Volume-dependent effect of supervised exercise training on fatty liver and visceral adiposity index in subjects with type 2 diabetes The Italian Diabetes Exercise Study (IDES). *Diabetes Res Clin Pract*. 2015;109(2):355-63.
9. Pletnikoff PP, Laukkanen JA, Tuomainen TP, Kauhanen J, Rauramaa R, Ronkainen K, et al. Cardiorespiratory fitness, C-reactive protein and lung cancer risk: A prospective population-based cohort study. *Eur J Cancer*. 2015;51(11):1365-70.
10. Sui X, Lee DC, Matthews CE, Adams SA, Hebert JR, Church TS, et al. Influence of cardiorespiratory fitness on lung cancer mortality. *Med Sci Sports Exerc*. 2010;42(5):872-8.
11. Blair SN, Kohl HW, 3rd, Paffenbarger RS, Jr., Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA*. 1989;262(17):2395-401.
12. Farrell SW, Fitzgerald SJ, McAuley PA, Barlow CE. Cardiorespiratory fitness, adiposity, and all-cause mortality in women. *Med Sci Sports Exerc*. 2010;42(11):2006-12.
13. Ortega FB, Lavie CJ, Blair SN. Obesity and Cardiovascular Disease. *Circ Res*. 2016;118(11):1752-70.
14. Barry VW, Baruth M, Beets MW, Durstine JL, Liu J, Blair SN. Fitness vs. fatness on all-cause mortality: a meta-analysis. *Prog Cardiovasc Dis*. 2014;56(4):382-90.
15. McAuley PA, Beavers KM. Contribution of cardiorespiratory fitness to the obesity paradox. *Prog Cardiovasc Dis*. 2014;56(4):434-40.
16. Lavie CJ, McAuley PA, Church TS, Milani RV, Blair SN. Obesity and cardiovascular diseases: implications regarding fitness, fatness, and severity in the obesity paradox. *J Am Coll Cardiol*. 2014;63(14):1345-54.

17. Lavie CJ, Ozemek C, Carbone S, Katzmarzyk PT, Blair SN. Sedentary Behavior, Exercise, and Cardiovascular Health. *Circ Res*. 2019;124(5):799-815.
18. Barry VW, Caputo JL, Kang M. The Joint Association of Fitness and Fatness on Cardiovascular Disease Mortality: A Meta-Analysis. *Prog Cardiovasc Dis*. 2018;61(2):136-41.
19. Oktay AA, Lavie CJ, Kokkinos PF, Parto P, Pandey A, Ventura HO. The Interaction of Cardiorespiratory Fitness With Obesity and the Obesity Paradox in Cardiovascular Disease. *Prog Cardiovasc Dis*. 2017;60(1):30-44.
20. Delextrat AA, Warner S, Graham S, Neupert E. An 8-Week Exercise Intervention Based on Zumba Improves Aerobic Fitness and Psychological Well-Being in Healthy Women. *J Phys Act Health*. 2016;13(2):131-9.
21. Ortega FB, Lee DC, Sui X, Kubzansky LD, Ruiz JR, Baruth M, et al. Psychological well-being, cardiorespiratory fitness, and long-term survival. *Am J Prev Med*. 2010;39(5):440-8.
22. Boots EA, Schultz SA, Oh JM, Larson J, Edwards D, Cook D, et al. Cardiorespiratory fitness is associated with brain structure, cognition, and mood in a middle-aged cohort at risk for Alzheimer's disease. *Brain Imaging Behav*. 2015;9(3):639-49.
23. Willis BL, Gao A, Leonard D, Defina LF, Berry JD. Midlife fitness and the development of chronic conditions in later life. *Arch Intern Med*. 2012;172(17):1333-40.
24. Zhu N, Jacobs DR, Jr., Schreiner PJ, Launer LJ, Whitmer RA, Sidney S, et al. Cardiorespiratory fitness and brain volume and white matter integrity: The CARDIA Study. *Neurology*. 2015;84(23):2347-53.
25. Garcia-Hermoso A, Caverio-Redondo I, Ramirez-Velez R, Ruiz JR, Ortega FB, Lee DC, et al. Muscular Strength as a Predictor of All-Cause Mortality in an Apparently Healthy Population: A Systematic Review and Meta-Analysis of Data From Approximately 2 Million Men and Women. *Arch Phys Med Rehabil*. 2018;99(10):2100-13.e5.
26. Garcia-Hermoso A, Ramirez-Velez R, Peterson MD, Lobelo F, Caverio-Redondo I, Correa-Bautista JE, et al. Handgrip and knee extension strength as predictors of cancer mortality: A systematic review and meta-analysis. *Scand J Med Sci Sports*. 2018;28(8):1852-8.
27. Kettunen O, Kyrolainen H, Santtila M, Vasankari T. Physical fitness and volume of leisure time physical activity relate with low stress and high mental resources in young men. *J Sports Med Phys Fitness*. 2014;54(4):545-51.
28. Beard JR, Officer A, De Carvalho IA, Sadana R, Pot AM, Michel J-P, et al. The World report on ageing and health: a policy framework for healthy ageing. *The Lancet*. 2016;387(10033):2145-54.
29. Andrasfay T. Changes in Physical Functioning as Short-Term Predictors of Mortality. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*. 2020;75(3):630-9.
30. Briggs AM, Cross MJ, Hoy DG, Sanchez-Riera L, Blyth FM, Woolf AD, et al. Musculoskeletal health conditions represent a global threat to healthy aging: a report for the 2015 World Health Organization world report on ageing and health. *The Gerontologist*. 2016;56(suppl_2):S243-S55.
31. Van Kan GA, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. *Journal of Nutrition Health & Aging*. 2009;13(10):881-9.

32. Cooper R, Strand BH, Hardy R, Patel KV, Kuh D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. *BMJ (Clinical research ed)*. 2014;348:2219.
33. Kelly T, Yang W, Chen C-S, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *International journal of obesity*. 2008;32(9):1431-7.
34. Meldrum DR, Morris MA, Gambone JC. Obesity pandemic: causes, consequences, and solutions—but do we have the will? *Fertility and sterility*. 2017;107(4):833-9.
35. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. *The Lancet*. 2017;390(10113):2627-42.
36. Medicine ACoS. ACSM's health-related physical fitness assessment manual: Lippincott Williams & Wilkins; 2013.
37. Corbin CB, Pangrazi RP, Franks BD. Definitions: Health, fitness, and physical activity. *President's Council on Physical Fitness and Sports Research Digest*. 2000.
38. Bandy WD, Sanders B. Therapeutic exercise: techniques for intervention: Lippincott Williams & Wilkins; 2001.
39. Goodway JD, Ozmun JC, Gallahue DL. Understanding motor development: Infants, children, adolescents, adults: Jones & Bartlett Learning; 2019.
40. España-Romero V, Artero E, Jimenez-Pavón D, Cuenca-Garcia M, Ortega F, Castro-Piñero J, et al. Assessing health-related fitness tests in the school setting: reliability, feasibility and safety; the ALPHA Study. *International journal of sports medicine*. 2010;31(07):490-7.
41. Castro-Piñero J, Artero EG, España-Romero V, Ortega FB, Sjöström M, Suni J, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *British journal of sports medicine*. 2010;44(13):934-43.
42. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. *Sports medicine*. 2008;38(4):297-316.
43. Docherty D. Field tests and test batteries. *Measurement in pediatric exercise science*. 1996:285-334.
44. Hopkins WG. Measures of reliability in sports medicine and science. *Sports medicine*. 2000;30(1):1-15.
45. Ortega FB, Cadenas-Sánchez C, Sánchez-Delgado G, Mora-González J, Martínez-Téllez B, Artero EG, et al. Systematic review and proposal of a field-based physical fitness-test battery in preschool children: the PREFIT battery. *Sports medicine*. 2015;45(4):533-55.
46. Ruiz JR, Castro-Piñero J, España-Romero V, Artero EG, Ortega FB, Cuenca MM, et al. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *British journal of sports medicine*. 2011;45(6):518-24.
47. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. *The Gerontologist*. 2013;53(2):255-67.
48. Suni JH, Oja P, Laukkanen RT, Mülunpalo SI, Pasanen ME, Vuori IM, et al. Health-related fitness test battery for adults: aspects of reliability. *Archives of physical medicine and rehabilitation*. 1996;77(4):399-405.

49. Canadian C. The Canadian Physical Activity, Fitness & Lifestyle Approach: CSEP–Health-Related Appraisal & Counseling Strategy. *Ottawa-ON: Health Canada*. 2003.
50. Rodríguez FA, Gusi N, Valenzuela A, Nàcher S, Nogués J, Marina M. Valoració de la condició física saludable en els adults (I): Antecedents i protocols de la bateria AFISAL-INEFC. *Apunts Educació física i esports*. 1998;2(52):54-77.
51. Oja P, Tuxworth B. Eurofit for adults: Assessment of health-related fitness: Council of Europe; 1995.
52. Cavanaugh EJ, Richardson J, McCallum CA, Wilhelm M. The Predictive Validity of Physical Performance Measures in Determining Markers of Preclinical Disability in Community-Dwelling Middle-Aged and Older Adults: A Systematic Review. *Physical Therapy*. 2018;98(12):1010-21.
53. Quan MH, Xun PC, Chen C, Wen J, Wang YY, Wang R, et al. Walking Pace and the Risk of Cognitive Decline and Dementia in Elderly Populations: A Meta-analysis of Prospective Cohort Studies. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2017;72(2):266-70.
54. Peel NM, Alapatt LJ, Jones LV, Hubbard RE. The Association Between Gait Speed and Cognitive Status in Community-Dwelling Older People: A Systematic Review and Meta-analysis. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2019;74(6):943-8.
55. Grande G, Triolo F, Nuara A, Welmer AK, Fratiglioni L, Vetrano DL. Measuring gait speed to better identify prodromal dementia. *Experimental Gerontology*. 2019;124.
56. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *Journal of Cachexia Sarcopenia and Muscle*. 2020;11(1):3-25.
57. Castro-Pinero J, Artero EG, Espana-Romero V, Ortega FB, Sjostrom M, Suni J, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med*. 2009;44(13):934-43.
58. Mayorga-Vega D, Aguilar-Soto P, Viciano J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med*. 2015;14(3):536-47.
59. Mayorga-Vega D, Viciano J, Cocca A, Merino-Marban R. Criterion-related validity of toe-touch test for estimating hamstring extensibility: A metaanalysis. *J Hum Sport Exerc*. 2014;9(1):188-200.
60. Mayorga-Vega D, Bocanegra-Parrilla R, Ornelas M, Viciano J. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: a systematic review and Meta-analysis. *PLoS One*. 2016;11(3):e0151671-e.
61. Mayorga-Vega D, Merino-Marban R, Viciano J. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: a meta-analysis. *J Sports Sci Med*. 2014;13(1):1-14.
62. Bennett H, Parfitt G, Davison K, Eston R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. *Sports Med*. 2016;46(5):737-50.
63. Fosbøl MØ, Zerahn B. Contemporary methods of body composition measurement. *Clinical physiology and functional imaging*. 2015;35(2):81-97.
64. Fedewa MV, Nickerson BS, Esco MR. Associations of body adiposity index, waist circumference, and body mass index in young adults. *Clinical Nutrition*. 2019;38(2):715-20.
65. Ortega FB, Sui X, Lavie CJ, Blair SN, editors. Body mass index, the most widely used but also widely criticized index: would a criterion standard measure of total body

fat be a better predictor of cardiovascular disease mortality? Mayo Clinic Proceedings; 2016: Elsevier.

66. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nature Reviews Endocrinology*. 2020;16(3):177-89.

67. Weir CB, Jan A. BMI classification percentile and cut off points. 2019.

68. Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao D, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *International journal of obesity*. 2002;26(6):789-96.

69. Pasco JA, Nicholson GC, Brennan SL, Kotowicz MA. Prevalence of obesity and the relationship between the body mass index and body fat: cross-sectional, population-based data. *PloS one*. 2012;7(1):e29580.

70. Bergman RN, Stefanovski D, Buchanan TA, Sumner AE, Reynolds JC, Sebring NG, et al. A better index of body adiposity. *Obesity (Silver Spring)*. 2011;19(5):1083-9.

71. Durnin JV, Womersley J. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *British journal of nutrition*. 1974;32(1):77-97.

72. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. *British journal of nutrition*. 1978;40(3):497-504.

73. Jackson AS, Pollock ML, Ward A. Generalized equations for predicting body density of women. *Medicine and science in sports and exercise*. 1980;12(3):175-81.

74. Castro-Piñero J, Marin-Jimenez N, Fernandez-Santos JR, Martin-Acosta F, Segura-Jimenez V, Izquierdo-Gomez R, et al. Criterion-Related Validity of Field-Based Fitness Tests in Adults: A Systematic Review. *Journal of clinical medicine*. 2021;10(16):3743.

75. Cuenca-Garcia M, Marin-Jimenez N, Perez-Bey A, Sánchez-Oliva D, Camiletti-Moiron D, Alvarez-Gallardo IC, et al. Reliability of Field-Based Fitness Tests in Adults: A Systematic Review. *Sports Medicine*. 2022:1-20.

HYPOTHESIS



HYPOTHESIS

The starting hypotheses of this Doctoral Thesis are listed below:

I. There are different levels of scientific evidence on the predictive validity, criterion-related validity and reliability of each of field-based physical fitness tests analyzed in the adult population.

II. The field-based physical fitness test evaluated, based on scientific evidence, to assess the different components of physical fitness related to health in adults are valid and/or reliable; and sex, age, and physical activity levels may play a role in it.

AIMS
[OBJETIVOS]

The overall aim of the present International Doctoral Thesis was: to propose a battery of field-based tests to assess physical fitness related to health in adults based on scientific evidence, valid (i.e. predictive validity and criterion-related validity) and reliable, depending on sex, age, and physical activity level: The ADULT-FIT project [Plan nacional I+D +i 2017-2020, (ref.: DEP2017-88043-R)].

The specific aims of the present International Doctoral Thesis are the following:

Chapter I. Predictive Validity of Field-Based Physical Fitness Physical Test in Adults: A Systematic Review (*study I*).

To comprehensively analyze the **predictive validity** of the existing motor fitness and flexibility tests in relation to several health outcomes in adults and older adults.

Chapter II. Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

To comprehensively study the **criterion-related validity** of the existing field-based physical fitness tests used in adults aged 18-64 years.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

To identify studies evaluating the **criterion-related validity** of existing field-based methods and equations for body composition estimation used in adults aged 18-64 years.

Chapter III. Reliability of Field-Based Physical Fitness Test in Adults: A (*study IV*).

To systematically review studies conducted to examine the **reliability** of field-based physical fitness tests used in adults aged 18-64 years.

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

To analyze the **criterion-related validity** and the **reliability** of the 2-km walk test and the 20-m SRT for evaluating cardiorespiratory fitness in the adult population, according to sex, age, and physical activity level.

Study VI: The standing long jump test.

To analyze the **criterion-related validity** and **reliability** of the SLJ test to assess the lower-body explosive muscular fitness in adults, according to sex, age, and physical activity levels.

OBJETIVOS

El objetivo general de la presente Tesis Doctoral Internacional fue: proponer una batería de test de campo para evaluar la condición física relacionada con la salud en adultos basada en la evidencia científica, válida (i.e. predictiva y de criterio) y fiable, en función del sexo, la edad y el nivel de actividad física: Proyecto ADULT-FIT [Plan nacional I+D +i 2017-2020, (ref.: DEP2017-88043-R)].

Los objetivos específicos de la presente Tesis Doctoral Internacional fueron:

Capítulo I. Validez Predictiva de Test de Campo en Adultos: Una Revisión Sistemática (*estudio I*).

Analizar la **validez predictiva**, de los test de campo existentes de capacidad motora y flexibilidad relacionados con diversos resultados de salud en adultos y adultos mayores.

Capítulo II. Validez de Criterio de Test de Campo en Adultos, y Métodos y Ecuaciones para estimar la Composición Corporal en Adultos: Revisiones Sistemáticas (*estudios II-III*).

Estudio II: Validez de Criterio de Test de Campo en Adultos.

Analizar la **validez de criterio** de los test de campo de condición física utilizados en adultos de 18-64 años.

Estudio III: Métodos y Ecuaciones para estimar la Composición Corporal en Adultos.

Identificar estudios que evalúen la **validez de criterio** de los métodos de campo y ecuaciones existentes para la estimación de la composición corporal utilizados en adultos de 18-64 años.

Capítulo III. Fiabilidad de Test de Campo en Adultos: Una Revisión Sistemática (*estudio IV*).

Revisar sistemáticamente los estudios realizados para examinar la **fiabilidad** de los test de campo de condición física usados en adultos de 18-64 años.

Capítulo IV. Estudios Metodológicos de Test de Campo en Adultos: Validez de Criterio y Fiabilidad (*estudios V-VI*).

Estudio V: Test de 2-km andando y test de 20-m ida y vuelta.

Analizar la **validez de criterio** y **fiabilidad** de los test de campo de 2-km andando y 20-m ida y vuelta para evaluar la capacidad cardiorrespiratoria en población adulta, según el sexo, la edad y el nivel de actividad física.

Estudio VI: Test de salto horizontal a pies juntos.

Analizar la **validez de criterio** y la **fiabilidad** del test de campo de salto horizontal a pies juntos para evaluar la fuerza explosiva muscular del tren inferior en adultos, según el sexo, la edad y el nivel de actividad física.

MATERIAL AND METHODS

MATERIAL AND METHODS

The present Doctoral Thesis is composed of six studies, four of them designed as systematic reviews, and two of them as methodological studies, derived from the ADULT-FIT project.

Participants

Studies I-IV included healthy adults aged 18-64 years. In addition, *study I* included older adults (>64 years old).

Studies I-IV included participants with a wide range of physical activity/fitness levels, weight status, and race-ethnicities.

Studies V and VI included participants from Cádiz (Spain), homogeneously distributed by sex (49.5% females and 50.5% males), age (18-34 years, 35-49 years, and 50-64 years), and physical activity levels (non-active and active). The flow chart of the participants is shown in **Figure 1**.

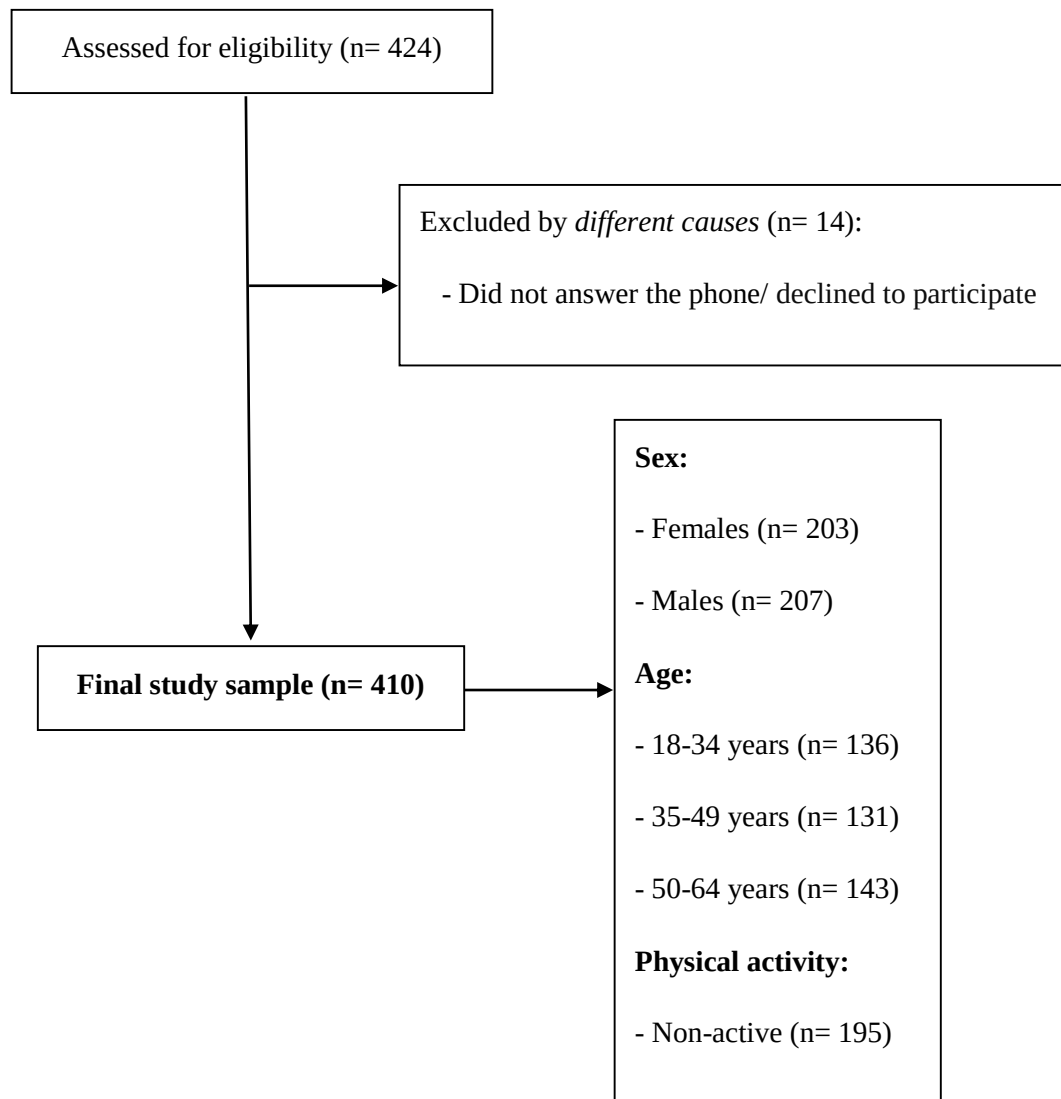


Figure 1. Flow-chart of the ADULT-FIT project.

In the studies included in the systematic reviews (*studies I-IV*), the sample size ranged from 13 to 6267 participants, and it was of 410 participants in *studies V* and *VI*.

The inclusion and exclusion criteria of the studies are shown in **Table 1**.

Table 1. Inclusion and exclusion criteria of the studies of the present Doctoral Thesis

Paper	Inclusion	Exclusion
I. Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review (<i>study I</i>)	(i) The study was an original full report published in a peer-reviewed journal, in English or Spanish; (ii) the study design was a longitudinal, prospective, retrospective study or prospective follow-up, or systematic review/meta-analysis (if there was more than one study on the same population dataset, the most recent study or the one with the most data best addressing the inclusion criteria was chosen); (iii) the study population was a healthy community based population (only those with cardiovascular risk factors were included), older than 18 years old; (iv) one or more motor fitness and flexibility tests were carried out; and (v) the outcome measure was one of the selected health outcomes.	(i) Studies examining the relationship between field-based motor fitness tests were excluded; (ii) studies that analysed the predictive validity of tests designed for exclusive use in sports or clinical settings were not included.
II. Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults: A Systematic Review (<i>study II</i>)	(i) Age criterion: adults (19–64 years old). During this review, we faced the problem that some studies sampled adults and older adults, or adults or adolescents together. In these cases, we observed whether these studies performed stratified analyses by age groups, isolating the adult population from the rest; if so, the study was included and information concerning the adult population reported. In contrast, when the authors analysed the whole sample together, we only included the study if the age of the sample was predominantly within our study age range; (ii) participants: the study population was based on a generally healthy population, who did not present any injury, physical and/or mental disabilities, irrespective of body mass index, diabetes or other cardiovascular risks (i.e., hypertension, hypercholesterolemia, lipid profiles, glucose levels, insulin sensitivity); and (iii) study design: original studies or systematic reviews/meta-analysis. The original studies that were selected for the	(i) Studies examining the relationship between field-based physical fitness tests were excluded; (ii) studies that analysed the criterion-related validity of tests designed for exclusive use in sports or clinical settings were not included.

analysis of their criterion-related validity but which were also included in the selected systematic reviews were excluded; (iv) language criterion: articles were only published in English or Spanish; (v) topic criterion: studies examining the criterion-related validity of the field-based physical fitness test.

III. Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults: A Systematic Review (*study III*)

(i) Age: adults (18-64 years old). During the review, we faced the problem that some studies sampled adults and older adults, or adults and adolescents together. In these cases, we observed whether these studies performed stratified analyses by age groups, isolating the adult population from the rest; if so, the study was included and information concerning the adult population was reported. In contrast, when the authors analyzed the whole sample together, we only included the study if the age of the sample was predominantly within our study age range; (ii) participants: the study population was based on a general healthy population, who did not present any injury, physical and/or mental disabilities, irrespective of BMI, diabetes or other cardiovascular risks (i.e., hypertension, hypercholesterolemia, lipid profiles, glucose levels, insulin sensitivity); (iii) study design: original studies; (iv) language: articles only published in English or Spanish and; (v) topic: studies examining the criterion-related validity of field-based body composition measures.

(i) Studies comparing field-based body composition measures with a non-gold standard criterion were excluded; (ii) studies that analyzed the criterion-related validity of tests designed for exclusive use in sports or clinical settings were not included.

IV. Reliability of Field-Based Physical Fitness Tests in Adults: A Systematic Review (*study IV*)

(i) Study design: original studies, reviews, and meta-analyses; (ii) Topic: studies examining the reliability of field-based physical fitness tests; (iii) Language: full reports published in English or Spanish; (iv) Age: adults aged 19–64 years. During the review, combined populations were observed (i.e., adults and older adults, adults and adolescents). In these cases, we noted whether these studies performed stratified analyzes by age groups, isolating the adult

(i) Studies conducted in professional athletes or designed for exclusive use in clinical settings were excluded.

population from the rest. If so, the study was included and information concerning the adult population reported. In contrast, when the authors analyzed the whole sample together, we only included a study if participant ages were predominantly within our established age range; (v) Participants: generally healthy adults with no significant health problems, diseases (e.g., heart failure, chronic obstructive pulmonary disease, cancer, or neurologic diseases) or conditions (e.g., cognitive impairment or intellectual disability, mobility problems or any injury).

V-VI. Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adult population: The Adult-Fit project / Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project (*studies V-VI*)

(i) Age: adults (18-64 years old); (ii) not having a physical or mental illness that prevents you from doing physical activity; (iii) intention to carry out all the tests that make up the study and; (iv) able to read and understand the informed consent as well as the object of the study.

(i) Acute or terminal illness; (ii) myocardial infarction three months before starting the study; (iii) unstable cardiovascular disease; (iv) medical prescription that prevents the performance of the tests; (v) injury or circumstance that makes it impossible to carry out the tests correctly.

BMI, Body Mass Index.

Procedures

Chapters I-III. Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study I*); Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults (*study II*), and Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults (*study III*); and Reliability of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study IV*).

Systematic reviews included in **Chapters I-III** (*studies I-IV*) were properly conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines¹. The specific review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO).

A systematic search was performed of the electronic databases MEDLINE via PubMed and Web of Science.

When using PubMed, we included Medical Subject Heading (MeSH) terms to enhance the power of the search. The same search strategy and combination of terms was repeated in Web of Science but without using MeSH terms or their equivalent as a similar option does not exist for this database. The specific search topics were combined with the Boolean operators ‘AND’ (inter topics) and ‘OR’ (intra topics).

The study selection process was conducted in three stages. In the first stage, records were identified through the PubMed and Web of Science databases and imported into MENDELEY software (version Desktop 1.19.4). Next, duplicate files were removed, firstly automatically by the software and secondly by visual checking. In the second stage, titles and abstracts of the search results were examined for eligibility. Finally, seemingly eligible full-text reports were read for their final inclusion or exclusion in the reviews. The outcomes of all stages were compared between two researchers. When there was no consensus between the two researchers (< 5%), a third researcher made the final decision about inclusion. Reasons for the exclusion of identified articles were recorded. Additionally, the reference lists of retrieved studies were examined to identify further articles.

The authors were not blinded to the articles selected, as the reviewers who performed the quality assessment were familiar with the literature.

Data extraction was independently double-checked by two researchers. The extracted data from the included studies were (i) study (i.e., first author’s name and year of publication); (ii) sample size used for the analysis and characteristics of the study

sample (i.e., sex, age, and health condition); (iii) physical/motor fitness test or body composition test/equation; (iv) study design; (v) statistical methods; and (vi) main outcome and conclusions.

The risk of bias was assessed by two independent researchers. Disagreements were resolved by consensus in a meeting with a third researcher.

Risks of bias of the original studies were determined according to a quality assessment list^{2, 3}. In *study I* this list included five items based on population, designs, methods, and report of the results. The items on the list were rated as “1” (positive), “0” (negative) or “?” (unclear) (see **Table 2**). For *study II* see **Table 3**; for *study III* see **Table 4**; and for *study IV* see **Table 5**. Likewise, for selected systematic reviews/meta-analysis was assessed using the assessment of multiple systematic reviews (AMSTAR) rating scale⁴.

To make final conclusions on the predictive validity of the field-based motor fitness and flexibility tests, and criterion-related validity/reliability of the field-based physical fitness or body composition tests/equations identified, three levels of evidence were established²: (i) strong evidence: consistent findings in three or more high-quality studies; (ii) moderate evidence: consistent findings in two high-quality studies; and (iii) limited evidence: consistent findings in multiple low-quality studies, inconsistent results in multiple high-quality studies, or results based on a single study. When strong or moderate evidence that a test was or was not valid/reliable, its level of validity/reliability was discussed.

Table 2. Quality assessment criteria for predictive validity studies (*study I*).

Grading system parameter	Grade	Criterion
Non-selective population selected with respect to physical health	1	yes
	0	no
	?	unclear
Selected health outcome clearly described	1	yes
	0	no
	?	unclear
The time between the measurement of physical fitness and the health outcome was at least one year	1	yes
	0	no
	?	unclear
Were the results adjusted for confounders?	1	yes
	0	no
	?	unclear
Standards errors and/or confidence intervals given for the estimates or information given to calculate it	1	yes
	0	no
	?	unclear
Rating for total score:		
High quality = 3-5		
Low quality = 2		
Very low quality = 1		

Table 3. Quality assessment criteria for physical fitness criterion-related validity studies (*study II*).

Grading system parameter	Grade	Criterion
Number of study subjects	0	$n \leq 10$
	1	$n = 11-50$
	2	$n \geq 51$
Description of the study population with respect to age, sex, health status, fitness levels, pubertal status, ethnicity, physical activity patterns, body composition, etc.	0	Less items than required for grade 1
	1	At least age, sex, health status, and fitness levels
	2	More items than required for grade 1
Statistical analysis included in the study	0	Those not included in 1
	1	Error indexes or regression analysis
	2	≥ 3 items or Bland-Altman plot and/or ANOVA for repeated measurements
Rating for total score:		
High quality = 5-6		
Low quality = 3-4		
Very low quality = 0-2		

Table 4. Quality assessment criteria for body composition criterion-related validity studies (*study III*).

Grading system parameter	Grade	Criterion
Number of study subjects	0	$n \leq 10$
	1	$n = 11-50$
	2	$n \geq 51$
Description of the study population with respect to age, sex, health status and body composition, menopausal status, ethnicity, physical activity patterns, fitness levels, etc.	0	Less items than required for grade 1
	1	At least age, sex, health status, and body composition
	2	More items than required for grade 1
Statistical analysis included in the study	0	Those not included in 1
	1	Error indexes or regression analysis
	2	≥ 3 items or Bland-Altman plot and/or ANOVA for repeated measurements
Rating for total score:		
High quality = 5-6		
Low quality = 3-4		
Very low quality = 0-2		

Table 5. Quality assessment criteria-related reliability studies (*study IV*).

Grading system parameter	Grade	Criterion
Description of the participants	0	Less items than required for grade 1
	1	At least age, sex, and fitness level
	2	Items required for grade 1 plus anthropometry or ethnicity
Description of time interval	0	Interval unknown
	1	Vague and imprecise information about interval
	2	Precise and complete description about interval
Description of the results	0	Less results presented than required for grade 1
	1	Description of test-retest results or description of the differences
	2	Description of test-retest results and description of the differences
Appropriateness of statistic	0	Only coefficient of variation
	1	Everything between grades 0 and 2 (normally – but not always – correlation plus an additional statistic)
	2	At least paired statistics, ANOVA for repeated measures (or non-parametrical corresponding tests) or Bland-Altman method
Rating for total score:		
High quality = 6–8		
Low quality = 2–5		
Very low quality = 0–1		

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability: The 2-km walk test and the 20-m shuttle run test (*study V*), and The standing long jump test (*study VI*).

Briefly, a total of 410 adults were recruited through leaflets, local newspapers, and social media from Cadiz (Spain). The total sample was homogeneously distributed by sex (49.5% females and 50.5% males), age (18-34 years, 35-49 years, and 50-64 years), and physical activity levels (50% non-active and 50% active). All interested volunteers provided written informed consent to participate in the present study. The study was approved by the Committee for Research of Cadiz, Spain.

After providing written informed consent and being informed of the protocol to be carried out, they signed the “Physical Activity Readiness Questionnaire” (PAR-Q) questionnaire⁵ to detect possible contraindications to the practice of physical exercise, a health status questionnaire regarding the Covid-19 disease, and a questionnaire to determine the physical activity levels of the participants⁶.

Participants were tested in three sessions during three weeks (one per week). In the first week, sociodemographic, anthropometric measurements, and a maximum treadmill test, laboratory lower-body muscular fitness tests (i.e., force platform and isokinetic dynamometer), and the SLJ test were carried out; in the second week the 2-km walk test, the 20-m SRT, and the retest SLJ were carried out; and, in the third week, participants performed the field-based cardiorespiratory fitness test again, in the same conditions as before. Furthermore, the SLJ was tested in a pre-test and post-test protocol with a three learning sessions carried out on alternate days for a week (*Monday, Wednesday and Friday*), to analyze whether the reliability of this test may depend on a previous learning period.

Before field-based testing sessions, all participants completed a standardized 10-minute warm-up. All the participants received comprehensive instructions for the tests and were encouraged to do their best in each test. Participants were instructed to rest 24 h before evaluations and to maintain their eating and hydration habits.

Measurements

Physical activity levels

Participants were initially classified as active/non-active when following/not following World Health Organization recommendations for adults⁷. The following self-

reported question was asked: *how many days (in a typical week) do you practice physical activity/exercise or some sport, of at least moderate intensity, lasting at least 50 minutes per day?*

Anthropometric assessment

Height, weight, triceps and subscapular SKF, and HC and WC were measured using the protocol described by the International Society for the Advancement of Kinanthropometry (ISAK)⁸. For the neck circumference (NC), the protocol established by the Center for Disease Control and Prevention was followed⁹. Measurements were always performed by the same trained evaluator (to avoid intra-evaluator variability), of the same sex as the participant.

All measurements were collected with bare feet, in light sports clothing, and with a 3-hour fast. Height was measured using a TANITA HR001 portable height rod (Tanita®, Illinois, USA; sensitivity, 1 mm). The margin of error that was established to make a third measurement was 1 cm. Weight was measured using an OMRON BF-400 electronic scale (Omron Healthcare Europe BV, Hoofddorp, The Netherlands; sensitivity, 100 gr). The established margin of error by which a third measurement should be made was a difference of 1 kg. The BMI was calculated as weight (kg) divided by squared height (m²).

Triceps and subscapular SKF were measured using the Harpenden Skinfold Caliper (Holtain, Dyfed, United Kingdom; range, 0-80 mm; sensitivity, 0.2 mm), and HC, WC and NC were assessed using the tape measure using SECA 201 (Seca Int, Hamburg, Germany; range, 0-205 cm; sensitivity, 0.1 cm). The margin of error for a third measurement was 1 millimeter for SKF and 1 cm for circumferences.

The %BF and lean mass (kg) were determined by bioimpedance Tanita MC 780-P MA (Tanita Co., Guangzhou, China), according to the protocol described by the National Institute of Health (NIH)¹⁰. For its correct evaluation, the participants were asked about their level of hydration¹¹.

Cardiorespiratory fitness (study V: The 2-km walk test and the 20-m shuttle run test)

Maximal treadmill test

Participants completed an incremental cardiopulmonary exercise test (CPET) on a treadmill (Lode Valiant, Groningen, Netherlands) for the determination of VO_{2max}

through indirect calorimetry (Jaeger MasterScreen CPX®, CareFusion, San Diego, USA). $\text{VO}_{2\text{max}}$ was recorded as absolute values ($\text{ml} \cdot \text{min}^{-1}$) and relative per kilograms of body weight ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$). Heart rate and peak respiratory exchange ratio (RER) were also collected.

Three different protocols, performed until volitional exhaustion, were used given the large heterogeneity of our study sample (i.e., physical activity/fitness level and age). The protocol was selected for each participant to reach their limit of tolerance in approximately 10-12 min. In all protocols, the test involved a walking warm-up, an incremental, and a recovery phase.

For the selection of the CPET protocol for each participant, the following criteria was used: (i) participants with a low physical activity level or who could not run on a treadmill, performed the Balke protocol, which maintained a constant speed of 4.8 km/h and increased the slope by 2.5% every two minutes¹². Those participants with no previous experience on a treadmill (both walking or running) underwent a brief familiarization period of 1-3 min, to feel confident enough with the test; (ii) participants with average physical activity/fitness warmed up at 4.8 km/h during 2 min, starting the test at 6 km/h and increasing by 1 km/h per minute; (iii) participants who presented a greater physical activity/fitness level, warmed up at 6 km/h during 2 min, starting the test at 8 km/h, increasing 1 km/h every minute. After exhaustion was attained, a 5 min recovery phase was performed in all the aforementioned protocols.

The main criteria used to determine the maximality of a test was the attainment of a VO_2 plateau. In our study we used the criteria of defining the plateau as a difference of ≤ 150 ml/min between consecutive stages¹³. To this aim, we compared the 2-final time-averaged periods of 30 seconds of the test^{14, 15}. Thus, when the difference in VO_2 between the final and the preceding 30-seconds period was ≤ 150 ml/min the plateau was established and the final VO_2 value was considered a real $\text{VO}_{2\text{max}}$.

When the plateau criteria was not met, at least 3 of the following secondary criteria had to be met to establish that the test was maximal¹⁶: (i) volitional exhaustion or the incapacity to maintain the treadmill speed despite verbal encouragement, (ii) heart rate within 10 bpm of the maximal age predicted, (iii) $\text{RER} \geq 1.1$, and (iv) rated perceived exertion (RPE) of ≥ 7 using the Borg scale from 0 to 10.

2-km walk test

The test consisted of walking 2-km at maximum speed in the shortest possible time¹⁷. The participants were instructed to walk at their maximum speed from the beginning to the end of the test in a 100-meter rectangular circuit. At the end of the test, the total time spent on the test (minutes and seconds) was recorded. Moreover, final heart rate was recorded using the activity bracelet, validated in adults, Xiaomi Mi Band 4 (Xiaomi Inc, Beijing, China)¹⁸. $\text{VO}_{2\text{max}}$ estimations were based on regression equations, established by Oja et al.¹⁷.

20-m SRT

The test is an incremental intermittent running test between two separate lines at a distance of 20 m¹⁹. The initial speed, marked by acoustic signals, is 8.5 km/h, increasing 0.5 km/h every minute; so, the time that the participants have to cover the distance of 20 m decreases over time. The test ends when the participants reach physical exhaustion or are unable to follow the set pace, or when they cannot cover the distance in the set time during two consecutive acoustic signals. Results were registered as fully completed stages. $\text{VO}_{2\text{max}}$ estimations were based on regression equations, established by Leger et al.¹⁹.

Lower-body muscular fitness (study VI: The standing long jump test)

Isokinetic test

Lower-body muscular fitness was analyzed using a HUMAC NORM isokinetic dynamometer (Cybex Isokinetic Dynamometer; Humac Norm, USA), being calibrated according to the operation manual.

The dominant and non-dominant leg were evaluated. Following the indications of other authors^{20, 21}, the participants were seated with a hip flexion of 110° with the trunk upright and leaning against the back of the chair, secured with safety belts at the level of the shoulders, abdomen and mid-thigh of the evaluated leg (to reduce awkward movements during contractions). The participant was fitted with a proximal shin pad 4-5 cm from the medial malleolus on the tibia of the tested leg, on which he/she exerted pressure²⁰. The axis of rotation of the dynamometer was aligned with the femoral condyle (axis of rotation of the knee joint).

At the beginning of the test, the participant's relative knee extension 180° was calculated, considering the absolute zero (point at which gravity acts vertically)²⁰ and the 90° flexion was calculated based on this measurement.

First, an isometric test of knee extension at 60° was developed in which the participant executed 3 maximal repetitions preceded by 3 warm-up repetitions²¹. The maximal isometric peak torque was recorded (Nm). Secondly, a 60 deg/s test was performed in which 3 maximum flexion-extensions were executed preceded by a warm-up of 5 progressive repetitions with the same gesture²². Maximal flexion and extension peak torque (Nm), and maximal flexion and extension peak torque, per repetition (Nm) were recorded.

A 2 min recovery period was always left between each test, and 1 min between warm-up and test. Verbal motivation was given throughout the test and participants were told that the effort should be maximal.

Standing Long Jump test

The SLJ is a simple test which assess the participant's ability to generate explosive strength. The participants performed the SLJ test twice, in two separate sessions. First time the SLJ test was tested in laboratory conditions using the force platform (criterion-related validity session), and the second one the SLJ test was carried out in field-based conditions (reliability session). The same procedure was carried out in both sessions. In both conditions, the participant stood behind the starting line with their feet apart equal to the width of their shoulders, and was instructed to push off vigorously and jump forward as far as possible. The participant had to land with the feet at the same time and to stay upright. The test was repeated twice, and the best score was used for analyses. A further attempt was allowed if the subjects fell backward or touched the floor with another part of the body³⁸.

For laboratory assessment, a software "Measurement, Analysis and Reporting 5.0" (MARS) was used, connected to a Kistler force platform (Type 9428A, Kistler Group; Winterthur, Switzerland). Maximal force (%BF), maximal horizontal and vertical force (%BF), maximal power (W/kg), maximal horizontal and vertical power (W/kg), jump take off velocity (m/s), horizontal and vertical jump take off velocity (m/s), push off force impulse (Ns), and horizontal and vertical push off force impulse (Ns) were recorded during the SLJ test³⁹. In both conditions (i.e., laboratory and field-based test

assessment) the distance was measured in centimeters (cm), from the take off line to the point where the back of the heel nearest to the take off line landed on the floor.

Statistical analyses

Study V: The 2-km walk test and the 20-m shuttle run test (study V)

Descriptive sample values, cardiorespiratory fitness tests are presented as mean $\pm$ SD. One-way analysis of variance (ANOVA) was performed to assess significant differences between age groups and t-test analysis for independent sample for sex and physical activity level groups.

Criterion-related validity

Bivariate correlations and simple linear regression were used to evaluate the agreement between cardiorespiratory fitness laboratory and field-based tests. When significant, the strength of the correlations was classified as follows: 0.00-0.25, very low; 0.26-0.49, low; 0.50-0.69, moderate; 0.70-0.89, high and; 0.90-1.00, very high²⁴.

Subsequently, the mean difference and the 95% limits of agreement [95% LoA (mean difference $\pm$ 1.96 SD of the difference)] were calculated using the Bland-Altman method^{25, 26} to analyze the agreement between measured and estimated $\text{VO}_{2\text{max}}$, whose difference was calculated using an ANOVA test for repeated measures. Where appropriate, Cohen d was computed to quantify the magnitude of the effect size. Cohen d values of 0.8, 0.5, and 0.2 represented large, medium, and small effect sizes, respectively²⁷.

Finally, in order to develop a more precise equation for the sample, stepwise linear regression model was used to estimated $\text{VO}_{2\text{max}}$ [relative $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$)]. To do this, the variables that presented a higher correlation were sequentially added, which were sex, age, physical activity level and %BF. Additional analyses were performed including BMI, or sum tricipital+subscapular SKF, WC, HC and lean mass instead of %BF.

Reliability

To investigate the reliability of the 2-km walk test and the 20-m SRT, we compared test and retest (hereafter called T1 and T2), through t-test and intraclass correlation coefficient (ICC). ICC is commonly used to describe *relative reliability* (i.e., the consistency of measurements on individuals in a group relative to others). An ICC

<0.8 were considered insufficient, values between 0.8 and 0.9 were considered moderate and values >0.9 were considered high²⁸.

Since any reliability study should not be based on a single statistic method, we also examined the differences between T1 and T2 using different error measures. Generally, the lower the error value, the lower the dispersion between T1 and T2 measurements.

The sum of squared errors (SSE) was calculated as follows:

$$SSE = \sum_{i=1}^N (y_i - \hat{y})^2$$

where n is the cases to evaluate the error measurements, $\hat{y}$ is the T2, and y is the T1.

The mean sum of squared errors (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

The root mean sum of squared errors (RMSE) was calculated by converting MSE into domain units by taking the root square:

$$RMSE = \sqrt{MSE}$$

The percentage error was calculated as follows:

$$\%Error = \frac{RMSE}{y_{max} - y_{min}} \times 100$$

The *absolute reliability* (consistency of repeated measurements for individuals) was analyzed by calculating the standard error of measurement (SEM) as percentage of the mean value of the measurements. The SEM quantifies the precision of individual scores in a test, and it is not influenced by variability among individuals (i.e., is considered a fixed characteristic of any measure, regardless of the sample of participants under investigation). A value $\leq 15\%$ is considered acceptable²⁹.

$\%SEM = \text{mean of the difference scores between 2 trials} \times 100 / \text{mean of the first trial}$.

The coefficient of variation (CV) as follow:

$$\%CV = \frac{\delta}{\bar{x}} \times 100$$

The CV method provides useful information in the presence of heteroscedasticity (assumes that greatest T1 and T2 variation occurs in individuals scoring the highest values in the test). A $CV \leq 10\%$ was considered as acceptable reliability³⁰.

The standard error of estimate was calculated as follows:

$$SEE = SD\hat{y}\sqrt{(1 - R^2)}$$

Finally, Bland-Altman plots were used to evaluate the reproducibility^{25, 26} of the field-based cardiorespiratory fitness tests, whose difference was calculated using an ANOVA test for repeated measures. Where appropriate, Cohen *d* was computed to quantify the magnitude of the difference between T1 and T2.

We also examined the difference and the magnitude of the measurement (i.e., heteroscedasticity) by conducting regression analysis.

We conducted the analyses for the whole sample, as well as separately by sex, age groups, and physical activity levels, for all criterion-related and reliability analysis.

All the analyses were performed using the Statistical Package for Social Sciences (IBM SPSS Statistics for Windows, version 26.0; Armonk, NY) and the level of significance was set at $p < 0.05$.

Study VI: The standing long jump test

Means and standard deviations were calculated for all the variables. One-way analysis of variance (ANOVA) was performed to assess significant differences between continuous variables and Chi-square test for categorical variables.

Criterion-related validity

Bivariate correlation was calculated to quantify the association between SLJ test and anthropometric variables and laboratory lower-body muscular tests. When significant, the strength of the correlations was classified as follows: 0.00-0.25, very low; 0.26-0.49, low; 0.50-0.69, moderate; 0.70-0.89, high and; 0.90-1.00, very high⁴⁰.

A stepwise linear regression model was conducted to examine the contribution of the different laboratory test variables to the explained variance of the SLJ distance. The SLJ distance was introduced as the dependent variable. Sex, age, physical activity levels, %BF and lean mass were introduced in the model, in addition to the laboratory variables that presented a higher correlation. Variables were excluded from the model when $p > 0.05$.

Reliability

Relative reliability of the field SLJ test was investigated through t-test and intraclass correlation coefficient (ICC). ICC ranges from 0 to 1, with values >0.9 being considered high reliability⁴¹.

We also examined the differences between test and retest (hereafter called T1 and T2) using different error measures. Generally, the lower the error value, the lower the dispersion between T1 and T2 measurements. The sum of squared errors (SSE) was calculated as follows: $SSE = \sum_{i=1}^N (y_i - \hat{y})^2$, where n is the cases to evaluate the error measurements, $\hat{y}$ is the T2, and y is the T1. The mean sum of squared errors (MSE): $MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$. The root mean sum of squared errors (RMSE) was calculated by converting MSE into domain units by taking the root square: $RMSE = \sqrt{MSE}$. The percentage error was calculated as follows: $\%Error = \frac{RMSE}{y_{max} - y_{min}} \times 100$.

The following methods were used to study absolute reliability: standard error of measurement (SEM) as percentage of the mean value of the measurements, standard error of estimate (SEE), the coefficient of variation (CV), and Bland-Altman plots^{42, 43}, whose difference was calculated using an ANOVA test for repeated measures. These measures were calculated as follows: $\%SEM = \text{mean of the difference scores between 2 trials} \times 100 / \text{mean of the first trial}$, with a value of $SEM \leq 15\%$ considered as acceptable⁴⁴; $SEE = SD \hat{y} \sqrt{(1 - R^2) y \hat{y}}$; $CV = \frac{\delta}{\bar{x}} \times 100$, with a $CV \leq 10\%$ considered as acceptable⁴⁵. In addition, the CV method provides useful information in the presence of heteroscedasticity (assumes that greatest T1 and T2 variation occurs in individuals scoring the highest values in the test). Heteroscedasticity was also identified by conducting regression analysis.

Finally, Cohen d was computed to quantify the magnitude of the difference between T1 and T2. Cohen d values of 0.8, 0.5, and 0.2 represented large, medium, and small effect sizes, respectively⁴⁶.

We conducted the analyses for the whole sample, as well as separately by sex, age groups, and physical activity levels, for all criterion-related and reliability analyses.

All the analyses were performed using the Statistical Package for Social Sciences (IBM SPSS Statistics for Windows, version 26.0; Armonk, NY) and the level of significance was set at $p < 0.05$.

The materials and methods section of the present Doctoral Thesis are summarized in **Table 6**. This table includes the most relevant methodological features from the scientific studies that compose the present Doctoral Thesis. For further information of any study, please check the material and methods sections of the corresponding paper.

Table 6. Summary of the main methodological features of the studies of the present Doctoral Thesis

Paper	Design	Participants		Measures/variables	Materials	Statistical analysis
		Number	Years			
I. Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review (<i>study I</i>)	Systematic review	100422 ($\approx$ 51717 females / $\approx$ 48705 males)	$\geq$ 18	- Motor fitness - Flexibility	N/A	N/A
II. Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults: A Systematic Review (<i>study II</i>)	Systematic review	10632 ($\approx$ 5539 females / $\approx$ 4722 males)	18-64	- Cardiorespiratory fitness - Muscular fitness - Motor fitness - Flexibility	N/A	N/A
III. Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults: A Systematic Review (<i>study III</i>)	Systematic review	103092 ($\approx$ 59161 females / $\approx$ 43931 males)	18-64	- Body fat mass - Body fat percentage	N/A	N/A
IV. Reliability of Field-Based Physical Fitness Tests in Adults: A Systematic Review (<i>study IV</i>)	Systematic review	10899 ($\approx$ 4971 females / $\approx$ 5067 males)	18-64	- Cardiorespiratory fitness - Muscular fitness - Motor fitness - Flexibility	N/A	N/A

V. Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adult population: The Adult-Fit project (<i>study V</i>)	Methodological study	410 adults (203 females / 207 males)	18-64	- Anthropometric assessment - Cardiorespiratory fitness	- TANITA HR001 portable height rod (Tanita®, Illinois, USA) - OMRON BF-400 electronic scale (Omron Healthcare Europe BV, Hoofddorp, The Netherlands) - Harpenden Skinfold Caliper (Holtain, Dyfed, United Kingdom) - Tape SECA 201 (Seca Int, Hamburg, Germany) - Tanita MC 780-P MA (Tanita Co., Guangzhou, China) - Treadmill (Lode Valiant, Groningen, Netherlands) - Indirect calorimetry (Jaeger MasterScreen CPX®, CareFusion, San Diego, USA) - Xiaomi Mi Band 4 (Xiaomi Inc, Beijing, China)	- Descriptive analysis (ANOVA and t-test) - Bivariate correlations - Simple linear regressions - Bland-Altman analysis (95% LoA) - multiple linear regression stepwise - Test-retest reliability (ICC) - Measurement errors (SSE, MSE, RMSE, %Error, SEM, CV, SEE) - Cohen <i>d</i> - Heteroscedasticity

VI. Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project (<i>study VI</i>)	Methodological study	Criterion-related validity and reliability: 410 adults (203 females / 207 males) Additional analyses on reliability (learning period): 66 adults (33 females / 33 males)	18-64	- Anthropometric assessment - Lower body muscular fitness	- TANITA HR001 portable height rod (Tanita®, Illinois, USA) - OMRON BF-400 electronic scale (Omron Healthcare Europe BV, Hoofddorp, The Netherlands) - Harpenden Skinfold Caliper (Holtain, Dyfed, United Kingdom) - Tape SECA 201 (Seca Int, Hamburg, Germany) - Tanita MC 780-P MA (Tanita Co., Guangzhou, China) - HUMAC NORM isokinetic dynamometer (Cybex Isokinetic Dynamometer; Humac Norm, USA) - Kistler force platform (Type 9428A, Kistler Group;	- Descriptive analysis (ANOVA and t-test) - Bivariate correlations - Simple linear regressions - Bland-Altman analysis (95% LoA) - multiple linear regression stepwise - Test-retest reliability (ICC) - Measurement errors (SSE, MSE, RMSE, %Error, SEM, CV, SEE) - Cohen <i>d</i> - Heteroscedasticity

Winterthur,
Switzerland)

ANOVA, Analysis of variance; CV, Coefficient of Variation; ICC, Intraclass Correlation Coefficient; LoA, Limits of Agreement; MSE, Mean sum of Squared Errors; N/A, Not Applicable; RMSE, Root Mean sum of Squared Errors; SEE, Standard Error of Estimate; SEM, Standard Error of Measurement; SSE, Sum of Squared Errors; %Error, Percentage error

REFERENCES

1. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery*. 2021;88:105906.
2. Castro-Piñero J, Artero EG, España-Romero V, Ortega FB, Sjöström M, Suni J, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *British journal of sports medicine*. 2010;44(13):934-43.
3. Artero E, España-Romero V, Castro-Piñero J, Ortega F, Suni J, Castillo-Garzon M, et al. Reliability of field-based fitness tests in youth. *International journal of sports medicine*. 2011;32(03):159-69.
4. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*. 2007;7(1):10.
5. Adams R. Revised Physical Activity Readiness Questionnaire. *Canadian Family Physician*. 1999;45:992.
6. Martínez-Gómez D, Martínez-De-Haro V, Del-Campo J, Zapatera B, Welk GJ, Villagra A, et al. Validez de cuatro cuestionarios para valorar la actividad física en adolescentes españoles. *Gaceta Sanitaria*. 2009;23(6):512-7.
7. Organization WH. Global recommendations on physical activity for health. *Global recommendations on physical activity for health*. 2010.
8. Marfell-Jones M, Olds T, Stewart A, Carter L. ISAK accreditation handbook. 2006.
9. Control CfD, Prevention. National health and nutrition examination survey: Anthropometry procedures manual. *Atlanta, GA, USA: CDC*. 2007.
10. Research NIOHOoMAo. Bioelectrical impedance analysis in body composition measurement: National Institutes of Health Technology Assessment Conference statement, December 12-14, 1994: NIH Office of Medical Applications of Research; 1994.
11. Armstrong LE, Soto JAH, Hacker FT, Casa DJ, Kavouras SA, Maresh CM. Urinary indices during dehydration, exercise, and rehydration. *International Journal of Sport Nutrition and Exercise Metabolism*. 1998;8(4):345-55.
12. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *United States Armed Forces Medical Journal*. 1959;10(6):675-88.
13. Taylor H, Henschel a, Buskirk E.(1955) Maximal Oxygen Make us ati Objective Cur&u-Respiratory Pefurmancel. *J Appl Physiol*. 1955:73-80.
14. Martin-Rincon M, González-Henríquez JJ, Losa-Reyna J, Perez-Suarez I, Ponce-González JG, de La Calle-Herrero J, et al. Impact of data averaging strategies on VO₂max assessment: Mathematical modeling and reliability. *Scandinavian journal of medicine & science in sports*. 2019;29(10):1473-88.
15. Gordon D, Wood M, Porter A, Vetrivel V, Gernigon M, Caddy O, et al. Influence of blood donation on the incidence of plateau at $\dot{V}_{\text{O}_2}$ max. *European Journal of Applied Physiology*. 2014;114(1):21-7.
16. Midgley AW, McNaughton LR, Polman R, Marchant D. Criteria for determination of maximal oxygen uptake. *Sports medicine*. 2007;37(12):1019-28.
17. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *International journal of sports medicine*. 1991;12(04):356-62.

18. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare informatics research*. 2015;21(4):315-20.
19. Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre tritium run test for aerobic fitness. *Journal of sports sciences*. 1988;6(2):93-101.
20. Francis P, Toomey C, Mc Cormack W, Lyons M, Jakeman P. Measurement of maximal isometric torque and muscle quality of the knee extensors and flexors in healthy 50-to 70-year-old women. *Clinical physiology and functional imaging*. 2017;37(4):448-55.
21. Harbo T, Brincks J, Andersen H. Maximal isokinetic and isometric muscle strength of major muscle groups related to age, body mass, height, and sex in 178 healthy subjects. *European journal of applied physiology*. 2012;112(1):267-75.
22. Croisier J-L, Ganteaume S, Binet J, Genty M, Ferret J-M. Strength imbalances and prevention of hamstring injury in professional soccer players: a prospective study. *The American journal of sports medicine*. 2008;36(8):1469-75.
23. Castro-Piñero J, González-Montesinos JL, Mora J, Keating XD, Girela-Rejón MJ, Sjöström M, et al. Percentile values for muscular strength field tests in children aged 6 to 17 years: influence of weight status. *The Journal of Strength & Conditioning Research*. 2009;23(8):2295-310.
24. Silva P, Franco J, GUSMÃO A, Moura J, Teixeira-Salmela L, Faria C. Trunk strength is associated with sit-to-stand performance in both stroke and healthy subjects. *Eur J Phys Rehabil Med*. 2015;51(6):717-24.
25. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*. 1986;327(8476):307-10.
26. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *The lancet*. 1995;346(8982):1085-7.
27. Cohen J. Statistical power analysis for the behavioral sciences: Routledge; 2013.
28. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Manual therapy*. 1999;4(2):87-93.
29. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research*. 2005;19:231-40.
30. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine*. 1998;26(4):217-38.

RESULTS



RESULTS

The results section of the present Doctoral Thesis is shown as a compilation of six scientific studies.

Chapter I. Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study I*).

Figure 2 presents the flow-chart of retrieved and selected studies and systematic reviews/meta-analysis. The electronic search strategy retrieved 1274 studies, from PubMed and Web of Science databases. Duplicate references were removed, resulting in 1102 studies screened. Of these, 929 records were excluded by not meeting title and abstract criteria, while 173 full-text studies were assessed for eligibility. Thereafter, 95 studies and two systematic reviews were further excluded for different reasons. The main reasons for rejection were: unable to source full text; no predictive validity; no language inclusion criteria; no healthy population; no fitness test assessed; and no complying with study designs criteria; original studies previously identified through selected systematic reviews/meta-analysis. Finally, 70 studies met the inclusion criteria and were included in the systematic review, six of them as systematic reviews/meta-analysis.

Risk of bias within studies

Table 7 shows the list of included longitudinal studies with quality scores. Overall, methodological quality was considered high, without any study scoring 2 or less, that is, with a low or very low quality. The quality assessment of included systematic reviews/meta-analysis is presented in **Table 8**. Five of them were classified as high quality¹⁻⁵, and the remaining one as medium quality⁶.

Characteristic of included studies

The overall number of participants in the selected studies⁷⁻⁷⁶ was 100422 (51.5% females), while in five studies, sex was not specified^{17, 26, 55, 56, 60}. The ages of the participants ranged from 35 to 105 years, with a mean age of 72. The length of follow-up ranged from 0.5 to 25 years (detailed characteristics and findings of the selected studies in **Tables 9–13**).

A total of three systematic reviews^{2,3,6}, and three meta-analyses^{1,4,5} were included. The overall sample size involved 274871 participants. Specific sex was not usually reported. The ages of the participants ranged from 54 to 108 years old. The follow-up period ranged from 11 days to 25 years (detailed characteristics and findings of the selected systematic reviews/meta-analysis in **Table 14**).

Tables 15 and 9–13 summarize the main characteristics, motor fitness and flexibility tests and health outcomes of the selected longitudinal studies and systematic

reviews/meta-analysis, as well as a broad picture of significant/non-significant associations found between motor fitness and flexibility tests with health outcomes.

Figure 3 presents the number of studies reporting predictive validity or lack of predictive validity of the included motor fitness and flexibility tests for different health outcomes. The highest number of predictive results were found for gait speed, followed by postural balance assessment and the multidimensional measurement test, Timed Up&Go test (TUG test).

A recap of the different cut-off points reported for gait speed and the TUG test and the risk of adverse health outcomes is shown in **Figures 4 and 5**.

The predictive validity of the different motor fitness and flexibility tests for diverse health outcomes are detailed in **Tables 9–13**.

Predictive validity for falls and fall-related outcomes

A total of 25 studies⁷⁷, and two systematic reviews^{2, 6} included falls or hip fracture as health outcomes (see **Table 9** and **Table 14**).

Gait speed tests

Gait speed was reported in 13 studies^{8, 20, 24, 27, 39-41, 45, 51, 58, 62, 66, 74}, and two systematic reviews^{2, 6}, the 6-m gait speed test being the most common.

There was strong evidence for slower gait predicting falls in adults over 60 years old, as a result of 7 selected studies^{20, 27, 45, 51, 58, 62, 66} plus seven studies derived from two systematic reviews^{2, 6}. Three studies did not find association between gait speed tests and falls^{8, 39, 41}. There was limited evidence for a short step length (<0.45m), assessed over 5m predicting risk of falling in older adults over 70 years⁴⁵.

There was moderate evidence indicating that slower gait speed predicts hip fracture in older females over 75 years^{24, 74}; and limited evidence, due to a low number of studies, in adults over 55 years old⁴⁰.

Postural balance tests

Balance was assessed in 13 studies^{12, 32, 33, 41, 45, 46, 49, 50, 54, 57, 68, 70, 74}, and a systematic review². Different protocols were used, such as the Berg Balance Scale^{32, 49}, the Tinetti Scale⁴⁵, Functional Reach Test (FRT)², tandem and semi-tandem stance or one-leg stance^{12, 41, 50, 70, 74}, with some of these tests performed in a force platform^{33, 46, 54, 57, 68}.

There was strong evidence reporting that impaired balance (worse performance or inability to complete the test) predicts falls in adults over 40 years old, as a result of 11

studies^{12, 32, 33, 41, 45, 46, 49, 54, 57, 68, 70}. A systematic review (with three studies)² and one study⁵⁰ did not find association between balance and falls.

There was limited evidence indicating that impaired balance predicts hip fracture in females over 75 years old⁷⁴.

Multidimensional measurement tests

The TUG is a multidimensional test that measures mobility skills, combining gait speed, balance and functional capacity⁷⁸.

The TUG was reported in 10 studies^{8, 11, 12, 22, 32, 39, 41, 45, 50, 54}, and a systematic review².

There was strong evidence for worse TUG performance predicting falls (and fear of falling) in adults over 40 years old, as a result of 9 studies^{8, 11, 12, 22, 32, 39, 41, 45, 54}. A systematic review (with three studies)² and a study did not find association between TUG and falls⁵⁰.

Flexibility tests

There was limited evidence, due to a limited number of studies (only one), reporting that a reduced hip and knee range of motion, were associated with falls in older adults over 70 years⁴⁵.

Predictive validity for cognitive decline and impairment

A total of 11 studies^{18, 19, 29, 30, 36, 56, 61, 62, 67, 69, 76}, two systematic reviews^{3, 6}, and two meta-analysis^{1, 4} included cognitive health outcomes (see **Table 10** and **Table 14**).

Gait speed tests

Gait speed was reported in eight studies^{19, 36, 56, 61, 62, 67, 69, 76}, and four included systematic reviews/meta-analysis^{1, 3, 4, 6}, where different distances were employed.

There was strong evidence for slower gait speed predicting cognitive decline and impairment (including develop of dementia and Alzheimer's disease) in adults over 55 years old^{1, 3, 4, 6, 19, 36, 56, 67, 69}. However, three selected studies found not association between gait speed tests and memory complaints⁶¹, or cognitive decline^{62, 76}.

Postural balance tests

One study reported balance assessment, determining limited evidence indicating that impaired standing balance predicts incident dementia in adults over 90 years old¹⁸.

Multidimensional measurements

Due to inconsistent results found in two high-quality studies^{29, 30}, there was limited evidence indicating that the TUG test predicts cognitive decline and impairment.

Predictive validity for depressive symptoms and well-being

A total of four studies^{17, 63, 71, 72} included depressive symptoms, and one study²⁵ included well-being as health outcomes (see **Table 11** and **Table 14**).

Gait speed tests

There was moderate evidence indicating that slower gait speed predicts depressive symptoms in adults over 50 years old^{17, 71}, and also in older males over 65 years^{63, 72}.

Multidimensional measurement tests

There was limited evidence indicating that worse TUG test performance predicts decline in well-being in older adults over 70 years²⁵.

Predictive validity for mobility limitations, disability and frailty

A total of 16 studies^{7, 9, 14, 16, 26, 28, 35, 38, 42, 47, 48, 52, 59, 64, 65, 73}, two systematic reviews^{2, 6}, and a meta-analysis⁵ included related-mobility disability and frailty as health outcomes (see **Table 12** and **Table 14**).

Gait speed tests

Gait speed was explored in 13 selected studies^{7, 9, 14, 16, 26, 28, 35, 38, 42, 47, 48, 52, 59}, and three reviews^{2, 5, 6}, where different distances were employed.

There was strong evidence for slower gait speed that predicts disability in instrumental activities of daily living (IADL) in adults over 54 years old, as a result of seven selected studies^{7, 9, 35, 42, 47, 48, 52}, and the three reviews (composed of 61 studies)^{2, 5, 6}.

There was strong evidence indicating that slower gait speed predicts mobility disability in adults over 50 years old, as a result of four selected studies^{14, 26, 28, 59}, and two systematic reviews^{2, 6} (composed of 10 and three studies, respectively).

There was limited evidence indicating that slower gait speed predicts frailty status in older adults over 65 years³⁸, and that faster gait speed predicts greater functional autonomy in older adults over 68 years, especially in females¹⁶.

Postural balance tests

Balance was assessed in three studies^{16, 47, 52}, a systematic review², and a meta-analysis⁵. Different protocols were used to assess balance.

There was strong evidence indicating that impaired balance predicts disability in IADL or mobility disability in adults over 40 years old^{2, 5, 52}. Makizako et al.⁴⁷ did not find balance predictive of disability in IADL in older adults over 68 years.

There was limited evidence indicating that balance could moderately predict functional decline in older adults over 68 years, especially in females¹⁶.

Multidimensional measurement tests

The TUG test was reported in two studies^{16, 64}, and two systematic reviews^{2, 5}.

Due to inconclusive results, limited evidence was found for TUG performance and disability in IADL or mobility disability in older adults. A meta-analysis⁵ found it predictive, but a systematic review² did not.

Due to a limited number of studies (only one per health outcome), there was limited evidence indicating that worse TUG performance identifies frailty status⁶⁴, and loss of functional autonomy in older adults over 65 years¹⁶.

Flexibility tests

There was limited evidence indicating that poor flexibility predicts mobility disability in older adults 65 years⁷³.

Other tests

Three studies included other tests^{9, 65, 73} to assess the predictive validity on mobility limitations, disability and frailty.

Due to a limited number of studies (only one per each test), limited evidence was found. Ward et al.⁷³ reported that speed of movement test predicts mobility disability in older adults over 65 years. Adachi et al.⁹ employed a maximum step length test to assess disability in IADL, and found that the test predicts disability in IADL, in older adults over 76 years. Finally, Schroll et al.⁶⁵ found that the stair mounting test predicts disability in IADL, in older males over 75 years.

Predictive validity for cardiovascular disease risk and mortality

A total of 19 studies^{7, 10, 13, 15, 21, 23, 31, 34-38, 43, 44, 53, 55, 60, 62, 75}, and a systematic review⁶ included CVD risk, stroke, all-cause mortality, and other-causes mortality as health outcomes (see **Table 13** and **Table 14**).

Gait speed tests

Gait speed was reported in 17 selected studies^{7, 10, 13, 21, 31, 34-38, 43, 44, 53, 55, 60, 62, 75}, and a systematic review⁶. The 2.4-m, 4-m, and 6-m gait speed were the most employed tests.

There was strong evidence indicating that slower gait speed predicted CVD risk in adults over 45 years old^{34, 35, 53, 55}. However, Lee et al.⁴³ found no interaction between gait speed and CVD in older adults.

There was strong evidence for slower gait speed that predicts all-cause mortality in adults over 35 years old^{6, 7, 10, 13, 31, 34, 36-38, 53, 55, 60, 62, 75}. Lee et al.⁴³ found not interaction between gait speed and all-cause mortality in older adults.

There was moderate evidence for slower gait speed that predicts other-causes mortality in adults over 50 years old^{44, 55}.

Regarding stroke, there was limited evidence indicating that slower gait speed was not associated with stroke risk in adults over 35 years old²¹.

Postural balance tests

Balance was reported in four studies^{13, 23, 37, 55}, where different protocols were used.

There was strong evidence indicating that impaired balance predicts all-cause mortality in adults over 53 years old^{13, 23, 37, 55}.

There was limited evidence indicating that impaired balance predicts CVD risk in older adults over 65 years⁵⁵. Moreover, there was limited evidence for impaired balance predicting other-causes mortality in older adults over 65 years⁵⁵.

Flexibility tests

There was limited evidence indicating that the toe-to-touch test does not predict all-cause mortality in adults over 60 years old¹⁵.

Predictive validity for institutionalization or hospitalization

One study⁷⁵, and two systematic reviews^{2, 6}, composed of two and five studies, respectively, included institutionalization or unplanned hospitalization as health outcomes (see **Table 13** and **Table 14**).

Gait speed tests

Gait speed was reported in the selected study⁷⁵, and the systematic reviews^{2, 6}, where the 6-m gait speed is the most common test.

There was strong evidence indicating that slower gait speed predicts risk of institutionalization or hospitalization in adults over 60 years old^{2, 6, 75}.

Postural balance tests

There was limited evidence indicating that impaired balance predicts risk of hospitalization in adults over 60 years old².

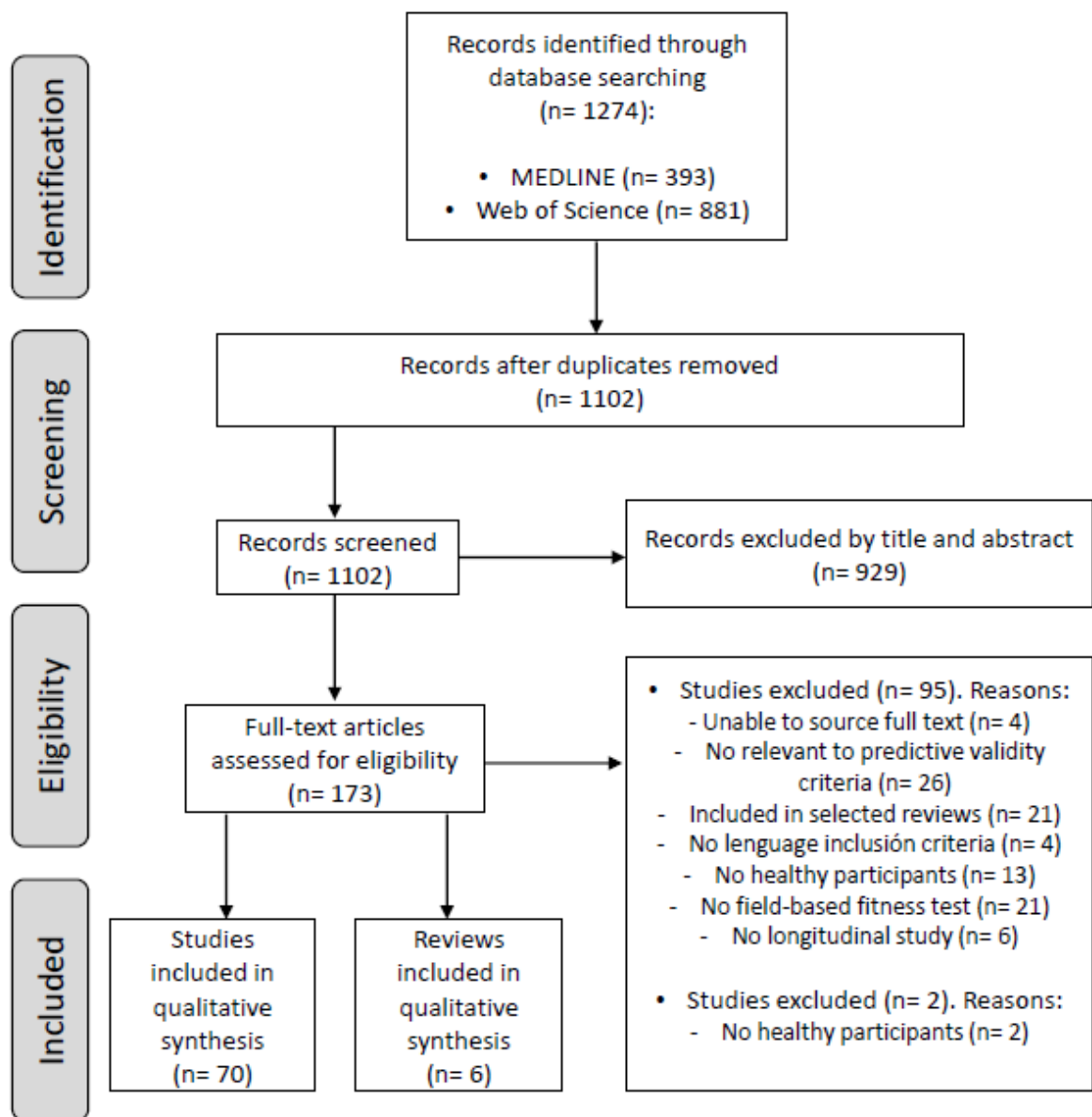


Figure 2. Flow-chart of the study selection process.

Table 7. List of included longitudinal studies with quality scores with reference to predictive value of motor fitness and flexibility tests for health outcomes in adults and older adults.

Study	Fitness Components	Non-Selective Population	Clear Health Outcomes	PF and Health Measurement >1 Year	Confounders	SE/CI Information	Total Score
Abu et al., 2018	Motor fitness (Speed)	1	1	0	0	1	3
Brach et al., 2012	Motor fitness (Speed)	1	1	0	0	1	3
Sakurai et al., 2017	Motor fitness (Speed)	1	1	0	0	1	3
Dargent-Molina et al., 1999	Motor fitness (Speed)	1	1	1	1	0	4
Doi et al., 2013	Motor fitness (Speed)	1	1	0	1	1	4
Kang et al., 2017	Motor fitness (Speed)	1	1	0	1	1	4
Kauppi et al., 2014	Motor fitness (Speed)	1	1	0	1	1	4
Laukkanen et al., 2000	Motor fitness (Speed)	1	1	1	0	1	4
Makizako et al., 2015	Motor fitness (Speed)	1	1	1	0	1	4
Abe et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Adachi et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Andrasfay et al., 2020	Motor fitness (Speed)	1	1	1	1	1	5
Blain et al., 2010	Motor fitness (Speed)	1	1	1	1	1	5
Briggs et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Buracchio et al., 2010	Motor fitness (Speed)	1	1	1	1	1	5
Callisaya et al., 2016	Motor fitness (Speed)	1	1	1	1	1	5
Camargo et al., 2016	Motor fitness (Speed)	1	1	1	1	1	5
Deshpande et al., 2013	Motor fitness (Speed)	1	1	1	1	1	5
Doi et al., 2020	Motor fitness (Speed)	1	1	1	1	1	5
Elbaz et al., 2013	Motor fitness (Speed)	1	1	1	1	1	5
Georgiopoulou et al., 2016	Motor fitness (Speed)	1	1	1	1	1	5
Heiland et al., 2018	Motor fitness (Speed)	1	1	1	1	1	5

Hoogendijk et al., 2020	Motor fitness (Speed)	1	1	1	1	1	5
Idland et al., 2013	Motor fitness (Speed)	1	1	1	1	1	5
Jung et al., 2018	Motor fitness (Speed)	1	1	1	1	1	5
Lee et al., 2017	Motor fitness (Speed)	1	1	1	1	1	5
Looker et al., 2015	Motor fitness (Speed)	1	1	1	1	1	5
Luukinen et al., 1995	Motor fitness (Speed)	1	1	1	1	1	5
Makizako et al., 2010	Motor fitness (Speed)	1	1	1	1	1	5
Muraki et al., 2013	Motor fitness (Speed)	1	1	1	1	1	5
Nakamoto et al., 2015	Motor fitness (Speed)	1	1	1	1	1	5
Niiranen et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Nofuji et al., 2016	Motor fitness (Speed)	1	1	1	1	1	5
Ojagbemi et al., 2015	Motor fitness (Speed)	1	1	1	1	1	5
Osuka et al., 2020	Motor fitness (Speed)	1	1	1	1	1	5
Quach et al., 2011	Motor fitness (Speed)	1	1	1	1	1	5
Rosso et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Sabia et al., 2014	Motor fitness (Speed)	1	1	1	1	1	5
Sanders et al., 2012	Motor fitness (Speed)	1	1	1	1	1	5
Sanders et al., 2016	Motor fitness (Speed)	1	1	1	1	1	5
Stenhagen et al., 2013	Motor fitness (Speed)	1	1	1	1	1	5
Stijntjes et al., 2017	Motor fitness (Speed)	1	1	1	1	1	5
Tian et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Veronese et al., 2017	Motor fitness (Speed)	1	1	1	1	1	5
Veronese et al., 2017b	Motor fitness (Speed)	1	1	1	1	1	5
Ward et al., 2016	Motor fitness (Speed)	1	1	1	1	1	5
Wihlborg et al., 2015	Motor fitness (Speed)	1	1	1	1	1	5
Zucchelli et al., 2019	Motor fitness (Speed)	1	1	1	1	1	5
Pajala et al., 2008	Motor fitness (Balance)	1	1	0	0	1	3
Austin et al., 2007	Motor fitness (Balance)	1	1	1	0	1	4

Ersoy et al., 2009	Motor fitness (Balance)	1	1	0	1	1	4
Frames et al., 2018	Motor fitness (Balance)	1	1	1	0	1	4
Maki et al., 1994	Motor fitness (Balance)	1	1	0	1	1	4
Muir et al., 2010	Motor fitness (Balance)	1	1	0	1	1	4
Mulasso et al., 2017	Motor fitness (Balance)	1	1	0	1	1	4
Swanenburg et al., 2010	Motor fitness (Balance)	1	1	0	1	1	4
Blain et al., 2010	Motor fitness (Balance)	1	1	1	1	1	5
Breton et al., 2014	Motor fitness (Balance)	1	1	1	1	1	5
Bullain et al., 2016	Motor fitness (Balance)	1	1	1	1	1	5
Cooper et al., 2014	Motor fitness (Balance)	1	1	1	1	1	5
Idland et al., 2013	Motor fitness (Balance)	1	1	1	1	1	5
Kwan et al., 2012	Motor fitness (Balance)	1	1	1	1	1	5
Luukinen et al., 1995	Motor fitness (Balance)	1	1	1	1	1	5
Makizako et al., 2010	Motor fitness (Balance)	1	1	1	1	1	5
Nakamoto et al., 2015	Motor fitness (Balance)	1	1	1	1	1	5
Nitz et al., 2013	Motor fitness (Balance)	1	1	1	1	1	5
Nofuji et al., 2016	Motor fitness (Balance)	1	1	1	1	1	5
Vellas et al., 1997	Motor fitness (Balance)	1	1	1	1	1	5
Wihlborg et al., 2015	Motor fitness (Balance)	1	1	1	1	1	5
Abu et al., 2018	Motor fitness (Speed–agility)	1	1	0	0	1	3
Sakurai et al., 2017	Motor fitness (Speed–agility)	1	1	0	0	1	3
Austin et al., 2007	Motor fitness (Speed–agility)	1	1	1	0	1	4
Clemson et al., 2015	Motor fitness (Speed–agility)	1	1	1	0	1	4
Davis et al., 2015	Motor fitness (Speed–agility)	1	1	0	1	1	4
Doi et al., 2013	Motor fitness (Speed–agility)	1	1	0	1	1	4
Ersoy et al., 2009	Motor fitness (Speed–agility)	1	1	0	1	1	4
Kang et al., 2017	Motor fitness (Speed–agility)	1	1	0	1	1	4
Mulasso et al., 2017	Motor fitness (Speed–agility)	1	1	0	1	1	4

Asai et al., 2020	Motor fitness (Speed–agility)	1	1	1	1	1	5
Breton et al., 2014	Motor fitness (Speed–agility)	1	1	1	1	1	5
Doi et al., 2019	Motor fitness (Speed–agility)	1	1	1	1	1	5
Donoghue et al., 2017	Motor fitness (Speed–agility)	1	1	1	1	1	5
Kwan et al., 2012	Motor fitness (Speed–agility)	1	1	1	1	1	5
Nitz et al., 2013	Motor fitness (Speed–agility)	1	1	1	1	1	5
Savva et al., 2013	Motor fitness (Speed–agility)	1	1	1	1	1	5
Schroll et al., 1997	Motor fitness (Speed–agility)	1	1	1	1	1	5
Bravell et al., 2017	Flexibility	1	1	1	1	1	5
Luukinen et al., 1995	Flexibility	1	1	1	1	1	5
Ward et al., 2016	Flexibility	1	1	1	1	1	5

PF, Physical Fitness component; SE, Standard Error; CI, Confidence Intervals. Bold values are defined as total score.

Table 8. Quality assessment of the systematic reviews included using the AMSTAR rating tool.

Reviews	1	2	3	4	5	6	7	8	9	10	11	Rating	Quality*
Van Kan et al. (2009)	Yes	Yes	No	No	No	Yes	Yes	Yes	No	N/A	Yes	6	Medium
Grande et al. (2019)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	N/A	Yes	8	High
Cavanaugh et al. (2018)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	N/A	Yes	9	High
Wang et al. (2020)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	9	High
Quan et al. (2017)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	10	High
Peel et al. (2019)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	10	High

AMSTAR contains 11 items to appraise the methodological aspects of the systematic reviews. All 11 items were scored as “Yes”, “No”, “Can’t Answer” (C/A), or “Not Applicable” (N/A). A total possible score of 11 was calculated, counting only for positive responses (“Yes”). Quality *, The final quality rates were computed by tertiles, where the first tertile ranged from 0 to 3 points, the second tertile from 4 to 7 points, and the third tertile from 8 to 11 points. Likewise, each tertile was treated as “low”, “medium”, or “high” quality, respectively. All 11 items were scored as “Yes”, “No”, “C/A”, or “N/A”. AMSTAR comprises the following items: 1. “a priori” design provided; 2. duplicate study selection/data extraction; 3. comprehensive literature search; 4. status of publication as inclusion criteria (i.e., grey or unpublished literature); 5. list of studies included/excluded provided; 6. characteristics of included studies documented; 7. scientific quality assessed and documented; 8. appropriate formulation of conclusions (based on methodological rigor and scientific quality of the studies); 9. appropriate methods of combining studies (homogeneity test, effect model used and sensitivity analysis); 10. assessment of publication bias (graphic and/or statistical test); and 11. conflict of interest statement.

Table 9. Longitudinal validity of motor fitness and flexibility tests for falls and hip fracture risk in adults and older adults.

Author	Participants	Sample age, years (Range)	Follow-up	Fitness test	Health outcomes	Main outcomes and conclusions
Kang et al. 2017	Females= 307 Males= 234	67.4 ± 5.6 (60-86)	1 year	Gait speed (4-m)	Falls	Slower gait speed (<0.9m/s): HR= 1.00 (0.99–1.01). The 4m gait test was not associated with falls.
Quach et al. 2011	Females= 487 Males= 276	78.1 ± 5.4	1.5 years	Gait speed (4-m)	Falls	Slower gait speed (<0.6m/s), indoor falls: IRR= 2.17 (1.33–3.55). Faster gait speed, outdoor falls: IRR= 2.11 (1.40–3.16). There was a nonlinear, U-shaped relationship between gait speed and falls. Those with faster and slower gait speeds were at highest risk of falls.
Callisaya et al. 2016	Females= 266 Males= 243	74.9 ± 6.8 (60-105)	2 years	Gait speed (4.2-m, 7-m, 8.5-m)	Falls	Slower gait speed (<0.8m/s): RR= 1.30 (1.14-1.47). Slower gait speed predicted falls.
Luukinen et al. 1995	Females= 640 Males= 376	76.1 ± 4.9 (70-92)	2 years	Step length (5-m)	Falls	Short step length (<0.45m): RR= 1.6 (1.04-2.46). A shorter step length predicted risk of falling.
Abu et al. 2018	Females= 180 Males= 145	67.67 ± 5.5 (60-89)	½ year	Gait speed (6-m)	Falls	Slower gait speed: Mean (SD)= 1.06 (0.25), p= 0.20. The 6m gait speed test was not associated with falls.
Kwan et al. 2012	Females= 120 Males= 160	74.9 ± 6.4 (65–91)	2 years	Gait speed (6-m)	Falls	Slower gait speed: IRR= 0.95 (0.60–0.94). Slower gait speed did not predict falls.
Muraki et al. 2013	Females= 1470 Males= 745	68.5 ± 11.3	3 years	Gait speed (6-m)	Falls	Slower gait speed (0.1m/s decrease), males: OR= 1.15 (1.09–1.23), and females: OR= 1.05 (1.01–1.10). Slower gait speed predicted falls.
Sanders et al. 2016	Females= 2179 Males= 1933	74.5 ± 5.8 (≥ 65)	13 years; 6 years; 21 years	Gait speed (6-m)	Falls	Slower gait speed (<0.71m/s): HR= 1.3 (1.0-1.5). Slow gait speed predicted falls.
Luukinen et al. 1995	Females= 640 Males= 376	76.1 ± 4.9 (70-92)	2 years	Gait speed (10-m)	Falls	Slower gait speed (<0.77m/s): RR= 2.3 (1.57-3.48).

						Slow gait speed was an independent risk factor for recurrent falling.
Doi et al. 2013	Females= 42 Males= 15	79.7 ± 8.2 (≥ 65)	1 year	Gait speed (10-m)	Falls	Slower gait speed (<0.63m/s): OR= 0.02 (0.00–0.23).
Stenhagen et al. 2013	Females= 840 Males= 923	(60-93)	3-6 years	Gait speed (15-m)	Falls	Gait speed predicted risk of falling. Slower gait speed: OR= 1.77 (1.28–2.46).
Dargent-Molina et al. 1999	Females= 5895	80.5 ± 3.8 (≥ 75)	2 years	Gait speed (6-m)	Hip fracture	Slow gait speed predicted falls. Slower gait speed: RR= 1.8 (1.5–2.0). Gait speed may be a predictor of risk of hip fracture.
Kauppi et al. 2014	Females= 1331 Males= 969	66.38 ± 8.24 (≥ 55)	10 years	Gait speed (6.1-m)	Hip fracture	Slower gait speed (<0.40m/s): HR= 0.70 (0.54–0.92).
Wihlborg et al. 2015	Females= 1044	75.2 ± 0.2	10 years	Gait speed (30-m)	Hip fracture	Gait speed predicted hip fracture. Slower gait speed (<1.3m/s): HR= 1.37 (1.14–1.64).
Ersoy et al. 2009	Females= 125	61.4 ± 7.9 (50-79)	½ year	Postural balance (Berg Balance Scale; one-leg stance, eyes open, ≥ 5s)	Falls	Gait speed predicted hip fracture. Balance (≤52 points): OR= 6.19 (1.32–28.94). Failing one-leg stance (<5s): 0.24 (0.06-1.03). Worse Berg Balance performance predicted falls.
Muir et al. 2010	Females= 33 Males= 57	79.7 ± 4.8 (60-90)	1 year	Postural balance (Berg Balance Scale)	Falls	Balance impairment: RR= 2.00 (1.13–3.56). Impairment balance was a predictor of falls.
Luukinen et al. 1995	Females= 640 Males= 376	76.1 ± 4.9 (70-92)	2 years	Postural balance (Tinetti Scale)	Falls	Balance (>7 points): RR= 2.9 (1.73-4.92). Impairment balance predicted falls.
Austin et al. 2007	Females= 1282	± 75.2 (70-85)	3 years	Postural balance (tandem stance, eyes open and closed, ≥ 10s)	Fear of falling	Poor balance, eyes open: OR= 1.78 (1.38–2.28). Poor balance, eyes closed: 1.87 (1.44–2.42). Poor balance predicted developing fear of falling.
Kwan et al. 2012	Females= 120 Males= 160	74.9 ± 6.4 (65–91)	2 years	Postural balance (near-tandem stance, eyes open and closed, ≥ 30s; one-leg stance, eyes open, ≥ 5s)	Falls	Worse near-tandem stance performance: 0.80 (0.65–0.99). Worse one-leg stance performance: IRR= 0.78 (0.62–0.99).
Vellas et al. 1997	Females= 156 Males= 111	72 ± 6.1 (≥ 60)	3 years	Postural balance (one-leg stance, eyes open, ≥ 5s)	Falls and injurious falls	Poor balance predicted risk of falls. Impaired one-leg stance: RR= 2.13 (1.04-4.34), for injurious falls. One-leg test predicted injurious falls but not falls.

Mulasso et al. 2017	Females= 119 Males= 73	73.0 ± 6.2 (≥ 65)	1 year	Postural balance (one-leg stance, eyes open, 60s)	Falls	One-leg stance: OR= 0.99 (0.97–1.01). The one-leg standing balance test was not associated with falls.
Nitz et al. 2013	Females= 449	59.3 ± 10.6 (40-80)	9 years	Postural balance (bipedal stance, eyes open and closed; one-leg stance, eyes open, 10s. Force platform)	Falls	Unsteady in bipedal stance, eyes closed: OR= 1.99 (1.18–3.38), for single falls. Unsteady in bipedal stance, eyes closed: OR= 4.21 (1.79–9.92), for multiple falls. Impairment balance detected potential fallers.
Swanenburg et al. 2010	Females= 225 Males= 45	73 ± 7 (60-90)	1 year	Postural balance (bipedal stance, eyes open and closed, 20s. Force platform)	Falls	Amplitude of in medial–lateral directions during single-task condition: OR= 21.8 (3.19-149.25). Impairment balance predicted multiple falls.
Maki et al. 1994	Females= 83 Males= 17	83 ± 6 (62-96)	1 year	Postural balance (bipedal stance, eyes open and closed, 30s. Force platform)	Falls	Lateral spontaneous-sway amplitude (eyes closed): Mean (SD)= 3.35 (2.29), p= 0.004. Postural balance predicted future risk of falling with moderate accuracy.
Pajala et al. 2008	Females= 434	(63–76)	1 year	Postural balance (side-by-side, eyes open and closed; tandem stance and semi-tandem stance, eyes open; 30s. Force platform)	Falls	Inability to complete the tandem stance: IRR= 4.33 (2.17–8.52). Impairment balance predicted risk of falling.
Frames et al. 2018	Females= 65 Males= 33	Non obese: 77.41 ± 8.49 Obese: 72.68 ± 7.40	2 years	Postural balance (bipedal stance, eyes open and closed, 60s. Force platform and an inertial measurement unit affixed at the sternum)	Falls	Obese fallers had higher sway range than non-fallers (p= 0.001, F=7.44), root mean square values (p= 0.002, F=6.62) and SD values (p=0.002, F= 6.62) from the force plate center of pressure time series. Obese fallers had higher sway area than non-fallers (ellipse area, p= 0.003, F= 5.89; and circular area p< 0.0002, F= 8.97), mean velocity (p=0.011, F=4.56), mean radius (p< 0.0001, F= 10.47) and mean path length of center of pressure (p= 0.011, F= 4.56). The postural balance test may help detecting fall risk in elderly persons who are obese.

Wihlborg et al. 2015	Females= 1044	75.2 ± 0.2	10 years	Postural balance (one-leg stance, eyes open and closed, ≥ 60s)	Hip fracture	Failing the balance test: HR= 1.98 (1.18–3.32). Balance predicted hip fracture.
Abu et al. 2018	Females= 180 Males= 145	67.7 ± 5.5 (60-89)	½ year	Timed Up&Go (TUG)	Falls	TUG (>8s): IQR: 4.8-17.5, with a sensitivity of 83.95%% (74.1-91.2) and specificity of 32.43% (26.3-39.0). The TUG test had predictive validity for risk of falls.
Asai et al. 2020	Females= 425 Males= 224	± 76.2 (≥ 60)	1 year	TUG	Falls	Worse TUG performance (>7s): OR= 1.14 (1.01-1.29), p= 0.032. Worse TUG predicted risk of falls.
Austin et al. 2007	Females= 1282	± 75.2 (70-85)	3 years	TUG	Fear of falling	TUG (>10s): mean= 9.5; IQR= 8.1-11.0, for fear of falling, and mean= 10.8; IQR= 8.7-21.3, for persistent fear. The TUG test predicted fear of falling.
Clemson et al. 2015	Females= 533 Males= 467	± 73.4 (65-94)	11 years	TUG	Falls	TUG (>12s): HR= 1.02 (1.01-1.04). Worse TUG performance predicted injurious falls.
Doi et al. 2013	Females= 42 Males= 15	79.7 ± 8.2 (≥ 65)	1 year	TUG	Falls	TUG (>21s): OR= 1.08 (1.02–1.15). The TUG test predicted risk of falling.
Ersoy et al. 2009	Females= 125	61.4 ± 7.9 (50-79)	½ year	TUG	Falls	TUG (>13s): OR= 4.55 1.95-10.61. The TUG test predicted falls.
Kang et al. 2017	Females= 307 Males= 234	67.4 ± 5.6 (60-86)	1 year	TUG	Falls	TUG (>16s): HR= 1.07 (1.00–1.14). The TUG test predicted falls.
Kwan et al. 2012	Females= 120 Males= 160	74.9 ± 6.4 (65–91)	2 years	TUG	Falls	TUG (>11s): IRR= 1.27 (1.04–1.56). The TUG test predicted risk of falling.
Mulasso et al. 2017	Females= 119 Males= 73	73.0 ± 6.2 (≥ 65)	1 year	TUG	Falls	TUG (>11s): OR= 1.076 (0.94–1.21) The TUG test was not associated with falls.
Nitz et al. 2013	Females= 449	59.3 ± 10.6 (40-80)	9 years	TUG	Falls	TUG (>8s): 1.32 (1.15, 1.51), for multiple falls. The TUG test predicted multiple falls.
Luukinen et al. 1995	Females= 640 Males= 376	76.1 ± 4.9 (70-92)	2 years	Range Of Motion (ROM) (Hip and knee ROM)	Falls	Hip motion reduced (<120°): RR= 1.4 (0.89-2.04). Knee motion reduced (<100°): 1.4 (0.97-2.04). Reduced ROM was not associated with falls.

HR, Hazard Ratio; IRR, Incidence Rate Ratio; IQR, Interquartile Range; OR, Odds Ratio; ROM, Range Of Motion; RR, Relative Risk; SD, Standard Deviation; TUG, Timed Up&Go test.

Table 10. Longitudinal validity of motor fitness and flexibility tests for cognitive outcomes in adults and older adults.

<i>Author</i>	<i>Participants</i>	<i>Sample age, years (Range)</i>	<i>Follow-up</i>	<i>Fitness test</i>	<i>Health outcomes</i>	<i>Main outcomes and conclusions</i>
<i>Ojagbemi et al. 2015</i>	2149 participants (unspecified sex)	(≥ 65)	2 years	Gait speed (3-m, 4-m)	Cognitive decline	Slower gait speed (>6.52s in 3m; >8.70s in 4m, in total): HR= 0.77 (0.43–1.2). Change in gait speed was a significant predictor of reduced cognitive performance.
<i>Hoogendijk et al. 2020</i>	Females= 2221 Males= 1999	± 72.0 (≥ 55)	9-25 years	Gait speed (4-m, 6-m, 30-m)	Cognitive decline	Slower gait speed (<0.82m/s): HR= 1.23 (1.00-1.50), p< 0.05. Slower gait speed predicted cognitive decline.
<i>Sanders et al. 2016</i>	Females= 2179 Males= 1933	74.5 ± 5.8 (≥ 65)	13 years; 6 years; 21 years.	Gait speed (6-m)	Persistent cognitive decline	Slower gait speed (<0.71m/s): HR= 2.1 (0.8-5.8). Slow gait speed did not predict persistent cognitive decline.
<i>Stijntjes et al. 2017</i>	Females= 1647 Males= 1332	(55-90)	5-12 years	Gait speed (6-m)	Cognitive decline	No association was found for the group of 55–64 years. Group of 65–74 years: slower gait speed (<0.86m/s) was associated with a steeper decline in global cognitive function (p< 0.05). Group of 75–85 years: slower gait speed (<0.69m/s) was associated with a steeper decline on each of the cognitive domains (p< 0.04). Slow gait speed was a marker for cognitive impairment in adults over 65years old.
<i>Tian et al. 2019</i>	Females= 93 Males= 108	79.3 ± 6.3	5.2 years	Gait speed (6-m, 400-m)	Cognitive impairment/ Alzheimer's disease	Slower gait speed (<1.02m/s): β= -0.22 (-0.47-0.04), p< 0.05. Slower gait speed predicted cognitive impairment/Alzheimer's disease.
<i>Buracchio et al. 2010</i>	Females= 118 Males= 86	79.0 ± 8.8 (≥ 65)	20 years	Gait speed (9.14-m)	Mild cognitive impairment	Gait speed rate of change before converters to mild cognitive impairment= -0.005 m/s/year (-0.010-0.000), p= 0.07. A decrease of 0.02 m/s/year (p= 0.001) in gait speed may predict mild cognitive impairment. Changes in gait speed may predict cognitive impairment.
<i>Sakurai et al. 2017</i>	Females= 106 Males= 117	72.6 ± 4.9 (65–85)	1 year	Gait speed (10-m fast gait speed)	Subjective memory complaints	Gait speed (<2.0m/s): OR= 0.84 (0.39–1.81).

<i>Bullain et al. 2016</i>	Females= 402 Males= 176	93.3 ± 2.6 (≥ 90)	2.6 years	Postural balance (one-leg stance, eyes open, 10s)	Dementia	Gait speed did not predict developing subjective memory complaints. Unable to complete balance test: HR= 3.20 (1.91–5.37).
<i>Doi et al. 2019</i>	Females= 2124 Males= 1962	± 72.0 (≥ 65)	3.6 years	Timed Up&Go (TUG)	Dementia	Poor balance may predict dementia. Worse TUG performance: HR= 1.54 (1.01-2.35), p= 0.04.
<i>Donoghue et al. 2017</i>	Females= 1169 Males= 1081	72.4 ± 6.0 (65–98)	5.9 years	TUG	Cognitive decline	Worse TUG predicted incidence of dementia. TUG (>10.5s): β= 1.00 (0.99-1.00). The TUG test was not significant.

β, Standardized Regression Coefficient; B, Non-Standardized Regression Coefficient; SE, Standard Error; HR, Hazard Ratio; OR, Odds Ratio; TUG, Timed Up&Go test.

Table 11. Longitudinal validity of motor fitness and flexibility tests for depressive symptoms and well-being in adults and older adults.

<i>Author</i>	<i>Participants</i>	<i>Sample age, years (Range)</i>	<i>Follow-up</i>	<i>Fitness test</i>	<i>Health outcomes</i>	<i>Main outcomes and conclusions</i>
<i>Veronese et al. 2017</i>	Females= 485 Males= 1247	69.3 ± 6.7 (≥ 50)	2 years	Gait speed (2.5-m)	Depressive symptoms	Slower gait speed: OR= 1.82 (1.00–3.32). Slow gait speed may predict depression.
<i>Briggs et al. 2019</i>	3615 participants (unspecified sex)	± 63.3 (≥ 50)	2-4 years	Gait speed (4-m)	Depressive symptoms	Slower gait speed (<1.3m/s): OR= 1.54 (1.08-2.19), <i>p</i> < 0.003. Slower gait speed predicted depression.
<i>Veronese et al. 2017^b</i>	Females= 533 Males= 437	72.5 ± 6.0 (65-96)	4 years	Gait speed (4-m)	Depressive symptoms	Slower gait speed (<0.80m/s), females: OR= 1.55 (0.90-2.68). Slower gait speed (<0.80m/s), males: OR= 2.22 (1.16-4.28). Slow walking speed predicted depression only in males.
<i>Sanders et al. 2012</i>	Females= 686 Males= 773	68.9 ± 8.5 (≥ 65)	16 years	Gait speed (6-m)	Depressive symptoms	Slower gait speed (<0.55m/s), females: HR= 1.22 (0.87–1.72). Slower gait speed (<0.55 m/s), males: HR= 1.67 (1.08–2.57).

<i>Davis et al. 2015</i>	Females= 196 Males= 112	82.5 ± 6.5 (≥ 70)	1 year	Timed Up&Go (TUG)	Well-being	Slow gait speed predicted depressive symptoms only in males. TUG (>20s): $\beta = -0.30(-0.45, -0.15)$. The TUG test was a predictor of well-being decrease.

β, Standardized Regression Coefficient; HR, Hazard Ratio; OR, Odds Ratio; TUG, Timed Up&Go test.

Table 12. Longitudinal validity of motor fitness and flexibility tests for mobility limitations and disability risk in adults and older adults.

Author	Participants	Sample age, years (Range)	Follow-up	Fitness test	Health outcomes	Main outcomes and conclusions
Makizako et al. 2015	Females= 463 Males= 485	78.5 ± 3.7 (≥ 75)	1.25 years	Gait speed (2.4-m)	Disability in instrumental activities of daily living (IADL)	Slower gait speed, females (<0.9m/s): OR= 0.97 (0.95–0.99). Slower gait speed, males (<1.1m/s): OR= 0.98 (0.96–0.99). Gait speed predicted disability in IADL.
Heiland et al. 2018	Females= 1154 Males= 602	70.6 ± 9.6 (≥ 60)	9 years	Gait speed (2.4-m or 6-m)	Disability in IADL	Gait speed limitation (<0.8m/s): HR= 3.11 (1.92–5.03). Gait speed predicted disability in IADL.
Abe et al. 2019	Females= 463 Males= 510	74.6 ± 5.5 (≥ 65)	4.4 years	Gait speed (5-m)	Disability in IADL	Slower gait speed (<1.25m/s): HR= 0.86 (0.18–0.92), $p < 0.001$. Slower gait speed predicted disability in IADL.
Makizako et al. 2010	Females= 146 Males= 119	78 ± 5.6 (68–96)	4 years	Gait speed (5-m)	Disability in IADL	Slower gait speed (≥ 5.6s in total), females: OR= 5.31 (1.63–17.22). Slower gait speed, males (≥ 4.6s in total): OR= 1.58 (0.94–2.63). Gait speed predicted disability in IADL only in older females.
Adachi et al. 2019	Females= 417 Males= 99	± 79 (76–82)	2 years	Gait speed (10-m)	Disability in IADL	Slower gait speed (<1.13 m/s): AUC= 0.80 (0.74–0.86). Slow gait speed predicted disability in IADL.
Laukkanen et al. 2000	Females= 261 Males= 127	75 and 80	5 years	Gait speed (10-m)	Disability in IADL	Slower gait speed: t -value=-0.23, SE=0.06. Reduced gait speed predicted disability in IADL.

Nakamoto et al. 2015	Females= 466 Males= 495	(40–79)	14 years	Gait speed (10-m)	Disability in IADL	Slower gait speed, females (>1 SD increase): OR= 0.68 (0.50-0.92). Slower gait speed, males (>1 SD increase): OR= 0.78 (0.59-1.04). Gait speed predicted disability in IADL.
Doi et al. 2020	Females= 1944 Males= 1752	71.2 ± 4.9 (≥ 65)	4.1 years	Gait speed (2.4-m)	Mobility disability	Slower gait speed (<1.1m/s): HR= 2.06 (1.65-2.57), p< 0.001.
Brach et al. 2012	Females= 336 Males= 216	79.4 ± 4.1 (≥ 65)	1 year	Gait speed (4-m)	Mobility disability	Slower gait speed predicted mobility disability. Slower gait speed (<1m/s): sensitivity, AUC= 0.61 (0.47-0.73), specificity, AUC= 0.60 (0.55-0.65). Slow gait may be a predictor of mobility disability.
Deshpande et al. 2013	622 participants (<i>unspecified sex</i>)	(50-85)	3 years	Gait speed (7-m)	Mobility disability	Slower gait speed (<1.2m/s): OR= 24.95 (6.03-103.13).
Rosso et al. 2019	Females= 171 Males= 166	78.5 ± 2.9 (70-79)	8 years	Gait speed (20-m)	Mobility Disability	Slower gait speed predicted mobility disability. Slower gait speed (<1.08m/s): HR= 0.82 (0.77-0.86).
Jung et al. 2018	Females= 746 Males= 602	± 76 (≥ 65)	5.38 years	Gait speed (4-m)	Frailty	Slower gait speed predicted mobility disability. Slower gait speed (<0.8m/s): β= -0.534, p= 0.001. Gait speed may predict frailty status.
Breton et al. 2014	Females= 652 Males= 613	± 73 (68-82)	5 years	Gait speed (4-m)	Functional autonomy	Faster gait speed, females (<1.5m/s): R ² = 0.37; p< 0.001, for greater mobility/autonomy. Faster gait speed, males (<1.7m/s): R ² = 0.28; p< 0.001, for with greater mobility/autonomy. Gait speed changes could moderately predict functional autonomy, especially in females.
Makizako et al. 2010	Females= 146 Males= 119	78 ± 5.6 (68-96)	4 years	Postural balance (one-leg stance, eyes open, 60s)	Disability in IADL	Balance, females: OR= 0.98 (0.95-1.02). Balance, males: OR= 0.98 (0.95-1.00). Balance did not predict disability in IADL.
Nakamoto et al. 2015	Females= 466 Males= 495	(40–79)	14 years	Postural balance (bipedal stance, eyes open, 60s. Force platform)	Disability in IADL	A 1-SD increase in sway area with eyes open, females: OR= 1.49 (1.17–1.90). A 1-SD increase in sway area with eyes open, males: OR= 0.92 (0.71–1.20). Sway area increases might predict disability in IADL, only in females.

Breton et al. 2014	Females= 652 Males= 613	± 73 (68-82)	5 years	Postural balance (one- leg stance, eyes open, 60s)	Functional autonomy	Balance, females: $R^2 = 0.37$; $p < 0.001$. Balance, males: $R^2 = 0.28$; $p < 0.001$. Balance could moderately predict functional autonomy, especially in females.
Savva et al. 2013	Females= 925 Males= 889	± 70 (≥ 65)	4 years	TUG	Frailty	TUG ($>16s$): AUC= 0.87 for frailty. TUG ($>12s$): AUC= 0.73 for prefrailty. The TUG test identified prefrailty and frailty.
Breton et al. 2014	Females= 652 Males= 613	± 73 (68-82)	5 years	TUG	Functional autonomy	TUG ($>11s$), females: β (SE)= 0.56 (0.13); $p < 0.001$. TUG ($>10.5s$), males: β (SE)= 0.42 (0.12); $p < 0.001$. The TUG test predicted decline in functional autonomy.
Ward et al. 2016	Females= 261 Males= 130	76.5 ± 7.1 (≥ 65)	1-2 years	Range Of Motion (ROM) (knee flexion- extension, ankle)	Mobility disability	Restricted knee flexion: OR= 2.03 (1.24-3.35). ROM predicted mobility disability.
Ward et al. 2016	Females= 261 Males= 130	76.5 ± 7.1 (≥ 65)	1-2 years	Speed of movement (leg velocity, reaction time, rapid leg coordination)	Mobility disability	Slower leg velocity: OR= 2.35 (1.21-4.58). Leg velocity predicted mobility disability.
Adachi et al. 2019	Females= 417 Males= 99	± 79 (76-82)	2 years	Maximum step length	Disability in IADL	Maximum step length: AUC= 0.81 (0.75-0.87). Maximum step length was a predictor for declined walking speed and predicted disability in IADL.
Schroll et al. 1997	Females= 146 Males= 113	(75-80)	5 years	Stair mounting (40- 50cm step)	Disability in IADL	Disability to complete the stair mounting test, females (40 cm): RR= 4.14, $p = 0.09$. Disability to complete the stair mounting test, males (50 cm): RR= 4.07, $p < 0.001$. Inability to mount a 50cm step was a predictor of disability in IADL in males.

β , Standardized Regression Coefficient; IADL, Disability in Instrumental Activities of Daily Living; AUC, Area Under the Curve; HR, Hazard Ratio; OR, Odds Ratio; ROM, Range Of Motion; RR, Relative Risk; SD, Standard Deviation; SE, Standard Error; TUG, Timed Up&Go test.

Table 13. Longitudinal validity of motor fitness and flexibility tests for CVD risk and mortality in adults and older adults.

Author	Participants	Sample age, years (Range)	Follow-up	Fitness test	Health outcomes	Main outcomes and conclusions
Heiland et al. 2018	Females= 1154 Males= 602	70.6 ± 9.6 (≥ 60)	9 years	Gait speed (2.4-m or 6-m)	CVD risk	Slower gait speed (<0.8m/s): HR= 1.09 (1.02–1.17), for group < 78 years. Slower gait speed (<0.8m/s): HR= 0.98 (0.92–1.03), for group ≥ 78 years. Slow gait speed predicted CVD risk only in adults <78 years.
Elbaz et al. 2013	Females= 1817 Males= 4450	(35-55)	6.4 years	Gait speed (2.4-m)	All-cause mortality	Slower gait speed, females (<1.09m/s), males (<1.26m/s): HR= 2.19 (1.54–3.11). Gait speed was a marker of mortality risk.
Zucchelli et al. 2019	Females= 2182 Males= 1181	74.7 ± 11.2 (≥ 60)	3-5 years	Gait speed (2.4-m, 6-m)	All-cause mortality and unplanned hospitalization	Slower gait speed (<0.8m/s): AUC= 0.85 (0.83-0.87), for 3-years mortality; AUC= 0.85 (0.83-0.86), for 5-years mortality. Slower gait speed (<0.8m/s): AUC= 0.70 (0.68-0.72) for 3-years hospitalization. Slower gait speed predicted mortality and unplanned hospitalization.
Andrasfay et al. 2020	Females= 430 Males= 457	70.1 ± 8.7 (≥ 60)	4 years	Gait speed (3-m)	All-cause mortality	A 1-SD increase in gait speed (<0.1m/s): HR= 0.70 (0.52-0.94), $p < 0.05$. Slower gait speed predicted mortality risk.
Camargo et al. 2016	Females= 1175 Males= 1001	62 ± 9 (35-84)	11 years	Gait speed (4-m)	Stroke	Slower gait speed (<1.0m/s): HR= 1.13 (0.88–1.45). Slow gait speed was not significantly associated with risk of stroke.
Jung et al. 2018	Females= 746 Males= 602	± 76 (≥ 65)	5.38 years	Gait speed (4-m)	All-cause mortality	Gait speed (<0.8m/s): HR= 3.37 (1.25–9.08). Gait speed may predict mortality.
Niiranen et al. 2019	Females= 1852 Males= 1601	54.7 ± 9.2 (45-74)	8 years	Gait speed (4-m)	All-cause mortality and Cardiovascular Disease (CVD) risk	Slower gait speed (<1.29m/s): females, HR= 3.72 (2.53-5.47), $p < 0.001$; men, HR= 2.57 (1.77-3.74), $p < 0.001$, for mortality. Slower gait speed (<0.99m/s): HR= 2.95 (2.26-3.85), $p < 0.05$, for CVD risk. Slower gait speed predicted incidence of CVD risk and mortality.

Hoogendijk et al. 2020	Females= 2221 Males= 1999	± 72.0 (≥ 55)	9-25 years	Gait speed (4-m, 6-m, 30-m)	All-cause mortality	Slower gait speed ($<0.82\text{m/s}$): HR= 1.70 (1.30-6.28), $p < 0.05$. Slower gait speed predicted mortality.
Abe et al. 2019	Females= 463 Males= 510	74.6 ± 5.5 (≥ 65)	4.5 years	Gait speed (5-m)	All-cause mortality	Slower gait speed ($<1.25\text{m/s}$): HR= 0.93 (0.88-1.00), $p = 0.038$. Slower gait speed mortality risk.
Nofuji et al. 2016	1085 participants (<i>unspecified sex</i>)	(65-89)	10.5 years	Gait speed (5-m)	All-cause mortality and cause-specific mortality ((CVD), cancer, other-causes and unknown causes))	Slower gait speed, females ($<0.94\text{m/s}$), males ($<1.10\text{m/s}$): HR= 1.95 (1.39-2.73), all-cause death. Slower gait speed, females ($<0.94\text{m/s}$), males ($<1.10\text{m/s}$): HR= 2.80 (1.55-5.05), CVD death. Slower gait speed females ($<0.94\text{m/s}$), males ($<1.10\text{m/s}$): HR= 2.08 (1.14-3.77), other-cause death. Slow gait speed predicted all-cause, CVD, and other-cause mortality, but not cancer mortality.
Lee et al. 2017	Females= 407 Males= 504	65.3 ± 9.3	4.1 years	Gait speed (3-m)	All-cause mortality and CVD risk	Slower gait speed ($<0.8\text{m/s}$): HR= 1.3 (0.6–2.5), for mortality. Slower gait speed ($<0.8\text{m/s}$): HR= 0.8 (0.2–3.3), for CVD risk. Not associations were found between gait speed and any studied variable.
Blain et al. 2010	Females= 1548	± 79 (77-81)	8 years	Gait speed (6-m)	All-cause mortality	Slower gait speed ($<0.8\text{m/s}$): HR= 1.33 (0.94-1.88). Slow gait speed predicted mortality.
Sabia et al. 2014	4016 participants (<i>unspecified sex</i>)	73.4 ± 4.7 (65–85)	12 years	Gait speed (6-m)	All-cause mortality	Slower gait speed decline/year ($<-0.08\text{m/s}$): HR= 1.40 (1.02-1.92). Fast gait speed decline predicted mortality.
Sanders et al. 2016	Females= 2179 Males= 1933	74.5 ± 5.8 (≥ 65)	13 years; 6 years; 21 years	Gait speed (6-m)	All-cause mortality	Slower gait speed ($<0.71\text{m/s}$): HR= 2.1 (1.6-2.6 45). Slow gait speed predicted mortality.
Georgiopolou et al. 2016	Females= 1529 Males= 1406	73.6 ± 2.9 (70-79)	10 years	Gait speed (20-m)	All-cause mortality and CVD risk	Slower gait speed ($<1.35\text{m/s}$): HR= 0.34 (0.24-0.48), mortality. Slower gait speed ($<1.35\text{m/s}$): HR= 0.36 (0.22-0.60), CVD. Slow gait speed test may predict mortality and CVD risk.

Idland et al. 2013	Females= 113	± 79.5 (75–92)	9 years	Gait speed (29-m)	All-cause mortality	Slower gait speed (<1.1m/s): OR= 0.05 (0.02-0.16). Slow gait speed predicted mortality risk.
Looker 2015	Females= 1475 Males= 1500	(≥ 50)	3 years	Gait speed (6.15-m)	Mortality risk by dysmobility	Slower gait speed (<1.0m/s): HR= 3.63 (2.69-4.90). Dysmobility (by slow gait speed) was associated with increased mortality risk.
Blain et al. 2010	Females= 1548	± 79 (77-81)	8 years	Postural balance (side- by-side, tandem stance, to complete 10-foot taps, eyes open, s)	All-cause mortality	Balance (≥ 4.6s): HR= 1.67 (1.21-2.31). Poor balance predicted mortality.
Cooper et al. 2014	Females= 1411 Males= 1355	53	13 years	Postural balance (one- leg stance, eyes closed, 30s)	All-cause mortality	Unable to perform standing balance test: HR= 3.40 (2.05-5.63). Inability to perform balance test was a strongly predictor of mortality.
Nofuji et al. 2016	1085 participants (<i>unspecified sex</i>)	(65-89)	10.5 years	Postural balance (one- leg stance, eyes open, 60s)	All-cause mortality and cause-specific mortality (CVD, cancer, other-causes and unknown causes)	One-leg stance (<18 score): HR= 1.91 (1.39-2.63), all-cause death. One-leg stance (<18 score): HR= 2.67 (1.54-4.62), CVD death. One-leg stance (<18 score): HR= 2.60 (1.45-4.69), other-causes death. Poor balance predicted all-cause, CVD, and other- cause mortality, but not cancer mortality.
Idland et al. 2013	Females= 113	± 79.5 (75–92)	9 years	Dynamic balance (Functional reach test, cm)	All-cause mortality	Poorer functional reach: OR= 0.92 (0.88-0.95). Poorer functional reach predicted mortality risk.
Bravell et al. 2017	Females= 343 Males= 242	66 ± 9 (60-91)	19 years	Flexibility (Toe-to- touch test)	All-cause mortality	Toe-to-touch test: HR= 0.99 (0.42-2.31). The predictive validity of this flexibility test was not significant.

CVD, Cardiovascular Disease; HR, Hazard Ratio; OR, Odds Ratio.

Table 14. Systematic reviews/meta-analysis on predictive validity of motor fitness tests for adverse health outcomes in adults and older adults.

AUTHOR	TYPE OF REVIEW; NUMBER OF STUDIES (K)	SAMPLE AGE	FITNESS TESTS	HEALTH OUTCOMES	MAIN RESULTS	CONCLUSIONS
FALLS						
CAVANAUGH ET AL. 2018	Systematic review K = 3	≥ 60	Gait speed (6-m) Dynamic balance (functional reach test, cm (FRT)) Timed Up&Go (TUG)	Falls	Slower gait speed (≥ 6s in total); RR= 1.8 (1.2-2.6). FRT: r= 0.013, p> 0.05. TUG: r= 0.143, p> 0.05.	SLOWER GAIT SPEED IDENTIFIED FALLS. NEITHER THE FRT NOR TUG TEST AND FALLS RISK WERE SIGNIFICANT.
VAN KAN ET AL. 2009	Systematic review K = 4	≥ 65	Gait speed (5-m, 6-m and 10m)	Falls	Slower gait speed (<0.67m/s): RR ranging from 0.2 to 5.4 (0.1-14.3).	SLOW GAIT SPEED WAS AN INDEPENDENT PREDICTOR OF FALLS.
COGNITIVE DECLINE AND IMPAIRMENT						
QUAN ET AL. 2017	Meta-analysis K = 17	≥ 60	Gait speed (2.4-m, 3-m, 4-m, 4.9-m, 6-m, 7-m, 8-m, 9-m, 10-m, 60-m and 400-m)	Cognitive decline and dementia	Slower gait speed: RR= 1.9 (1.5–2.3), for cognitive decline. Slower gait speed: RR= 1.7 (1.4–1.9), for dementia. Every 1 m/s decrement in gait speed: RR= 1.1 (1.1–1.2), for dementia.	SLOW OR DECREASED GAIT SPEED PREDICTED ELEVATED RISK OF COGNITIVE DECLINE AND DEMENTIA.

VAN KAN ET AL. 2009	Systematic review K = 7	≥ 65	Gait speed (2.4-m, 4-m, 4.6-m, 5-m and 6-m)	Cognitive decline	Slower gait speed (<1m/s): RR ranging from 0.4 to 5.6 (0.2-12.6).	GAIT SPEED AT USUAL PACE PREDICTED COGNITIVE IMPAIRMENT.
GRANDE ET AL. 2019	Systematic review K = 39	62-93	Gait speed (2.4m, 3m, 4m, 4.5m, 4.8m, 5m, 5.6m, 6m, 7m, 7.6m, 9m, 15m, 40m)	Cognitive decline and/or dementia	Slower gait speed (<0.8m/s): RR ranging from 1.1 to 5.0 (0.53-48.2).	SLOW OR DECREASED GAIT SPEED PREDICTED ELEVATED RISK OF COGNITIVE DECLINE AND/OR DEMENTIA.
PEEL ET AL. 2019	Systematic review K = 36 Meta-analysis K = 27	≥ 60	Gait speed (2.4-m, 3.7-m, 4-m, 4.6-m, 5-m, 5.3-m, 5.6-m, 6-m, 7-m, 7.6-m, 7.9-m, 8-m, 10-m, 20-m, 30-m)	Cognitive decline and dementia	Parameter estimate= -0.1 (-0.1, -0.1) m/s, p< 0.001, for cognitive decline. Parameter estimate= -0.2 (-0.2, -0.2) m/s, p< 0.001, for mild dementia. Parameter estimate= -0.4 (-0.5, -0.3) m/s, p< 0.001, for moderate dementia.	SLOW OR DECREASED GAIT SPEED PREDICTED ELEVATED RISK OF COGNITIVE DECLINE AND DEMENTIA.
MOBILITY LIMITATIONS AND DISABILITY						
CAVANAUGH ET AL. 2018	Systematic review K = 6	≥ 60	Gait speed (2.4-m, 4-m, 6-m and 15.24-m)	Disability in instrumental daily activities (IADL); mobility limitations	Usual gait speed (<1m/s): HR= 1.5 (1.3-1.7), for persistent mobility disability; and: HR= 1.5 (1.2-1.8), for persistent severe mobility disability. Slower gait speed, 2.4m (≥ 9s in total): OR= 3.4 (1.8–6.5); 6-8s: OR= 2.6 (1.4–4.9); 4-5s: OR= 2.1 (1.2–4.0), for mobility limitations.	SLOWER GAIT SPEED WAS A PREDICTOR OF DISABILITY IN IADL AND MOBILITY LIMITATIONS.

			Postural balance (semi-tandem stance, tandem stance one-leg stance, 30s/test (0-90s); side-by-side, semi-tandem stance, tandem stance, 10s/test)		Fastest gait speed: OR= 7.9, p= 0.04, for mobility limitations. Gait speed, 2.4m ($\geq$ 9s in total): OR= 5.4 (1.2–23.6), for IADL. Gait speed, 4m ($<$ 0.6m/s): 12 months, OR= 0.5 (0.02-16.00); 18 months, OR= 0.2 (0.01-8.79), for IADL. Total balance test ($<$ 53s): HR= 1.6 (1.4-1.8), for persistent mobility disability; and HR= 1.85 (1.6-2.2), for persistent severe mobility disability. Tandem stance ($>$ 2- $<$ 10s): OR= 1.6 (1.1-2.6), for mobility disability. Tandem stance ($>$ 2- $<$ 10s): OR= 2.5 (1.2-4.8), for IADL. TUG: OR= 1.1, p= 0.24, for mobility disability.	BALANCE TESTS IDENTIFIED PRECLINICAL DISABILITY DUE TO DISABILITY IN IADL AND MOBILITY DISABILITY.
			TUG			THE TUG TEST DID NOT IDENTIFY PRECLINICAL DISABILITY IN MOBILITY AND IADL.
VAN KAN ET AL. 2009	Systematic review K = 10	≥ 65	Gait speed (2.4-m, 4-m, 4.6-m, 4.9-m, 6-m, 11-m and 400-m)	IADL; mobility limitations	Slower gait speed ($<$ 1m/s): RR ranging from 0.7 to 6.1 (0.5-23.6), for IADL. Slower gait speed ($<$ 1m/s): RR ranging from 0.6 to 2.1 (0.9-6.5), for mobility limitations.	SLOW GAIT SPEED WAS FOUND TO PREDICT DISABILITY IN IADL AND MOBILITY LIMITATIONS.
WANG ET AL. 2020	Systematic review K = 62	54-86	Gait speed (<i>unreported distances</i>)	IADL	Slower gait speed ($<$ 1m/s): OR= 4.4 (1.3–14.5).	WORSE PHYSICAL PERFORMANCE IN GAIT SPEED, ONE-LEG BALANCE OR TUG PREDICTED IADL.

	Meta-analysis K = 45		Postural balance (Berg Balance Scale, one-leg stance (5s), side-by-side, semi-tandem stance, tandem stance) TUG		One-leg balance stance (<5s): OR= 2.7 (1.3–5.7). Worse TUG performance (>12s): OR= 3.4 (1.9–6.3).	
HOSPITALIZATION/ INSTITUTIONALISATION						
CAVANAUGH ET AL. 2018	Systematic review K = 2	≥ 60	Gait speed (6-m) Postural balance ((semi-tandem stance, tandem stance one-leg stance, 30s/test (0-90s))	Hospitalization	Slower gait speed (<1m/s): HR= 1.3 (1.0–1.6). Total balance (<53s): HR= 1.4 (1.1-1.7).	SLOW GAIT SPEED AND IMPAIRMENT BALANCE IDENTIFIED HOSPITALIZATION.
VAN KAN ET AL. 2009	Systematic review K = 5	≥ 65	Gait speed (4-m, 4.9-m, 6-m, 10-m)	Institutionalisation or hospitalization	Slower gait speed (<1m/s): RR ranging from 1.0 to 5.9 (1.0-18.5).	SLOW GAIT SPEED WAS FOUND TO PREDICT INSTITUTIONALISATION OR HOSPITALIZATION.
MORTALITY						
VAN KAN ET AL. 2009	SYSTEMATIC REVIEW K = 9	≥ 65	GAIT SPEED (2.4-M, 4-M, 4.6-M, 4.9-M AND 6-M)	ALL-CAUSE MORTALITY	SLOWER GAIT SPEED (<1M/S): RR RANGING FROM 0.7 TO 7.4 (0.5-14.5).	SLOW GAIT SPEED WAS FOUND TO PREDICT MORTALITY.

B, Non-Standardized Regression Coefficient; FRT, Functional Reach Test; HR, Hazard Ratio; IADL, Disability in Instrumental Activities of Daily Living; OR, Odd Ratio; RR, Relative Risk; TUG, Timed Up&Go test.

Table 15. Main characteristics and findings of longitudinal studies investigating the predictive validity of motor fitness and flexibility tests for health outcomes in adults and older adults.

Author	N (% female sex)	Mean (range) age	Fitness test					Health outcomes						
			Gait speed	Balance	TUG	Flexibility	Other	Falls/ Hip fracture	Cognitive decline/ impairment	Depression/ Well-being	Related-mobility disability/ Frailty	CVD/ Stroke	All-cause mortality	Other-causes mortality
Kang et al. 2017	541 (57)	67 (60-86)												
Quach et al. 2011	763 (64)	78												
Callisaya et al. 2016	509 (52)	75 (60-105)												
Abu et al. 2018	325 (55)	68 (60-89)												
Kwan et al. 2012	280 (43)	75 (65–91)												
Muraki et al. 2013	2215 (66)	68												
Sanders et al. 2016	4112 (53)	74 (≥ 65)												
Luukinen et al. 1995	1016 (63)	76 (70-92)												
Doi et al. 2013	57 (74)	80 (≥ 65)												
Stenhagen et al. 2013	1763 (48)	76 (60-93)												
Dargent-Molina et al. 1999	5895 (100)	80 (≥ 75)												
Kauppi et al. 2014	2300 (58)	66 (≥ 55)												
Wihlborg et al. 2015	1044 (100)	75												
Ersoy et al. 2009	125 (100)	61 (50-79)												
Muir et al. 2010	90 (37)	80 (60-90)												
Austin et al. 2007	1282 (100)	75 (70-85)												
Vellas et al. 1997	267 (58)	73 (≥ 60)												
Mulasso et al. 2017	19 (62)	73 (≥ 65)												
Nitz et al. 2013	449 (100)	59 (40-80)												
Swanenburg et al. 2010	270 (83)	73 (60-90)												

Maki et al. 1994	100 (83)	83 (62-96)
Pajala et al. 2008	434 (100)	70 (63-76)
Frames et al. 2018	98 (66)	75
Asai et al. 2020	649 (65)	76 (≥ 60)
Clemson et al. 2015	1000 (53)	73 (65-94)



Author	N (% female sex)	Mean (range) age	Fitness test					Health outcomes						
			Gait speed	Balance	TUG	Flexibility	Other	Falls/ Hip fracture	Cognitive decline/ impairment	Depression/ Well-being	Related-mobility disability/ Frailty	CVD/ Stroke	All-cause mortality	Other-causes mortality
Doi et al. 2020	3696 (53)	71 (≥ 65)	●								●			
Makizako et al. 2015	948 (49)	78 (≥ 75)	●								●			
Heiland et al. 2018	1756 (66)	71 (≥ 60)	●								●	*	●	
Abe et al. 2019	973 (48)	75 (≥ 65)	●								●		●	
Makizako et al. 2010	265 (55)	79 (68-96)	●	●						*	●			
Adachi et al. 2019	516 (81)	79 (76-82)	●				●				●			
Laukkanen et al. 2000	388 (67)	77 (75-80)	●								●			
Nakamoto et al. 2015	961 (48)	60 (40-79)	●	*	●						●			
Brach et al. 2012	552 (61)	79 (≥ 65)	●								●			
Deshpande et al. 2013	622 (?)	67 (50-85)	●								●			
Rosso et al. 2019	337 (51)	78 (70-79)	●								●			
Jung et al. 2018	1348 (55)	76 (≥ 65)	●								●		●	
Breton et al. 2014	1265 (52)	73 (68-82)	●	●	●						●			
Ward et al. 2016	391 (67)	77 (≥ 65)				●	●				●			
Savva et al. 2013	1814 (51)	70 (≥ 65)			●						●			
Schroll et al. 1997	259 (56)	77 (75-80)					●			*	●			

Author	N (% female sex)	Mean (range) age	Fitness test					Health outcomes						
			Gait speed	Balance	TUG	Flexibility	Other	Falls/ Hip fracture	Cognitive decline/ impairment	Depression/ Well-being	Related-mobility disability/ Frailty	CVD/ Stroke	All-cause mortality	Other-causes mortality
Elbaz et al. 2013	6267 (29)	45 (35-55)	●										●	
Zucchelli et al. 2019	3363 (65)	75 (≥ 60)	●										●	
Andrasfay et al. 2020	887 (48)	70 (≥ 60)	●										●	
Niiranen et al. 2019	3453 (54)	55 (45-74)	●										●	
Hoogendijk et al. 2020	4220 (53)	72 (≥ 55)	●						●				●	
Nofuji et al. 2016	1085 (?)	77 (65-89)	●	●								●	●	●
Lee et al. 2017	911 (45)	65	●									●	●	
Blain et al. 2010	1548 (100)	79 (77-81)	●	●									●	
Sabia et al. 2014	4016 (?)	73 (65-85)	●										●	
Georgiopoulou et al. 2016	2935 (52)	74 (70-79)	●									●	●	
Idland et al. 2013	113 (100)	79 (75-92)	●	●								●	●	
Camargo et al. 2016	2176 (54)	63 (35-84)	●									●		
Looker 2015	2975 (49)	≥ 50	●										●	
Cooper et al. 2014	2766 (51)	53		●									●	
Bravell et al. 2017	585 (59)	67 (60-91)				●						●		
Ojagbemi et al. 2015	2179 (?)	≥ 65	●											
Stijntjes et al. 2017	2979 (55)	72 (55-90)	●											
Osuka et al. 2020	725 (62)	77 (73-80)	●											
Tian et al. 2019	201 (46)	79	●											
Buracchio et al. 2010	204 (58)	79 (≥ 65)	●											
Sakurai et al. 2017	223 (48)	73 (65-85)	●											
Bullain et al. 2016	578 (70)	93 (≥ 90)		●										
Doi et al. 2019	4086 (52)	72 (≥ 65)			●									
Donoghue et al. 2017	2250 (52)	72 (65-98)			●									
Briggs et al. 2019	3615 (?)	63 (≥ 50)	●							●				
Veronese et al. 2017	1732 (28)	69 (≥ 50)	●							●				

Veronese et al. 2017 ^b	970 (55)	72 (65-96)
Sanders et al. 2012	1459 (47)	69 (≥ 65)
Davis et al. 2015	308 (63)	≥ 70



N, sample size; TUG, Timed Up&Go test. “Other” includes: maximum step length, speed of movement and stair mounting test. CVD, Cardiovascular Disease. *Partial predictive values due to sex or age range (see Table S8–S12).

● = significant association; ● = not associated or non-significant association; ● = association found only in some outcomes. ? = female and male sample size are presented together.

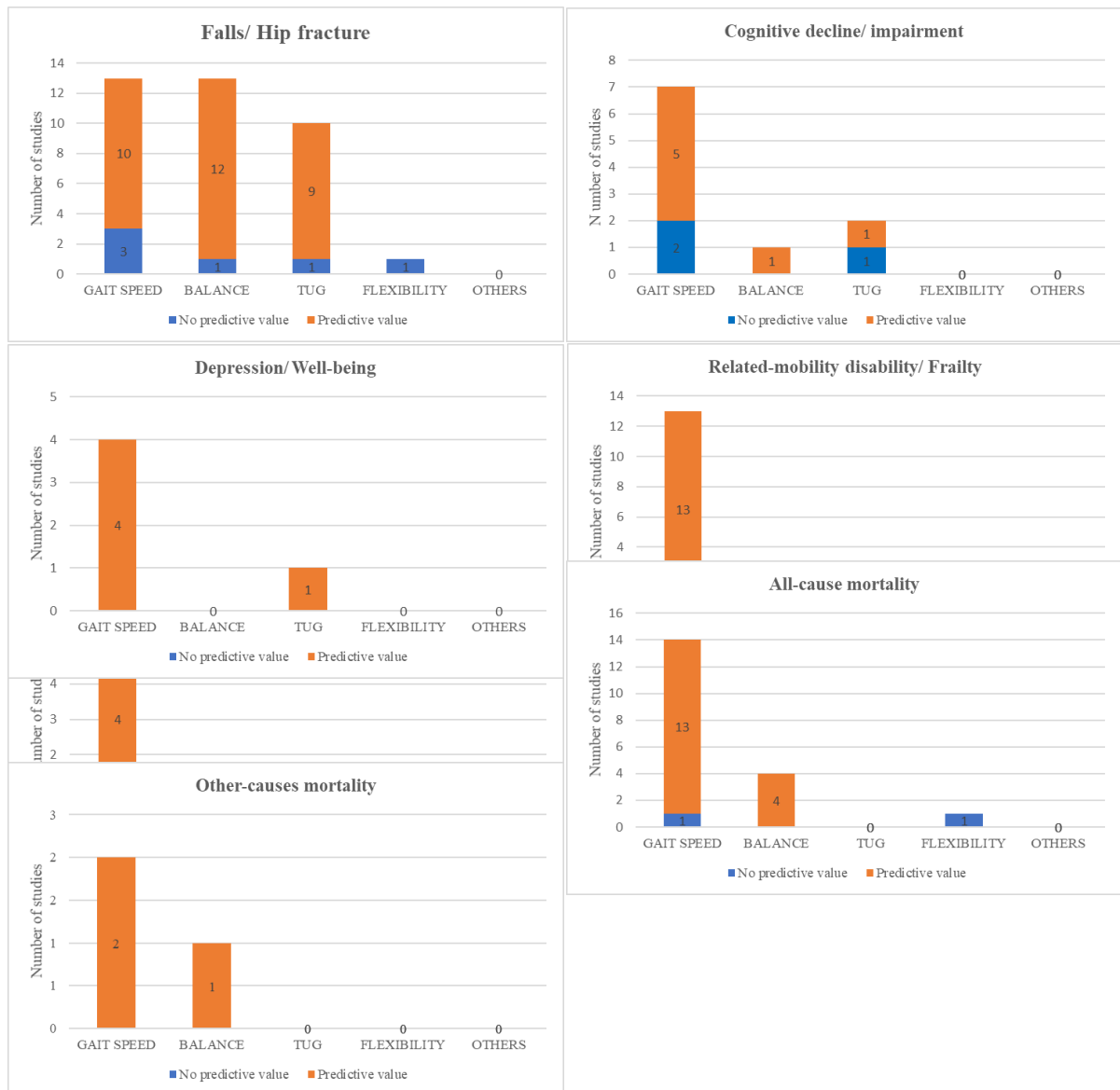
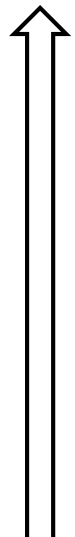


Figure 3. Number of studies reporting predictive validity or lack of predictive validity of the included motor fitness and flexibility tests for different health outcomes.

CVD, Cardiovascular Disease; TUG, Timed Up&Go test; “Other” includes: maximum step length, speed of movement and stair mounting test.

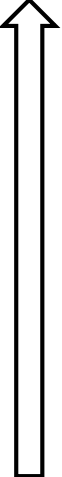
Figure 4. Cut points of gait speed at usual pace and risk of adverse outcomes found in the selected studies and systematic reviews/meta-analysis.



<1.3m/s	Hip fracture; Depressive symptoms; CVD risk; All-cause mortality
<1.2m/s	Mobility disability; IADL; All-cause mortality
<1.1m/s	Mobility disability; IADL; CVD risk; All-cause mortality; Other-causes mortality
<1.0m/s	Mobility disability; IADL; Cognitive decline/impairment/Alzheimer's disease; CVD risk; Hospitalization/Institutionalization; All-cause mortality; Other-causes mortality
<0.9m/s	IADL; CVD risk; All-cause mortality; Other-causes mortality
<0.8m/s	Falls; IADL; Frailty; Cognitive decline/Dementia; Depressive symptoms; CVD risk; Hospitalization; All-cause mortality
<0.7m/s	Falls; Cognitive decline; All-cause mortality
<0.6m/s	Falls; IADL
<0.5m/s	Depressive symptoms
<0.4m/s	Hip fracture

CVD, Cardiovascular Disease risk; IADL, Instrumental Disability in Daily Activities.

Figure 5. Cut points of timed up&go test and risk of adverse outcomes found in the selected studies and systematic reviews/meta-analysis.



>7s	Falls
>8s	Falls
>10s	Fear of falling
>11s	Falls; Functional autonomy decline
>12s	Falls; IADL
>13s	Falls
>16s	Falls; Frailty
>20s	Well-being decline
>21s	Falls
>7s	Falls

IADL, Instrumental Disability in Daily Activities

Chapter II: Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

The literature search yielded 9202 and 27 additional records were identified through other sources (see the PRISMA flow-chart in **Figure 6**). After removal of duplicate references (1805 studies), and screening of titles and abstracts (7233 studies), we excluded 9038 studies. A total of 191 full-text studies were assessed for eligibility, and 85 studies (six systematic reviews) were excluded due to reasons indicated in **Figure 6**.

Finally, a total of 101 original studies (see **Table 16**) addressed the criterion-related validity of field-based physical fitness tests in adults aged 18-64 years. The sample size involved 10632 participants (see **Table 17**). Eighty-six and 78 original studies reported female (n= 5539) and male (n= 4722) sample proportions, respectively; however, in seven studies sex was not specified.

A total of four meta-analyses and one systematic review were included in the present systematic review (see **Table 18**). The sample size involved 9985 participants with ages ranging from 19 to 64 years (see **Table 19**).

Quality assessment

Of the 101 original studies included in the present systematic review, 11 and 40 studies were classified as very low (a total score less than 2) and low quality (a total score of 3 and 4), respectively (see **Table 16**). A total of 50 original studies were classified as high quality (a total score higher than 4). Of these, 40, nine and one analysed the criterion-related validity of cardiorespiratory fitness, muscular fitness and flexibility field-based physical fitness tests in adults, respectively. No study of those classified as high quality, analysed the criterion-related validity of motor fitness (i.e. speed, agility, balance and coordination).

Two meta-analyses^{79, 80}, and one systematic review⁸¹ were ranked as high quality (all 8 points), and two meta-analyses^{82, 83} were ranked as medium quality (both 7 points) (see **Table 18**). Three of them assessed the criterion-related validity of field-based cardiorespiratory fitness tests: the 20-m SRT⁷⁹, distance and time-based run/walk tests⁸⁰,

and the step tests⁸¹; and two of them studied the criterion-related validity of the sit-and-reach⁸² and toe-to-touch tests⁸³.

Table 20 shows a summary of the different levels of evidence found for the criterion-related validity of cardiorespiratory fitness tests.

Cardiorespiratory fitness

Distance and time-based run/walk tests

Seventeen high quality studies examined the criterion-related validity of the distance run/walk or walk tests (see **Table 17**). Four and two studies showed that the 2-km walk⁸⁴⁻⁸⁷ and 1.5-mile run/walk^{88, 89} tests, respectively, were valid for assessing cardiorespiratory fitness ($r = 0.80$ – 0.93 , all $p < 0.05$). Four studies⁹⁰⁻⁹³ observed that the 1-mile walk test was an accurate test for estimating VO_{2max} ($r = 0.81$ – 0.88 , all $p < 0.05$), while another two studies^{94, 95} showed that it was not a valid test ($r = 0.69$, 13.3% E, $p < 0.05$; mean differences range from 2.360 to 9.131 mL/kg/min, all $p < 0.001$, respectively). The treadmill jogging test reported contradictory results: one study⁹⁶ found it to have high validity for assessing cardiorespiratory fitness ($r = 0.84$, both $p < 0.001$); whereas another study⁸⁹ revealed that it was not a valid test ($r = 0.50$, $p < 0.05$).

Five high quality studies investigated the criterion-related validity of the time-based run/walk or walk tests (see **Table 17**). These studies showed that the 3-minute walk⁹⁷, 6-minute walk⁹⁸⁻¹⁰⁰, and the 12-minute run/walk⁸⁹ tests were valid for assessing cardiorespiratory fitness ($r = 0.70$ – 0.95 , all $p < 0.05$). Additionally, one original high quality study reported that the University Montreal test¹⁰¹ was valid for estimating cardiorespiratory fitness ($r = 0.71$, $p < 0.001$; mean difference = 0.025 ± 7.445 mL/kg/min., $p > 0.05$).

A meta-analysis⁸⁰ consisting of 102 studies on adults determined that the criterion-related validity of the distance run/walk field tests for estimating cardiorespiratory fitness ranged from low to high, with the 1.5-mile ($r_p = 0.80$; 95% CI: 0.72–0.80) and 12-minute run/walk tests ($r_p = 0.79$; 95% CI: 0.71–0.87) being the best predictors (see **Table 19**).

One very low quality study suggested that the University Montreal test¹⁰² was valid for estimating VO_{2max} ($r = 0.96$, $p < 0.05$). Low quality studies reported that the 400-m¹⁰³ ($p = -0.56$, $p < 0.001$) and 6-minute walk¹⁰⁴ tests ($r = 0.85$, $p < 0.05$) had high validity; the 1500-m fast walk test¹⁰⁵ had moderate validity ($r = 0.52$, $p < 0.01$); and the 2.4-km run¹⁰⁶, 1-mile walk¹⁰⁶, and 4-minute run¹⁰⁷ tests had low validity ($r = 0.42$, $p = 0.13$; $r =$

0.18, $p = 0.52$; mean difference = 1.6 ± 3.6 mL/kg/min, $p < 0.05$, respectively) to assess $\text{VO}_{2\text{max}}$ (see **Table 17**).

20-m shuttle run test

Nine high quality studies analysed the criterion-related validity of the 20-m SRT^{89, 108-112} or modifications of it^{109, 111, 113-115} (see **Table 17**). Four studies^{89, 109-111} reported that the 20-m SRT was a valid test for assessing cardiorespiratory fitness ($r = 0.82\text{--}0.94$, all $p < 0.05$). However, one study¹¹² concluded that this test was not valid for assessing cardiorespiratory fitness (mean differences range from -0.54 ± 6.23 to -2.94 ± 6.55 mL/kg/min, all $p < 0.01$). Two studies^{113, 114} proved that the incremental shuttle walk test was not valid ($r = 0.72$, 19% E, both $p < 0.001$), while one study¹¹⁵ found that this test was valid for assessing cardiorespiratory fitness (mean difference = 0.14 ± 9.27 mL/kg/min, $p > 0.05$). Moreover, two studies^{109, 111} reported that the 20-m square SRT was valid ($r = 0.95$, both $p < 0.001$).

A meta-analysis⁷⁹ which included 24 studies in adults found that the 20-m SRT had a moderate-to-high criterion-related validity for estimating $\text{VO}_{2\text{max}}$ ($r_p = 0.79\text{--}0.94$; 95% CI: 0.56–1.00) (see **Table 19**).

One low quality study¹¹⁶ investigated the criterion-related validity of the 20-m SRT, reporting that this test was valid to assess cardiorespiratory fitness ($r = 0.90$, $p < 0.001$) (see **Table 17**).

Step tests

Eleven high quality studies analysed the criterion-related validity of the step tests (see **Table 17**). Four studies observed that the Danish step¹¹⁷, the Queen's College step¹¹⁸, and the 2-minute step¹¹⁹ tests were not valid for estimating $\text{VO}_{2\text{max}}$ ($r = 0.034\text{--}0.72$, all $p < 0.05$). However, another eight studies proved the validity of the modified Canadian aerobic fitness¹²⁰, 6-minute single 15 cm step¹²¹, YMCA step¹²²⁻¹²⁶, Tecumseh step¹²⁵, and modified Harvard step¹²⁷ tests ($r = 0.80\text{--}0.91$, all $p < 0.05$).

A systematic review⁸¹, comprised of 11 studies on adults investigated the criterion-related validity of the step tests (see **Table 19**). Validity measures were varied, and a broad range of correlation coefficients were reported across the 11 studies ($r = 0.469\text{--}0.95$; all $p < 0.005$) with conflicting results in most of the step test protocols. The study concluded that the Chester step test was the best predictor for assessing cardiorespiratory fitness.

Three low quality studies revealed that the 6-minute step^{128, 129}, and 3-minute single step¹³⁰ tests, were valid ($r = 0.72\text{--}0.92$, both $p < 0.05$) (see **Table 17**). However, 2-

level step¹³¹, the USO step¹³², modified Harvard step¹³³, and Danish step¹³⁴ tests reported low validity ($r = 0.47-0.66$, all $p < 0.01$).

Muscular fitness

Table 21 shows a summary of the different levels of evidence found for the criterion-related validity of muscular fitness, flexibility and motor fitness tests.

Maximal isometric muscular fitness

Four high quality studies assessed the criterion-related validity of maximal isometric muscular fitness, using the handgrip tests (see **Table 17**). Three reported that the TKK dynamometer¹³⁵⁻¹³⁷ was valid (mean difference range -0.20 , $p > 0.05$ to 2.02 kg $p < 0.001$) ($r = 0.98$, $p < 0.001$). While three studies showed inconclusive results about the validity of the DynEx dynamometer^{135, 137, 138}, and two studies observed that the Jamar dynamometer^{135, 138} was less accurate than the TKK and DynEx dynamometer for estimating maximal isometric muscular fitness.

Two and one low quality studies observed that the Jamar^{139, 140}, and Rolyan¹³⁹ dynamometer, respectively, were valid ($r = 0.99$, both $p < 0.05$). Moreover, one very low quality study concluded that the modified flex-hang arm test¹⁴¹ was valid to assess upper-body relative isometric muscular fitness ($r = 0.88$, $p < 0.001$) (see **Table 17**).

Endurance muscular fitness

Four high quality studies assessed the criterion-related validity of trunk endurance muscular fitness (see **Table 17**). Two studies^{142, 143} suggested that the Biering-Sørensen ($r = 0.84-0.98$, $p < 0.01$) test was valid, whereas another study¹⁴⁴ reported acceptable validity ($r = 0.60-0.71$, $p < 0.05$). One study showed that the prone bridging test¹⁴⁵ was valid to assess trunk endurance muscular fitness (no mean difference, $p > 0.05$).

Thirteen low quality studies examined the criterion-related validity of endurance muscular fitness (see **Table 17**). The Baumgartner modified pull-up test¹⁴⁶ turned out to be valid to assess upper-body endurance muscular fitness ($r = 0.60-0.85$, $p < 0.05$), while the back-leg-chest test¹⁴⁷ had moderate-high validity ($r = 0.56-0.81$, all $p < 0.01$). The standard push-up test¹⁴⁸ had moderate validity to assess relative upper-body endurance muscular fitness ($r = 0.7$, $p < 0.001$) but not the hand-release push-up test¹⁴⁸ ($r = 0.39$, $p < 0.05$). Three studies reported that the curl-up test or its modifications¹⁴⁹⁻¹⁵¹ were not valid to assess trunk endurance muscular fitness ($r = 0.32-0.68$, $p < 0.05$). While three studies¹⁵²⁻¹⁵⁴ showed that the Biering-Sørensen test had moderate ($r = 0.75$, $p < 0.001$) to high ($r = 0.92-0.94$, $p < 0.05$) validity to assess trunk endurance muscular fitness. Four

studies¹⁵⁵⁻¹⁵⁷ found inconclusive results when analysing the validity of the sit-to-stand test different versions to assess lower-body endurance muscular fitness.

Two very low quality studies observed that the flexed-arm hang¹⁵⁸, and bent-knee push-up and revised push-up¹⁵⁹ tests had acceptable validity to assess upper-body endurance muscular fitness ($r = 0.67-0.72$, $p < 0.05$); and four very low quality studies concluded that different versions of the curl-up test¹⁶⁰⁻¹⁶³ were not valid to assess trunk endurance muscular fitness ($r = 0.04-0.51$, $p < 0.05$) (see **Table 17**).

Explosive muscular fitness

Only one high quality study assessed the criterion-related validity of the explosive muscular fitness (see **Table 17**). This study concluded that the Sargent test¹⁶⁴ was not valid (mean difference: 4.4 ± 5.1 , $p < 0.001$) for estimating lower-body explosive muscular fitness.

One very low quality study reported that the medicine ball put test¹⁶⁵ is valid to assess upper-body explosive muscular fitness in females and males ($r = 0.86$ and 0.79 , respectively, both $p < 0.01$) (see **Table 17**).

Flexibility

Only one study¹⁶⁶ that examined the criterion-related validity of flexibility tests (sit-and-reach tests) was classified as high quality (see **Table 17**). They found that the sit-and-reach was not a valid test ($r = 0.44-0.48$, $p < 0.05$).

A meta-analysis⁸² which includes 28 studies in adults, found that the sit-and-reach test and its different versions had moderate validity for estimating hamstring extensibility (r_p ranged from 0.49 ; 95%CI: $0.29-0.68$ to 0.68 ; 95% CI: $0.55-0.80$), but a low validity for estimating lumbar extensibility (r_p ranged from 0.16 ; 95% CI: $-0.10-0.41$ to 0.35 ; 95% CI: $0.15-0.54$). Moreover, another meta-analysis⁸³ carried out in adults (six studies), reported that the toe-to-touch test had moderate validity to assess hamstring extensibility ($r_p = 0.66$; 95% CI: $0.56-1.00$) (see **Table 19**).

Six low quality studies determined that the sit-and-reach test¹⁶⁷⁻¹⁷² and its different versions^{171, 172} had low-moderate validity to assess low back and hamstring flexibility ($r = 0.12-0.82$, $p < 0.05$). However, the toe-to-touch test^{173, 174} showed high validity to assess hamstring flexibility ($r = 0.81-0.85$, both $p < 0.05$). One low quality studies reported that the horizontal and vertical hip joint angle tests¹⁷⁵ had moderate validity ($r = 0.79$, $p < 0.01$).

Two studies classified as very low quality observed that the sit-and-reach¹⁷⁶ ($r = 0.90$, $p < 0.05$), and the adapted Kraus-Weber¹⁷⁷ tests ($r = 0.74-0.82$, $p < 0.05$) were valid to assess hip and hamstring and low back flexibility, respectively (see **Table 17**).

Motor fitness

No study investigating criterion-related validity of motor fitness tests was classified as high quality (see **Table 17**).

Two low quality studies found that the ten step test^{178, 179} had moderate validity to estimate agility ($r = 0.35-0.68$, $p < 0.01$). Another low quality study concluded that the Romberg test¹³¹ was not valid to assess the balance ($r = -0.18-0.23$, $p > 0.05$) (see **Table 17**).

The levels of evidence found about the criterion-related validity of field-based physical fitness tests in adults is summarized in **Figure 7**.

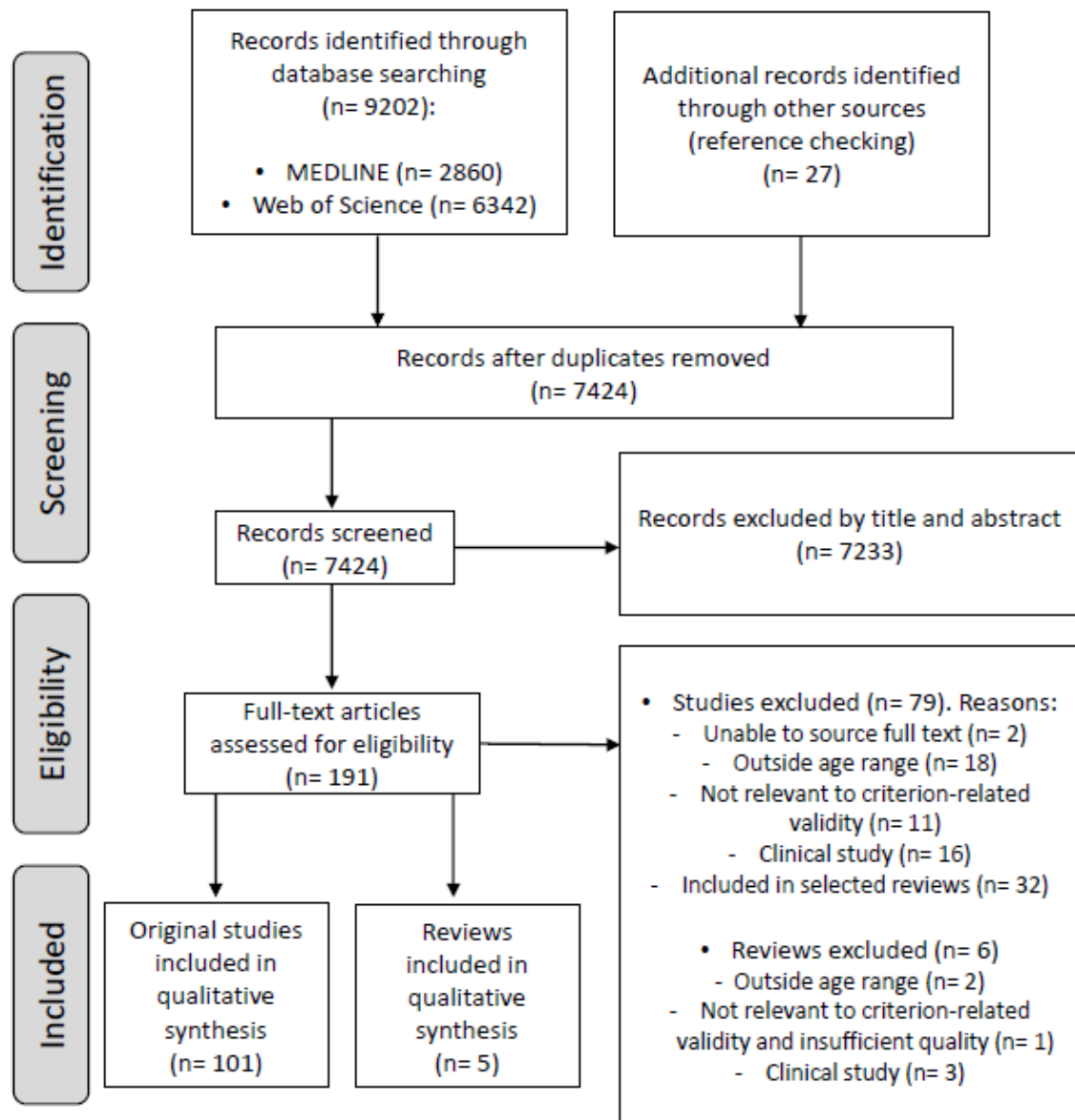


Figure 6. Flow-chart of retrieved and selected articles.

Table 16. Quality assessment of fitness criterion-related validity studies.

Study	Fitness Component	Number of study subjects	Description of the study population	Statistical analysis	Total score
Leger & Boucher 1980	Cardiorespiratory fitness	1	0	1	2
Santa María et al. 1976	Cardiorespiratory fitness	1	2	0	3
Siconolfi et al. 1985	Cardiorespiratory fitness	1	1	1	3
Ritchie et al. 2005	Cardiorespiratory fitness	1	2	0	3
Azmi & Sulaiman, 2017	Cardiorespiratory fitness	1	2	0	3
McGawley 2017	Cardiorespiratory fitness	0	1	2	3
Giacomantonio et al. 2020	Cardiorespiratory fitness	1	2	0	3
Leger et al. 1988	Cardiorespiratory fitness	2	1	1	4
Flouris et al. 2010	Cardiorespiratory fitness	1	2	1	4
Gabriel et al. 2010	Cardiorespiratory fitness	2	2	0	4
Hansen et al. 2011	Cardiorespiratory fitness	2	2	0	4
Burr et al. 2011	Cardiorespiratory fitness	1	2	1	4
Mikawa & Senjyu et al. 2011	Cardiorespiratory fitness	2	2	0	4
Arcuri et al. 2015	Cardiorespiratory fitness	2	2	0	4
Jurgensen et al. 2015	Cardiorespiratory fitness	2	2	0	4
Lerche et al. 2017	Cardiorespiratory fitness	2	2	0	4
Siconolfi et al. 1982	Cardiorespiratory fitness	2	2	1	5
Kline et al. 1987	Cardiorespiratory fitness	2	2	1	5
Oja et al. 1991	Cardiorespiratory fitness	2	2	1	5
Laukkanen et al. 1992	Cardiorespiratory fitness	2	2	1	5
Weller et al. 1992	Cardiorespiratory fitness	2	2	1	5

Laukkanen et al. 1993	Cardiorespiratory fitness	2	2	1	5
George et al. 1993	Cardiorespiratory fitness	2	2	1	5
George et al. 1998	Cardiorespiratory fitness	2	2	1	5
McNaughton et al. 1998	Cardiorespiratory fitness	1	2	2	5
Greenhalgh et al. 2001	Cardiorespiratory fitness	1	2	2	5
Larsen et al. 2002	Cardiorespiratory fitness	2	2	1	5
Flouris et al. 2004	Cardiorespiratory fitness	1	2	2	5
Cooper et al. 2005	Cardiorespiratory fitness	1	2	2	5
Kumar et al. 2012	Cardiorespiratory fitness	1	2	2	5
Seneli et al. 2013	Cardiorespiratory fitness	1	2	2	5
Carvalho et al. 2015	Cardiorespiratory fitness	1	2	1	5
Teren et al. 2016	Cardiorespiratory fitness	2	2	1	5
Guo et al. 2018	Cardiorespiratory fitness	1	2	2	5
Manttari et al. 2018	Cardiorespiratory fitness	2	2	1	5
Ricci et al. 2019	Cardiorespiratory fitness	1	2	2	5
Bonet et al. 2020	Cardiorespiratory fitness	1	2	2	5
Kieu et al. 2020	Cardiorespiratory fitness	1	2	2	5
Dolgener et al. 1994	Cardiorespiratory fitness	2	2	2	6
Laukkanen et al. 2000	Cardiorespiratory fitness	2	2	2	6
Flouris et al. 2005	Cardiorespiratory fitness	2	2	2	6
Metsios et al. 2008	Cardiorespiratory fitness	2	2	2	6
Kim et al. 2011	Cardiorespiratory fitness	2	2	2	6
Aadahl et al. 2012	Cardiorespiratory fitness	2	2	2	6
Cao et al. 2013	Cardiorespiratory fitness	2	2	2	6

Lunt et al. 2013	Cardiorespiratory fitness	2	2	2	6
Beutner et al. 2015	Cardiorespiratory fitness	2	2	2	6
Di Thommazo-Luporini et al. 2015	Cardiorespiratory fitness	2	2	2	6
Di Thommazo-Luporini et al. 2016	Cardiorespiratory fitness	2	2	2	6
Hansen et al. 2016	Cardiorespiratory fitness	2	2	2	6
Jamnick et al. 2016	Cardiorespiratory fitness	2	2	2	6
Jurio-Iriarte et al. 2017	Cardiorespiratory fitness	2	2	2	6
Jurio-Iriarte et al. 2018	Cardiorespiratory fitness	2	2	2	6
Hong et al. 2019	Cardiorespiratory fitness	2	2	2	6
Lee et al. 2019	Cardiorespiratory fitness	2	2	2	6
Lima et al. 2019	Cardiorespiratory fitness	2	2	2	6
Berger 1966	Musculoskeletal fitness	1	0	0	1
Wells & Dillon 1952	Musculoskeletal fitness	2	0	0	2
DeWitt 1944	Musculoskeletal fitness	2	0	0	2
Mathews 1957	Musculoskeletal fitness	2	0	0	2
Harvey & Scott 1967	Musculoskeletal fitness	2	0	0	2
Knudson & Johnston 1995	Musculoskeletal fitness	1	1	0	2
Clemons et al. 2004	Musculoskeletal fitness	2	0	0	2
Wood & Baumgartner 2004	Musculoskeletal fitness	2	0	0	2
Clemons et al. 2010	Musculoskeletal fitness	1	1	0	2
Clemons et al. 2014	Musculoskeletal fitness	1	0	1	2
Cotten 1972	Musculoskeletal fitness	2	1	0	3
Jackson & Langford 1989	Musculoskeletal fitness	2	0	1	3

Kipper & Parker 1987	Musculoskeletal fitness	1	2	0	3
Liemohn et al. 1994	Musculoskeletal fitness	1	1	1	3
Diener et al. 1995	Musculoskeletal fitness	1	2	0	3
Mannion et al. 1997	Musculoskeletal fitness	1	0	2	3
Mathiowetz 2002	Musculoskeletal fitness	2	1	0	3
Baumgartner & Gaunt 2005	Musculoskeletal fitness	2	1	0	3
Ritchie et al. 2005	Musculoskeletal fitness	1	2	0	3
López-Miñarro et al. 2015	Musculoskeletal fitness	2	0	1	3
Silva et al. 2015	Musculoskeletal fitness	1	2	0	3
Broer & Galles 1958	Musculoskeletal fitness	2	2	0	4
Hall et al. 1992	Musculoskeletal fitness	2	2	0	4
Harkonen 1993	Musculoskeletal fitness	N/A	N/A	1	Low
Sparto et al. 1997	Musculoskeletal fitness	0	2	2	4
Knudson et al. 2001	Musculoskeletal fitness	1	2	1	4
Youdas et al. 2008	Musculoskeletal fitness	2	2	0	4
Bohannon et al. 2010	Musculoskeletal fitness	2	1	1	4
Ayala et al. 2012	Musculoskeletal fitness	1	2	1	4
Mier et al. 2013	Musculoskeletal fitness	2	2	0	4
Taylor et al. 2013	Musculoskeletal fitness	1	2	1	4
Ten Hoor et al. 2016	Musculoskeletal fitness	1	2	1	4
Clemons et al. 2018	Musculoskeletal fitness	1	2	1	4

Applegate et al. 2019	Musculoskeletal fitness	1	2	1	4
Mannion et al. 1994	Musculoskeletal fitness	2	2	1	5
Kankaanpää et al. 1998	Musculoskeletal fitness	2	2	1	5
Shechtman et al. 2005	Musculoskeletal fitness	2	2	1	5
Coorevits et al. 2008	Musculoskeletal fitness	1	2	2	5
Bui et al. 2015	Musculoskeletal fitness	1	2	2	5
De Blaiser et al. 2018	Musculoskeletal fitness	1	2	2	5
Kawano et al. 2010	Musculoskeletal fitness	2	2	2	6
España-Romero et al. 2010	Musculoskeletal fitness	N/A	N/A	2	High
Cadenas-Sánchez et al. 2016	Musculoskeletal fitness	N/A	N/A	2	High
Kolimechkov et al. 2020	Musculoskeletal fitness	N/A	N/A	2	High
Ritchie et al. 2005	Motor fitness	1	2	0	3
Miyamoto et al. 2008 ^a	Motor fitness	2	1	1	4
Miyamoto et al. 2008 ^b	Motor fitness	2	2	0	4

N/A indicates not applicable because the criterion-related validity of the study is carried out on the dynamometer, not in subjects.

Total score indicates high quality = 5-6; low quality = 3-4; very low quality = 0-2.

Table 17. Overview of field-based physical fitness tests criterion-related validity studies in adults.

Author	Participants	Age	Field-based Test	Gold standard	Statistical methods	Main outcome	Conclusion
<i>Very low quality studies</i>							
<i>Cardiorespiratory fitness</i>							
Leger & Boucher 1980	Adults= 25 (Healthy)	24.4 ± 2.8	The University Montreal track (UM- track)	Gas analyzer in maximal treadmill test (Modified Balke protocol)	Correlation coefficient (r), SEE	r= 0.96 (p< 0.05), SEE= 2.81 ml/kg/min, between UM-track test and VO _{2max} .	The UM-track test is accurate, valid and safe for young and middle-age adults, males and females, whether they are trained or not.
<i>Musculoskeletal fitness (isometric)</i>							
Clemons et al. 2014	Females= 31 (Healthy)	19-36	Modified flexed-arm hang	1RM (Absolute and relative isometric muscular fitness)	Correlation coefficient (r), linear regression	r= 0.10 (p= 0.61) between modified flexed arm hang and absolute isometric muscular fitness. r= 0.88 (p< 0.001) between modified flexed arm hang and relative isometric muscular fitness.	Modified flexed-arm hang scores were not related to absolute muscular fitness, however, is a valid estimate of the construct of relative isometric muscular fitness.
<i>Musculoskeletal fitness (endurance)</i>							
Clemons et al. 2004	Females= 60 (Healthy)	20.7 ± 2.15	Flexed-arm hang	1RM (absolute muscular fitness) 1RM/mass (relative muscular fitness)	Correlation coefficient (r)	r ranges from -0.14 to 0.72 between flexed-arm hang and the different criterion measures.	No significant relationship was found between any of the flexed-arm hang variations and absolute muscular fitness or muscle endurance;

				Repetitions at 70% of 1RM (muscle endurance)			
Wood & Baumgartner 2004	Females= 87 (Healthy)	College	Bent-knee push-up Revised push- up	1RM bench press	Correlation coefficient (r)	r= 0.72 (p< 0.01) between U90° variation and relative muscular fitness. r= 0.67 between bent-knee push-up and bench press. r= 0.68 between revised push-up and bench press.	however, all flexed-arm hang variations correlated significantly with relative muscular fitness. The validity for the bent- knee push-up test is acceptable, however, somewhat more questionable. Because the validity coefficient for both the bent-knee and the revised push-up tests are similar, either test can be used.
Harvey & Scott 1967	Females= 60 (Healthy)	College	Modified curl- up (30 s) Curl-up	Dynamometer (Elgin Multiple Angle)	Correlation coefficient (r)	r= 0.44 (p< 0.05) between modified curl-up and dynamometer scores. r= 0.32 (p< 0.05) between curl-up and dynamometer scores.	Both tests have low validity.
Berger 1966	Adults= 47 (Healthy)	College	2-minute sit- up	1 RM sit-up	Correlation coefficient (r)	r= 0.51 (p< 0.05) between 2- minute sit-up and 1-RM sit- up.	The 2-minute sit-up test is not valid to assess endurance of the abdominal muscles.
Knudson & Johnston 1995	Females= 10 Males= 10 (Healthy)	19-32	Modified curl- up (2 min, legs elevated on a bench, hands interlocked on the abdomen)	Cibex isokinetic dynamometer	Correlation coefficient (r)	r= 0.38 (p=0.07) between the modified curl-up and isokinetic dynamometer.	Modified curl-up test has low validity.

DeWitt 1944	Males= 102 (Healthy)	College	1° Modified curl-up (touch knees with opposites elbow) 2° Modified curl-up (from setting, hands interlocked behind the neck, touch knees with opposites elbow) 3° Modified curl-up (held feet)	Dynamometer	Correlation coefficient (r)	1° modified curl-up: muscular fitness, r= 0.04; endurance, r= 0.25 with dynamometer. 2° modified curl-up: muscular fitness, r= 0.16; endurance, r= 0.37 with dynamometer. 3° modified curl-up: muscular fitness, r= 0.14; endurance, r= 0.26 with dynamometer.	Modified curl-up variations have low validity.
<i>Musculoskeletal fitness (explosive)</i>							
Clemons et al. 2010	Females= 24 Males= 19 (Healthy)	22.1 ± 1.82 23.1 ± 2.95	Medicine ball put	Bench press power	Correlation coefficient (r)	r= 0.86 and 0.79 (p< 0.01 for both) for female and male, respectively, between bench press power and medicine ball put.	The medicine ball put test is a valid method to use in assessing upper-body power for both college-age males and females.
<i>Musculoskeletal fitness (flexibility)</i>							
Wells & Dillon 1952	Adults=100 (Healthy)	College	Sit-and-reach	The standing bobbing test	Correlation coefficient (r)	r= 0.90 between the standing bobbing and sit-and-reach tests.	The sit-and-reach is a valid test of back and leg flexibility.

Mathews 1957	Females= 66 (Healthy)	College	Adapted Kraus-Weber floor touch test and Wells sit-and-reach	Leighton flexometer	Correlation coefficient (r)	r= 0.82 between adapted Kraus-Weber floor touch and Leighton flexometer. r= 0.74 between Wells sit-and-reach and Leighton flexometer.	Adapted Kraus-Weber floor touch and Wells sit-and-reach tests are valid test to assess hip flexibility.

Low quality studies

Cardiorespiratory fitness

Ritchie et al. 2005	Females= 10 Males= 13 (Healthy)	55-70	2-level step	Gas analyzer in submaximal cycle test	Correlation coefficient (r)	r= -0.60 (p< 0.001) between step test and submaximal exercise test.	The 2-level step test was found to be a valid test.
Giacomantonio et al. 2020	Females= 16 Males= 12 (Risk of cardiovascular disease)	>17	6-minute step	Gas analyzer in maximal treadmill test	Correlation coefficient (r), CI	r= 0.88 (CI: 0.57-0.97) between 6-minute step test and VO _{2peak} .	The 6-minute step test seems to be a valid option for assessing cardiorespiratory fitness of people at risk of cardiovascular disease.
Arcuri et al. 2015	Females= 49 Males= 42 (Healthy)	18-86	6-minute single 20-cm step	6-minute walk test	Correlation coefficient (r)	r= 0.72 (p< 0.001)	6-minute single 20-cm step test is a valid test.
Siconolfi et al. 1985	Females= 19 Males= 29 (Healthy)	19-70	3-minute single 25.4-cm step	Gas analyzer in submaximal cycle test (Modified Astrand-Rhyming protocol)	Correlation coefficient (r), SEE	r= 0.92 (p= 0.01) between step test and directly measured VO _{2 max} . SEE= 0.30 ml/kg/min.	The 3-minute single 25.4-cm step test protocol provides valid estimates of cardiorespiratory fitness over a wide age range.

Santa María et al. 1976	Males= 21 (Healthy)	22.54 ± 4.12	USO step	Gas analyzer in submaximal cycle test (Astrand method)	Correlation coefficient (r)	r= 0.47 between USO step test and VO _{2max} .	USO step test seems to have a poor validity.
Hansen et al. 2011	Females= 60 Males= 53 (Healthy)	23-75	Modified Harvard step	Gas analyzer in maximal cycle test	Paired t-tests	r= 0.54 (p< 0.01) between modified Harvard step test and VO _{2max} .	Standardized fixed-rate modified Harvard step test elicited vigorous exercise intensities, especially in small, obese, and/or physically deconditioned subjects.
Lerche et al. 2017	Females= 65 Males= 60 (Healthy)	20-63	Danish step	Gas analyzer in maximal treadmill test	Correlation coefficient (r)	r= 0.56 and 0.66 (p< 0.001 for both) between Danish step test and VO _{2max} for males and females, respectively.	The Danish step test is a valid assessment method for females, however not for males (under estimation of cardiorespiratory fitness).
Leger et al. 1988	Females= 24 Males= 53 (Healthy)	18-50	20-m shuttle run	Douglas bag methods	Multiple regression, SEE	r= 0.90 between VO _{2max} and maximal speed. Y=31.025+3.238*speed+3.248*age+0.1536*speed*age (r= 0.71; SEE=5.1 ml/kg/min)	20-m SRT is a valid method to measure VO _{2max} in adult population.
Jurgensen et al. 2015	Females= 46 (Obese)	18-46	Incremental shuttle walking	Gas analyzer in maximal treadmill test	Correlation coefficient (r)	r= 0.64 (p< 0.05).	The incremental shuttle walking test had acceptable validity in this population.
Mikawa & Senjyu et al. 2011	Males= 68 (Healthy)	40-59	15-m incremental shuttle walk	Gas analyzer in maximal cycle test	Linear regression	r= 0.86 (p< 0.01) between 15-m incremental shuttle walk and run test and	The 15-m incremental shuttle walk and run test is highly recommended

			and run 1500-m fast walk			VO _{2max} . r= 0.52 (p< 0.01) between 1500-m fast walk test and VO _{2max} .	as a field test for evaluating cardiorespiratory fitness in middle-aged adults. The 1500-m fast walk test is not valid test in this population.
Flouris et al. 2010	Males= 45 (Healthy)	18-29	15-m square shuttle run	Gas analyzer in maximal treadmill test	Correlation coefficients (r), linear regression	r= 0.87-0.92 (p< 0.001) between 15-m square shuttle run test and VO _{2max} .	15-m square shuttle run test is a highly valid predictive test for VO _{2max} .
McGawley 2017	Females= 5 Males= 5 (Healthy)	32 ± 7	4-minute running time trial	Gas analyzer in STEP treadmill test + VER phase	Bland-Altman method, Two-way repeated measures ANOVA	Significant difference between STEP treadmill test and 4-minute running time trial (p= 0.008). Mean difference= 1.6 ± 3.6 ml/kg/min (p= 0.459) for STEP treadmill test vs 4- minute running time trial.	VO _{2max} values attained during 4-minute running time trial were significantly lower than those attained during STEP.
Azmi & Sulaiman, 2017	Females= 15 (Healthy)	20-24	Harvard step 20-m shuttle run 2.4-km 1-mile walk	Gas analyzer in maximal cycle test (Bruce protocol test)	Correlation coefficient (r)	r= 0.31 (p= 0.27), between Harvard Step test and VO _{2max} . r= 0.64 (p< 0.05), between 20-m SRT and VO _{2max} . r= 0.42 (p= 0.13), between 2.4-km run test and VO _{2max} . r= 0.18 (p= 0.52), between 1-mile walk test and VO _{2max} .	The most suitable cardiorespiratory test to measure VO _{2max} in healthy females adults is the 20-m SRT.

Gabriel et al. 2010	Females= 66 (Healthy)	45-65	400-m walk	Gas analyzer in maximal treadmill test	Spearman correlation coefficient (ρ)	$\rho = -0.43$ ($p < 0.001$), between 400-m walk test and estimated VO_{2max} . $\rho = -0.56$ ($p < 0.001$), between 400m walk test and measured VO_{2max} .	The 400-m walk test is valid for estimating cardiorespiratory fitness in healthy middle-aged females.
Burr et al. 2011	Females= 21 Males= 23 (Healthy)	25-59	6-minute walk	Gas analyzer in maximal treadmill test (Modified Bruce protocol test) Douglas bag methods	Spearman correlation coefficient (r), multiple linear regression	$r = 0.49$ ($p < 0.001$), between distance walked and VO_{2max} . $R^2 = 0.72$ ($p < 0.05$), between prediction equation and VO_{2max} .	The 6-minute walk test is valid for estimating cardiorespiratory fitness.

Musculoskeletal fitness (isometric)

Mathiowetz 2002	Females= 30 Males= 30 (Healthy)	20-50	Rolyan dynamometer Jamar dynamometer	Known weights	Correlation coefficient (r), paired t-test	$r = 0.99$ ($p < 0.05$), between Rolyan and Jamar dynamometers with known weights. Differences between Rolyan and Jamar dynamometer $< 2.3\%$ ($p > 0.05$).	Rolyan and Jamar dynamometer measure equivalently for practical purposes and are valid test to assess isometric muscular fitness.
Harkonen 1993	NA	NA	Jamar dynamometer	Known weights	Errors	Error > 2 kg was found for dynamometers that were purchased in the years 1985, 1986 and 1987.	Jamar dynamometers seem to be rather accurate.

Musculoskeletal fitness (endurance)

Baumgartner & Gaunt 2005	Females= 40 Males= 42 (Healthy)	21.1 ± 3.8 21.0 ± 1.6	Baumgartner modified pull-up	1 RM bench press Modified lateral pull-down	Correlation coefficient (r)	r= 0.61 females and r= 0.67 males, between Baumgartner modified pull-up and bench press. r= 0.60 females and r= 0.85 males, between Baumgartner modified pull-up and modified lateral pull-down.	Baumgartner modified pull-up test is valid to assess arm and shoulder girdle muscular fitness and endurance in college males and females.
Clemons et al. 2018	Males= 31 (Healthy)	18-29	Standard push-up Hand-release push-up	1 RM bench press	Correlation coefficient (r), multiple linear regression	r= 0.25 (p= 0.18) and r= 0.7 (p< 0.001) between standard push-up with absolute and relative 1RM. rho= -0.05 (p= 0.787) and rho= 0.39 (p< 0.05) between hand-release push-up with absolute and relative 1RM.	Standard push-up has moderate validity to assess relative upper-body endurance muscular fitness but not absolute. Hand-release push-up is not valid test to assess absolute and relative upper-body endurance muscular fitness.
Bohannon et al. 2010	Females= 111 Males= 70 (Healthy)	14-85	5-repetition sit-to-stand	Isokinetic dynamometer muscular fitness	Correlation coefficient (r), multiple linear regression	r= -0.54 and -0.56 (p< 0.001 for both) between isokinetic muscular fitness and 5-repetition sit-to-stand test for 14-85 years and 50-85 years respectively. R ² = 0.30 isokinetic muscular fitness predicting 5-repetition sit-to-stand test in 14-85 and 50-85 group.	There is a significant relationship between knee extension muscular fitness and 5-repetition sit-to-stand test. Nevertheless, knee extension muscular fitness alone provides an insufficient explanation of 5-repetition sit-to-stand test performance. Age, body weight, and stature also influence 5-

							repetition sit-to-stand performance.
Silva et al. 2015	Stroke= 18 Healthy= 18	≥ 20	5-repetition sit-to-stand	Isokinetic dynamometer	Correlation coefficient (r)	r range from -0.41 to -0.51 (all $p < 0.05$) between total duration of 5-repetition sit-to-stand test and muscular fitness of the trunk	5-repetition sit-to-stand test obtained from low to moderate correlation with trunk muscular fitness.
Taylor et al. 2013	Females= 15 Males= 15 (Healthy)	23 ± 1	8-repetition maximum test of knee extensors	Isokinetic dynamometer	Correlation coefficient (r), linear regression	$R^2 = 0.73, 0.73$ and 0.51 (all $p < 0.001$) between for peak torques at 60, 120 and 240 degrees per second, respectively, between 8-repetition and isokinetic muscular fitness.	8-repetition maximum testing of knee extensors can be used as a valid measure of lower-body muscular fitness.
Ten Hoor et al. 2016	Females= 22 Males= 23 (Healthy)	18-35	Back-leg-chest	Isokinetic dynamometer	Correlation coefficient (r), stepwise linear regression	r range from 0.56 to 0.81 (all $p < 0.01$) and from 0.60 to 0.71 (all $p < 0.01$) between back-leg-chest and criterion measures in males and females respectively. $R^2 = 0.74$ for dominant knee extensor and flexor muscular fitness predicting back-leg-chest muscular fitness. $R^2 = 0.86$ for predicting back-leg-chest muscular fitness.	Back-leg-chest test can be used to evaluate changes in muscular fitness in adults.
Diener et al. 1995	Females= 21 Males= 15	17-68	Modified curl-up (reach 8.89cm with	Isometric spine flexion muscular fitness (strength	Correlation coefficient (r)	$r = 0.44$ ($p < 0.01$) between the modified curl-up test and	Modified curl-up test is not valid to measure abdominal muscular

	(Healthy)		fingertip)	table)		isometric abdominal muscular fitness.	fitness and endurance.
Ritchie et al. 2005	Females= 10 Males= 13 (Healthy)	55-70	Sit-to-stand Lift and reach	1RM bench press, 1RM leg press,	Correlation coefficient (r)	r= 0.68 (p< 0.05) between sit-to-stand test and 1RM leg press. r= 0.43 (p>0.05) between lift and reach and 1RM bench press.	Only sit-to-stand was found to be valid when compared 1RM test. Single timed chair raise is not a valid test.
Knudson et al. 2001	Females= 22 Males= 22 (Healthy- active)	25.8 ± 5.0	Curl-up Modified curl- up (90s, legs elevated on a bench, hands interlocked on the abdomen)	Isokinetic dynamometer	Correlation coefficient (r), linear regression, SEE	r= 0.36, SEE= 0.10 and r= 0.21, SEE= 0.11 (p>0.05 for both) between the curl-up and isokinetic muscular fitness for males and females, respectively. r= 0.50, SEE=0.08 and r= 0.46, SEE=0.08 (p< 0.05 for both) between the curl-up and isokinetic muscular endurance for males and females, respectively. r= 0.07, SEE=0.10 and r= - 0.19, SEE=0.10 (p>0.05 for both) between the modified curl-up and isokinetic muscular fitness for males and females, respectively. r= 0.23, SEE= 0.09 and r= 0.10, SEE= 0.09 (p> 0.05 for both) between the modified curl-up and isokinetic muscular endurance for	The curl-up test and the modified curl-up have weak validity to assess abdominal muscular fitness and endurance.

						males and females, respectively.	
Hall et al. 1992	Females= 28 Males= 23 (Healthy)	22.2 ± 4.6 23.1 ± 7.4	1° Modified curl-up (knees extend, hand interlocked behind the neck, held feet, flex up to 90°) 2° Modified curl-up (arm folded across the chest, feet held, flex up to 90°) Curl-up	Isokinetic dynamometer	Correlation coefficient (r)	1° Modified Curl-up: r= 0.42 and 0.40 (p< 0.01, for both) in males, for concentric and eccentric peak torque, respectively; and r= -0.18 and -0.21 (p> 0.05 for both) in females for concentric and eccentric peak torque, respectively. 2° Modified Curl-up: r= - 0.07 and -0.08 (p>0.05 for both) in males, for concentric and eccentric peak torque, respectively; and r= -0.41 and -0.38 (p< 0.01, for both) in females for concentric and eccentric peak torque, respectively. Curl-up: r= 0.27 (p>0.05) and 0.32 (p< 0.01) in males, for concentric and eccentric peak torque, respectively; and r= -0.25 and -0.28 (p> 0.05 for both) in females for concentric and eccentric peak torque, respectively.	The use of timed curl-up test is not a valid method of estimating isokinetic abdominal muscular fitness.
Mannion et al. 1997	Females= 17 Males= 17 (Healthy)	19-48	Modified Biering- Sørensen	Electromyography	Simple and multiple linear regression, ANOVA/ANCOVA	R ² = 0.85 and 0.57, (p< 0.05 for both) between time endurance and median frequency of left and right lumbar region.	The Biering-Sørensen test is valid for measuring back muscle fatigue.

						No difference between the greatest power spectrum and endurance time ($p > 0.05$).	
Sparto et al. 1997	Males= 10 (Healthy)	20-32	Biering-Sørensen	Electromyography	Linear regression, MANOVA for repeated measurements	$R^2 = 0.88$ ($p < 0.05$) between median frequency and time test. There was not significant level among lumbar level, muscle location and endurance time ($p > 0.05$).	The Biering-Sørensen test is valid for measuring back muscle fatigue.
Applegate et al. 2019	Females= 24 Males= 24 (Healthy and low back pain (50%))	29.2 ± 2.2 (Healthy) 24.3 ± 1.5 (Low back pain)	Biering-Sørensen	Electromyography	Hierarchical and stepwise linear regression	$R^2 = 0.56$ ($p < 0.001$) between time test with normalized median power frequency slope of the erector spinae ($\beta = 0.350$, $p < 0.01$), the biceps femoris ($\beta = 0.375$, $p < 0.01$), and self-efficacy ($\beta = 0.437$, $p < 0.01$).	The Biering-Sørensen test has moderate validity for measuring back muscle fatigue in healthy adults. In adults with low back pain self-efficacy and trunk mass appears to be the best predictor.
<i>Musculoskeletal fitness (flexibility)</i>							
Jackson & Langford 1989	Females= 52 Males= 52 (Healthy)	20-45	Sit-and-reach	Manual goniometer	Correlation coefficients (r)	$r = 0.70$ between the sit-and-reach test and hamstring flexibility; and $r = 0.12$ with low back flexibility, in female respectively. $r = 0.89$ between the sit-and-reach test and hamstring flexibility; and $r = 0.59$ with low back flexibility, in male respectively.	The sit-and-reach test appears to possess moderate validity as a measure of hamstring flexibility, but is not a valid measure of low back flexibility.

Youdas et al. 2008	Males= 106 Females= 106 (Healthy)	20-79	Sit-and-reach	Passive straight leg raise	Correlation coefficient (r)	r= 0.59 (p< 0.01) between sit-and-reach test and passive straight leg raise.	Sit-and-reach test is not a valid method to measure hip joint angle.
Mier et al. 2013	Females= 35 Males= 35 (Healthy)	23 ± 4	Sit-and-reach	Pelvic angle during sit-and-reach and hip joint angle during passive straight leg raise.	Correlation coefficient (r), paired t-test	r= 0.82 (CI: 0.70-0.89) and 0.83 (CI: 0.72-0.90) between sit-and-reach scores and pelvic angle for males and females, respectively. r= 0.60 (CI: 0.39-0.75) and 0.56 (CI: 0.33-0.73) between sit-and-reach scores and passive straight leg raise hip joint angle for males and females, respectively.	Sit-and-reach score does not adequately assess spine and pelvic flexibility.
López-Miñarro et al. 2015	Females= 120 (Healthy)	62,34 ± 8,75	Sit-and-reach	Passive straight leg raise test	Correlation coefficient (r), linear regression	R ² = 0.50, 0.45 and 0.49 (all p< 0.001) between sit-and- reach and straight leg raise test for left, right and both legs, respectively.	The validity of sit-and- reach test as measure of hamstring muscle extensibility on older females is moderate.
Liemohn et al 1994	Females= 20 Males= 20 (Healthy)	25.1 ± 6.2 24.0 ± 4.6	Sit-and-reach, Modified sit- and-reach (Cailliet)	Lumbar spine range with Inclinator	Correlation coefficient (r)	r= 0.56 and 0.68 (p< 0.01, for both) as measures of low back flexibility in females for sit-and-reach and modified tests, respectively. r= 0.31 and 0.33 (p> 0.05, for both) as measures of low back flexibility in males for sit-and-reach and modified tests, respectively.	The sit-and-reach test and its modified, do not have criterion-related validity as a field test of low back flexion ROM.

Cotten 1972	Females= 38 Males= 37 (Healthy)	College	Sit-and-reach Modified sit-and-reach Cureton test Scott-French bobbing test	Leighton flexometer	Correlation coefficient (r)	r= 0.65 between sit-and-reach and flexometer. r= 0.66 between Modified sit-and-reach and flexometer. r= 0.43 between Cureton test and flexometer. r= 0.52 between Scott-French bobbing test and flexometer.	None of these tests should be considered interchangeable with the Leighton flexometer.
Broer & Galles 1958	Females= 100 (Healthy)	18-31	Toe-to-touch	Leighton Flexometer	Correlation coefficient (r)	r= 0.81 between Toe-to-touch test and flexometer measures.	A study of a group which is known to include more extreme body types is needed for a more detailed validation of the toe-to-touch test.
Kipper & Parker 1987	Females= 16 Males= 17 (Healthy)	22.1 ± 3.8 21.1 ± 3.0	Toe-to-touch	Trunk, hip and vertical flexion, were assessed with a camera	Correlation coefficient (r)	r= 0.85 (p< 0.001) between vertical fingertip-floor distances and trunk flexion. r= 0.79 (p< 0.001) between vertical fingertip-floor distances and hip flexion. r= 0.10 between vertical fingertip-floor distances and vertebral flexion.	The vertical fingertip-floor distances a valid measure of hip flexion with the knees extended, although it should not be considered an indication of hip flexion also if the knees are allowed to flex, relieving hamstring muscle tension.
Ayala et al. 2012	Males= 70 (Active recreationally)	21.3 ± 2.5	Horizontal hip joint angle Vertical hip joint angle	Straight leg raise test with an inclinometer	Multiple linear regression, SEE	Between horizontal hip joint angle test and passive straight leg raise test β = 0.78, SEE= 0.10, R^2 = 0.62, p< 0.01.	The validity of horizontal hip joint angle test and vertical hip joint angle test is moderate. Furthermore, the horizontal hip joint angle

Between vertical hip joint angle test and passive straight leg raise test $\beta=0.80$, SEE= 0.10, $R^2= 0.63$, $p< 0.01$.
test cutoff scores should not be used for vertical hip joint angle test for the detection of short hamstring muscles in young adults.

Motor fitness (balance)

Miyamoto et al. 2008a	Females= 103 Males= 49 (Healthy)	71 ± 13	Ten step	Motor reaction time Knee muscular fitness Functional reach test	Correlation coefficient (r)	r= 0.59 ($p< 0.01$) between ten step test and motor reaction time. r= -0.35 ($p< 0.01$) between ten step test and knee muscular fitness. r= -0.42 ($p< 0.01$) between ten step test and FRT.	Ten step test is a more useful measurement of agility than previous test batteries.
Miyamoto et al. 2008b	Females= 562 Males= 266 (Healthy)	22-99	Ten step	Motor reaction time, Single leg standing time	Correlation coefficient (r)	r= 0.68 and 0.59 ($p< 0.01$ for both) between ten step test and time supine-to-stand and between ten step test and motor reaction time respectively.	Ten step test is a valid tool to assess agility.
Ritchie et al. 2005	Females= 10 Males= 13	55-70	Romberg test	Change centre of pressure	Correlation coefficient (r)	r ranges from -0.18 to 0.23 ($p>0.05$ for all) between balance test and change to centre of pressure.	Romberg test were found to be a not valid test to assess static balance.

High quality studies

Cardiorespiratory fitness

Weller et al. 1992	Females= 78 Males= 76 (Healthy)	15-69	Modified Canadian aerobic fitness	Gas analyzer in maximal treadmill test	Linear regression, ICC	$R^2 = 0.75$ ($p < 0.05$) and $ICC = 0.81$, between prediction equation and VO_{2max} .	Modified Canadian aerobic fitness test offers an improved method for estimating VO_{2max} .
Carvalho et al. 2015	Females= 31 (Sedentary obese and lean)	20-45	6-minute single 15-cm step	Gas analyzer in submaximal treadmill test	Correlation coefficients (r), multiple linear regression, SEE	$r = 0.80, 0.79$ and 0.31 (all $p < 0.05$) between number of step cycles and VO_{2peak} in total sample, lean and obese females. The final model explained 83% of the total variance in VO_{2peak} during treadmill test. The reference equation obtained was: $VO_{2peak} = 35.335 - (0.328 \times BMI) + (0.069 \times NSC) - (0.298 \times age)$, with a $SEE = 2.9$ ml/kg/min (Durbin-Watson test = 2.2; $p < 0.01$).	The step cycles of 6-minute single 15-cm step test with BMI, and age, accurately predict VO_{2peak} in both, obese and lean sedentary females.
Teren et al. 2016	Females= 38 Males= 48 (Healthy)	45-66	YMCA step	Gas analyzer in maximal treadmill test	Correlation coefficient (r), linear regression	$r = 0.64$ ($p < 0.001$) between VO_2 YMCA step test and VO_{2peak} . Age, gender, BMI and YMCA step test explained the 91% of the variance of VO_{2peak} ($p < 0.05$).	YMCA step test equation proved be convenient for estimating individual VO_{2peak} .

Beutner et al. 2015	Females= 56 Males= 55 (Healthy)	22-79	YMCA step	Gas analyzer in maximal cycle test (Bruce protocol)	Correlation coefficient (r), linear regression, SEE	r= 0.86 (p< 0.001), SEE= 5.5 ml/kg/min	The YMCA step test is appropriate for estimating individual VO_{2peak} .
Lee et al. 2019	Females and Males= 568 (Healthy)	20-66	YMCA step	Gas analyzer in maximal treadmill test (Bruce protocol)	Multiple linear regression, SEE, repeated ANOVA	r= -0.44 (p< 0.001) between VO_{2max} and heart rate recovery. Equation of YMCA step test: $R^2= 0.60$ (p< 0.05), SEE= 4.76 ml/kg/min, mean difference= 0.14 ± 1.23 ml/kg/min (p> 0.05).	The YMCA step test is accurate for estimating individual VO_{2peak} .
Hong et al. 2019	Females= 36 Males= 37 (Healthy)	32.3 ± 8.9 29.4 ± 9.5	YMCA step Tecumseh step 6-minute walk	Gas analyzer in maximal treadmill test (Modified Bruce protocol)	Multiple linear regression, SEE, paired t-test, Bland-Altman method	r= 0.5281 (p< 0.001) between VO_{2max} and heart rate 30s in Tecumseh step test. Equation of Tecumseh step test: $R^2= 0.734$ (p< 0.05), SEE= 4.7 ml/kg/min, no mean difference (p> 0.05). r= 0.682 (p< 0.001) between VO_{2max} and heart rate 30s in YMCA step test. Equation of YMCA step test: $R^2= 0.722$ (p< 0.05), SEE= 4.7 ml/kg/min, no mean difference (p> 0.05). r= 0.671 (p< 0.001) between VO_{2max} and distance walked	The YMCA step, Tecumseh step and 6-minute walk tests are valid for estimating VO_{2max} in young healthy adults.

						in 6-minute test. Equation of 6-minute walk test: $R^2 = 0.744$ ($p < 0.05$), SEE = 4.6 ml/kg/min, no mean difference ($p > 0.05$).	
Kieu et al. 2020	Females and Males = 30 (Healthy)	19-35	YMCA step	Gas analyzer in maximal treadmill test (Bruce protocol)	Correlation coefficient paired t-test, Bland-Altman method	$r = 0.80$ ($p < 0.001$) between VO_{2max} measured and VO_{2ma} predicted. Mean difference = 2.26 ± 5.47 ml/kg/min ($p < 0.001$).	The YMCA step test, is valid for estimating VO_{2max} in young healthy adults.
Hansen et al. 2016	Females = 59 Males = 53 (Healthy)	18-75	Modified Harvard step	Gas analyzer in maximal cycle test	Correlation coefficient (r), linear regression, Bland-Altman method, ICC	$r = -0.48$ and 0.52 ($p < 0.001$ for both) between VO_{2max} and heart rate and step test duration. Equation of modified Harvard step test: $R^2 = 0.78$ ($p < 0.00$) Mean difference = 0.0 ± 0.4 L/min ($p = 1.000$), between measured and predicted VO_{2max} ICC = 0.94 ($p < 0.001$) between measured and predicted VO_{2max} .	The equation developed is valid to estimate VO_{2max} in healthy adults. However, only the test performance (heart rate) is not valid for estimating individual VO_{2max} .
Aadahl et al. 2012	Females = 449 Males = 346 (Healthy)	47.0 ± 8.0 46.7 ± 8.4	Danish step	Gas analyzer in maximal cycle test (Watt-max protocol)	Correlation coefficient (r), Bland-Altman method	$r = 0.69$ and 0.77 ($p < 0.001$ both) between Danish step test and VO_{2max} , for males and females, respectively. Mean difference = 2.25 ± 13.45 and 3.06 ± 10.68	The Danish step test slightly overestimated VO_{2max} compared to the watt-max test, more so in females than in males.

						ml/kg/min, for males and females, respectively.	
Kumar et al. 2012	Males= 19 (Healthy-sedentary)	20-30	Queen's College step	Gas analyzer in maximal cycle test	Correlation coefficient (r), linear regression, paired t-test, Bland-Altman method	r= 0.034 (p< 0.05) between Queen's College step test and VO _{2max} . Mean difference= 20 ± 15.31 ml/kg/min.	Queen's College step test is not a valid test for this population.
Ricci et al. 2019	Females= 31 (Obese with comorbidities and morbidly)	18-60	2-minute step	Gas analyzer in maximal treadmill test (Bruce protocol)	Correlation coefficient (r), stepwise multiple linear regression, student t-test, Bland-Altman method	r= 0.55 (p< 0.01) between up-and-down step cycles and VO _{2max} . Equation of 2-minute step test: R ² = 0.744 (p< 0.001). Mean difference= 3.1 ± 3.0 ml/kg/min (p< 0.05).	The 2-minute step test underestimated VO _{2max} .
Cooper et al. 2005	Males= 30 (Healthy active)	21.8 ± 3.6	20-m shuttle run	Gas analyzer in maximal treadmill test	Paired t-test, Bland-Altman method,	Mean difference= 1.8 ± 6.3 ml/kg/min (p= 0.004)	The 20-m SRT is not a valid test for estimating individual VO _{2max} in active males.
Flouris et al. 2004	Males= 50 (Healthy)	21.6 ± 1.6	20-m shuttle run 20-m square shuttle run	Gas analyzer in maximal treadmill test (Modified Bruce protocol)	linear regression, SEE, CV, one-way ANOVA	Equation of the 20-m SRT: r= 0.61 (p< 0.05), SEE= 4.23 ml/kg/min, mean difference= -2.0 ± 2.9 ml/kg/min (p< 0.001), CV= 7.9%. Equation of 20-m square shuttle run test: r= 0.88 (p< 0.001), SEE= 2.51 ml/kg/min, mean difference= -0.7 ± 0.3	The 20-m square shuttle run test had a higher agreement with the gold standard laboratory test in predicting VO _{2max} .

						ml/kg/min ($p < 0.001$), CV= 6.0%	
Flouris et al. 2005	Males= 110 (Healthy)	21.6 $\pm$ 2.5	20-m shuttle run	Gas analyzer in maximal treadmill test (Modified Bruce protocol)	Correlation coefficient (r), linear regression, Bland-Altman method, one-way ANOVA	r= 0.94 and $R^2 = 0.89$ ($p < 0.001$, for both), SEE= 1.94 ml/kg/min between measured equation and predicted VO_{2max} . Mean difference= 0.1 $\pm$ 0.6 ml/kg/min ($p > 0.05$) between measured equation and predicted VO_{2max} .	The prediction model increases the efficacy of the 20-m SRT for predicting VO_{2max} .
Metsios et al. 2008	Males= 74 (Healthy)	21 $\pm$ 2	20-m shuttle run 20-m square shuttle run	Gas analyzer in maximal treadmill test (Modified Bruce protocol)	Correlation coefficient (r), 95% LoA, CV, one repeated measures ANOVA	Equation of 20-m shuttle run test: $r = 0.63$ ($p < 0.001$), 95%LOA= 4.23 $\pm$ 7.25 ($p < 0.001$, CV= 7.37%) Equation of 20-m square shuttle run test: $r = 0.95$ ($p < 0.001$), 95%LOA= -0.28 $\pm$ 3.25 ($p > 0.05$), CV= 3.46%	The 20-m square shuttle run test is a valid and reproducible assessment tool and a more efficacious test for predicting VO_{2max} in healthy adult males compared to the 20-m SRT.
Kim et al. 2011	Females= 155 Males= 158 (Healthy)	28.3 $\pm$ 4.1	20-m shuttle run	Gas analyzer in maximal treadmill test (Modified Bruce protocol)	Correlation coefficient (r), Bland-Altman method, one-way repeated measures ANOVA	r range from 0.86 to 0.95 (all $p < 0.001$) between 20-m shuttle run test and VO_{2max} Mean differences range from -0.54 $\pm$ 6.23 to -2.94 $\pm$ 6.55 (all $p < 0.01$) between 20-m shuttle run test and VO_{2max} .	Equations based on the 20-m SRT result, should not be applied for predicting VO_{2max} in Korean adults.
Jurio-Iriarte et al. 2017	Females= 75 Males= 181 (Overweight-obese/primary)	53.9 $\pm$ 8.1	Incremental shuttle walk (15-level)	Gas analyzer in maximal cycle test	Correlation coefficient (r), linear regression, t-test, Bland-Altman method, SEE, E	r= 0.72 and $R^2 = 0.52$ ($p < 0.001$, for both), SEE= 4.35 ml/kg/min (19% E) between equation and predicted VO_{2max} .	Modified shuttle walk test (15-level) does not accurately predict VO_{2max} in overweight/obese people

	hypertension)					Mean difference= 0.2 ± 8.4 ml/kg/min ($p= 0.48$) between equation and predicted VO_{2max} .	with primary hypertension.
Jurio-Iriarte et al. 2018	Females= 74 Males= 174 (Overweight-obese/ primary hypertension)	54.0 ± 7.3	Incremental shuttle walk (15-level)	Gas analyzer in maximal cycle test	Correlation coefficient (r), linear regression, SEE, ICC, Bland-Altman method	$r= 0.72$ and $R^2= 0.58$ ($p< 0.001$, for both), $SEE= 4.9$ ml/kg/min between measured and predicted VO_{2max}. ICC= 0.69 ($0.34-0.82$) between equation and predicted VO_{2max}. Mean difference=3.7 ± 9.0 ml/kg/min, $SEE= 4.35$ ml/kg/min, between equation and predicted VO_{2max}.	Modified shuttle walk test (15-level) does not accurately predict VO_{2max} in overweight/obese people with primary hypertension.
Lima et al. 2019	Females= 54 (Sedentary)	18-45	Incremental shuttle walking (15-level)	Gas analyzer in maximal treadmill test	Correlation coefficient (r), stepwise multiple linear regression, paired t-test, Bland-Altman method,	$r= 0.57$ ($p< 0.001$) between VO_{2peak} measured and VO_{2peak} predicted. $R^2= 0.363$ ($p< 0.05$) between prediction equation and VO_{2peak} measured. Mean difference= -0.14 ± 9.27 ml/kg/min ($p>0.05$).	Incremental shuttle walking test is valid to assess cardiorespiratory fitness in sedentary females.
Oja et al. 1991	Females= 80 Males= 79 (Healthy)	20-65	2-km walk	Gas analyzer in maximal treadmill test (Modified Pennsylvania State University protocol)	Correlation coefficient (r), multivariate linear regression; SEE	$r= -0.65$ and -0.74 ($p< 0.001$, for both), between walking time VO_{2max}, for males and females, respectively. $R^2= 0.75$, $SEE= 5.1$ ml/kg/min ($p< 0.05$) for	2-km walk test is an accurate alternative for predicting VO_{2max} in healthy adults.

						prediction model in males. $R^2 = 0.73$, SEE= 3.3 ml/kg/min ($p < 0.05$) for prediction model in females.	
Laukkanen et al. 1992	Females= 32 Males= 36 (Overweight/ obese)	20-65	2-km walk	Gas analyzer in maximal treadmill test	Correlation coefficient (r), E	$r = -0.49$ ($p < 0.01$) and -0.72 ($p < 0.001$) between walking time and VO_{2max} , for males and females, respectively. $r = 0.75$ and 0.77 ($p < 0.001$) for prediction model in males and females, respectively. E= 6.9 ml/kg/min in males and 3.2 ml/kg/min in females.	2-km walk test is valid predicting VO_{2max} in overweight/obese individuals, although slightly underestimate in overweight/obese amles.
Laukkanen et al. 1993	Females= 32 Males= 79 (Moderate active and highly active healthy)	35-45	2-km walk	Gas analyzer in maximal treadmill test	Correlation coefficient (r), linear regression, TE	$r = -0.73$ ($p < 0.05$), -0.52 ($p < 0.05$), and -0.31 ($p < 0.05$), between walking time and VO_{2max} , for moderate active males and females, and highly active males, respectively. $r = 0.80$ ($p < 0.05$), TE= 5.7 ml/kg/min for prediction model in moderate active males. $r = 0.55$ ($p < 0.05$), TE= 5.0 ml/kg/min for prediction model in moderate active females. $r = 0.60$ ($p < 0.05$), TE= 7.7	2-km walk test is a reasonably valid field test for estimating the cardiorespiratory fitness of healthy males and females, except for people with very high maximal aerobic power.

						ml/kg/min for prediction model in moderate active males.	
Laukkanen et al. 2000	Females= 61 Males= 55 (Healthy)	30-55	2-km walk	Gas analyzer in maximal treadmill test	Two-way ANOVA, TE	Mean difference= 0.9 ± 4.4 ml/kg/min, TE= 4.5 ml/kg/min in males. Mean difference= 2.2 ± 3.5 ml/kg/min, TE= 4.1 ml/kg/min in females.	2-km walk test has a reasonably accuracy to predict VO_{2max} in untrained healthy adults.
Kline et al. 1987	Females= 178 Males= 165 (Healthy)	30-69	1-mile walk	Gas analyzer in maximal treadmill test	Correlation coefficient (r), multiple linear regression, SEE	$r = 0.64$ ($p < 0.05$) between time test and VO_{2max} . $r = 0.88$, SEE= 5.0 l/min ($p < 0.05$) for prediction model.	1-mile walk test is valid to assess VO_{2max} in this population.
Dolgener et al. 1994	Females= 145 Males= 129 (Healthy)	19.3 ± 2.32 19.4 ± 3.08	1-mile walk	Gas analyzer in maximal treadmill test	Correlation coefficient, (r), multiple linear regression, SEE, E	$r = 0.69$ ($p < 0.05$), SEE= 5.5 ml/kg/min, E= 13.26 between Kline's equation and VO_{2max} . New predicted equation (l/min): $r = 0.86$ ($p < 0.05$), SEE= 0.383, E= 0.953. New predicted equation (ml/kg/min): $r = 0.69$ ($p < 0.05$), SEE= 5.500, E= 13.26.	1-mile walk test over predicts VO_{2max} in college students and should not be used with this population.
George et al. 1998	Females= 59 Males= 39 (Healthy)	18-29	1-mile walk	Gas analyzer in maximal treadmill test	Correlation coefficients (r), SEE, E, paired t-test,	Kline's equation: $r = 0.84$ ($p < 0.05$), SEE= 3.61ml/kg/min, E=6.16, overpredicted ± 4.98 ml/kg/min ($p < 0.05$).	The Dolgener's equation seems to be more accurate that Kline's equation to predict VO_{2max} .

						Dolgener's equation: r= 0.85 (p< 0.05), SEE= 3.48 ml/kg/min, E= 3.74, underestimated $\pm$ 1.24 ml/kg/min (p< 0.05).	
Seneli et al. 2013	Females= 10 Males= 13 (Healthy)	19-44	1-mile walk	Gas analyzer in maximal treadmill test	Correlation coefficient (r), Bland-Altman method, one-way repeated measures ANOVA	r range from 0.71 to 0.84 (all p< 0.05) between equations and VO _{2max} . Mean differences range from 2.360 to 9.131 ml/kg/min (all p< 0.001, except for 2.360 where p= 0.064).	Despite high correlations, Kline's and Dolgener's equations of 1-mile walk test underestimate VO _{2max} . It would be beneficial if a new equation were created specifically for the curved treadmill.
Lunt et al. 2013	Females= 44 Males= 162 (Healthy)	18-39	1-mile walk	Gas analyzer in maximal treadmill test	Stepwise linear regression, SEE, one-way repeated measures ANOVA,	R ² = 0.68 (p< 0.05), SEE= 4.6 ml/kg/min, -0.5%, mean difference= 1.2 ml/kg/min (p>0.05), between new equation and VO _{2peak} .	The new equation derived from this study gave the most accurate estimate of VO _{2peak} .
Greenhalgh et al. 2001	Females= 24 Males= 21 (Healthy)	18-29	1-mile walk 1/4-mile walk	Gas analyzer in maximal treadmill test	Correlation coefficient (r), SEE, TE, % $\pm$ 4.5 ml/kg/min	r= 0.81 (p< 0.05), SEE= 4.33 ml/kg/min, TE= 4.8, % 67.5 and r= 0.84 (p< 0.05), SEE= 4.03 ml/kg/min, TE= 4.12, % 75.7 between Kline's equation and 1/4-mile walk test and 1-mile walk test, respectively. r= 0.84 (p< 0.05), SEE= 4.03 ml/kg/min, TE= 7.07, % 43.52 and r= 0.85 (p< 0.05), SEE= 3.93 ml/kg/min, TE= 47.93, % 18.9 between	The 1/4-mile walk test predicts VO _{2max} with similar accuracy as the 1-mile walk test.

						Dolgener's equation and 1/4-mile walk test and 1-mile walk test, respectively.	
Larsen et al. 2002	Females= 47 Males= 52 (Healthy)	18-26	1.5-mile run/walk	Gas analyzer in maximal treadmill test	Multiple linear regression, SEE	$R^2 = 0.86$ ($p < 0.05$), $SEE = 3.37$ ml/kg/min, between new equation and VO_{2max} .	1.5-mile test is valid without measuring heart rate.
McNaughton et al. 1998	Males= 32 (Active healthy)	20 ± 0.3	20-m shuttle run 12-minute run 1.5-mile run Treadmill jogging	Gas analyzer in maximal treadmill test	Correlation coefficient (r), SE, one-way ANOVA	$r = 0.82$ ($p < 0.05$), $SE = 1.0$, mean difference= 1.62 ml/kg/min ($p > 0.05$), between the 20-m SRT and VO_{2max} . $r = 0.87$ ($p < 0.05$), $SE = 0.7$, mean difference= -1.02 ml/kg/min ($p > 0.05$), between 12-minute run test and VO_{2max} . $r = 0.87$ ($p < 0.05$), $SE = 1.2$, mean difference= 0.74 ml/kg/min ($p > 0.05$), between 1.5-mile run and VO_{2max} . $r = 0.50$ ($p < 0.05$), $SE = 1.5$, mean difference= 5.58 ml/kg/min ($p > 0.05$), between treadmill jogging test and VO_{2max} .	The 12-minute run test is the best predictor of VO_{2max} in young adult males with active to trained levels of fitness. The submaximal treadmill jogging test has low-moderate validity to predict VO_{2max} .
George et al. 1993	Females= 45 Males= 84 (Healthy-	18-29	Submaximal single-stage treadmill jogging	Gas analyzer in maximal treadmill test	Multiple linear regression, SEE	$r = 0.84$, ($p < 0.001$), $SEE = 3.2$ ml/kg/min between prediction equation and VO_{2max} .	Submaximal single-stage treadmill jogging test is valid for estimating VO_{2max} .

	active)						
Cao et al. 2013	Females= 140 Males= 143 (Healthy)	20-69	3-minute walk	Gas analyzer in maximal cycle test	Multiple linear regression, SEE, %SEE	r= 0.55 (p< 0.001) between time of 3-minute walk test and VO _{2max} . Best equation %Fat: R ² = 0.70 (p< 0.001), SEE= 4.565 ml/kg/min, % SEE= 13.316.	The 3-minute walk test is a valid tool for estimating VO _{2max} in Japanese adult males and females.
Di Thommazo-Luporini et al. 2015	Females= 51 (Sedentary obese)	20-45	6-minute walk Incremental shuttle walking	Gas analyzer in maximal treadmill test (Bruce protocol)	Correlation coefficient (r), linear regression, Bland-Altman method	r= 0.70 and 0.62 (all p< 0.001) between VO _{2max} and 6-minute walk and incremental shuttle walking tests There was agreement between Bruce protocol and both the incremental shuttle walking test: VO _{2max} = 6.1 (CI: 4.9–7.3), and heart rate=36.2 (CI: 32.1–40.3); and 6-minute walk test: VO _{2max} = 6.9 (CI: 5.7–8.1), and heart rate=37.0 (CI: 33.3–40.7).	6-minute walk and incremental shuttle walking tests are valid to assess cardiorespiratory fitness in obese females.
Di Thommazo-Luporini et al. 2016	Females= 56 (Sedentary obese)	35 ± 7	6-minute walk	Gas analyzer in maximal treadmill test (Bruce protocol test)	Correlation coefficient (r), stepwise multiple linear regression, Bland-Altman method	r= 0.56 (p< 0.001) between VO _{2max} and up and down step cycles 6-minute walk. Mean difference: 5.1 ± 3.6 ml/kg/min.	6-minute walk test is valid to assess cardiorespiratory fitness in obese females.
Manttari et al. 2018	Females= 36 Males= 39	19-75	6-minute walk	Gas analyzer in maximal treadmill test	Linear regression, SEE, Bland-Altman method	R ² = 0.85(p< 0.001) and SEE= 3.9 ml/kg/min, between prediction equation	6-minute walk test is an accurate and valid tool for estimating VO _{2max} .

	(Healthy)			Douglas bag methods		and VO _{2max} , in males. R ² = 0.90 (p< 0.001) and SEE= 3.1 ml/kg/min, between prediction equation and VO _{2max} , in females. Mean difference: 0 ± 5.4 ml/kg/min.	
Jamnack et al. 2016	Females= 26 Males= 31 (Healthy)	18-39	YMCA cycle Mankato submaximal exercise	Gas analyzer in maximal treadmill test	Linear regression, ICC, SEE, TE, CV, one-way repeated measures ANOVA,	r= 0.646 (p< 0.05), ICC= 0.271, SEE=6.91 ml/kg/min, TE= 5.57, CV= 15.17%, mean difference=3.96 ml/kg/min (p< 0.01), between YMCA test and VO _{2max} . r= 0.72 (p< 0.05), ICC= 0.722, SEE=6.20 ml/kg/min, TE=4.62, CV= 11.28%, mean difference= -1.54 ml/kg/min (p>0.05), between Mankato submaximal exercise test and VO _{2max} .	The Mankato submaximal exercise test yield better estimates of VO _{2max} . The YMCA test underestimates VO _{2max}
Siconolfi et al. 1982	Females= 28 Males= 35 (Healthy)	20-70	Modified Astrand- Ryhming	Gas analyzer in maximal cycle test	Linear regression, SEE, paired t-test	r= 0.86 (p< 0.05), SEE= 0.359 l/min, mean difference= 0.07 ml/kg/min (p>0.05), between prediction equation and VO _{2max} in males. r= 0.97 (p< 0.05), SEE= 0.199 l/min, mean difference= -0.04 ml/kg/min (p>0.05), between prediction	Modified Astrand- Ryhming test is valid for estimating VO _{2max} .

						equation and $VO_{2\max}$ in females.	
Bonet et al. 2020	Females= 42 (Healthy-active)	35-45	The University Montreal track (UM- track)	Gas analyzer in treadmill performing UM- track test	Correlation coefficient (r), linear regression, paired t-test, Cohen's d effect size, Bland-Altman method	$r = 0.93$ ($p < 0.001$) between time in UM-track test and UM-treadmill test. $r = 0.71$ ($p < 0.001$) between time in UM-track test and $VO_{2\max}$. Mean difference= 0.025 ± 7.445 ml/kg/min ($p > 0.05$), Cohen's d < 0.5.	UM- track test valid for estimating $VO_{2\max}$ in middle-aged active females.
Guo et al. 2018	Females= 22 Males= 18 (Healthy)	18-64	Ruffier	Gas analyzer in treadmill (modified Balke protocol)	Multiple linear regression, RMSE, sensitivity and specificity, kappa	$R^2 = 0.637$ ($p < 0.05$), RMSE= 0.517, sensitivity= 0.79 and specificity= 0.56, $k = 0.60$ between prediction equation and $VO_{2\max}$.	Ruffier test has moderate validity for estimating $VO_{2\max}$.
<i>Musculoskeletal fitness (isometric)</i>							
Shechtman et al. 2005	Females= 50 Males= 50 (Healthy)	20-40	DynEx dynamometer	Known weights Jamar dynamometer	Correlation coefficient (r), measurement error (%), repeated measures ANCOVA	$r = 0.99$ ($p < 0.05$) between DynEx and Jamar dynamometers. Average measurement error= 1.63% and 7.74% for DynEx and Jamar dynamometer, respectively, in relation known weights. Significant differences in grip muscular fitness scores between DynEx and Jamar dynamometers ($F = 6.222$, $p =$	The DynEx dynamometer is valid for assessing grip muscular fitness.

0.014).

España-Romero et al. 2010	NA	NA	TKK, DynEx and Jamar dynamometers	Known weights	Bland-Altman method	Negative systematic error for the Jamar and DynEx dynamometer (-1.92 ± 1.92 and -1.43 ± 3.56 kg, respectively, $p < 0.05$). Positive systematic bias for the TKK dynamometer (0.49 ± 1.32 kg, $p < 0.05$).	TKK dynamometer is the most appropriate dynamometer to assess grip muscular fitness.
Cadenas-Sánchez et al. 2016	NA	NA	TKK dynamometers (old and new digitals and analog)	Known weights	One-sample t-test, Bland-Altman method	Negative systematic error for the new digital (-2.64 to -2.02 kg), new analogic (-2.15 kg to -1.51 kg), old digital and analogic (-0.94 kg and -2.29 kg, respectively). All $p < 0.001$.	TKK dynamometer is valid to assess grip muscular fitness.
Kolimechkov et al. 2020	NA	NA	TKK and DynEx and dynamometers	Known weights	Paired t-test, Bland-Altman method	Non-significant systematic error of -0.20 ± 0.62 kg ($p > 0.05$) for the TKK, and a significant systemic error of -0.42 ± 0.46 kg ($p < 0.001$) for the DynEX dynamometer.	TKK dynamometer showed to be more valid than DynEX dynamometer.
<i>Musculoskeletal fitness (endurance)</i>							
Mannion et al. 1994	Females= 208 Males= 21	26.9 ± 4.2	Biering-Sørensen	Electromyography	Linear regression	$R^2 = 0.69$ ($p = 0.001$), between time endurance and changes in median frequency	The Biering-Sørensen test is valid for measuring back muscle fatigue.

	(Healthy)						
Kankaanpää et al. 1998	Females= 153 Males= 100 (Healthy)	38.9 ± 7.6	Biering-Sørensen	Electromyography	Correlation coefficients (r), multiple linear regression	r ranges from 0.60 to 0.71 (p< 0.05), between NMF _{slope} and the endurance time. BMI contributed most of the variance in females: R ² = 0.29 (p< 0.001) and BMI and age in men: R ² = 0.12 (p< 0.05).	The validity of this test to assess lumbar muscle endurance is acceptable.
Coorevits et al. 2008	Females= 13 Males= 7 (Healthy-active)	23.5 ± 1.1 32.9 ± 14.3	Biering-Sørensen	Electromyography	Correlation coefficients (r), multiple stepwise linear regression, two-way ANOVA	r ranges from 0.46 to 0.71 (p< 0.05), between NMF _{slope} and the endurance time. The most predictive muscle was iliocostalis lumborum pars thoracis: R ² = 0.477 (p< 0.001). Difference between both methods to estimated end point test (± 16 sec., p=0.22); but highly correlated: r ranges 0.94 to 0.98 (p< 0.01).	The Biering-Sørensen test is valid for measuring back muscle fatigue.
De Blaiser et al. 2018	Females= 15 Males= 14 (Healthy-active)	25.5 ± 2.1	Prone bridging	Electromyography	Correlation coefficients (r), multiple backward linear regression, one-way repeated measures ANOVA,	r= 0.59 (p< 0.01) between NMF _{slope} and the endurance time. Endurance time is the best	The prone bridging test is valid for evaluating core muscle fatigue.

						predictor: $R^2 = 0.261$ ($p < 0.05$). NMF_{slope} values showed non-significant differences between the different muscles ($p > 0.05$).	
<i>Musculoskeletal fitness (explosive)</i>							
Bui et al. 2015	Females= 18 Males= 23 (Healthy)	19-40	Sargent jump	Contact mat Optical system	Linear regression, repeated measures ANOVA, Bland-Altman method	$R^2 = 0.8$ ($p < 0.05$) between contact mat and Sargent jump. $R^2 = 0.76$ ($p < 0.05$) between optical sensor and Sargent jump. Positive systematic error = 4.4 ± 5.1 and 4.8 ± 5.7 cm ($p < 0.001$ for both) between the Sargent jump with contact mat and optical sensor, respectively.	Sargent jump test overestimates the height of a vertical jump and its accuracy is reduced as jump height increase.
<i>Musculoskeletal fitness (flexibility)</i>							
Kawano et al. 2010	Females= 100 Males= 100 (Healthy-active)	18-25	Sit-and- reach (Hip joint angle)	Angular kinematic analysis	Correlation coefficient (r), one-way repeated measures ANOVA, 95%CI	$r = 0.48$ ($p < 0.05$) between sit-and- reach and kinematic analysis with ankle dorsiflexed. $r = 0.44$ ($p < 0.05$) between sit-and- reach and kinematic analysis with ankle plantarflexed. The hip joint angle values for males presented	Sit-and- reach test is not valid to assess hip joint angle.

significant statistical differences when comparing dorsiflexed (=83.0°) with plantarflexed (87.3°), 95% CI of mean difference (3.6-4.9; $p < 0.0001$).

These were similar to those found for females (dorsiflexed=107.9°; plantarflexed=112.9°), 95% CI of mean difference (3.5-4.9; $p < 0.001$).

ANOVA indicates Analysis of Variance; ANCOVA, Analysis of Covariance; BMI, Body Mass Index; CI, Coefficient of Interval; CV, Coefficient of Variation; E, Error; ICC, Intraclass Correlation Coefficient; MANOVA, Multiple Analysis of Variance; NMF_{slope} , Normalized Median Frequency slope; RM, Repetition Maximum; RMSE, Root Mean Square Error; SEE, Standard Error Estimate; TE, Total Error; VO_{2max} , Maximum Oxygen Uptake; VO_{2peak} , Peak Oxygen Uptake.

Table 18. Quality assessment of the included meta-analysis/systematic reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) rating tool.

Reviews	1	2	3	4	5	6	7	8	9	10	11	Rating	Quality*
Mayorga-Vega et al. 2014	Yes	No	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	7	Medium
Mayorga-Vega et al. 2014	Yes	No	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	7	Medium
Bennet et al. 2016	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	N/A	N/A	Yes	8	High
Mayorga-Vega et al. 2015	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	8	High
Mayorga-Vega et al. 2016	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	8	High

AMSTAR contains 11-items to appraise the methodological aspects of the systematic reviews. All 11-items were scored as “Yes”, “No”, “Can’t Answer” (C/A) or “Not Applicable” (N/A). A total possible score of 11 was calculated, counting only for positive responses (“Yes”).

*The final quality rates were computed by tertiles, where the first tertile ranged from 0 to 3 points, the second tertile from 4 to 7 points and the third tertile from 8 to 11 points. Likewise, each tertile were treated as “low”, “medium” or “high”, quality, respectively.

Table 19. Overview of field-based physical fitness tests criterion-related validity of included meta-analysis/systematic reviews.

Author, date and years covered	Type of review; number of studies (K)	Sample for current analysis	Field-based tests	Criterion measure	Main results	Conclusion
Mayorga-Vega et al. 2016 1982 - 2014	Meta-analysis and systematic review K= 102	N= 3988 18–65 years	5000-m, 3-miles, 2-miles, 3000-m, 1.5-mile, 1-mile, 1000-m, ½-mile, 600-m, 600-yd, ¼-mile, 15-minute, 12-minute, 9-minute, 6-minute	Gas analyzer in maximal treadmill test	r_p from 0.38 (CI: 0.13–0.64) to 0.80 (CI: 0.72–0.80) for the ¼-mile and 1.5-mile tests, respectively.	The overall criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness ranged from low to moderate.
Mayorga-Vega et al. 2015 1982 - 2014	Meta-analysis and systematic review K= 24	N= 1022 18–65 years	20-m shuttle run test protocols: Léger, Eurofit, Queen’s university Belfast, Dong-Ho	Gas analyzer in maximal treadmill test	r_p = 0.69 (CI: 0.56–1.00) for Eurofit’s protocols. r_p = 0.94 (CI:0.87–1.00) for Leger’s protocol.	The 20-m SRT has a moderate-to-high criterion-related validity for estimating VO_{2max} .
Bennet et al. 2016 1950 - 2015	Systematic review K= 11	N= 422 18–65 years	Chester step, personalised step, STEP tool step, Queen’s College step, Skubic and Hodgkins step, height-adjusted, rate-specific, single-state step, Astrand-Ryhming step and a modified YMCA 3-minutes step	Gas analyzer in maximal test	r from 0.64 between the Skubic and Hodgkins step test and VO_{2max} ; to 0.95 between Queen’s college step test and VO_{2max} .	Overall, step tests provide a simple, effective and valid method of submaximal assessing cardiorespiratory fitness.

Mayorga-Vega et al. 2014	Meta-analysis and systematic review	N= 3568 18–65 years	Classic sit-and-reach, modified sit-and-reach, back-saver sit-and-reach, modified back-saver sit-and-reach, V sit-and-reach, Modified v sit-and-reach, unilateral sit-and-reach, chair sit-and-reach	Hamstring extensibility: passive straight leg raise test, active knee extension test, active straight leg raise test Lumbar extensibility: Macraae & Wright method, single inclinometer method, American Academy of Orthopedic, spinal mouse method	Hamstring extensibility: r_p range from 0.49 (CI:0.29–0.68), to 0.68 (CI:0.55–0.80) for chair sit-and-reach and classic sit-and-reach, respectively. Lumbar extensibility: r_p from 0.16 (CI: -0.10–0.41), to 0.35 (CI: 0.15–0.54) for back-saver sit-and-reach and unilateral sit-and-reach, respectively.	Three sit-and-reach test and its different versions have a moderate validity for estimating hamstring extensibility, but a low validity for estimating lumbar extensibility.
1985 - 2012	K= 28					
Mayorga-Vega et al. 2014	Meta-analysis and systematic review	N= 985 18–65 years	Toe-to-touch	Passive straight leg raise test	r_p = 0.66 (CI: 0.56–1.00), between hamstring extensibility and passive straight leg raise test.	Toe-to-touch test has moderate criterion-related validity for estimating hamstring extensibility.
	K= 6					

CI indicates Coefficient Intervals; r_p , the overall weight mean of r corrected for sampling error and measurement error of the field tests.

Table 20. Levels of evidence of cardiorespiratory fitness tests.

Field-based physical fitness test	Strong	Moderate	Limited
<i>Shuttle run tests</i>			
20-m shuttle run	●		
20-m square shuttle		●	
Incremental shuttle walk			■
<i>Distance and time-based run/walk test</i>			
1.5-mile run/walk	●		
12-minute run/walk	●		
5000-m run/walk	◐		
3-miles run/walk	◐		
2-miles run/walk	◐		
3,000-m run/walk	◐		
1000-m run/walk	◐		
600-m run/walk	◐		
600-yd run/walk	◐		
½-mile run/walk	◐		
¼-mile run/walk	◐		
9-minute run/walk			●
2-km walk	●		
6-minute walk	●		
1-mile walk			■
¼-mile walk			●
3-minute walk			●
Treadmill jogging			■
Mankato submaximal exercise			●
Modified Astrand–Ryhming			●
University-Montreal			●
Ruffier			○
<i>Step tests</i>			
YMCA step	●		
Chester step			■
Modified Harvard step			●
6-minute single 15cm-step			●
Modified Canadian aerobic fitness step			●
Tecumseh step			●
Astrand–Ryhming step			●
Danish step			○
Queen’s College step			■
2-minute step			○

● Indicates high validity; ◐, moderate validity; ○, low/null validity; ■, inconclusive validity.

Table 21. Levels of evidence of muscular fitness, flexibility and motor fitness tests.

Field-based physical fitness test	Strong	Moderate	Limited
<i>Maximal isometric muscular fitness</i>			
Handgrip test (TKK)	●		
Handgrip test (Jamar)		◐	
Handgrip test (DynEx)			■
<i>Hip/back endurance muscular fitness</i>			
Biering-Sørensen	●		
<i>Abdominal endurance muscular fitness</i>			
Prone bridging			●
Original/Modifications curl-up			○
<i>Lower-body endurance muscular fitness</i>			
Sit-to-stand			■
<i>Lower-body explosive muscular fitness</i>			
Sargent jump			●
<i>Upper-body endurance muscular fitness</i>			
Original/Modification flexed-arm hang			■
Baumgartner modified pull-up			■
Standard push-up			■
Hand-release push-up			■
Bent-knee push-up			■
Revised push-up			■
<i>Low back flexibility</i>			
Original/Modifications sit-and-reach	○		
<i>Hamstring flexibility</i>			
Original/Modifications sit-and-reach	◐		
Toe-to-touch	◐		
<i>Agility</i>			
Ten step			◐
<i>Balance</i>			
Romberg test			■

● Indicates high validity; ◐, moderate validity; ○, low/null validity; ■, inconclusive validity.

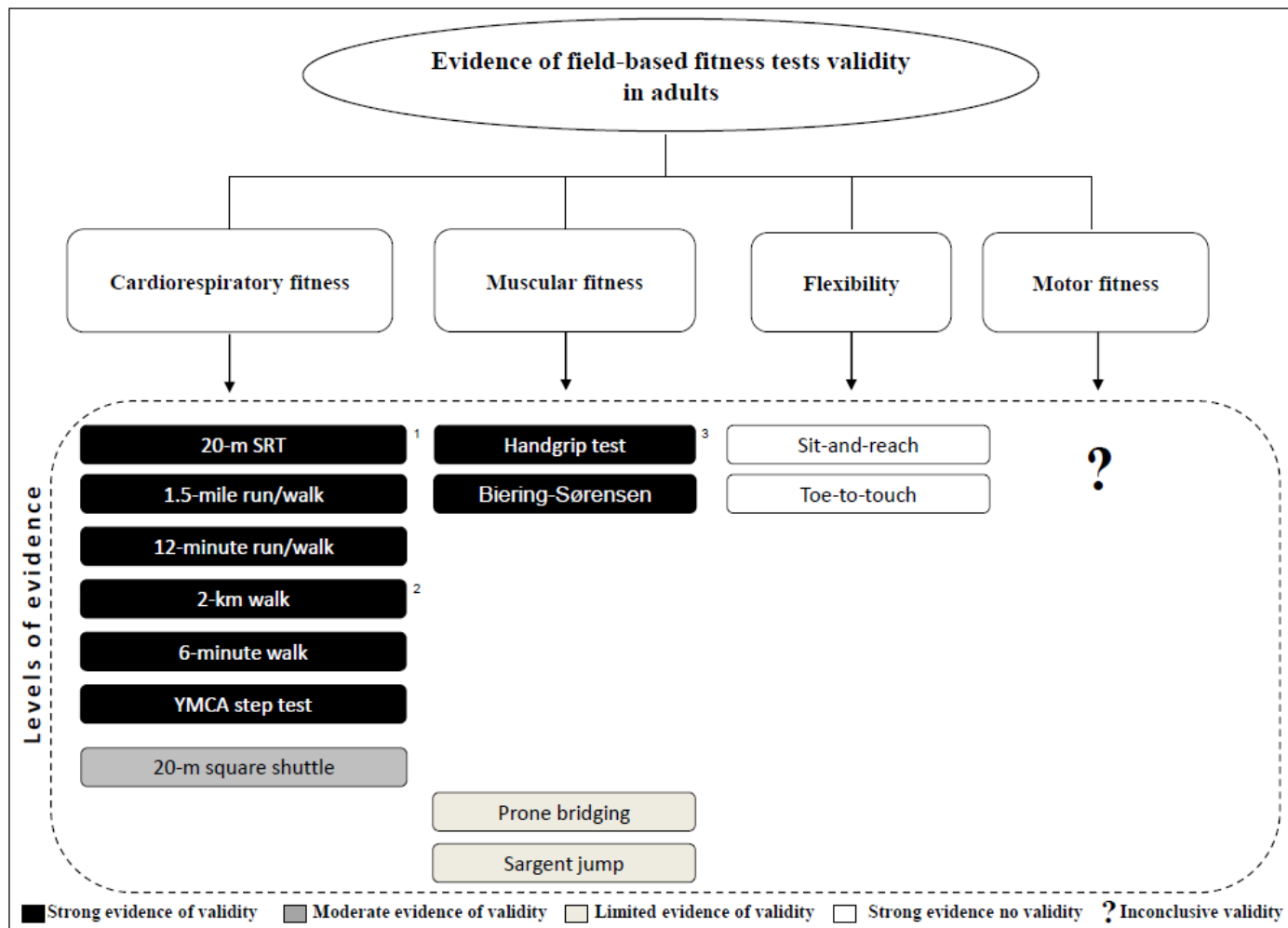


Figure 7. Levels of evidence found about the criterion-related validity of field-based physical fitness tests in adults.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

In total, 15517 references were found, 12743 in MEDLINE and 2737 in the Web of Science. Additionally, 37 records were identified through other sources (i.e., reference checking). After removing duplicates and applying inclusion and exclusion criteria, 100 studies were selected. Of them, 10 as field-based studies and 90 as estimation equations studies (see **Figure 8**).

Of the 100 original studies included in the present systematic review, one field-based study¹⁸⁰ and 12 estimation equations studies¹⁸¹⁻¹⁹² studies were classified as low quality (a total score of 3 and 4). A total of 87 original studies were classified as high quality (a total score higher than 4) (See **Table 22** and **Table 22**).

The studies included a highly variable number of participants, from 30 to 18198. A total of 59161 female and 43931 male adults were included (average age of 39.9 years). BMI status range from normal weight to obese class 3. Non-standard groupings were used to classify participants, such as Caucasian/^{180, 185, 188, 190, 193-230} (n= 43), African-American/Black^{194, 195, 198-202, 204, 205, 207, 214, 218, 219, 221, 223, 225, 229-232} (n= 20), Asian^{198, 200, 204, 211, 219, 224, 233-241} (n= 15), Mexican-American/Hispanic^{194, 199-202, 204, 205, 207, 214, 223, 232, 242, 243} (n= 13), African-Indian²⁴⁴⁻²⁴⁶ (n= 3), Brazilian^{247, 248} (n= 2), Colombian^{249, 250} (n= 2), Anglo-Celtic Australian²⁵¹ (n= 1), Chinese Australian²⁵² (n= 1), Australian²⁵³ (n= 1), European-American²⁵⁴ (n= 1), Syrian²⁵⁵ (n= 1), Guatemalan²⁵⁶ (n= 1), Polynesian²²⁸ (n= 1), Chilean²⁵⁷ (n= 1), Canadian²⁵⁸ (n= 1), Portuguese²⁵⁹ (n= 1), Tasmanian²⁶⁰ (n= 1), or Swedish²⁶¹ (n= 1). Whenever possible, these race-ethnicities have been grouped into four large groups²⁶² (i.e., Caucasian/White, African-American/Black, Hispanic, Asian) to discuss findings. However, in other cases race-ethnicity or native country have not been specifically reported^{181-184, 186, 187, 189, 191, 192, 263-277} (n= 24).

One field-based study²⁷⁸ and 10 estimation equations studies^{202, 205, 210, 213, 215, 250, 257, 263, 265, 279} were published in the last five years (see **Table 22** and **Table 23**).

Criterion-related validity of field-based body composition assessment in adults

An overview of field-based body composition estimation studies in adults is shown in **Table 24**.

The criterion methods used to compare with field-based methods were DXA^{180, 193-195, 234, 247, 260, 278, 280} (n= 9) and UWW¹⁹⁶ (n= 1). The field-based methods used were

BMI^{193-195, 234, 260, 280} (n= 6), WC^{194, 234, 260, 278, 280} (n= 4), BAI (n= 3)^{195, 234, 247}, HC^{234, 260} (n= 2), SKF¹⁹⁶ (n= 1), BIA²⁷⁸ (n= 1), WHR²⁶⁰ (n= 1), and WHtR²⁷⁸ (n= 1).

The following field-based methods showed **very high/high validity**, compared with criterion methods: BMI (derived from five studies, $r = 0.71-0.95$; all, $p < 0.05$)^{193-195, 234, 260}; WC (derived from three studies, $r = 0.72-0.89$; all, $p < 0.05$)^{194, 260, 278}; HC (derived from one study, $r = 0.77-0.85$, $p < 0.0001$)²⁶⁰; SKF (derived from one study, $r = 0.88-0.95$, $p < 0.001$)¹⁹⁶; BAI (derived from one study, $r = 0.70-0.81$, $p < 0.05$)¹⁹⁵; BIA (derived from one study, $r = 0.99$, $p < 0.001$)²⁷⁸ and; WHtR (derived from one study, $r = 0.82$, $p < 0.001$)²⁷⁸.

The following field-based methods showed **moderate validity**, compared with criterion methods: WC (derived from two studies, $r = 0.57-0.63$; both, $p < 0.001$)^{234, 278}; BAI (derived from two studies, $r = 0.58-0.68$; both, $p < 0.001$)^{234, 247}; BMI (derived from two studies, $r = 0.54-0.69$; both, $p < 0.01$)^{195, 234}; HC (derived from one study, $r = 0.66$, $p < 0.0001$)²³⁴; BIA (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸; WHR (derived from one study, $r = 0.51-0.63$, $p < 0.001$)²⁶⁰ and; WHtR (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸.

The following field-based methods showed **very low/low validity**, compared with criterion methods: WC (derived from one study, $r = 0.23$, $p < 0.05$)²⁸⁰; BMI (derived from one study, $r = 0.22$, $p < 0.005$)²⁸⁰ and; WHR (derived from one study, $r = 0.46$, $p < 0.05$)²⁶⁰. There was a balanced distribution in the female-male ratio between studies and the validity of the results does not change significantly between sexes.

Age groups

Regarding age-specific groups, the mean age was 44.6 years old. To compare the results, three tertiles were constructed (*first group*: 19-33, *second group*: 34-49, and *third group*: 50-64 years old).

The 19-33 years old group: WC showed **high validity** (derived from two studies, $r = 0.83-0.86$; both, $p < 0.05$)^{194, 278}, and **moderate validity** (derived from one study, $r = 0.57$, $p < 0.001$)²⁷⁸; WHtR showed **high validity** (derived from one study, $r = 0.75$, $p < 0.001$)²⁷⁸, and **moderate validity** (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸; BIA showed **very high validity** (derived from one study, $r = 0.99$, $p < 0.001$)²⁷⁸, and **moderate validity** (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸ and; BMI showed **high validity** (derived from one study, $r = 0.79-0.84$, $p < 0.05$)¹⁹⁴.

The 34-49 years old group: WC showed **high validity** (derived from one study, $r = 0.76-0.77$, $p < 0.05$)¹⁹⁴, and **low validity** (derived from one study, $r = 0.23$, $p < 0.05$)²⁸⁰;

BMI showed **high validity** (derived from one study, $r = 0.75-0.79$, $p < 0.05$)¹⁹⁴, and **very low validity** (derived from one study, $r = 0.22$, $p < 0.05$)²⁸⁰ and; SKF showed **very high/high validity** (derived from one study, $r = 0.88-0.95$, $p < 0.05$)¹⁹⁶.

The 50-64 years old group: BMI showed **very high/high validity** (derived from five studies, $r = 0.71-0.95$; all, $p < 0.05$)^{193-195, 234, 260}; WC showed **very high/high validity** (derived from three studies, $r = 0.72-0.90$; all, $p < 0.05$)^{194, 234, 260}, and **moderate validity** (derived from one study, $r = 0.63$, $p < 0.0001$)²³⁴; BAI showed **high validity** (derived from one study, $r = 0.72$, $p < 0.01$)¹⁹⁵ and **moderate validity** (derived from two studies, $r = 0.55-0.65$; both, $p < 0.01$)^{234, 247}; HC showed **high validity** (derived from one study, $r = 0.77-0.85$, $p < 0.05$)²⁶⁰, and **moderate validity** (derived from one study, $r = 0.66$, $p < 0.0001$)²³⁴ and; WHR showed **moderate validity** (derived from one study, $r = 0.46-0.63$, $p < 0.05$)²⁶⁰.

Weight status groups

Regarding BMI status, we observed a mean BMI of 25.7 kg/m². Participants were classified by BMI status, according to the World Health Organization²⁸¹, in three groups: normal weight, $\geq 18.5-24.9$ kg/m²; overweight, $25-29.9$ kg/m²; obesity ≥ 30 kg/m².

Group with normal weight: WC showed **high validity** (derived from one study, $r = 0.72$, $p < 0.0001$)²³⁴, **moderate validity** (derived from two studies, $r = 0.57-0.63$; both, $p < 0.001$)^{234, 278}, and **low validity** (derived from one study, $r = 0.23$, $p < 0.05$)²⁸⁰; BMI showed **high validity** (derived from two studies, $r = 0.70-0.84$; both, $p < 0.05$)^{193, 234}, and **very low** (derived from one study, $r = 0.22$, $p < 0.05$)²⁸⁰; SKF showed **high/very high validity** (derived from one study, $r = 0.88-0.95$, $p < 0.001$)¹⁹⁶; HC showed **moderate validity** (derived from one study, $r = 0.66$, $p < 0.0001$)²³⁴; WHtR showed **moderate validity** (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸; BAI showed **moderate validity** (derived from one study, $r = 0.58-0.62$, $p < 0.001$)²³⁴ and; BIA showed **moderate validity** (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸.

Group with overweight: WC showed **very high/high validity** (derived from two studies, $r = 0.84-0.90$; both, $p < 0.001$)^{260, 278}; BMI showed **very high/high validity** (derived from two studies, $r = 0.71-0.95$; both, $p < 0.05$)^{195, 260}; BAI showed **high validity** (derived from one study, $r = 0.72$, $p < 0.01$)¹⁹⁵ and **moderate validity** (derived from two studies, $r = 0.55-0.65$; both, $p < 0.001$)^{195, 247}; HC showed **high validity** (derived from one study, $r = 0.77-0.85$, $p < 0.001$)²⁶⁰; WHtR showed **high validity** (derived from one study, $r = 0.75$, $p < 0.001$)²⁷⁸; BIA showed **very high validity** (derived from one study, $r = 0.99$, $p < 0.001$)²⁷⁸ and; WHR showed **moderate validity** (derived from one study, $r = 0.46-0.63$, $p < 0.001$)²⁶⁰.

Group with obesity: WC showed **high validity** (derived from one study, $r = 0.84$, $p < 0.001$)²⁷⁸ and; WHtR showed **high validity** (derived from one study, $r = 0.75$, $p < 0.001$)²⁷⁸.

Race-ethnicity groups

Race-ethnicity was classified, in an attempt to discuss the findings of the present systematic review, following the latest report for Medical and Science Journals²⁶².

Caucasian/White adults: BMI showed **high validity** (derived from two studies, $r = 0.78-0.84$; both, $p < 0.05$)^{193, 194} and **moderate validity** (derived from one study, $r = 0.54$, $p < 0.01$)¹⁹⁵; WC showed **high validity** (derived from two studies, $r = 0.83-0.84$; both, $p < 0.05$)^{194, 278} and **moderate validity** (derived from one study, $r = 0.57$, $p < 0.001$)²⁷⁸; WHtR showed **high validity** (derived from one study, $r = 0.75$, $p < 0.001$)²⁷⁸ and **moderate validity** (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸; BIA showed **very high/high validity** (derived from one study, $r = 0.99$, $p < 0.001$)²⁷⁸ and **moderate validity** (derived from one study, $r = 0.59$, $p < 0.001$)²⁷⁸; SKF showed **very high/high validity** (derived from one study, $r = 0.88-0.95$, $p < 0.05$)¹⁹⁶ and; BAI showed **high validity** (derived from one study, $r = 0.70$, $p < 0.01$)¹⁹⁵.

African-American/Black adults: WC showed **high validity** (derived from one study, $r = 0.84$, $p < 0.05$)¹⁹⁴; BAI showed **high validity** (derived from one study, $r = 0.81$, $p < 0.01$)¹⁹⁵; BMI showed **high validity** (derived from one study, $r = 0.78$, $p < 0.05$)¹⁹⁴ and **moderate validity** (derived from one study: $r = 0.62$, $p < 0.01$)¹⁹⁵.

Other different races-ethnicities: In Tasmanian adults, WC, HC and BMI showed **very high/high validity** ($r = 0.77-0.90$, $p < 0.0001$)²⁶⁰; in Hispanic, WC and BMI showed **high validity** ($r = 0.74-0.78$, $p < 0.05$)¹⁹⁴; in Chinese adults, WC and BMI showed **moderate to high validity** ($r = 0.63-0.72$, $p < 0.0001$)²³⁴, and HC and BAI showed **moderate validity** ($r = 0.58-0.66$, $p < 0.0001$)²³⁴; in Brazilian adults, BAI showed **moderate validity** ($r = 0.65$, $p < 0.001$)²⁴⁷; in Korean adults, WC and BMI showed **low validity** ($r = 0.23-0.39$, $p < 0.05$)²⁸⁰.

Levels of evidence

According of these results, we established the **levels of evidence** (see Table 25).

Regarding the **whole group**: (i) Strong evidence indicated that WC, BAI, and BMI are valid to estimate body adiposity; (ii) moderate evidence showed that HC is valid to estimate body adiposity and; (iii) limited evidence was found for SKF, BIA, WHR and WHtR.

Age groups

In **the 19-33 years old group**: (i) moderate evidence showed that WC is valid to estimate body adiposity; (ii) there was limited evidence about the validity of WHtR, BIA, and BMI. **In the 34-49 years old group**: (i) inconclusive evidence was found for WC and BMI, and there was limited evidence about the validity of SKF. **In the 50-64 years old group**: (i) strong evidence indicated that WC, BMI and BAI are valid to estimate total body adiposity; (ii) moderate evidence showed that HC is valid to estimate total body adiposity; (iii) there was limited evidence about the validity of WHR.

Weight status groups

Group with normal weight: (i) inconclusive evidence was found for WC and BMI; (ii) there was limited evidence about the validity of HC, SKF, BAI, BIA, and WHtR. **Group with overweight**: (i) moderate evidence showed that WC, BMI, and BAI are valid to estimate body adiposity; (ii) there was limited evidence about the validity of HC, BIA, WHR, and WHtR. **Group with obesity**: (i) there was limited evidence about the validity of WC and WHtR.

Race-ethnicity groups

In **Caucasian/White** adults: (i) strong evidence indicated that BMI is valid to estimate body adiposity; (ii) moderate evidence showed that WC is valid to estimate body adiposity; (iii) there was limited evidence about the validity of SKF, BAI, BIA, and WHtR.

Due to heterogeneity found in other races-ethnicities no further level of evidence could be established.

Criterion-related validity of estimation equations for body composition in adults.

The Durnin and Womersley equations, Jackson, Pollock and Ward equations, and Jackson and Pollock generalized equations (and Siri and Brozek converted formulas for body fat calculation) were the most common (see **Table 26-27**).

An overview of estimation equations for body composition in adults is shown in **Table 28**.

The criterion methods used to compare with field-based methods, in the estimation equations studies were: DXA^{184-186, 197-202, 205-215, 217-219, 222, 224, 230, 232, 236, 238-240, 242, 245, 248, 251, 253, 254, 257-259, 263, 265, 279} (n= 44), UWW^{181-183, 187-189, 191, 192, 203, 213, 216, 218, 219, 223, 225-227, 229, 233, 235, 237, 246, 249, 250, 256, 267-276} (n= 35), D₂O^{218, 219, 229, 241, 244, 252, 255} (n= 7),

ADP^{190, 220, 231, 243, 266, 277} (n= 6), isotope dilution^{221, 228, 261} (n= 3), ³H₂O^{218, 219} (n= 2), K40^{221, 261} (n= 2), MRI^{204, 264} (n= 2), and photon absorptiometry²²⁹ (n= 1). The field-based methods used were: SKF^{181-185, 187-192, 197-202, 206, 207, 209-217, 221, 224, 226-229, 231, 233, 235-238, 241, 244-246, 248-252, 255, 256, 259, 261, 263-275, 277, 279} (n= 66), WC^{182, 189, 198, 199, 201-206, 208-212, 215, 217, 218, 221, 224, 226, 228, 232, 236, 238-240, 243, 245, 248, 250, 253, 255-257, 259, 261, 263, 270, 273, 274, 279} (n= 42), HC^{189, 191, 192, 198, 203, 204, 206, 208-212, 215, 217, 218, 223, 226, 228, 232, 236-239, 243, 245, 249, 250, 253, 255, 256, 259, 272} (n= 32), BMI^{199, 202, 210, 212-214, 217, 218, 220, 222, 224-226, 228, 230, 233, 236, 239-241, 248, 253, 254, 258, 261, 276} (n= 26), arm circumference^{181-183, 188, 189, 191, 192, 201-203, 211, 215, 221, 223, 224, 226, 237, 250, 256, 259, 261, 269, 272-274} (n= 25), thigh circumference^{181-183, 188, 191, 192, 201-203, 206, 211, 215, 221, 223, 224, 226, 237, 245, 250, 256, 259, 269, 272-274} (n= 25), AbC^{181, 186, 188, 191, 192, 201, 211, 215, 223, 228, 250, 253, 256, 263, 268, 272-274} (n= 18), chest circumference^{181, 182, 188, 189, 191, 192, 201, 203, 206, 211, 215, 221, 223, 224, 268, 272-274} (n= 18), BIA^{184, 197, 205, 216, 233, 235, 242-244, 246, 248, 251, 252, 258, 264, 265, 267, 279} (n= 18), calf circumference^{181, 183, 188, 201-203, 211, 215, 223, 237, 245, 250, 256, 269, 272-274} (n= 17), forearm circumference^{181, 188, 191, 192, 201, 211, 215, 270, 272-274} (n= 11), NC^{181, 186, 189, 191, 192, 201, 203, 215, 223, 263} (n= 10), bi-acromial diameter^{183, 188, 191, 192, 221, 223, 272-274} (n= 9), bi-iliac diameter^{183, 188, 191, 192, 221, 223, 272-274} (n= 9), WHR^{208, 217, 218, 220, 226, 228, 243, 248} (n= 8), ankle circumference^{188, 191, 192, 221, 223, 272-274} (n= 8), shoulder circumference^{181, 188, 191, 192, 223, 272-274} (n= 8), chest diameter^{188, 191, 192, 221, 223, 272-274} (n= 8), bitrochanteric diameter^{188, 191, 192, 223, 272-274} (n= 7), knee diameter^{188, 191, 192, 223, 272-274} (n= 7), wrist diameter^{188, 191, 192, 223, 272-274} (n= 7), gluteal circumference^{182, 188, 261, 271, 273, 274} (n= 6), wrist circumference^{181, 188, 221, 223, 273, 274} (n= 6), WHtR^{199, 220, 253, 257, 258} (n= 5), BAI^{220, 232, 239, 248, 254} (n= 5), head circumference^{188, 191, 192, 223} (n= 4), ankle diameter^{188, 191, 192, 223} (n= 4), elbow diameter^{188, 191, 192, 223} (n= 4), knee circumference^{188, 191, 192} (n= 3), bideltoid diameter^{191, 192, 223} (n= 3), deltoid circumference^{191, 192} (n= 2), arm diameter^{191, 192} (n= 2), foot length^{191, 192} (n= 2), hand length^{191, 192} (n= 2), arm span²²⁶ (n= 1), bone breadths²⁰⁶ (n= 1), cup size²⁷⁴ (n= 1), leg length²²⁶ (n=1) and tibia/total body length²¹⁸ (n=1). Whenever applicable, the total body density was converted to %BF by Siri^{181-184, 187, 190-192, 197-201, 207, 209, 223, 225, 229, 231, 241, 244, 248-250, 256, 264-266, 269-271, 273-276} (n= 35), by Brozek^{182, 183, 188, 189, 191, 192, 198, 209, 216, 227, 228, 235, 237, 241, 264, 267, 269-274} (n= 22), and by Rathbun and Pace^{182, 191, 192, 268} (n= 4).

Among the studies analyzing the criterion-related validity of the equations proposed, those by Durnin/Womersley^{181, 184, 185, 190, 197, 199-201, 206, 211, 213, 226, 228, 231, 233, 237, 238, 241, 244, 249-252, 264-267, 269} (n= 28), by Jackson, Pollock and Ward^{190, 197, 198, 201, 213, 226, 237, 246, 248-250, 271} (n= 12), and by Jackson/Pollock^{197, 213, 216, 226, 237, 248, 270} (n= 7) were the most

common. Due to the large number of proposed equations, only the most common ones, at least cited in three studies, were analysed for validity.

The following classical equations showed **very high/high validity** to estimate total body fat mass or BF%, compared with criterion methods: Durnin/Womersley equation (derived from 18 studies, $r = 0.71-0.95$; all, $p < 0.05$)^{190, 197, 199-201, 211, 226, 228, 233, 237, 241, 244, 250, 252, 264, 265, 267, 269}; Jackson, Pollock and Ward equation (derived from 10 studies, $r = 0.71-0.97$; all, $p < 0.05$)^{197, 198, 213, 226, 237, 246, 248-250, 271} and; Jackson/Pollock equation (derived from six studies, $r = 0.71-0.98$; all, $p < 0.05$)^{197, 213, 216, 226, 248, 270}.

Moreover, the following estimation equations and indexes showed **very high/high validity** to estimate total body fat mass or BF%, compared with criterion methods: BMI equations (derived from 16 studies, $r = 0.71-0.90$; all, $p < 0.05$)^{199, 209, 217-220, 222, 225, 230, 233, 239, 248, 253, 254, 261, 276}; BIA equations (derived from 14 studies, $r = 0.71-0.99$; all, $p < 0.05$)^{197, 208, 216, 233, 235, 242-244, 248, 252, 258, 264, 265, 267}; equations implying sum of SKF (derived from 10 studies, $r = 0.71-0.95$; all, $p < 0.05$)^{197, 199, 211, 227, 229, 235, 246, 268, 275, 277}; equations combining circumferences+SKF (derived from nine studies, $r = 0.75-0.97$; all, $p < 0.05$)^{202, 206, 210, 212, 238, 259, 261, 272, 273}; equations based on WC (derived from seven studies, $r = 0.72-0.89$; all, $p < 0.05$)^{199, 202, 204, 205, 248, 255, 257}; equations combining WC+SKF (derived from five studies, $r = 0.80-0.99$; all, $p < 0.05$)^{200, 206, 226, 245, 279}; BAI equations (derived from five studies, $r = 0.72-0.85$; all, $p < 0.05$)^{220, 232, 239, 248, 254}; equations combining various circumferences (derived from four studies, $r = 0.80-0.93$; all, $p < 0.05$)^{202, 210, 223, 228}; WHtR equations (derived from four studies, $r = 0.72-0.87$; all, $p < 0.05$)^{199, 253, 255, 257}; equations combining AbC+SKF (derived from three studies, $r = 0.84-0.97$; all, $p < 0.05$)^{211, 215, 250}; equations combining WC+SKF+BMI (derived from three studies, $r = 0.82-0.91$; all, $p < 0.05$)^{208, 217, 236}; equations combining circumferences+diameters+SKF (derived from three studies, $r = 0.75-0.99$; all, $p < 0.05$)^{221, 273, 274}; equations combining SKF+BMI (derived from three studies, $r = 0.81-0.93$; all, $p < 0.05$)^{214, 224, 226}; equations combining abdominal circumference (AbC)+NC (derived from two studies, $r = 0.80-0.90$; both, $p < 0.05$)^{223, 263} and; WHR equations (derived from one study, $r = 0.82-0.84$, $p < 0.05$)²²⁰.

The following classical equations showed **moderate validity** to estimate total body fat mass or BF%, compared with criterion methods: Durnin/Womersley equation (derived from six studies, $r = 0.50-0.67$; all, $p < 0.001$)^{199, 206, 231, 249, 264, 265}; and Jackson/Pollock equation (derived from one study, $r = 0.67$, $p < 0.01$)²³⁷. Moreover, the following estimation equations and indexes showed **moderate validity** to estimate total

body fat mass or BF%, compared with criterion methods: BMI equations (derived from seven studies, $r = 0.50-0.68$; all, $p < 0.05$)^{199, 205, 213, 218, 219, 233, 276}; BIA equations (derived from three studies, $r = 0.54-0.69$, all, $p < 0.01$)^{233, 251, 264}; equations based on WC (derived from two studies, $r = 0.56-0.64$; both, $p < 0.05$)^{205, 209}; equations implying sum of SKF (derived from one study, $r = 0.52$, $p < 0.05$)²⁰⁹; BAI equations (derived from one study, $r = 0.56-0.64$; all, $p < 0.05$)²⁰⁵ and; WRH equations (derived from one study, $r = 0.55-0.65$, $p < 0.05$)²⁴⁸.

The following classical equation showed **very low/low validity** to estimate total body fat mass or BF%, compared with criterion methods: Durnin/Womersley equation (derived from four studies, $r = 0.00-0.44$; all, $p < 0.14$)^{199, 251, 264, 266}. Moreover, the following estimation equations and indexes showed **very low/low validity** to estimate total body fat mass or BF%, compared with criterion methods, WHR equations (derived from two studies, $r = 0.21-0.30$; both, $p < 0.05$)^{217, 255}; BAI equations (derived from one study, $r = 0.35$, $p < 0.05$)²²⁰; BMI equations (derived from one study, $r = 0.47$, $p < 0.05$)²⁷⁶ and; WHtR equations (derived from one study, $r = 0.47-0.49$, $p < 0.05$)²²⁰.

Different validity has been found between female and male populations, specific age groups (19-33 years, 34-49 years, and 50-64 years), BMI status (group with normal weight, with overweight and with obesity), and race-ethnicity.

Sex groups

Female adults: Durnin/Womersley equation showed **very high/high validity** (derived from 14 studies, $r = 0.71-0.95$; all, $p < 0.05$)^{197, 200, 201, 211, 226, 228, 233, 237, 241, 250, 252, 264, 267, 269}, **moderate validity** (derived from three studies, $r = 0.50-0.63$; all, $p < 0.05$)^{231, 249, 265}, and **low validity** (derived from two studies, $r = 0.00-0.44$; both, $p < 0.05$)^{251, 266}; Jackson, Pollock and Ward equation showed **very high/high validity** (derived from 10 studies, $r = 0.71-0.93$; all, $p < 0.05$)^{197, 198, 213, 226, 237, 246, 248-250, 271}; BMI equations showed **very high/high validity** (derived from 15 studies, $r = 0.71-0.90$; all, $p < 0.05$)^{199, 209, 213, 218-220, 222, 225, 233, 239, 241, 248, 253, 254, 276} and **moderate validity** (derived from two studies, $r = 0.55-0.68$; both, $p < 0.05$)^{205, 219}; BIA equations showed **very high/high validity** (derived from 11 studies, $r = 0.76-0.99$; all, $p < 0.05$)^{197, 208, 233, 242, 243, 248, 252, 258, 264, 265, 267} and **moderate validity** (derived from one study, $r = 0.56$, $p < 0.0001$)²⁵¹; equations combining circumferences+SKF showed **high validity** (derived from eight studies, $r = 0.82-0.97$; all, $p < 0.05$)^{202, 206, 210, 212, 238, 259, 261, 272}; equations implying sum of SKF showed **very high/high validity** (derived from six studies, $r = 0.75-0.92$; all, $p < 0.05$)^{197, 199, 229, 246, 275, 277} and **moderate validity** (derived from one study, $r = 0.52$, $p < 0.05$)²⁰⁹; BAI equations

showed showed **high validity** (derived from five studies, $r = 0.76-0.85$; all, $p < 0.05$)^{220, 232, 239, 248, 254} and **moderate validity** (derived from one study, $r = 0.56$, $p < 0.05$)²⁰⁵; equations based on WC showed **high validity** (derived from four studies, $r = 0.74-0.84$; all, $p < 0.05$)^{199, 202, 248, 257} and **moderate validity** (derived from two studies, $r = 0.56-0.64$; both, $p < 0.05$)^{205, 209}; equations combining various circumferences showed **very high/high validity** (derived from four studies, $r = 0.80-0.93$; all, $p < 0.05$)^{202, 210, 223, 228}; equations combining AbC+SKF showed **high validity** (derived from three studies, $r = 0.85-0.94$; all, $p < 0.05$)^{211, 215, 250}; equations combining SKF+BMI showed **high validity** (derived from three studies, $r = 0.85-0.89$; all, $p < 0.05$)^{214, 224, 226}; WHtR equations showed **high validity** (derived from three studies, $r = 0.74-0.85$; all, $p < 0.05$)^{199, 253, 257}; equations combining WC+SKF showed **very high/high validity** (derived from two studies, $r = 0.93-0.98$; both, $p < 0.05$)^{245, 279}; equations combining WC+SKF+BMI showed **high validity** (derived from two studies, $r = 0.87-0.90$; both, $p < 0.05$)^{208, 217}; equations combining circumferences+diameters+SKF showed **very high validity** (derived from two studies, $r = 0.92-0.99$; both, $p < 0.05$)^{221, 274}; WHR equations showed **very high/high validity** (derived from one study, $r = 0.82-0.84$, $p < 0.05$)²²⁰ and **moderate validity** (derived from one study, $r = 0.55$, $p < 0.05$)²⁴⁸ and; equations combining AbC+NC showed **very high validity** (derived from one study, $r = 0.81$, $p < 0.05$)²⁶³.

Male adults: Durnin/Womersley equation showed **very high/high validity** (derived from 11 studies, $r = 0.70-0.95$; all, $p < 0.05$)^{197, 200, 211, 226, 233, 237, 241, 244, 265, 267, 269} and **low validity** (derived from two studies, $r = 0.00-0.44$; both, $p < 0.05$)^{251, 266}; Jackson/Pollock equation showed **very high/high validity** (derived from six studies, $r = 0.86-0.92$; all, $p < 0.05$)^{197, 213, 216, 226, 248, 270} and **moderate validity** (derived from one study, $r = 0.67$, $p < 0.01$)²³⁷; BMI equations showed **very high/high validity** (derived from 12 studies, $r = 0.72-0.90$; all, $p < 0.05$)^{199, 218-220, 222, 225, 239, 241, 248, 253, 254, 276} and **moderate validity** (derived from six studies, $r = 0.61-0.68$; all, $p < 0.05$)^{199, 205, 213, 218, 219, 233}; equations implying sum of SKF showed **very high/high validity** (derived from nine studies, $r = 0.70-0.95$; all, $p < 0.05$)^{197, 199, 211, 227, 229, 235, 268, 275, 277}; BIA equations showed **very high/high validity** (derived from nine studies, $r = 0.71-0.99$; all, $p < 0.05$)^{197, 216, 235, 243, 244, 248, 258, 265, 267} and **moderate validity** (derived from two studies, $r = 0.56-0.69$; both, $p < 0.05$)^{233, 251}; equations based on WC showed **high validity** (derived from seven studies, $r = 0.73-0.89$; all, $p < 0.05$)^{199, 202, 204, 205, 248, 255, 257}; equations combining circumferences+SKF showed **high validity** (derived from five studies, $r = 0.82-0.96$; all, $p < 0.05$)^{202, 210, 238, 259, 273}; BAI equations showed **high validity** (derived from four studies,

$r = 0.72-0.85$; all, $p < 0.05$)^{232, 239, 248, 254}, **moderate validity** (derived from one study, $r = 0.64$, $p < 0.05$)²⁰⁵, and **low validity** (derived from one study, $r = 0.35$, $p < 0.05$)²²⁰; equations combining WC+SKF showed **very high/high validity** (derived from four studies, $r = 0.86-0.93$; all, $p < 0.05$)^{206, 226, 245, 279}; WHtR equations showed **high validity** (derived from four studies, $r = 0.72-0.87$; all, $p < 0.05$)^{199, 253, 255, 257}; equations combining AbC+NC showed **very high validity** (derived from two studies, $r = 0.88-0.90$; both, $p < 0.05$)^{223, 263}; equations combining WC+SKF+BMI showed **high validity** (derived from two studies, $r = 0.82-0.91$; both, $p < 0.05$)^{217, 236}; equations combining various circumferences showed **very high/high validity** (derived from two studies, $r = 0.86-0.92$; both, $p < 0.05$)^{202, 210}; equations combining circumferences+diameters+SKF showed **very high validity** (derived from two studies, $r = 0.94-0.99$; both, $p < 0.05$)^{221, 273}; equations combining SKF+BMI showed **high validity** (derived from two studies, $r = 0.82-0.84$; both, $p < 0.05$)^{214, 224}; WHR equations showed **high validity** (derived from one study, $r = 0.84$, $p < 0.05$)²²⁰, **moderate validity** (derived from one study, $r = 0.65$, $p < 0.05$)²⁴⁸, and **low validity** (derived from one study, $r = 0.33$, $p < 0.05$)²⁵⁵ and; equations combining AbC+SKF showed **high validity** (derived from one study, $r = 0.97$, $p < 0.001$)²¹⁵.

Age groups

The 19-33 years old group: Durnin/Womersley equation showed **very high/high validity** (derived from 11 studies, $r = 0.71-0.95$; all, $p < 0.05$)^{197, 201, 211, 228, 233, 237, 250, 252, 264, 267, 269} and **moderate validity** (derived from two studies, $r = 0.50-0.62$; both, $p < 0.05$)^{231, 249}; Jackson, Pollock and Ward equation showed **very high/high validity** (derived from seven studies, $r = 0.71-0.92$; all, $p < 0.05$)^{197, 198, 213, 237, 249, 250, 271}; Jackson/Pollock equation showed **very high/high validity** (derived from four studies, $r = 0.77-0.92$; all, $p < 0.05$)^{197, 213, 216, 270}; BIA equations showed **very high/high validity** (derived from seven studies, $r = 0.74-0.96$; all, $p < 0.05$)^{197, 216, 233, 235, 252, 264, 267} and **moderate validity** (derived from one study, $r = 0.69$, $p < 0.05$)²³³; equations implying sum of SKF showed **very high/high validity** (derived from six studies, $r = 0.71-0.95$; all, $p < 0.05$)^{211, 227, 229, 235, 268, 275}; BMI equations showed **very high/high validity** (derived from three studies, $r = 0.89-0.92$; all, $p < 0.05$)^{222, 230, 276}, **moderate validity** (derived from three studies, $r = 0.51-0.66$; all, $p < 0.05$)^{205, 213, 276}, and **low validity** (derived from one study, $r = 0.47$, $p < 0.05$)²⁷⁶; equations based on WC showed **high validity** (derived from three studies, $r = 0.73-0.85$; all, $p < 0.05$)^{205, 255, 257} and **moderate validity** (derived from one study, $r = 0.64$, $p < 0.05$)²⁰⁵; equations combining AbC+SKF showed **high validity** (derived from two studies, $r = 0.85-0.92$; both, $p < 0.05$)^{211, 250}; equations combining

various circumferences showed **very high/high validity** (derived from two studies, $r=0.90-0.93$; both, $p < 0.05$)^{223, 228}; equations combining circumferences+diameters+SKF showed **very high validity** (derived from two studies, $r=0.92-0.94$; both, $p < 0.05$)^{273, 274}; WHtR equations showed **high validity** (derived from two studies, $r=0.72-0.87$; both, $p < 0.05$)^{255, 257}; equations combining AbC+NC showed **very high validity** (derived from one study, $r=0.90$, $p < 0.05$)²²³; equations combining WC+SKF showed **very high/high validity** (derived from one study, $r=0.88-0.99$, $p < 0.05$)²⁷⁹; equations combining WC+SKF+BMI showed **high validity** (derived from one study, $r=0.82$, $p < 0.05$)²³⁶; equations combining circumferences+SKF showed **high validity** (derived from one study, $r=0.82$, $p < 0.05$)²⁷²; equations combining SKF+BMI showed **high validity** (derived from one study, $EST=0.17-0.19$, $p < 0.005$)²¹⁴; BAI equations showed **moderate validity** (derived from one study, $r=0.56-0.64$, $p < 0.05$)²⁰⁵ and; WHR equations showed **low validity** (derived from one study, $r=0.30$, $p < 0.05$)²⁵⁵.

The 34-49 years old group: Durnin/Womersley equation showed **very high/high validity** (derived from 11 studies, $r=0.75-0.95$; all, $p < 0.05$)^{199, 200, 211, 226, 233, 237, 241, 244, 250, 265, 269}, **moderate validity** (derived from one study, $r=0.63$, $p < 0.01$)²⁶⁵ and **very low/low validity** (derived from one study, $r=0.00-0.39$, $p < 0.05$)²⁶⁶; Jackson, Pollock and Ward equation showed **high validity** (derived from five studies, $r=0.77-0.89$; all, $p < 0.05$)^{226, 246, 248, 250, 271}; Jackson/Pollock equation showed **very high/high validity** (derived from three studies, $r=0.86-0.92$; all, $p < 0.05$)^{226, 248, 270} and **moderate validity** (derived from one study, $r=0.67$, $p < 0.05$)²³⁷; BMI equations showed **very high/high validity** (derived from 11 studies, $r=0.72-0.90$; all, $p < 0.05$)^{199, 218-220, 222, 225, 239, 248, 253, 254, 276} and **moderate validity** (derived from three studies, $r=0.66-0.68$; all, $p < 0.05$)^{199, 218, 219}; equations combining circumferences+SKF showed **very high validity** (derived from seven studies, $r=0.75-0.97$; all, $p < 0.05$)^{202, 206, 210, 238, 259, 261, 273}; equations implying sum of SKF showed **very high/high validity** (derived from six studies, $r=0.70-0.95$; all, $p < 0.05$)^{199, 211, 246, 268, 275, 277}; BIA equations showed **very high/high validity** (derived from six studies, $r=0.75-0.98$; all, $p < 0.05$)^{233, 243, 244, 248, 258, 265} and **moderate validity** (derived from one study, $r=0.69$, $p < 0.05$)²³³; BAI equations showed **high validity** (derived from five studies, $r=0.72-0.85$; all, $p < 0.05$)^{220, 232, 239, 248, 254} and **low validity** (derived from one study, $r=0.35$, $p < 0.05$)²²⁰; equations based on WC showed **high validity** (derived from four studies, $r=0.74-0.89$; all, $p < 0.05$)^{199, 202, 204, 248}; equations combining AbC+SKF showed **very high/high validity** (derived from three studies, $r=0.84-0.97$; all, $p < 0.05$)^{211, 215, 250}; equations combining WC+SKF showed **very high**

validity (derived from three studies, $r = 0.80-0.93$; all, $p < 0.05$)^{200, 226, 245}; WHtR equations showed **high validity** (derived from two studies, $r = 0.74-0.82$; both, $p < 0.05$)^{199, 253} and **moderate validity** (derived from one study, $r = 0.47-0.49$, $p < 0.05$)²²⁰; equations combining WC+SKF+BMI showed **very high/high validity** (derived from two studies, $r = 0.82-0.91$; both, $p < 0.05$)^{217, 236}; equations combining various circumferences showed **very high/high validity** (derived from two studies, $r = 0.83-0.92$; both, $p < 0.05$)^{202, 210}; equations combining circumferences+diameters+SKF showed **very high validity** (derived from two studies, $r = 0.92-0.95$; both, $p < 0.05$)^{221, 274}; WHR equations showed **high validity** (derived from one study, $r = 0.82-0.84$, $p < 0.05$)²²⁰ and **moderate validity** (derived from one study, $r = 0.55-0.65$, $p < 0.05$)²⁴⁸; equations combining AbC+NC showed **high validity** (derived from one study, $r = 0.81-0.88$, $p < 0.05$)²⁶³ and; equations combining SKF+BMI showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²²⁶.

The 50-64 years old group: Durnin/Womersley equation showed **very high/high validity** (derived from two studies, $r = 0.84-0.95$; both, $p < 0.05$)^{200, 269} and **low validity** (derived from one study, $r = 0.44$, $p < 0.05$)²⁵¹; Jackson/Pollock equation showed **very high validity** (derived from one study, $r = 0.91-0.92$; all, $p < 0.05$)²⁷⁰; BMI equations showed **very high/high validity** (derived from four studies, $r = 0.71-0.90$; all, $p < 0.05$)^{209, 218, 222, 276} and **moderate validity** (derived from two studies, $r = 0.50-0.68$; both, $p < 0.05$)^{218, 276}; equations implying sum of SKF showed **high validity** (derived from two studies, $r = 0.80-0.85$; both, $p < 0.05$)^{275, 277} and **moderate validity** (derived from one study, $r = 0.52$, $p < 0.05$)²⁰⁹; BIA equations showed **very high/high validity** (derived from two studies, $r = 0.80-0.91$; both, $p < 0.001$)^{208, 242} and **moderate validity** (derived from one study, $r = 0.56$, $p < 0.05$)²⁵¹; equations combining WC+SKF+BMI showed **high validity** (derived from two studies, $r = 0.82-0.89$; both, $p < 0.05$)^{208, 236}; equations combining circumferences+SKF showed **very high validity** (derived from two studies, $r = 0.95-0.96$; both, $p < 0.05$)^{212, 261}; equations combining WC+SKF showed **high validity** (derived from one study, $r = 0.80-0.89$, $p < 0.05$)²⁰⁰; equations combining SKF+BMI showed **high validity** (derived from one study, $r = 0.82-0.87$, $p < 0.05$)²²⁴ and; equations based on WC showed **moderate validity** (derived from one study, $r = 0.56$, $p < 0.05$)²⁰⁹.

Weight status groups

Group with normal weight: Durnin/Womersley equation showed **very high/high validity** (derived from nine studies, $r = 0.71-0.94$; all, $p < 0.05$)^{199-201, 233, 241, 244, 250, 252, 264}, **moderate validity** (derived from three studies, $r = 0.50-0.62$; all, $p < 0.05$)^{199, 231, 249}, and **low validity** (derived from one study, $r = 0.26$, $p < 0.05$)²⁶⁶; Jackson, Pollock

and Ward equation showed **very high/high validity** (derived from three studies, $r = 0.77-0.92$; all, $p < 0.05$)^{198, 249, 250}; BMI equations showed **high validity** (derived from seven studies, $r = 0.75-0.93$; all, $p < 0.05$)^{217-219, 222, 230, 233, 276}, **moderate validity** (derived from three studies, $r = 0.55-0.62$; all, $p < 0.05$)^{199, 205, 233}, and **low validity** (derived from one study, $r = 0.42$, $p < 0.05$)¹⁹⁹; BIA equations showed **very high/high validity** (derived from four studies, $r = 0.76-0.96$; all, $p < 0.05$)^{197, 233, 244, 252} and **moderate validity** (derived from two studies, $r = 0.54-0.69$; both, $p < 0.05$)^{233, 264}; equations based on WC showed **high validity** (derived from three studies, $r = 0.71-0.85$; all, $p < 0.05$)^{199, 205, 257}, **moderate validity** (derived from one study, $r = 0.64$, $p < 0.05$)²⁰⁵, and **low validity** (derived from one study, $r = 0.46$, $p < 0.05$)¹⁹⁹; equations combining WC+SKF showed **very high/high validity** (derived from three studies, $r = 0.80-0.93$; all, $p < 0.05$)^{200, 226, 245}; equations combining circumferences+SKF showed **very high validity** (derived from three studies, $r = 0.94-0.97$; all, $p < 0.05$)^{212, 238, 261}; equations combining WC+SKF+BMI showed **very high/high validity** (derived from two studies, $r = 0.82-0.91$; both, $p < 0.05$)^{217, 236}; equations implying sum of SKF showed **high validity** (derived from two studies, $r = 0.77-0.88$; both, $p < 0.05$)^{197, 199}; equations combining SKF+BMI showed **high validity** (derived from two studies, $r = 0.82-0.89$; both, $p < 0.05$)^{224, 226}; WHtR equations showed **high validity** (derived from two studies, $r = 0.71-0.87$; both, $p < 0.05$)^{199, 257}, and **low validity** (derived from one study, $r = 0.46$, $p < 0.05$)¹⁹⁹; equations combining AbC+NC showed **very high validity** (derived from one study, $r = 0.90$, $p < 0.05$)²²³; equations combining AbC+SKF showed **high validity** (derived from one study, $r = 0.85$, $p < 0.05$)²⁵⁰; equations combining various circumferences showed **very high validity** (derived from one study, $r = 0.90$, $p < 0.05$)²²³; BAI equations showed **moderate validity** (derived from one study, $r = 0.56-0.64$, $p < 0.05$)²⁰⁵ and; WHR equations showed **very low/low validity** (derived from one study, $r = 0.21-0.27$, $p < 0.05$)²¹⁷.

Group with overweight: Durnin/Womersley equation showed **very high/high validity** (derived from three studies, $r = 0.80-0.92$; all, $p < 0.05$)^{200, 211, 265}, **moderate validity** (derived from three studies, $r = 0.59-0.65$; all, $p < 0.05$)^{199, 264, 265}, and **low validity** (derived from three studies, $r = 0.17-0.44$; all, $p < 0.05$)^{199, 251, 266}; Jackson, Pollock and Ward equation showed **high validity** (derived from two studies, $r = 0.82-0.89$; both, $p < 0.05$)^{246, 248}; Jackson/Pollock equation showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²⁴⁸; BMI equations showed **very high/high validity** (derived from 11 studies, $r = 0.71-0.94$; all, $p < 0.05$)^{209, 218, 219, 222, 225, 230, 239, 248, 253, 254, 276}, **moderate validity** (derived from one study, $r = 0.66$, $p < 0.05$)²¹⁸, and **low validity** (derived from one study,

$r = 0.14-0.45$, $p < 0.05$)¹⁹⁹; BIA equations showed **very high/high validity** (derived from six studies, $r = 0.75-0.98$; all, $p < 0.05$)^{208, 243, 248, 258, 264, 265} and **moderate validity** (derived from one study, $r = 0.56$, $p < 0.05$)²⁵¹; equations combining circumferences+SKF showed **very high/high validity** (derived from five studies, $r = 0.79-0.96$; all, $p < 0.05$)^{202, 206, 210, 212, 238}; equations based on WC showed **high validity** (derived from three studies, $r = 0.81-0.89$; all, $p < 0.05$)^{202, 204, 248}, **moderate validity** (derived from two studies, $r = 0.56-0.62$; both, $p < 0.05$)^{199, 209}, and **low validity** (derived from one study, $r = 0.20$, $p < 0.05$)¹⁹⁹; equations combining WC+SKF showed **very high/high validity** (derived from three studies, $r = 0.80-0.99$; all, $p < 0.05$)^{200, 206, 279}; equations combining various circumferences showed **very high validity** (derived from three studies, $r = 0.83-0.92$; all, $p < 0.05$)^{202, 210, 223}; BAI equations showed **high validity** (derived from three studies, $r = 0.72-0.83$; all, $p < 0.05$)^{239, 248, 254}; equations implying sum of SKF showed **very high/high validity** (derived from two studies, $r = 0.82-0.92$; both, $p < 0.05$)^{211, 246}, **moderate validity** (derived from two studies, $r = 0.52-0.65$; both, $p < 0.05$)^{199, 209}, and **low validity** (derived from one study, $r = 0.17$, $p < 0.05$)¹⁹⁹; equations combining AbC+NC showed **very high/high validity** (derived from two studies, $r = 0.81-0.90$; both, $p < 0.05$)^{223, 263}; equations combining AbC+SKF showed **very high/validity** (derived from two studies, $r = 0.94-0.97$; both, $p < 0.05$)^{211, 215}; equations combining SKF+BMI showed **very high/high validity** (derived from two studies, $r = 0.81-0.93$; both, $p < 0.05$)^{214, 224}; WHtR equations showed **high validity** (derived from one study, $r = 0.74-0.82$, $p < 0.05$)²⁵³, **moderate validity** (derived from one study, $r = 0.62$, $p < 0.05$)¹⁹⁹, and **low validity** (derived from one study, $r = 0.20$, $p < 0.05$)¹⁹⁹; equations combining WC+SKF+BMI showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²⁰⁸ and; WHR equations showed **moderate validity** (derived from one study, $r = 0.55-0.65$, $p < 0.05$)²⁴⁸.

Group with obesity: Durnin/Womersley equation showed **very high/high validity** (derived from one study, $r = 0.85-0.92$, $p < 0.05$)²²⁸; **moderate validity** (derived from one study, $r = 0.60$, $p < 0.05$)¹⁹⁹, and **low validity** (derived from for studies, $r = 0.00-0.44$; all, $p < 0.05$)^{199, 251, 264, 266}; BMI equations showed **very high/high validity** (derived from four studies, $r = 0.77-0.90$; all, $p < 0.05$)^{220, 222, 230, 276}, **moderate validity** (derived from one study, $r = 0.54$, $p < 0.05$)¹⁹⁹, and **low validity** (derived from one study, $r = 0.28$, $p < 0.05$)¹⁹⁹; BIA equations showed **very high/high validity** (derived from three studies, $r = 0.80-0.92$; all, $p < 0.05$)^{208, 242, 264} and **moderate validity** (derived from one study, $r = 0.56$, $p < 0.05$)²⁵¹; equations combining circumferences+SKF showed **very high validity** (derived from two studies, $r = 0.94-0.97$; both, $p < 0.05$)^{212, 238}; BAI equations showed **high**

validity (derived from two studies, $r = 0.81-0.85$; both, $p < 0.05$)^{220, 232} and **low validity** (derived from one study, $r = 0.35$, $p < 0.05$)²²⁰; equations based on WC showed **moderate validity** (derived from one study, $r = 0.64$, $p < 0.05$)¹⁹⁹ and **low validity** (derived from one study, $r = 0.35$, $p < 0.05$)¹⁹⁹; equations implying sum of SKF showed **moderate validity** (derived from one study, $r = 0.65$, $p < 0.05$)¹⁹⁹ and **low validity** (derived from one study, $r = 0.17$, $p < 0.05$)¹⁹⁹; and WHtR equations showed **moderate validity** (derived from one study, $r = 0.64$, $p < 0.05$)¹⁹⁹ and **low validity** (derived from two studies, $r = 0.35-0.49$; both, $p < 0.05$)^{199, 220}; equations combining SKF+BMI showed **high validity** (derived from one study, $r = 0.82-0.87$, $p < 0.05$)²²⁴; WHR equations showed **high validity** (derived from one study, $r = 0.82-0.84$, $p < 0.05$)²²⁰; equations combining WC+SKF+BMI showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²⁰⁸ and; equations combining various circumferences showed **very high validity** (derived from one study, $r = 0.90-0.93$, $p < 0.05$)²²⁸.

Race-ethnicity groups

Caucasian/White adults: Durnin/Womersley equation showed **very high/high validity** (derived from five studies, $r = 0.73-0.92$; all, $p < 0.05$)^{199-201, 211, 228}; Jackson/Pollock equation showed **very high/high validity** (derived from two studies, $r = 0.88-0.90$; both, $p < 0.05$)^{213, 216}; Jackson, Pollock and Ward equation showed **very high/high validity** (derived from two studies, $r = 0.88-0.92$; both, $p < 0.05$)^{198, 213}; BMI equations showed **very high/high validity** (derived from 10 studies, $r = 0.71-0.93$; all, $p < 0.05$)^{199, 209, 217-220, 222, 225, 230, 254} and **moderate validity** (derived from two studies, $r = 0.55-0.64$; both, $p < 0.05$)^{205, 213}; equations combining circumferences+SKF showed **very high/high validity** (derived from seven studies, $r = 0.75-0.96$; all, $p < 0.05$)^{188, 202, 206, 210, 212, 259, 261}; equations implying sum of SKF showed **very high/high validity** (derived from five studies, $r = 0.73-0.95$; all, $p < 0.05$)^{197, 199, 211, 227, 229} and **moderate validity** (derived from one study, $r = 0.52$, $p < 0.05$)²⁰⁹; equations based on WC showed **high validity** (derived from four studies, $r = 0.72-0.89$; all, $p < 0.05$)^{199, 202, 204, 205} and **moderate validity** (derived from two studies, $r = 0.56-0.64$; both, $p < 0.05$)^{205, 209}; equations combining various circumferences showed **very high validity** (derived from four studies, $r = 0.83-0.93$; all, $p < 0.05$)^{202, 210, 223, 228}; equations combining WC+SKF showed **very high/high validity** (derived from three studies, $r = 0.80-0.93$; all, $p < 0.05$)^{200, 206, 226}; equations combining SKF+BMI showed **very high/high validity** (derived from three studies, $r = 0.81-0.93$; all, $p < 0.05$)^{214, 224, 226}; BIA equations showed **very high/high validity** (derived from three studies, $r = 0.74-0.92$; all, $p < 0.05$)^{197, 208, 216}; BAI equations showed **high**

validity (derived from two studies, $r = 0.72-0.81$; both, $p < 0.05$)^{220, 254}, **moderate validity** (derived from one study, $r = 0.56-0.64$, $p < 0.05$)²⁰⁵, and **low validity** (derived from one study, $r = 0.35$, $p < 0.05$)²²⁰; equations combining AbC+SKF showed **very high validity** (derived from two studies, $r = 0.94-0.97$; both, $p < 0.05$)^{211, 215}; equations combining WC+SKF+BMI showed **very high/high validity** (derived from two studies, $r = 0.89-0.90$; both, $p < 0.05$)^{208, 217}; WHR equations showed **high validity** (derived from one study, $r = 0.82-0.84$, $p < 0.05$)²²⁰ and **low validity** (derived from one study, $r = 0.21-0.27$, $p < 0.05$)²¹⁷; WHtR equations showed **high validity** (derived from one study, $r = 0.72-0.82$, $p < 0.05$)¹⁹⁹ and **low validity** (derived from one study, $r = 0.47-0.49$, $p < 0.05$)²²⁰; equations combining AbC+NC showed **high validity** (derived from one study, $r = 0.88$, $p < 0.05$)²²³ and; equations combining circumferences+diameters+SKF showed **very high validity** (derived from one study, $r = 0.96-0.99$, $p < 0.05$)²²¹.

African-American/Black adults: Durnin/Womersley equation showed **high validity** (derived from two studies, $r = 0.74-0.89$; both, $p < 0.05$)¹⁹⁹⁻²⁰¹ and **moderate validity** (derived from one study, $r = 0.50-0.53$, $p < 0.05$)²³¹; Jackson, Pollock and Ward equation showed **very high/high validity** (derived from one study, $r = 0.89-0.92$, $p < 0.05$)¹⁹⁸; BMI equations showed **very high/high validity** (derived from five studies, $r = 0.72-0.93$; all, $p < 0.05$)^{199, 218, 219, 225, 230} and **moderate validity** (derived from one study, $r = 0.55-0.61$, $p < 0.05$)²⁰⁵; equations based on WC showed **high validity** (derived from four studies, $r = 0.70-0.89$; all, $p < 0.05$)^{199, 202, 204, 205} and **moderate validity** (derived from one study, $r = 0.64$, $p < 0.05$)²⁰⁵; equations combining various circumferences showed **high validity** (derived from two studies, $r = 0.83-0.86$; both, $p < 0.05$)^{202, 223}; equations implying sum of SKF showed **very high/high validity** (derived from two studies, $r = 0.74-0.90$; both, $p < 0.05$)^{199, 229}; BAI equations showed **high validity** (derived from one study, $r = 0.85$, $p < 0.05$)²³² and **moderate validity** (derived from one study, $r = 0.56-0.64$, $p < 0.05$)²⁰⁵; equations combining AbC+NC showed **very high validity** (derived from one study, $r = 0.90$, $p < 0.05$)²²³; equations combining WC+SKF showed **high validity** (derived from one study, $r = 0.80-0.89$, $p < 0.05$)²⁰⁰; equations combining circumferences+SKF showed **high validity** (derived from one study, $r = 0.79-0.89$, $p < 0.05$)²⁰²; equations combining circumferences+diameters+SKF showed **high validity** (derived from one study, $r = 0.96-0.99$, $p < 0.05$)²²¹; equations combining SKF+BMI showed **high validity** (derived from one study, $EST = 0.17-0.19$, $p < 0.005$)²¹⁴ and; WHtR equations showed **high validity** (derived from one study, $r = 0.70-0.84$, $p < 0.05$)¹⁹⁹.

Hispanic adults: Durnin/Womersley equation showed **high validity** (derived from four studies, $r = 0.75-0.89$; all, $p < 0.05$)^{199-201, 250} and **moderate validity** (derived from two studies, $r = 0.62-0.68$; both, $p < 0.05$)^{199, 249}; Jackson, Pollock and Ward equation showed **very high/high validity** (derived from four studies, $r = 0.71-0.92$; all, $p < 0.05$)^{198, 248-250}; Jackson/Pollock equation showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²⁴⁸; equations based on WC showed **high validity** (derived from six studies, $r = 0.73-0.89$; all, $p < 0.05$)^{199, 202, 204, 205, 248, 257} and **moderate validity** (derived from two studies, $r = 0.64-0.68$; both, $p < 0.05$)^{199, 205}; BIA equations showed **very high/high validity** (derived from three studies, $r = 0.75-0.98$; all, $p < 0.05$)^{242, 243, 248}; BAI equations showed **high validity** (derived from two studies, $r = 0.72-0.85$; both, $p < 0.05$)^{232, 248} and **moderate validity** (derived from one study, $r = 0.56-0.64$, $p < 0.05$)²⁰⁵; BMI equations showed **high validity** (derived from two studies, $r = 0.70-0.82$; both, $p < 0.05$)^{199, 248} and **moderate validity** (derived from one study, $r = 0.55-0.61$, $p < 0.05$)²⁰⁵; WHtR equations showed **high validity** (derived from two studies, $r = 0.81-0.87$; both, $p < 0.05$)^{199, 257} and **moderate validity** (derived from one study, $r = 0.68$, $p < 0.05$)¹⁹⁹; equations combining WC+SKF showed **very high/high validity** (derived from two studies, $r = 0.80-0.99$; both, $p < 0.05$)^{200, 279}; equations combining various circumferences showed **very high validity** (derived from two studies, $r = 0.83-0.92$; both, $p < 0.05$)^{202, 223}; equations implying sum of SKF showed **high validity** (derived from one study, $r = 0.79$, $p < 0.05$)¹⁹⁹ and **moderate validity** (derived from one study, $r = 0.68$, $p < 0.05$)¹⁹⁹; equations combining AbC+NC showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²²³; equations combining AbC+SKF showed **high validity** (derived from one study, $r = 0.85$, $p < 0.05$)²⁵⁰; equations combining circumferences+SKF showed **high validity** (derived from one study, $r = 0.79-0.89$, $p < 0.05$)²⁰²; equations combining SKF+BMI showed **high validity** (derived from one study, $EST = 0.17-0.19$, $p < 0.005$)²¹⁴ and; WHR equations showed **moderate validity** (derived from one study, $r = 0.55-0.65$, $p < 0.05$)²⁴⁸.

Asian adults: Durnin/Womersley equation showed **very high/high validity** (derived from six studies, $r = 0.70-0.94$; all, $p < 0.05$)^{200, 233, 237, 241, 244, 252}; Jackson, Pollock and Ward equation showed **very high/high validity** (derived from three studies, $r = 0.82-0.93$; all, $p < 0.05$)^{198, 237, 246}; Jackson/Pollock equation showed **moderate validity** (derived from one study, $r = 0.67$, $p < 0.05$)²³⁷; BIA equations showed **very high/high validity** (derived from four studies, $r = 0.71-0.96$; all, $p < 0.05$)^{233, 235, 244, 252} and **moderate validity** (derived from one study, $r = 0.69$, $p < 0.05$)²³³; BMI equations showed **very high/high validity** (derived from three studies, $r = 0.82-0.94$; all, $p < 0.05$)^{219, 233, 239} and

moderate validity (derived from one study, $r = 0.62$, $p < 0.05$)²³³; equations combining WC+SKF showed **very high/high validity** (derived from two studies, $r = 0.80$ - 0.93 ; both, $p < 0.05$)^{200, 245}; equations implying sum of SKF showed **high validity** (derived from two studies, $r = 0.71$ - 0.82 ; both, $p < 0.05$)^{235, 246}; equations based on WC showed **high validity** (derived from one study, $r = 0.89$, $p < 0.05$)²⁰⁴; equations combining WC+SKF+BMI showed **high validity** (derived from one study, $r = 0.82$, $p < 0.05$)²³⁶; equations combining circumferences+SKF showed **very high validity** (derived from one study, $r = 0.94$ - 0.97 , $p < 0.05$)²³⁸; equations combining SKF+BMI showed **high validity** (derived from one study, $r = 0.82$ - 0.85 , $p < 0.05$)²²⁴ and; BAI equations showed **high validity** (derived from one study, $r = 0.83$, $p < 0.05$)²³⁹.

Other different races-ethnicities: In Australian adults, WHtR and BMI equations showed **high validity** ($r = 0.83$ - 0.87 , $p < 0.01$)²⁵³; in Canadian adults, BIA equations showed **high validity** ($r = 0.88$, $p < 0.05$)²⁵⁸; in Polynesian females, Durnin/Womersley equation showed **high validity** ($r = 0.85$, $p < 0.05$)²²⁸, and equations combining various circumferences showed **very high validity** ($r = 0.90$, $p < 0.05$)²²⁸; in Syrian males, equations based on WC and WHtR showed **very high/high validity** ($r = 0.72$ - 0.74 , $p < 0.01$)²⁵⁵, and WHR **low validity** ($r = 0.33$, $p < 0.001$)²⁵⁵; in Anglo-Celtic Australian adults, BIA equations showed **moderate validity** ($r = 0.56$, $p < 0.0001$)²⁵¹, and Durnin/Womersley equation showed **low validity** ($r = 0.44$, $p < 0.0001$)²⁵¹.

Levels of evidence

According to these results, we constructed the **levels of evidence** (see **Table 29**).

Regarding the **whole group**: (i) Strong evidence indicated that Durnin/Womersley equation, BMI equations, BIA equations, equations implying sum of SKF, equations combining circumferences+SKF, equations based on WC, equations combining WC+SKF, BAI equations, equations combining various circumferences, WHtR equations, equations combining AbC+SKF, equations combining WC+SKF+BMI, equations combining circumferences+diameters+SKF, and equations combining SKF+BMI are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining AbC+NC, and WHR are valid to estimate total body fat mass or %BF.

Sex groups

The levels of evidence in **female adults** were: (i) strong evidence indicated that Durnin/Womersley equation, Jackson, Pollock and Ward equation, BMI equations, BIA

equations, equations combining circumferences+SKF, equations implying sum of SKF, BAI equations, equations based on WC, equations combining various circumferences, equations combining AbC+SKF, equations combining SKF+BMI, and WHtR equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining WC+SKF, equations combining WC+SKF+BMI, equations combining circumferences+diameters+SKF, and WHR equations are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining AbC+NC. **In male adults:** (i) strong evidence indicated that Durnin/Womersley equation, Jackson/Pollock equation, BMI equations, equations implying sum of SKF, BIA equations, equations based on WC, equations combining circumferences+SKF, BAI equations, equations combining WC+SKF, and WHtR equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining AbC+NC, equations combining WC+SKF+BMI, equations combining various circumferences, equations combining circumferences+diameters+SKF, and equations combining SKF+BMI are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining AbC+SKF (derived from a single study); (iv) inconclusive evidence was found for WHR equations.

Age groups

The levels of evidence in **the 19-33 years old group** were: (i) strong evidence indicated that Durnin/Womersley equation, Jackson/Pollock equation, Jackson, Pollock and Ward equation, BIA equations, equations implying sum of SKF, BMI equations, and equations based on WC are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining AbC+SKF, equations combining various circumferences, equations combining circumferences+diameters+SKF, and WHtR equations are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining AbC+NC, equations combining WC+SKF, equations combining WC+SKF+BMI, equations combining circumferences+SKF, equations combining SKF+BMI, BAI equations, and WHR equations (derived from a single study). **In the 34-49 years old group:** (i) strong evidence indicated that Durnin/Womersley equation, Jackson/Pollock equation, Jackson, Pollock and Ward equation, BMI equations, equations combining circumferences+SKF, equations implying sum of SKF, BIA equations, BAI equations, equations based on WC, equations combining AbC+SKF, equations combining WC+SKF, and WHtR equations are valid to

estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining WC+SKF+BMI, equations combining various circumferences, equations combining circumferences+diameters+SKF, and WHR equations are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining AbC+NC, and equations combining SKF+BMI (derived from a single study).

In the 50-64 years old group: (i) strong evidence indicated that BMI equations, equations implying sum of SKF, and BIA equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining WC+SKF+BMI, and equations combining circumferences+SKF are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of Jackson/Pollock equation, equations combining WC+SKF, equations combining SKF+BMI, and equations based on WC (derived from a single study); (iv) inconclusive evidence was found for Durnin/Womersley equation.

Weight status groups

The levels of evidence in the **group with normal weight** were: (i) strong evidence indicated that Durnin/Womersley equation, Jackson, Pollock and Ward equation, BMI equations, BIA equations, equations based on WC, equations combining WC+SKF, and equations combining circumferences+SKF are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining WC+SKF+BMI, equations implying sum of SKF, and equations combining SKF+BMI are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining AbC+NC, equations combining AbC+SKF, equations combining various circumferences, BAI equations, and WHR equations (derived from a single study); (iv) inconclusive evidence was found for WHtR equations.

Group with overweight: (i) strong evidence indicated that Durnin/Womersley equation, BMI equations, BIA equations, equations combining circumferences+SKF, equations based on WC, equations combining WC+SKF, equations combining various circumferences, BAI equations, and equations implying sum of SKF are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that Jackson, Pollock and Ward equation, equations combining AbC+NC, equations combining AbC+SKF, and equations combining SKF+BMI are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of Jackson/Pollock equation, equations combining WC+SKF+BMI, and WHR equations (derived from a single study); (iv) inconclusive evidence was found for WHtR equations. **Group with obesity:** (i) strong

evidence indicated that Durnin/Womersley equation is not valid to estimate total body fat mass or %BF; BMI equations and BIA equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining circumferences+SKF are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining SKF+BMI, WHR equations, equations combining WC+SKF+BMI, and equations combining various circumferences (derived from a single study); (iv) **inconclusive evidence** was found for BAI equations, equations based on WC, equations implying sum of SKF, and WHtR equations.

Race-ethnicity groups

The levels of evidence in **Caucasian/White adults** were: (i) strong evidence indicated that Durnin/Womersley equation, BMI equations, equations combining circumferences+SKF, equations implying sum of SKF, equations based on WC, equations combining various circumferences, equations combining WC+SKF, equations combining SKF+BMI, BIA equations, and BAI equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that Jackson/Pollock equation, Jackson, Pollock and Ward equation, equations combining AbC+SKF, and equations combining WC+SKF+BMI are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of equations combining AbC+NC, and equations combining circumferences+diameters+SKF (derived from a single study); (iv) inconclusive evidence was found for WHR equations and WHtR equations. **In African-American/Black adults:** (i) strong evidence indicated that Durnin/Womersley equation, BMI equations, and equations based on WC are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining various circumferences, equations implying sum of SKF, and BAI equations are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of Jackson, Pollock and Ward equation, equations combining AbC+NC, equations combining WC+SKF, equations combining circumferences+SKF, equations combining circumferences+diameters+SKF, equations combining SKF+BMI, and WHtR equations (derived from a single study). **In Hispanic adults:** (i) strong evidence indicated that Durnin/Womersley equation, Jackson, Pollock and Ward, equations based on WC, BIA equations, BAI equations, BMI equations, and WHtR equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining WC+SKF, equations combining various circumferences, and equations implying sum of SKF; (iii) there was limited evidence about the validity of Jackson/Pollock equation, equations combining AbC+NC,

equations combining AbC+SKF, equations combining circumferences+SKF, equations combining SKF+BMI, and WHR equations (derived from a single study). **In Asian adults:** (i) strong evidence indicated that Durnin/Womersley equation, Jackson, Pollock and Ward equation, BIA equations, and BMI equations are valid to estimate total body fat mass or %BF; (ii) moderate evidence showed that equations combining WC+SKF, and equations implying sum of SKF are valid to estimate total body fat mass or %BF; (iii) there was limited evidence about the validity of Jackson/Pollock equation, equations based on WC, equations combining WC+SKF+BMI, equations combining circumferences+SKF, equations combining SKF+BMI, and BAI equations (derived from a single study). Due to heterogeneity found in other races-ethnicities no further levels of evidence could be established.

Due to heterogeneity found in other races-ethnicities no further levels of evidence could be established.

A broader overview of the most common methods for body composition assessment, and the main advantages and limitations of the “gold-standard” methods, and valid field-based methods and equations used for the assessment of body composition in adults have been depicted in **Figure 9** and **Table 30**.

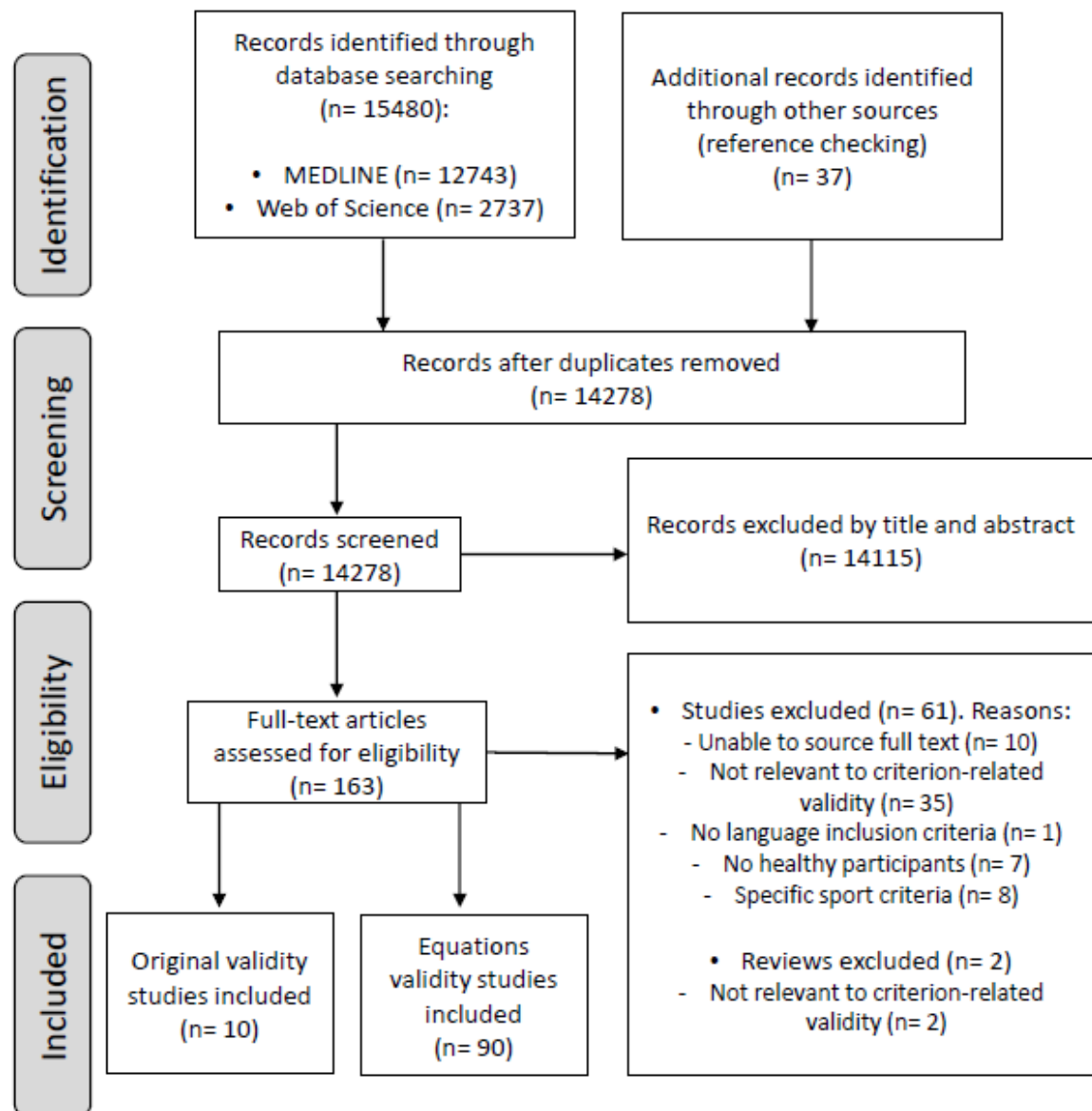


Figure 8. Flow-chart of retrieved and selected articles.

Table 22. Quality assessment of validity of field-based body composition estimation in adults.

Study	Measurement of body composition	Gold Standard	Number of study subjects	Description of the study population	Statistical analysis	Total score
Sun et al. 2013	BAI, BMI	DXA	2	2	0	4
Day et al. 2018	BIA, WC, WHtR	DXA	1	2	2	5
Kim et al. 2015	BMI, WC	DXA	2	2	1	5
Morabia et al. 1999	BIA, SKF	DXA	2	2	1	5
Shaw et al. 2007	BMI, HC, WC, WHR	DXA	2	1	2	5
Sun et al. 2010	BMI, SKF	DXA	2	2	1	5
Cerqueira et al. 2013	BAI	DXA	2	2	2	6
Chang et al. 2014	BAI, BMI	DXA	2	2	2	6
McLean et al. 1992	BMI, NIR, SKF	UWW	2	2	2	6
Zhang et al. 2014	BAI, BMI, HC, WC	DXA	2	2	2	6

BAI, Body Adiposity Index; BIA, Bioelectrical Impedance Analysis; BMI, Body Mass Index; DXA, Dual-energy X-ray Absorptiometry; HC, Hip Circumference; NIR, Near-Infrared Interactance; SKF, Skinfolts; UWW, Under Water Weighing; WC, Waist Circumference; WHR, Waist to Hip Ratio; WHtR, Waist to Height Ratio.

Table 23. Quality assessment of validity of estimation equations for body composition in adults.

Study	Measurement of body composition, equations	Gold Standard	Number of study subjects	Description of the study population	Statistical analysis	Total score
Hodgdon & Beckett 1984	AbC, AC, CC, ChC, FC, NC, ShC, SKF, TC, WrC	UWW	2	0	1	3
Sloan, Burt & Blyth 1962	AC, ChC, GLC, TC, WC, SKF	UWW	1	1	1	3
Durnin & Rahaman 1967	AC, Bi-acromial D, Bi-iliac D, CC, SKF, TC	UWW	2	1	1	4
Erselcan et al. 2000	BIA, SKF	DXA	1	1	2	4
Eston et al. 2005	BIA, SKF	DXA	2	2	0	4
Friedl et al. 1997	AbC, NC	DXA	2	1	1	4
Haisman 1970	SKF	UWW	2	1	1	4
Katch & McArdle 1973	AbC, AC, AkC, AkD, Bi- acromial D, Bi-iliac D, Bitrochanteric D, CC, ChC, ChD, ED, FC, GLC, HdC, KC, KD, ShC, SKF, TC, WrC, WrD	UWW	2	1	1	4
Katch & McArdle 1975	AC, ChC, HC, NC, SKF, WC	UWW	2	1	1	4
Temple et al. 2014	SKF	ADP	1	2	1	4
Wilmore & Behnke 1969	AbC, AC, AD, AkC, AkD, Biacromial D, Bideloid D, Bi-iliac D, Bi-trochanter D,	UWW	2	1	1	4

Wilmore & Behnke 1970	ChC, ChD, Deltoid C, KD, ED, FC, Foot length, HC, HdC, Hand length, KC, NC, ShC, SKF, TC, WrD AbC, AC, AD, AkC, AkD, Biacromial D, Bideltoïd D, Bi-iliac D, Bi-trochanter D, ChC, ChD, Deltoid C, KD, ED, FC, Foot length, HC, HdC, Hand length, KC, NC, ShC, SKF, TC, WrD	UWW	2	1	1	4
Aandstad et al. 2014	BIA, SKF	DXA	2	1	2	5
Aristizabal et al. 2008	HC, SKF	UWW	2	1	2	5
Aristizabal et al. 2018	AbC, AC, CC, HC, TC, SKF, WC	UWW	2	2	1	5
Balas-Nakash et al. 2010	BIA	DXA	1	2	2	5
Ball et al. 2004	HC, SKF, WC	DXA	2	1	2	5
Brozek & Keys 1951	AbC, ChC, SKF	UWW	2	2	1	5
Cui et al. 2014	BMI, SKF, WC, WHtR	DXA	2	2	1	5
Davidson et al. 2011	SKF	DXA	2	2	1	5
Demura et al. 2002	BIA, SKF	UWW	1	2	2	5
Dioum et al. 2005	SKF	ADP	2	1	2	5
Durnin & Womersley 1974	AC, CC, SKF, TC	UWW	2	2	1	5

Friedl et al. 2001	AC, AbC, CC, ChC, FC, NC, SKF, TC, WC	DXA	2	2	1	5
Gallagher et al. 1996	BMI, HC, Tibia/total body length, WC, WHR	DXA, D ₂ O, ³ H ₂ O, UWW	2	2	1	5
Gallagher et al. 2000	BMI	DXA, D ₂ O, ³ H ₂ O, UWW	2	2	1	5
Goel et al. 2008	CC, HC, SKF, TC, WC	DXA	2	2	1	5
Gómez-Ambrosi 2011	BAI, BMI, CUN-BAE, Rohrer index, WHR, WHtR	ADP	2	2	1	5
Gómez et al. 2018	WC, WHtR	DXA	2	1	2	5
Jackson & Pollock 1978	FC, SKF, WC	UWW	2	2	1	5
Jackson, Pollock & Ward 1980	GIC, SKF	UWW	2	2	1	5
Johnson et al. 2012	BAI, BMI	DXA	2	2	1	5
Ketel et al. 2007	BMI, HC, SKF, WC, WHR	DXA	2	2	1	5
Lahav et al. 2018	AbC, NC, SKF, WC	DXA	2	1	2	5
Lanham et al. 2001	BIA, SKF	D ₂ O	1	2	2	5
Lee et al. 2017	AC, BMI, CC, SKF, TC, WC	DXA	2	2	1	5
Lukaski et al. 1986	BIA, SKF	UWW	2	1	2	5
Noopa et al. 1979	AC, BMI, GIC, WC, SKF	K ⁴⁰ , Isotope dilution	2	2	1	5
Pasco et al. 2012	BMI, WC	DXA	2	2	1	5
Pollock et al. 1975	AbC, AC, AkC, biacromial D, bi-iliac D, bitrochanter D, CC, ChC, ChD, FC, GIC,	UWW	2	2	1	5

Pollock et al. 1976	KD, ShC, ShD, SKF, TC, WC, WrC, WrD AbC, AC, AkC, biacromion D, bideltoid D, bi-iliu D, bitrochanter D, CC, ChC, chest width, FC, GlC, KD, ShC, SKF, TC, WC, WrC, WrD	UWW	2	2	1	5
Pongchaiyakul et al. 2005	BMI, HC, SKF, WC	DXA	2	1	2	5
Simoes et al. 2016	AC, HC, SKF, TC, WC	DXA	2	1	2	5
Smith & Boyce 1977	AC, AbC, AkC, Biacromial D, Bi-iliac D, Bi-trochanteric D, CC, ChC, ChD, FC, HC, KD, ShC, SKF, TC, WrD	UWW	2	2	1	5
Steinkamp et al. 1965	AC, AkC, ChC, Biacromial D, Bi-iliac D, Chest D, SKF, TC, WC, WrC	K ⁴⁰ , Isotope dilution	2	2	1	5
Thomas et al. 1998	BIA, SKF	MRI	2	1	2	5
Tucker et al. 2001	AC, CC, ChC, FA, HC, NC, TC, WC	UWW	2	2	1	5
Vogel et al. 1988	AbC, AC, AkC, AkD, Biacromial D, Bi-iliac D, Bi- trochanteric D, CC, ChC, ChD, DD, ED, HC, HdC,	UWW	2	2	1	5

	KD, NC, ShC, TC, WtC,					
	WrD					
Wang et al. 1994	AC, ChC, IC, SKF, TC, WC,	DXA	2	2	1	5
	BMI					
Wang et al. 1996	BIA, BMI, SKF	UWW	2	1	2	5
Womersley & Durnin 1977	SKF	UWW	2	2	1	5
Al-Bachir et al. 2016	HC, SKF, WC	D ₂ O	2	2	2	6
Al-Gindan et al. 2015	HC, WC	MRI	2	2	2	6
Benito et al. 2019	BIA, SKF	DXA	2	2	2	6
Bergman et al. 2011	BAI, HC, WC	DXA	2	2	2	6
Bhat et al. 2005	BIA, BMI, SKF	D ₂ O	2	2	2	6
Deurenberg et al. 1991	BMI	UWW	2	2	2	6
Eston et al. 1995	AC, CC, HC, SKF, TC	UWW	2	2	2	6
Fedewa et al. 2019	BIA, WC	DXA	2	2	2	6
Garcia et al. 2005	Bone breadths, ChC, chest depth, HC, SKF, TC, WC	DXA	2	2	2	6
Hassager et al. 1986	SKF	ADP	2	2	2	6
Hicks et al. 2000	BIA, SKF	UWW	2	2	2	6
Jackson et al. 2002	BMI	UWW	2	2	2	6
Jackson et al. 2009	SKF	DXA	2	2	2	6
Kagawa, Byrne & Hills 2008	AbC, BMI, HC, WC, WHtR	DXA	2	2	2	6
Kanellakis et al. 2010	SKF, WC, WHR	DXA	2	2	2	6
Kanellakis et al. 2012	HC, SKF, WC	DXA	2	2	2	6

Kanellakis et al. 2017	BMI, CC, HC, FC, NC, SKF, WC	DXA	2	2	2	6
Kholi et al. 200	HC, SKF, WC	DXA	2	2	2	6
Lam et al. 2013	BAI, BMI, HC, WC	DXA	2	2	2	6
Leahy et al. 2013	AbC, AC, CC, ChC, FC, HC, SKF, TC, WC	DXA	2	2	2	6
Lean, Han & Deurenberg 1996	AC, Arm span, BMI, HC, leg length, SKF, TC, WC, WHR	UWW	2	2	2	6
Lee et al. 2021	BMI, WC	DXA	2	2	2	6
Macias et al. 2007	BIA, HC, WC, WHR	ADP	2	2	2	6
Manios et al. 2012	BMI, HC, SKF, WC	DXA	2	2	2	6
Nickerson et al. 2018	BMI, SKF	UWW, DXA	2	2	2	6
O'Connor et al. 2010	SKF	DXA	2	2	2	6
Pascale et al. 1956	SKF	UWW	2	2	2	6
Ramirez-Zea et al. 2006	AbC, AC, CC, HC, SKF, TC, WC	UWW	2	2	2	6
Ramos-Jiménez et al. 2018	SKF, WC	DXA	2	2	2	6
Rush et al. 1997	AbC, BMI, HC, SKF, WC, WHR	Isotope dilution	2	2	2	6
Segheto et al. 2017	BAI, BIA, BMI, SKF, WC, WHR	DXA	2	2	2	6
Shafer et al. 2010	SKF	ADP	2	2	2	6

Skoufas et al. 2018	AbC, AC, CC, ChC, FC, HC, NC, SKF, TC, WC	DXA	2	2	2	6
Slaughter et al. 1988	SKF	D ₂ O, photon absorptiometry, UWW	2	2	2	6
Stout et al. 1994	BIA, SKF	UWW	2	1	2	6
Sun et al. 2005	BIA, BMI, WHtR	DXA	2	2	2	6
Wattanapenpaiboon et al. 1998	BIA, SKF	DXA	2	2	2	6
Yao et al. 2002	BMI, SKF	D ₂ O	2	2	2	6
Zanovec et al. 2009	BMI	DXA	2	2	2	6

AbC, Abdominal Circumference; AC, Arm Circumference; AD, Arm Diameter; ADP, Air-Displacement Plethysmography; AkC, Ankle Circumference; AkD, Ankle Diameter; BAI, Body Adiposity Index; BIA, Bioelectrical Impedance Analysis; BMI, Body Mass Index; C, Circumference; CC, Calf Circumference; ChC, Chest Circumference; ChD, Chest Diameter; D, Diameter; DD, Deltoid Diameter; DXA, Dual energy X-ray Absorptiometry; D₂O, Deuterium Oxide; ED, Elbow Diameter; FC, Forearm Circumference; GlC, Gluteal Circumference; HC, Hip Circumference; HdC, Head Circumference; ³H₂O, Tritium Water; IC, Iliac Circumference; KC, Knee Circumference; KD, Knee Diameter; K⁴⁰, Potassium-40; MRI, Magnetic Resonance Imaging; NC, Neck Circumference; ShC, Shoulder Circumference; SKF, Skinfolts; TC, Thigh Circumference; UWW, Under Water Weighing; WC, Waist Circumference; WHR, Waist to Hip Ratio; WHtR, Waist to Height Ratio; WrC, Wrist Circumference; WrD, Wrist Diameter.

Table 24. Overview of field-based body composition estimation studies in adults.

Author	Participants	Age, range/ SD	Field test	Gold standard	Statistical methods	Results	Conclusions
Low quality studies							
Sun et al. 2013	Females= 1939	(≥ 20)	- BAI	DXA	Pearson correlation coefficient (r)	Field test correlations vs DXA:	The BAI method showed moderate validity to assess %BF, compared with DXA. The BMI showed very low validity to assess %BF, in groups with normal weight and overweight, while in the group with obesity showed low validity, compared with DXA.
	Males= 662	39.5 ± 14.4	- BMI			Group with normal weight:	
	Healthy	43.3 ± 12.4				- BAI: r= 0.65 (p< 0.001)	
	BMI= 26.5 (entire sample) normal weight, overweight and obesity		Group with overweight:			- BMI: r= 0.05 (p≥ 0.05)	
	Caucasian		- BAI: r= 0.62 (p< 0.001)			- BMI: r= -0.09 (p< 0.001)	
				Group with obesity:	- BAI: r= 0.68 (p< 0.001)		
					- BMI: r= 0.34 (p< 0.001)		
High quality studies							
Day et al. 2018	Females= 16	(18-65)	- WC, WHtR	DXA	Spearman's Rank correlations, Bland-Altman method	Field test correlations vs DXA:	In the group with normal weight, WC, WHtR and BIA showed moderate validity to assess %BF, compared with DXA. In the group with overweight/obesity, WC and BIA showed very high validity and WHtR high validity to assess %BF, compared with DXA.
	Males= 14	29.9 ± 11.2	- BIA			Group with normal weight:	
	Healthy					- WC: r _s = 0.57 (p< 0.001)	
	BMI= 29.2 (entire sample) normal weight, overweight/ obesity		- WHtR: r _s = 0.59 (p< 0.001)			- BIA: r _s = 0.59 (p< 0.001)	
	Caucasian					Group with overweight/obesity:	

						- WHtR: $r_s = 0.75$ ($p < 0.001$) - BIA: $r_s = 0.99$ ($p < 0.001$) Agreement: - BIA vs DXA, MD ($\pm$ 95% LoA)= 0.32 (-3.8 to 4.4)	
Kim et al. 2015	Females= 10.320 Males= 7.878 Healthy BMI= 23.6 Korean	(19-59) 	- WC - BMI	DXA	Pearson correlation coefficient (r), linear regression	Field test correlations vs DXA: - WC: $r = 0.23$ ($p < 0.05$) - BMI: $r = 0.39$ ($p < 0.05$) Regression analysis: - WC: $R^2 = 0.15$ ($p < 0.05$) - BMI: $R^2 = 0.05$ ($p < 0.05$)	WC and BMI showed very low and low validity to assess %BF, compared with DXA.
Morabia et al. 1999	Females= 153 Males= 73 Healthy BMI= 22.0 Caucasian	(15-86) 	- BMI	DXA	Pearson correlation coefficient (r), multiple regression	Field test correlations vs DXA: - Females: $r = 0.81$ ($p < 0.05$) - Males: $r = 0.84$ ($p < 0.05$) Regression analysis: - Females: $R^2 = 0.41$ ($p < 0.05$) - Males: $R^2 = 0.44$ ($p < 0.05$)	BMI showed high validity to assess %BF, compared with DXA.
Shaw et al. 2007	Females= 307 Males= 268 Healthy 2 age-groups: 50-59yrs and 60-69yrs BMI= 28.0	(50-69)	- HC, WC, WHR - BMI	DXA	Pearson correlation coefficient (r), Bland-Altman method	Field test correlations vs DXA: Females: - HC: $r = 0.81$ ($p < 0.0001$), in 50-59yrs group; $r = 0.77$ ($p < 0.0001$), in 60-69yrs group	HC, WC and BMI showed high to very high validity to assess %BF, compared with DXA. WHR showed low to moderate validity to assess %BF, compared with DXA.

Tasmanian

- WC: $r = 0.89$ ($p < 0.0001$), in 50-59yrs group; $r = 0.87$ ($p < 0.0001$), in 60-69yrs group

- WHR: $r = 0.51$ ($p < 0.0001$), in 50-59yrs group; $r = 0.46$ ($p < 0.0001$), in 60-69yrs group

- BMI: $r = 0.95$ ($p < 0.0001$), in 50-59yrs group; $r = 0.93$ ($p < 0.0001$), in 60-69yrs group

Males:

- HC: $r = 0.84$ ($p < 0.0001$), in 50-59yrs group; $r = 0.85$ ($p < 0.0001$), in 60-69yrs group

- WC: $r = 0.90$ ($p < 0.0001$), in 50-59yrs group; $r = 0.90$ ($p < 0.0001$), in 60-69yrs group

- WHR: $r = 0.63$ ($p < 0.0001$), in 50-59yrs group; $r = 0.63$ ($p < 0.0001$), in 60-69yrs group

- BMI: $r = 0.89$ ($p < 0.0001$), in 50-59yrs group; $r = 0.90$ ($p < 0.0001$), in 60-69yrs group

Agreement of body composition measures vs DXA:

**Mean difference was < 0.0001 for all comparisons; all comparisons demonstrated no tendency for agreement between measures to vary with the magnitude of measurement.*

Sun et al. 2010	Females= 2.126	(≥ 20)	- WC	DXA	Pearson correlation coefficient (r)	Field test correlations vs DXA:	WC and BMI showed high validity to assess %BF (and trunk fat mass), compared with DXA, independently of sex, age or race-ethnicity.
	Males= 2.260	42.3 ± 0.3	- BMI			Females (entire sample):	
	Healthy	41.0 ± 0.3				- WC, trunk fat mass (%): $r = 0.83$ ($p < 0.05$), in 20-39yrs group; $r = 0.80$ ($p < 0.05$), in 40-59yrs group; $r = 0.76$ ($p < 0.05$), in ≥ 60 yrs group	
	3 age-groups: 20-39yrs, 40-59yrs, ≥ 60 yrs					- BMI, %BF: $r = 0.84$ ($p < 0.05$), in 20-39yrs group; $r = 0.79$ ($p < 0.05$), in 40-59yrs group; $r = 0.78$ ($p < 0.05$), in ≥ 60 yrs group	
	<i>BMI not described</i>					Males (entire sample):	
	Non-Hispanic white (NHW), non-Hispanic Black (NHB), Hispanic (H)					- WC, trunk fat mass (%): $r = 0.86$ ($p < 0.05$), in 20-39yrs group; $r = 0.72$ ($p < 0.05$), in 40-59yrs group; $r = 0.77$ ($p < 0.05$), in ≥ 60 yrs group	
						- BMI, %BF: $r = 0.79$ ($p < 0.05$), in 20-39yrs group; $r = 0.75$ ($p < 0.05$), in 40-59yrs group; $r = 0.74$ ($p < 0.05$), in ≥ 60 yrs group	
						Ethnicity (not sex/age distinction):	
						NHW:	
						- WC, trunk fat mass (%): $r = 0.83$ ($p < 0.05$)	
						- BMI, %BF: $r = 0.78$ ($p < 0.05$)	
						NHB:	
						- WC, trunk fat mass (%): $r = 0.84$ ($p < 0.05$)	
						- BMI, %BF: $r = 0.78$ ($p < 0.05$)	

						H: - WC, trunk fat mass (%): $r = 0.78$ ($p < 0.05$) - BMI, %BF: $r = 0.74$ ($p < 0.05$)	
Cerqueira et al. 2013	Females= 102 Healthy BMI= 26.9 Brazilian	(35-83) 60.3 ± 9.8	- BAI	DXA	Pearson correlation coefficient (r), LCCC (Pc), Bland-Altman method	Field test correlations vs DXA: - BAI: $r = 0.65$ ($p < 0.001$) Concordance: - Pc= 0.73 Agreement: MD ($\pm$ 95% LoA)= 3.3 (- 5.9 to 12.5)	BAI showed moderate validity to assess %BF, compared with DXA.
Chang et al. 2014	Females= 471 Males= 483 Healthy BMI= 27.2 White Black	(≥ 55) 70.4 ± 9.5	- BAI - BMI	DXA	Pearson correlation coefficient (r), LCCC (Pc), Linear regression, Bland-Altman method	Field test correlations vs DXA: Females: - BAI: $r = 0.72$ ($p < 0.01$) - BMI: $r = 0.80$ ($p < 0.01$) Males: - BAI: $r = 0.55$ ($p < 0.01$) - BMI: $r = 0.71$ ($p < 0.01$) White (not sex distinction): - BAI: $r = 0.70$ ($p < 0.01$) - BMI: $r = 0.54$ ($p < 0.01$) Black (not sex distinction): - BAI: $r = 0.81$ ($p < 0.01$) - BMI: $r = 0.62$ ($p < 0.01$) Regression analysis:	BAI and BMI showed high validity to assess %BF, compared with DXA, in Females and White and Black populations. In males, BAI showed moderate validity. However, BAI was not accurate in people with extremely low or high %BF.

						- BAI: $R^2 = 0.56$ ($p < 0.01$) - BMI: $R^2 = 0.55$ ($p < 0.01$) Concordance: - BAI: $P_c = 0.55$ ($p < 0.01$) - BMI: $P_c = 0.30$ ($p < 0.01$) Agreement: - BAI vs DXA, MD ($\pm$ 95% LoA)= -5.17 (-16.94 to 6.59) - BMI vs DXA, MD ($\pm$ 95% LoA)= -7.6 (-21.78 to 6.58)	
McClean et al. 1992	Females= 31 Males= 30 Healthy BMI= 24.3 Caucasian	(18-45) 32 ± 8	- SKF: abdominal, subscapular, suprailiac, thigh, triceps - NIR - % BF (from BD by Brozek)	UWW	Pearson correlation coefficient (r), SEE, ANOVA	Field test correlations vs UWW: Females: - SKF: $r = 0.88$, SEE= 2.9 ($p < 0.001$) - NIR: $r = 0.63$, SEE= 4.7 ($p < 0.001$) Males: - SKF: $r = 0.95$, SEE= 2.6 ($p < 0.001$) - NIR: $r = 0.80$, SEE= 4.9 ($p < 0.001$) ANOVA: Females: - SKF: $F = 33.4$ ($p < 0.001$) - NIR: $F = 7.6$ ($p = 0.025$) Males: - SKF: $F = 69.8$ ($p < 0.001$) - NIR: $F = 53.4$ ($p < 0.001$)	SKF showed very high validity in the total group to assess %BF, compared with UWW.

Zhang et al. 2013	Females= 1707	(51-77)	- HC, WC	DXA	Pearson correlation coefficient (r), Bland-Altman method	Field test correlations vs DXA:	BMI showed high validity in Females, and WC in males to assess %BF, compared with DXA.
	Males= 680	59.6 ± 4.7	- BAI			Females:	
	BMI= 23.6	62.2 ± 5.2	- BMI			- HC: r= 0.66 (p< 0.0001)	
	Chinese					- WC: r= 0.63 (p< 0.0001)	
						- BAI: r= 0.62 (p< 0.0001)	
						- BMI: r= 0.70 (p< 0.0001)	
						Males:	
						- HC: r= 0.66 (p< 0.0001)	
						- WC: r= 0.72 (p< 0.0001)	
						- BAI: r= 0.58 (p< 0.0001)	
						- BMI: r= 0.69 (p< 0.0001)	
						Agreement:	
						- Females: BAI vs DXA, MD (± 95% LoA)= 5.8 (-1.5 to 13.0)	
						- Males: BAI vs DXA, MD (± 95% LoA)= 0.3 (-7.0 to 6.0)	

BAI, Body Adiposity Index; BD, Body Density; BIA, Bioelectrical Impedance Analysis; BMI, Body Mass Index; DXA, Dual-energy X-ray Absorptiometry; HC, Hip Circumference; NIR, Near-infrared Interactance; SKF, Skinfolks; UWW, Under Water Weighing; WC, Waist Circumference; WHR, Waist-Hip Ratio; WHtR, Waist Height Ratio; %BF, body fat percentage.

ANOVA, indicates analysis of variance; CI, Coefficient of Interval; LCCC, Lin's Concordance Coefficient; LoA, Limits of Agreement; RMSE, Root Mean Square Error; SD, Standard Deviation; SE, Standard Error; SEE, Standard Error Estimate.

Table 25. Levels of evidence of field-based body composition estimation studies in adults.

FIELD TEST	WHOLE GROUP	19-33 YRS	34-49 YRS	50-64 YRS	GROUP WITH NORMAL WEIGHT	GROUP WITH OVERWEIGHT	GROUP WITH OBESITY	Caucasian/ White
WC	●	◐	□	●	□	◐	■	◐
BAI	●	N/A	N/A	●	■	◐	N/A	■
BMI	●	■	□	●	□	◐	N/A	●
HC	◐	N/A	N/A	◐	■	■	N/A	N/A
SKF	■	N/A	■	N/A	■	N/A	N/A	■
BIA	■	■	N/A	N/A	■	■	N/A	■
WHR	■	N/A	N/A	■	N/A	■	N/A	N/A
WHtR	■	■	N/A	N/A	■	■	■	■

● Indicates strong evidence of validity; ◐ moderate evidence of validity; ■ limited evidence of validity*; □ inconclusive evidence of validity**.

*Limited evidence of validity derived from one single study; **Inconclusive evidence of validity derived from inconsistent results.

HC, hip circumference; WC, waist circumference; SKF, skinfolds; BAI, body adiposity index; BIA, bioelectrical impedance analysis; BMI, body mass index; WHR, waist to hip ratio; WHtR, waist to height ratio; N/A, not applicable.

Table 26. Durnin and Womersley (1974) generalized equations and, Siri and Brozek converted formulas for body fat calculation in adults.

Durnin and Womersley 1974		
Age (years)	Females	Males
< 17	BD= 1.1369 – (0.0598 X Log)	BD= 1.1533 – (0.0643 X Log)
17-19	BD= 1.1549 – (0.0678 X Log)	BD= 1.1620 – (0.0630 X Log)
20-29	BD= 1.1599 – (0.0717 X Log)	BD= 1.1631 – (0.0632 X Log)
30-39	BD= 1.1423 – (0.0632 X Log)	BD= 1.1422 – (0.0544 X Log)
40 -49	BD= 1.1333 – (0.0612 X Log)	BD= 1.1620 – (0.0700 X Log)
> 50	BD= 1.1339 – (0.0645 X Log)	BD= 1.1715 – (0.0779 X Log)

BD, Body Density; Log, Log of the Sum of 4 skinfolds.

Skinfolds: Bicep, Subscapular, Suprailiac, Tricep.

*Siri formula: Body Fat (%)= (4.95 / Body Density) – 4.50

**Brozek formula: Body Fat (%)= (4.57 / Body Density) – 4.142

no significant differences arose from the use of both body density formulas

Table 27. Jackson, Pollock and Ward (1980), and Jackson and Pollock (1978) generalized equations and, Siri and Brozek converted formulas for body fat calculation in adults.

Age (years)	Jackson, Pollock and Ward 1980 (females)
18-55	BD= 1.0994921 – (0.0009929 x sum of 3 skinfolds) + (0.0000023 x square of the sum of 3 skinfolds) – (0.0001392 x age)
18-55	BD= (0.29669 x sum of 4 skinfolds) – (0.00043 x square of the sum of 4 skinfolds) + (0.02963 x age) + 1.4072
18-55	BD= 1.097 – (0.00046971 x sum of 7 skinfolds) + (0.00000056 x square of the sum of 7 skinfolds) – (0.00012828 x age)
Age (years)	Jackson and Pollock 1978 (males)
18-61	BD= 1.10938 – (0.0008267 x sum of skinfolds) + (0.0000016 x square of the sum of 3 skinfolds) – (0.0002574 x age)
18-61	BD= (0.29288 x sum of 4 skinfolds) – (0.0005 x square of the sum of 4 skinfolds) + (0.15845 x age) – 5.76377
18-61	BD= 1.112 – (0.00043499 x sum of 7 skinfolds) + (0.00000055 x square of the sum of 7 skinfolds) – (0.00028826 x age)

BD, Body Density.

Skinfolds Females: Suprailiac, Thigh, Tricep (Sum 3); Abdominal, Suprailiac, Thigh, Tricep (Sum 4);

Abdominal, Axila, Chest, Subscapular, Suprailiac, Thigh, Tricep (Sum 7).

Skinfolds males: Abdominal, Pectoral, Thigh (Sum 3); Abdominal, Suprailiac, Thigh, Tricep (Sum 4);

Abdominal, Axila, Chest, Subscapular, Suprailiac, Thigh, Tricep (Sum 7).

*Siri formula: Body Fat (%)= (4.95 / Body Density) – 4.50

**Brozek formula: Body Fat (%)= (4.57 / Body Density) – 4.142

no significant differences arose from the use of both body density formulas

Table 28. Overview of estimation equations for body composition in adults.

Author	Participants	Age, range/ SD	Field test - Equations	Gold standard	Statistical methods	Results	Conclusion
<i>Low quality studies</i>							
Erselcan et al. 2000	Females= 37 Healthy, BMI= 22.2 normal weight, 34.5 obese <i>Race-ethnicity not described</i>	33.9 ± 12.2	- SKF: biceps, subscapular, suprailiac, triceps - BIA - %BF equations (from BD by Siri): 1. SKF by Durnin/ Womersley (DW Sum 4SKF)	DXA	Linear regression, SEE, Bland-Altman method	Regression analysis of field test vs DXA: Group with normal weight: - DW Sum 4SKF: R ² = 0.89, SEE=1.9 (p< 0.0001) - BIA: R ² = 0.93, SEE=1.5 (p< 0.0001) Group with obesity: - DW Sum 4SKF: R ² = 0.75, SEE=5.1 (p< 0.0001) - BIA: R ² = 0.84, SEE=4.0 (p< 0.0001) Agreement: Group with normal weight: - DW Sum 4SKF, MD (± 95% LoA)= 0.8 (1.3 to 3.6) - BIA, MD (± 95% LoA)= - 1.5 (1.6 to 3.1) Group with obesity: - DW Sum 4SKF, MD (± 95% LoA)= 2.5 (2.5 to 10.3)	DW Sum 4SKF and BIA showed very high validity to assess %BF, compared with DXA, in the group with normal weight. DW Sum 4SKF and BIA showed high and very high validity to assess %BF, compared with DXA, in the group with obesity. However, DXA should be chosen in the group with obesity.

Hodgdon & Beckett 1984	Females= 214	(18-44)	- AbC, AC, CC, ChC, FC, NC, ShC, TC, WrC	UWW	Pearson correlation coefficient (r), SEE	- BIA, MD ($\pm$ 95% LoA)= 2.5 (1.7 to 9.5)	Navy equation (based on AbC + NC + HC) showed high validity to assess %BF. In this specific population, the new equation should be chosen over the previous ones compared.
	Healthy status not described %BF= 27 Race-ethnicity not described		- SKF: abdominal, biceps, chest, midaxillary, subscapular, suprailiac, thigh, triceps - %BF equations (from BD by Siri): - %BF equations: 1. Wright et al. 2. Navy equation (present study) Cross-validation: 3. Durnin/ Womersley (DW Sum 4SKF) 4. Berrers et al. 5. Katch & McArdle 5. Brennen			Correlation of equations vs %BF (from BD): 1. Wright et al.: r= 0.80, SEE= 4.19 %BF (p< 0.05) 2. Navy equation: r= 0.85, SEE= 3.72 %BF (p< 0.05) Cross-validation (n= 146), Navy equation: r= 0.80-0.87, SEE= 4.36-4.04 %BF (p< 0.05) Cross-correlations using other equations: 1. DW Sum 4SKF: r= 0.81, SEE= 4.13 %BF (p< 0.05) 2. Berrers et al.: r= 0.78, SEE= 4.38 %BF (p< 0.05) 3. Wright et al: r= 0.80, SEE= 4.19 %BF (p< 0.05) 4. Katch & McArdle: r= 0.75- 0.81, SEE= 4.89-4.21 %BF (p< 0.05) 5. Brennen: r= 0.76, SEE= 4.56 %BF (p< 0.05)	
Sloan, Burt & Blyth 1962	Females= 50 Healthy %BF= 12-37	(17-25)	- AC, ChC, GlC, TC, WC - SKF: back arm, chest, iliac crest, thigh	UWW	Pearson correlation coefficient (r), SEE	Correlation of body composition measures vs %BF (from BD): - SKF, iliac crest: r= -0.71 (p< 0.05)	SKF, especially iliac crest and back arm, showed high validity to assess %BF, compared with UWW.

	<i>Race-ethnicity not described</i>		- %BF equations (from BD): 1. Brozek 2. Siri 3. Rathbun & Pace			- SKF, back arm: $r = -0.68$ ($p < 0.05$) - SKF, iliac crest+back arm: $r = -0.74$ ($p < 0.05$) Regression equation: $SEE = \pm 0.0082$ g/ml BD to %BF: 1. Brozek: 12.13-31.70 % (MD 20.06, SD ± 4.63) ($p < 0.05$) 2. Siri: 13.18-36.63 % (MD 22.91, SD ± 5.58) ($p < 0.05$) 3. Rathbun & Pace: 11.70-37.40% (MD 22.13, SD ± 7.08) ($p < 0.05$) Regression equations: 1. Brozek: ± 3 %BF ($p < 0.05$) 2. Siri: ± 3.7 %BF ($p < 0.05$) 3. Rathbun & Pace: ± 3.5 %BF ($p < 0.05$)	
Durnin & Rahaman 1967	Females= 45 Males= 60 Healthy %BF= 6-46 <i>Race-ethnicity not described</i>	(18-29) (18-34)	- AC, CC, TC - Bi-acromial D, Bi-iliac D - SKF: biceps, subscapular, suprailiac, triceps - %BF equations (from BD):	UWW	Pearson correlation coefficient (r)	Correlation of SKF vs %BF (from BD): - Females: $r = -0.78$ ($p < 0.001$) - Males: $r = -0.83$ ($p < 0.001$) Regression equation (values shown only by Siri): 2. Siri: ± 3.0 -3.5%BF ($p < 0.05$)	SKF showed high validity to assess %BF, compared with UWW.

1. Brozek							
2. Siri							
Eston et al. 2005	Females= 31	(18-36)	- SKF: abdominal,	DXA	Forward stepwise	Regression analysis of SKF vs	Lower limb SKF showed
	Males= 21	20.9 ± 2.0	biceps, calf, iliac crest,		regression, SEE	DXA:	high validity to assess
	Healthy, fit-active	22.3 ± 5.5	subscapular, thigh,			Females:	%BF, compared with
	BMI= 23.6		triceps			- Calf, R ² = 0.04, SEE=2.1 (p< 0.001)	DXA.
	Caucasian		SKF by Durnin/ Womersley (DW Sum 4SKF)			- Iliac crest, R ² = 0.03, SEE=2.1 (p< 0.001)	Conventional DW Sum
						- Thigh, R ² = 0.82, SEE=2.1 (p< 0.001)	4SKF is improved by
						- DW Sum 4SKF + Sum calf+thigh, R ² = 0.75, SEE=2.4 (p< 0.001)	adding thigh and calf SKF,
						Males:	independently or in
						- Abdomen, R ² = 0.03, SEE=1.8 (p< 0.001)	combination.
						- Iliac crest, R ² = 0.09, SEE=1.8 (p< 0.001)	
						- Thigh, R ² = 0.79, SEE=1.8 (p< 0.001)	
						- DW Sum 4SKF + Sum calf+thigh, R ² = 0.80, SEE=1.9 (p< 0.001)	
Friedl et al. 1997	Males= 496	(<40)	- AbC, NC	DXA	Linear regression, SEE	Regression analysis of equations vs %BF:	All equations showed very
	Healthy		- %BF equations:			- Marine Corps equation: R ² = 0.80, SEE= 3.27 (p< 0.01)	high validity to assess
	BMI= 24.8						%BF, being the new
							equation (based on AbC +

	<i>Race-ethnicity not described</i>		1. Marine Corps (Wright, Dotson, and Davis) 2. Navy (Hodgdon/Beckett) 3. Army (Vogel et al.) 4. New regression equation: $\%BF = (0.771 \times AbC - NC) (0.132 \times height) + 4.29$			- Navy equation: $R^2 = 0.82$, SEE= 3.13 ($p < 0.01$) - Army equation: $R^2 = 0.81$, SEE= 3.15 ($p < 0.01$) - New regression equation: $R^2 = 0.82$, SEE= 3.10 ($p < 0.01$)	NC) the most valid and accurate. It is recommended to add AbC in the equations because it has a good relationship with the %BF.
Haisman 1970	Males= 55 Healthy %BF= 12.5 <i>Race-ethnicity not described</i>	22.6 ± 2.2	- SKF: abdominal, biceps, chest, subscapular, suprailiac, triceps - %BF equations (from BD): 1. Siri - %BF equations: 1. Edwards & Whyte 2. Adam et al. 3. Durnin & Rahaman 4. Chinn & Allen 5. Pascale et al. 6. Brozek & Keys 7. Best 8. Equation of the present study	UWW	Pearson correlation coefficient (r), SEE	Correlation of SKF vs %BF (from BD): - Single/combination SKF: $r = 0.63-0.77$, SEE= 0.008-0.006 ($p < 0.05$) SKF equations: 1. Edwards & Whyte: $r = 0.71$ ($p < 0.05$) 2. Adam et al.: $r = 0.65$ ($p < 0.05$) 3. Durnin & Rahaman: $r = 0.76$ ($p < 0.05$) 4. Chinn & Allen: $r = 0.71$ ($p < 0.05$) 5. Pascale et al.: $r = 0.76$ ($p < 0.05$) 6. Brozek & Keys: $r = 0.69$ ($p < 0.05$)	SKF (Sum 4SKF or 9SKF), showed high validity to assess %BF, compared with UWW. Equations by Haisman (present study, Sum 4SKF), Durnin & Rahaman, Pascale et al. and Best showed high validity to assess %BF.

7. Best: $r = 0.75$ ($p < 0.05$) 8. Equation of the present study: $r = -0.81$, $SEE = 0.007$ ($p < 0.05$)							
Katch & McArdle 1973	Females= 69 Males= 53 Healthy %BF= 26 Females; 15 males Caucasian	20.3 ± 1.8 19.3 ± 1.5	- AbC, AC, AkC, CC, ChC, FC, GlC, HdC, KC, ShC, TC, WrC - AkD, Bi-acromial D, Bi-iliac D, Bitrochanteric D, ChD, ED, KD, WrD - SKF: subscapular, suprailiac, thigh, triceps, umbilicus - %BF equations (from BD, by Brozek)	UWW	Pearson correlation coefficient (r), SEE	Correlation of equations vs %BF (from BD): Females: - SKF, $r = 0.77$ ($p < 0.05$) - Circumferences, $r = 0.80$ ($p < 0.05$) - SKF+circumferences, $r = 0.78$ ($p < 0.05$) - SKF (scapula and iliac)+ED+TC, $r = 0.84$, $SEE = \pm 3.6\%BF$ ($p < 0.05$) Males: - SKF, $r = 0.86$ ($p < 0.05$) - Circumferences: $r = 0.86$ ($p < 0.05$) - SKF (triceps and scapula)+AbC+FC, $r = 0.89$; $SEE = \pm 2.9\%BF$ ($p < 0.05$)	All measures showed high validity to assess total %BF, compared with UWW. In Females, equation based on SKF (scapula and iliac) + ED + TC best predicted %BF. In males, equation based on SKF (triceps and scapula) + AbC + FC best predicted %BF.
Katch & McArdle 1975	Males= 53 Healthy %BF= 26 Females; 15 males	(17-24) 19.3 ± 1.5	- AbC, ChC, NC - Bi-iliac D - SKF: midanterior, subscapular, suprailiac, thigh, triceps, umbilicus	UWW	Pearson correlation coefficient (r), SEE, CE	Correlation of equations vs %BF (from BD): 1. Sloan: $r = 0.81$, $SEE = 0.008$ ($p < 0.05$) 2. Wilmore & Behnke: $r = 0.83$, $SEE = 0.008$ ($p < 0.05$)	Both equations (based on AbC + ChC +NC + bi-iliac D + abdomen, thigh and scapula SKF) showed high validity to assess total %BF.

	<i>Race-ethnicity not described</i>		- %BF equations (from BD, by Brozek): 1. Sloan 2. Wilmore & Behnke - %BF equations: 1. Steinkamp et al. 2. Yuhasz			Equation correlations: - Steinkamp et al.: $r = 0.81$, SEE= 3.27, CE= -0.71 ($p < 0.01$) - Yuhasz: $r = 0.84$, SEE= 3.02, CE= 1.12 ($p < 0.01$)	
Wilmore & Behnke 1969	Males= 133 Healthy %BF= 14 <i>Race-ethnicity not described</i>	(17-37) 22.0 ± 3.1	- AbC, AC, AkC, ChC, Deltoid C, FC, HC, HdC, KC, NC, ShC, TC - AD, AkD, Biacromial D, Bideltoid D, Bi-iliac D, Bi-trochanter D, ChD, KD, ED, WrD - Foot length, Hand length - SKF: abdominal, midaxillary, thigh, triceps - %BF equations (from BD): 1. Brozek 2. Siri 3. Rathbun & Pace - %BF equation: 1. Yuhasz	UWW	Pearson correlation coefficient (r), SEE	Correlations of SKF vs %BF (from BD): - Sum SKF: $r = -0.78$ ($p < 0.05$) - Diameters: $r = -0.04$ to -0.42 ($p < 0.05$) - Circumferences: $r = -0.16$ to -0.73 ($p < 0.05$) Equation correlations: - Sum SKF: $r = 0.80-0.81$, SEE= 0.008-0.007 ($p < 0.05$) - Diameters+circumferences: $r = 0.79-0.85$, SEE= 0.007 ($p < 0.05$) - SKF+diameters+circumferences: $r = 0.80-0.87$, SEE= 0.007-0.006 ($p < 0.05$) - Yuhasz equation vs equations in this study: $r = 0.78$ ($p < 0.05$)	All measures showed high validity to assess total %BF, compared with UWW. Equations based on SKF + diameters + circumferences best predicted %BF.

			2. Equation of the present study				
Wilmore & Behnke 1970	Females= 128 Healthy %BF= 26 <i>Race-ethnicity not described</i>	(18-48) 21.4 ± 3.8	- AbC, AC, AkC, ChC, Deltoid C, FC, HC, HdC, KC, NC, ShC, TC - AD, AkD, Biacromial D, Bideloid D, Bi-iliac D, Bi-trochanter D, ChD, KD, ED, WrD - Foot length, Hand length - SKF: abdominal, midaxillary, thigh, triceps %BF equations (from BD): 1. Brozek 2. Siri 3. Rathbun & Pace - %BF equations: 1. Sloan 2. Katch & Michael 3. Equation of the present study	UWW	Pearson correlation coefficient (r), SEE	Correlation of body composition measures vs %BF (from BD): - SKF: $r = -0.17$ to -0.58 ($p < 0.05$) - Diameters: $r = -0.07$ to -0.40 ($p < 0.05$) - Circumferences: $r = -0.02$ to -0.61 ($p < 0.05$) Equation correlations: - Sum SKF: $r = 0.68$, $SEE = 0.007$ ($p < 0.05$) - Diameters+circumferences: $r = 0.74$, $SEE = 0.007$ ($p < 0.05$) - SKF+diameters+circumferences: $r = 0.75$, $SEE = 0.007$ ($p < 0.05$) 1. Sloan equation vs equations in this study: $r = 0.58$ vs $r = 0.74$ ($p < 0.05$) 2. Katch & Michael vs equations in this study: $r = 0.64$ vs $r = 0.70$ ($p < 0.05$)	All measures showed very low to moderate validity to assess total %BF, compared with UWW. Equations based on SKF + diameters + circumferences best predicted %BF.

Temple et al. 2014	Females= 43 Healthy BMI= 26.2 Total sample, normal weight, overweight Caucasian	(21-55) 37.3 ± 9.7	- SKF: abdominal, biceps, chest, midaxillary, subscapular, suprailiac, thigh, triceps - %BF equations (from BD by Siri): 1. Jackson, Pollock and Ward, three, four and seven-SKF equations (JPW3, JPW4 and JPW7, respectively) 2. Durnin/ Womersley (DW Sum 4SKF)	ADP	Linear regression (bias change-point, %), CVRSME	Regression analysis of equations vs %BF (Bias change-point, %): - DW Sum 4SKF: 32%, *bias=0, for total sample; 35% for group with normal weight; 34% for group with overweight. - JPW3: 19% for total sample; 16% for group with normal weight; 29% for group with overweight. - JPW4: 18% for total sample; 15% for group with normal weight; 27% for group with overweight. - JPW7: 15% for total sample; 14% for group with normal weight; 25% for group with overweight. <i>*Bias data is only shown for this outcome.</i> Accuracy: - DW Sum 4SKF: CVRSME= 0.16 for total sample; 0.29 for group with normal weight; and 0.35 for group with overweight. - JPW3: CVRSME= 0.29 for total sample; 0.24 for group with normal weight; and 0.31 for group with overweight.	All measurements showed a positive bias, which decreased as %BF increased. The DW Sum 4SKF equation is the most accurate equation for the estimation of %BF in both Females with normal-weight and with overweight.

- JPW4: CVRSME= 0.35 for total sample; 0.26 for group with normal weight; and 0.37 for group with overweight.

- JPW7: CVRSME= 0.35 for total sample; 0.30 for group with normal weight; and 0.36 for group with overweight.

High quality studies

Aandstad et al. 2014	Females= 26 Males= 39 Healthy BMI= 23.3 Caucasian	(18-30) 21 ± 4 (19-27) 22 ± 2	- SKF: abdominal, biceps, chest, subscapular, suprailiac, thigh, triceps, in males. - 6 sites in Females, except chest. - %BF equations (from BD by Siri), SKF: 1. Durnin/ Womersley (DW Sum 4SKF) 2. Jackson/Pollock (JP Sum 3SKF), in males 3. Jackson, Pollock and Ward (JPW Sum 3SKF), in Females	DXA	Pearson correlation coefficient (r), ICC, Bland-Altman method	Correlation of equations vs %BF: Females: - SKF equations: r= 0.77 to 0.88; ICC (95% CI)= 0.73 to 0.88 (0.49-0.95) - BIA equations: r= 0.84 to 0.92; ICC (95% CI)= 0.80 to 0.92 (0.60-0.96) - SKF and BIA equation: r= 0.86; ICC (95% CI)= 0.86 (0.71-0.93) Males:	Both, SKF equations and SKF-BIA equations showed high/very high validity to assess %BF.

4. Lohman, in Females and males (Sum 5SKF)	- SKF equations: $r = 0.86$ to 0.88 ; ICC (95% CI)= 0.82 to 0.87 (0.69-0.93)
5. Slaughter et al., in Females and males (Sum 6SKF)	- BIA equations: $r = 0.76$ to 0.87 ; ICC (95% CI)= 0.66 to 0.84 (0.43-0.91)
- %BF equations, BIA:	
1. Deurenberg et al., in Females and males	- SKF and BIA equation: $r =$ 0.92 ; ICC (95% CI)= 0.69 (0.48-0.82)
2. Gray et al., in Females and males	Agreement:
3. Kotler et al., in Females and males	Females:
4. Kyle et al., in Females and males	- SKF equations: MD ($\pm$ 95% LoA)= 0.4 to 6.6% (4.7 to 6.6)
5. Lohman, in Females and males	- BIA equations: MD ($\pm$ 95% LoA)= -3.2 to 3.2% (4.0 to 8.0)
6. Lukaski et al., in Females and males	Males:
7. Segal et al. generalized equation	- SFK equations: MD ($\pm$ 95% LoA)= -4.8 to 3.4% (3.5 to 5.2)

			(GEN), in Females and males				- SKF-BIA equations: MD (± 95% LoA)= -2.2 to 4.4% (4.4 to 8.5)	
			8. Segal et al. fatness specific equation (FSE)					
			9. Sun et al., in Females and males					
			10. van Loan et al., in Females and males					
			- %BF equations (SKF and BIA)					
			1. Guo et al., in males					
			2. Yannakoulia et al., in Females					
Aristizabal et al. 2008	Females= 52 Healthy BMI= 22.3 Colombian	(18-40) 28.0 ± 6.6	- SKF: abdominal, biceps, iliocrestal, subscapular, supraspinal, thigh, triceps - %BF equations (from BD by Siri):	UWW	Pearson correlation coefficient (r), Bland-Altman method	Correlation of equations vs %BF: - DW Sum 4SKF equation: r= 0.62 (p< 0.001) - JPW Sum 3SKF equation: r= 0.71 (p< 0.001) - Ramirez/Torun equation: r= 0.67 (p< 0.001)	The DW Sum 4SKF and Ramirez/Torun equations showed moderate validity to predict %BF, while JPW Sum 3SKF equation showed high validity.	

			1. Durnin/ Womersley (DW Sum 4SKF) 2. Jackson, Pollock and Ward (JPW Sum 3SKF) 3. Ramirez/Torun ((%BF= -15,471 + (tricipital SKF*0,332) + (subescapular SKF*0,154) + (abdominal SKF*0,119) + (HC*0,356))			Agreement: - DW Sum 4SKF equation: MD (± 95% LoA)= -4.67 (- 13.25 to 3.90) - JPW Sum 3SKF equation: MD (± 95% LoA)= 3.81 (- 4.44 to 12.06) - Ramirez/Torun equation: MD (± 95% LoA)= -2.49 (- 11.00 to 6.03)	
Aristizabal et al. 2018	Females= 151 Healthy BMI= 23.6 ± 3.2 Colombian Development group, validation group	(18-59)	- AbC, AC, CC, HC, TC, WC - SKF: abdominal, axillar, biceps, calf, iliocrestal, subscapular, thigh, triceps - %BF equations (from BD by Siri):	UWW	Pearson correlation coefficient (r), regression analysis, SEE, ICC	Equations development: - Equation 1: R ² = 0.72, SEE= 3.12 (p< 0.001) - Equation 2: R ² = 0.72, SEE= 3.08 (p< 0.001) - Equation 3: R ² = 0.66, SEE= 3.44 (p< 0.001) Equations validation:	All new equations showed high validity to assess total %BF. Equation 1 (based on triceps and thigh SKF + AbC) was the most accurate to assess total %BF.

1. Durnin/ Womersley (DW Sum 4SKF)	- Equation 1: $R^2 = 0.71$, SEE = 2.81 ($p < 0.001$)	DW Sum 4SKF and Ramirez/Torun equations overestimated %BF, while JPW Sum 3SKF equation underestimated %BF in this population.
2. Jackson, Pollock and Ward (JPW Sum 3SKF)	- Equation 2: $R^2 = 0.67$, SEE = 3.06 ($p < 0.001$)	
3. Ramirez/Torun	- Equation 3: $R^2 = 0.55$, SEE = 3.55 ($p < 0.001$)	
4. Equations of the present study (3 equations)	Correlation of equations vs %BF: - DW Sum 4SKF equation: $r = 0.75$, ICC = 0.51 ($p < 0.001$) - JPW Sum 3SKF equation: $r = 0.77$, ICC = 0.53 ($p < 0.001$) - Ramirez/Torun equation: $r = 0.77$, ICC = 0.75 ($p < 0.001$) - Equation 1: $r = 0.81$, ICC = 0.77 ($p < 0.001$) - Equation 2: $r = 0.79$, ICC = 0.76 ($p < 0.001$) - Equation 3: $r = 0.77$, ICC = 0.73 ($p < 0.001$)	

Balas-Nakash et al. 2010	Females= 86 Healthy BMI= 30.7 Mexican	(42-72) 54.8 ± 6.9	- BIA - %BF equations: 1. NHANES III 2. Macias, et al.	DXA	Pearson correlation coefficient (r), CE, SEE, TE Bland- Altman method	Correlation of equations vs %BF: - BIA- NHANES III equations: r= 0.80 (p=0.001), CE= -2.61, SEE= 3.75, TE= 94.98 - BIA-Macias, et al. equations: r= 0.80 (p=0.001), CE=- 2.61, SEE= 3.75, TE= 94.98 Agreement: - BIA- NHANES III equations: MD (± 95% LoA)= -2.0 (-8.0 to 4.0) - BIA-Macias, et al. equations: MD (± 95% LoA)= -2.0 (-8.5 to 4.5)	Both BIA equations showed high validity to assess %BF.
Ball et al. 2004	Females= 150 Healthy BMI= 23.0	(18-55) 28.7 ± 8.6	- SKF: abdominal, chest, midaxillary, subscapular, suprailiac, thigh, triceps	DXA	Pearson correlation coefficient (r), regression analysis, SEE, Bland-Altman method	Correlation of equations vs %BF: - JPW 7SKF: r= 0.92 (p< 0.01)	The DXA criterion equation, JPW and JP (Sum 7SKF and Sum 3SKF) equations showed very high validity to assess %BF.

Caucasian, Hispanic, African American, Asian			- %BF equations (from BD by Siri and Brozek): 1. Jackson, Pollock and Ward, seven-SKF (JPW 7SKF) 2. Jackson, Pollock and Ward, three-SKF (JPW 3SKF) 4. DXA criterion		- JPW 3SKF: $r = 0.92$ ($p < 0.01$) - JP 3SKF: $r = 0.89$ ($p < 0.01$) Regression analysis for DXA criterion equation: - $R^2 = 0.86$, $SEE = 2.5$ ($p < 0.01$) Cross-validation of DXA criterion equation: - $R^2 = 0.93$, $SEE = 2.1$ ($p < 0.01$) Agreement: - MD ($\pm$ 95% LoA) = 0.2 (- 4.8 to 4.9)		
Brozek & Keys 1951	Males = 255 Healthy %BF = 10 college, 24 middle-aged <i>Race-ethnicity not described</i>	(18-55) 20.3 ± 1.9 49.0 ± 2.8	- AbC, ChC - SKF: abdominal, chest, scapular, thigh, triceps - %BF equations (from BD):	UWW	Pearson correlation coefficient (r), SEE	Correlation of SKF vs %BF: - Abdominal: $r = -0.84$, $SEE = 0.008$ in college; $r = -0.60$, $SEE = 0.010$ in middle-aged (both, $p < 0.05$)	Chest SKF was higher correlated with %BF (from BD), compared with UWW. Both SKF equations showed high validity to assess %BF.

2 age groups
(college students
and middle-aged
men)

1. Rathbun & Pace

- Chest: $r = -0.86$, $SEE = 0.007$
in college; $r = -0.68$, $SEE =$
 0.009 in middle-aged (both,
 $p < 0.05$)

- Scapular: $r = -0.81$, $SEE =$
 0.009 in college; $r = -0.68$,
 $SEE = 0.009$ in middle-aged
(both, $p < 0.05$)

- Triceps: $r = -0.83$, $SEE =$
 0.008 in college; $r = -0.65$,
 $SEE = 0.010$ in middle-aged
(both, $p < 0.05$)

- Thigh: $r = -0.75$, $SEE = 0.009$
in college; $r = -0.63$, $SEE =$
 0.010 in middle-aged

Best equations:

- College students, Equation 3:
 $r = 0.87$, $SEE = 0.007$ ($p < 0.05$)

- Middle-aged men, Equation
4: $r = 0.74$, $SEE = 0.008$ ($p <$
 0.05)

Cui et al. 2014	Females= 4666 Males= 5268 Healthy BMI= 26.2 Normal weight (NW), overweight (OW) and obese (OB) Mexican Americans (MA), Non-Hispanic Whites (NHW), Non-Hispanic Blacks (NHB)	(20-84) 46.6 ± 0.4 44.3 ± 0.3	- WC, WHtR - SKF: subscapular, triceps - BMI - %BF equations (from BD by Siri): 1) SKF: Durnin/Womersley (DW Sum 4SKF); Hassager et al.; Slaughter et al. 2) Height and weight and/or BMI: Gallagher et al.; Gomez-Ambrosi et al.; Heitmann; Jackson/Pollock; Noppa et al. (females); Pasco et al.; Rush et al. (females); Smith and Boyce (females); Visser et al.	DXA	Regression analysis, RMSE, MSD	Regression analysis of equations vs %BF: Females (entire sample): 1) SKF, R ² = 0.63 to 0.66 (p< 0.05); RMSE= 3.6 to 4.0; MSD= -4.5 to -3.1 2) BMI, R ² = 0.55 to 0.72 (p< 0.05); RMSE= 3.3 to 4.3; MSD= -8.5 to 3.3 3) WC or WHtR, R ² = 0.55 to 0.56 (p< 0.05); RMSE= 4.4; MSD= 0.2 to 0.5 4) Multiple, R ² = 0.46 to 0.69 (p< 0.05); RMSE= 3.4 to 4.2; MSD= -9.9 to 1.3 Males (entire sample): 1) SKF, R ² = 0.61 to 0.77 (p< 0.05); RMSE= 3.0 to 3.4; MSD= -5.3 to -0.9	In Females, equations using BMI, showed higher validity to assess %BF than the other equations. In males, equations using SKF and WC, showed higher validity to assess %BF than the other equations. Equations using BMI or WC underestimated %BF in adults with normal weight but slightly overestimate it in adults with obesity. Equations using SKF underestimated %BF more in Females with obesity than in females with non- obesity. These equations did not perform better in one race-

3) WC or WHtR:
Kagawa et al.; Lean et
al.

4) Multiple
anthropometric
variables: Chapman et
al.; Ramirez-Zea et al.;
Svendsen et al.;
Wilmore & Behnke

2) BMI, $R^2 = 0.47$ to 0.60 ($p < 0.05$); RMSE= 3.4 to 3.7;
MSD= -4.1 to 3.7

3) WC or WHtR, $R^2 = 0.68$ ($p < 0.05$); RMSE= 3.1; MSD= 0.5
to 1.1

4) Multiple, $R^2 = 0.46$ to 0.76
($p < 0.05$); RMSE= 2.7 to 3.4;
MSD= -7.7 to 1.1

BMI status (not sex
distinction):

1) SKF, $R^2 = 0.33$ to 0.60 for
NW; $R^2 = 0.03$ to 0.42 for
OW; $R^2 = 0.02$ to 0.36 for OB
(all, $p < 0.05$)

2) BMI, $R^2 = 0.18$ to 0.38 for
NW; $R^2 = 0.02$ to 0.20 for
OW; $R^2 = 0.08$ to 0.30 for OB
(all, $p < 0.05$)

3) WC or WHtR, $R^2 = 0.21$ to
 0.51 for NW; $R^2 = 0.04$ to 0.39

ethnicity than that in
others.

for OW; $R^2 = 0.12$ to 0.41 for
OB (all, $p < 0.05$)

4) Multiple, $R^2 = 0.30$ to 0.60
for NW; $R^2 = 0.05$ to 0.50 for
OW; $R^2 = 0.04$ to 0.51 for OB

Sex-race/ethnicity-specific
(not sex distinction):

1) SKF, $R^2 = 0.46$ to 0.63 for
MA; $R^2 = 0.53$ to 0.68 for
NHW; $R^2 = 0.55$ to 0.73 for
NHB (all, $p < 0.05$)

2) BMI, $R^2 = 0.49$ to 0.62 for
MA; $R^2 = 0.51$ to 0.70 for
NHW; $R^2 = 0.52$ to 0.67 for
NHB (all, $p < 0.05$)

3) WC or WHtR, $R^2 = 0.47$ to
 0.66 for MA; $R^2 = 0.52$ to 0.68
for NHW; $R^2 = 0.48$ to 0.71
for NHB (all, $p < 0.05$)

4) Multiple, $R^2 = 0.43$ to 0.69
for MA; $R^2 = 0.56$ to 0.73 for

					NHW; R ² = 0.54 to 0.75 for NHB (all, p< 0.05)		
Davidson et al. 2011	Females= 1002	(18-110)	- WC	DXA	Linear regression, SEE	Females:	New sex-race/ethnicity-specific prediction equations (based on Durnin/ Womersley Sum 4SKF + WC, age, height and weight) showed high validity. However, overestimated body fat by DXA in all but C males and AA Females.
	Males= 673	50.7 ± 19.1	- SKF: biceps,			- Logarithm of Sum SKF: 35.884 logSF - 33.709, R ² = 0.75, SEE= 4.56 (p< 0.05)	
	Healthy	47.7 ± 18.8	subscapular, suprailiac, triceps			- New DW equation sex-race/ethnicity-specific prediction equations:	
	BMI= 23.2 (C), 27.3 (AA), 27.8 (H), 21.9 (A)		- %BF equations (from BD by Siri):			R ² = 0.80, SEE= 3.94 (p< 0.05) for C	
	Caucasian (C), African, American (AA), Hispanic (H), Asian (A)		1. Durnin/ Womersley (DW Sum 4SKF), age-specific			R ² = 0.80, SEE= 3.98 (p< 0.05) for AA	
			2. DW all-age			R ² = 0.73, SEE= 3.71 (p< 0.05) for H	
			3. Sex and race: DW equation+age, height, weight and WC			R ² = 0.73, SEE= 3.43 (p< 0.05) for A	
						Males:	
						- Logarithm of Sum SKF: 30.729 logSF - 31.122, R ² = 0.71, SEE= 4.12 (p< 0.05)	

						- New DW equation sex- race/ethnicity-specific prediction equations: R ² = 0.77, SEE= 3.53 (p< 0.05) for C R ² = 0.79, SEE= 4.00 (p< 0.05) for AA R ² = 0.80, SEE= 3.43 (p< 0.05) for H R ² = 0.64, SEE= 3.45 (p< 0.05) for A	
Demura et al. 2002	Males= 50 Healthy %BF= <i>see results</i> Japanese	(18-27)	- SKF: subscapular, triceps - BIA -%BF equations (from BD by Brozek): 1. BIA, hand-foot method (H-F) 2.BIA, foot-foot method (F-F)	UWW	Linear regression, Bland-Altman method	%BF values: - BIA, H-F= 17.8 ± 3.8 - BIA, F-F= 17.7 ± 4.8 - SKF= 12.6 ± 2.7 - UWW= 14.5 ± 5.5 Prediction equations vs %BF: - BIA, H-F: R ² = 0.92 (p< 0.05)	Hand-foot BIA analysis and SKF showed very high validity to assess %BF, compared with UWW.

3. SKF method				- BIA, F-F: $R^2 = 0.50$ ($p < 0.05$) - SKF: $R^2 = 0.51$ ($p < 0.05$) Agreement: - H-F, MD ($\pm$ 95% LoA)= 3.3 (-0.03 to 1.58) - F-F, MD ($\pm$ 95% LoA)= 3.2 (0.23 to 1.84) - SKF, MD ($\pm$ 95% LoA)= 1.9 (-1.61 to 0.63)			
Dioum et al. 2005	Females= 196 Healthy BMI= 22.5 African	(18-56) 29.5 $\pm$ 8.7	- SKF: biceps, subscapular, suprailiac, triceps - %BF equations (from BD by Siri): 1. Durnin/ Womersley (DW Sum 4SKF) equation coupled with Siri equation (DWS) 2. Durnin/ Womersley (DW Sum 4SKF) equation coupled with	ADP	Pearson correlation coefficient (r), Pure Error (PE), Bland Altman method	Correlation of equations vs %BF: - DWS equation: $r = 0.50$, PE= 7.9 ($p < 0.00001$) - DWB equation: $r = 0.53$, PE= 6.5 ($p < 0.00001$) Agreement: - DWS equation, MD ($\pm$ 95% LoA)= 6.5 (-2.3 to 15.3) - DWB equation, MD ($\pm$ 95% LoA)= 4.6 (-4.2 to 13.4)	Both DW Sum 4SKF equations showed moderate validity and underestimate %BF.

			Black-specific equation (DWB)				
Durnin & Womersley 1974	Females= 272 Males= 209 Healthy %BF= 14-52 Females; 7-50 males <i>Race-ethnicity not described</i> Age-groups: 16- 19, 20-29, 30-39, 40-49 years	(16-72)	- AC, CC, TC - SKF: biceps, subscapular, suprailiac, triceps - %BF equations (from BD): 1. Brozek 2. Siri Different Sum SKF	UWW	Regression analysis, SEE	SKF vs %BF (from BD): - Females: R ² = 0.70-0.90, SEE= 0.012-0.009 (p< 0.05) - Males: R ² = 0.70-0.90, SEE= 0.010-0.007 (p< 0.05)	Durnin/ Womersley equations showed high to very high validity to assess %BF, compared with UWW. Although there is little error between even one single measured SKF or two and three SKF, the likelihood of a large error can be reduced by using the Sum 4SKF. In addition, the best-fit regression is derived from Sum 4SKF.
Friedl et al. 2001	Females= 150 Healthy BMI= 23.3 Black (B), Hispanic (H),	(17-33) 21.4 ± 3.6	- AC, AbC, CC, ChC, FC, NC, TC, WC - SKF: abdominal, biceps, calf, chest, midaxillary,	DXA	ME (residual SD), F ratio of PE, sensitivity, specificity	Correlation of body composition measures vs %BF: - Marines: ME= 6.7 (2.8) B, 8.3 (3.0) H, 7.1 (2.8) NHW; PE (all combined)= 8.1%; F	All the equations had relatively low sensitivity and specificity. The Navy and DW Sum 4SKF equations had the lowest error in the

Non-Hispanic White (NHW)	subscapular, suprailiac, thigh, triceps	ratio (all combined)= 8.08 (p< 0.05)	prediction of %BF.
	- %BF equations (from BD by Siri):	- Navy: ME= 1.3 (2.1) B, 0.9 (1.9) NHW, 2.4 (1.9) H; PE (all combined)= 3.9%; F ratio (all combined)= 5.95 (p< 0.05)	
	1. Marines		
	2. Navy		
	3. Army	- Army: ME= 2.0 (2.3) B, 3.0 (2.1) H, 2.8 (2.2) NH; PE (all combined)= 4.8%; F ratio (all combined)= 4.80 (p< 0.05)	
	4. Durnin/ Womersley (DW Sum 4SKF)		
	5. Jackson, Pollock and Ward (JPW Sum 7SKF)	- DW Sum 4SKF: ME= -0.8 (2.0) B, -0.1 (1.8) H, -0.1 (1.8) NHW; PE (all combined)= 3.5%; F ratio (all combined)= 3.90 (p< 0.05)	
		- JPW Sum 7SKF: ME= 5.0 (2.4) B, 5.9 (2.5) H, 5.1 (2.4) NHW; PE (all combined)= 5.9%; F ratio (all combined)= 3.49 (p< 0.05)	
		Sensitivity: 12% to 55%	

					Specificity: 48% to 86%		
Gallagher et al. 1996	Females= 394	(20-94)	- HC, WC, WHR	DXA,	Pearson correlation coefficient (r), multiple regression analysis	Correlation of age and body composition measures correlations vs BMI: - Females, Whites: r= 0.21-0.87 (all, p< 0.05); Blacks: r= 0.23-0.89 (all, p< 0.05), height, r= -0.07 (p>0.05) - Males, Whites: r= 0.44-0.84 (all, p< 0.05), age and height, r= -0.04 to 0.07 (p< 0.05); Blacks: r= 0.25-0.85 (all, p< 0.05), height, r= 0.02 (p>0.05) Regression analysis, BF vs age and BMI: - Females, Whites: R ² = 0.56 (p< 0.001); Blacks: R ² = 0.58 (p< 0.05) - Males, Whites: R ² = 0.44 (p< 0.001); Blacks: R ² = 0.52 (p< 0.05)	BMI showed moderate to high validity to assess BF. Nevertheless, BMI is age and sex dependent as an indicator of BF, but is ethnicity independent in White and Black adults.
	Males= 312	50.1 ± 17.1	- Tibia/total body length	D ₂ O,			
	Healthy	47.7 ± 17.5	- BMI	³ H ₂ O,			
	Whites (Females, BMI= 23.3; males, BMI= 25.2)		- BF equation: 2.513 x body volume - 0.739 x total body water + 0.947 x total body bone mineral mass - 1.79 x body weight	UWW (4C)			

Gallagher et al. 2000	Females= 1013 Males= 613 Healthy White (BMI= 25.5), African American (BMI= 23.1), and Asian (BMI= 26.5)	48.1 ± 16.7 48.7 ± 17.4	- BMI - %BF equations: 1. Equation for White and African American 2. Equations for Asian Females and Asian males	DXA, D ₂ O, ³ H ₂ O, UWW (4C)	Pearson correlation coefficient (r), multiple regression analysis	Regression analysis, BF vs age, sex and BMI:	BMI showed moderate to high validity to assess %BF. BMI is ethnicity dependent, especially in Asian population.
						- Whites: R ² = 0.64 (p< 0.001) - Blacks: R ² = 0.72 (p< 0.001) Regression analysis, BF vs age, sex, ethnicity and BMI: R ² = 0.67 (p< 0.001) Correlation of BMI vs %BF: - Whole sample: r= 0.68-0.89 (all, p< 0.001) Regression analysis of %BF equations vs DXA/ 4C: - Asian, Females: R ² = 0.88 (p< 0.05), SEE= 2.91 %BF - Asian, males: R ² = 0.77 (p< 0.05), SEE= 3.49 %BF - White and African American adults: R ² = 0.86 (p< 0.05), SEE= 4.98 %BF	

Goel et al. 2008	Females= 76 Males= 95 Healthy BMI= 22.9 Asian Indians	(18-50) 32.2 ± 9.5	- WC, HC, Mid-thigh C, CC - SKF: anterior axillary, biceps, calf, subscapular, suprailiac, thigh, triceps - %BF equation= 42.42 + 0.003 × age + 7.04 × gender + 0.42 × triceps SKF + 0.29 × WC – 0.22 × weight – 0.42 × height	DXA	Pearson correlation coefficient (r), multiple regression analysis, ME	Correlation of body composition measures vs %BF: - Triceps: r= 0.82 (p< 0.05) - Thigh: r= 0.78 (p< 0.05) - Calf: r= 0.69 (p< 0.05) - WC: r= 0.53 (p< 0.05) Regression analysis of the predictive equation: - R ² = 0.86 (p< 0.05), ME= - 0.60, CI= -9.60 to 7.82	This equation (based on triceps SKF + WC) showed very high validity to assess %BF.
Gómez-Ambrosi 2011	Females= 4356 Males= 2154 Healthy BMI= 32.0 White	(18-80) 44 ± 14 47 ± 14	- WHR, WHtR - BMI - BAI - CUN-BAE, Rohrer index - New equation (CUN- BAE):	ADP	Pearson correlation coefficient (r), SEE, Bland-Altman method	Correlation of body composition measures correlations vs %BF: Females (all, p< 0.001): - WHR: r= 0.82 - WHtR: r= 0.49 - BAI: r= 0.81 - BMI: r= 0.84	The CUN-BAE showed high validity to assess %BF. In Females, all measures (except for WHtR) showed high validity to assess %BF. In males, WHR and BMI high validity, while WHtR

			$-44.988 + (0.503 \times \text{age}) + (10.689 \times \text{sex}) + (3.1723 \text{BMI}) - (0.026 \times \text{BMI}^2) + (0.181 \times \text{BMI} \times \text{sex}) - (0.02 \times \text{BMI} \times \text{age}) - (0.005 \times \text{BMI}^2 \times \text{sex}) + (0.000213 \text{BMI}^2 \times \text{age})$			- CUN-BAE: $r = 0.89$, $\text{SEE} = 4.36\% \text{BF}$ Males (all, $p < 0.001$): - WHR: $r = 0.84$ - WHtR: $r = 0.47$ - BAI: $r = 0.35$ - BMI: $r = 0.77$ - CUN-BAE: $r = 0.81$, $\text{SEE} = 5.20\% \text{BF}$ Agreement, $\% \text{BF}$ CUN-BAE vs ADP: MD ($\pm 95\% \text{LoA}$) = -0.64 (-9.86 to 8.58)	and BAI low validity to assess $\% \text{BF}$.
Gomez et al. 2017	Females= 3359 Males= 3354 Healthy BMI= 23.9 Chilean	(17-27) 20.7 ± 2.0 20.6 ± 2.4	- WC, WHtR - Equations for fat mass (FM): Females: 1. $\text{FM} = -13,216.917 + 461,302 \times \text{weight} + 91,898 \times \text{WC}$	DXA	Regression analysis, SEE, Bland-Altman method	Regression analysis of equations: Equations for Females: - 1. $R^2 = 0.70$, $\text{SEE} = 4.6$, $p < 0.05$ - 2. $R^2 = 0.70$, $\text{SEE} = 4.6$, $p < 0.05$	The four equations (based on WC and WHtR) showed high validity to assess fat mass in young population.

$$2. FM = -14,144,220 + 464,061 \times \text{weight} + 16189,297 \times \text{WHtR}$$

Males:

$$1. FM = -35,997,486 + 232,285 \times \text{weight} + 432,216 \times \text{WC}$$

$$2. FM = -37,671,303 + 309,539 \times \text{weight} + 66028,109 \times \text{WHtR}$$

Equations for males:

$$- 1. R^2 = 0.73, SEE = 4.1, p < 0.05)$$

$$- 2. R^2 = 0.76, SEE = 3.8, p < 0.05)$$

Agreement:

- Females:

$$1. MD (\pm 95\% LoA) = 0.1 (-7.3 \text{ to } 7.0)$$

$$2. MD (\pm 95\% LoA) = -0.1 (-7.1 \text{ to } 6.8)$$

- Males:

$$1. MD (\pm 95\% LoA) = 0.0 (-9.0 \text{ to } 9.1)$$

$$2. MD (\pm 95\% LoA) = 0.0 (-9.0 \text{ to } 9.0)$$

Jackson & Pollock
1978

Males= 308

(18-61)

- FC, WC

UWW

Regression
analysis, SEE

Correlation between Sum
7SKF and Sum 3SKF: $r = 0.98$
($p < 0.01$)

Both equations (composed
by Sum 3SKF or Sum
7SKF) showed very high
validity to assess %BF.

Healthy

32.6 ± 10.8

- SKF: abdomen,
axilla, triceps, chest,

%BF= 17.8

	<i>Race-ethnicity not described</i>		subscapula, supra-iliac, thigh			Regression analysis of equations vs %BF (from BD):	
			- %BF equations (from BD):			- Sum 7SKF: r= 0.92, SEE= 0.007 g/ml (p< 0.01)	
			1. Brozek			- Sum 3SKF: r= 0.91, SEE= 0.007 g/ml (p< 0.01)	
			2. Siri			Cross-validation of these equations:	
			Different Sum SKF			- All equations: r= 0.90, SEE<0.0900 g/ml (p< 0.01)	
Jackson, Pollock & Ward 1980	Females= 249	(18-55)	- GIC	UWW	Regression analysis, SEE	Correlation between Sum 7SKF, Sum 4SKF and Sum 3SKF: r= 0.97 (p< 0.01)	All equations (composed by Sum 3SKF, Sum 4SKF or Sum 7 SKF) showed very high validity to assess %BF.
	Healthy	31.4 ± 10.8	- SKF: abdomen, axilla, chest, subscapula, suprailiac, thigh, triceps			Regression analysis of equations vs %BF (from BD):	
	%BF= 24.1		- %BF equations (from BD):			- Sum 7SKF: r= 0.87, SEE= 0.007 g/ml (p< 0.01)	
	<i>Race-ethnicity not described</i>		1. Brozek			- Sum 4SKF: r= 0.87, SEE= 0.008 g/ml (p< 0.01)	
			2. Siri			- Sum 3SKF: r= 0.85, SEE= 0.008 g/ml (p< 0.01)	
			Different Sum SKF				

					Cross-validation of these equations:		
					- All equations: $r = 0.80$, $SEE < 0.0100$ g/ml ($p < 0.01$)		
Johnson et al. 2011	Females= 332	(20-50)	- HC, WC	DXA	Pearson correlation coefficient (r), bias correction factor (c), LCCC ((Pc(SD)))	Correlations of BAI equation vs %BF:	BAI and BMI equations showed high validity to assess %BF.
	Males= 291	± 34.9	- %BF equation:			- Females: $r = 0.75$, $c = 0.89$ ($p < 0.05$), LCCC ((Pc(SD)))= 0.676 (0.026)	
	Healthy		1. BAI= HC (cm)/height (m) $1.5-18$				
	BMI= 25.7		2. BAIFels= $1.26 \times$ (HC (cm)/height (m) $1.4-32.85$			- Males: $r = 0.72$, $c = 0.77$ ($p < 0.05$), LCCC ((Pc(SD)))= 0.559 (0.028)	
	European-American		3. BMI			Correlation of BAIFels equation vs %BF: - Females: $r = 0.76$, $c = 0.98$ ($p < 0.05$), LCCC ((Pc(SD)))= 0.748 (0.024) - Males: $r = 0.73$, $c = 0.91$ ($p < 0.05$),	

						LCCC ((Pc(SD)))= 0.668 (0.028)	
						Correlation of BMI equation vs %BF:	
						- Females: r= 0.74, c= 0.54 (p< 0.05), LCCC ((Pc(SD)))= 0.406 (0.026)	
						- Males: r= 0.72, c= 0.76 (p< 0.05), LCCC ((Pc(SD)))= 0.561 (0.030)	
Ketel et al. 2007	Females= 200 Males= 176 Healthy BMI= 24.0 Caucasian	(36-37) 36.1 ± 0.7 (35-36) 36.0 ± 0.8	- HC, WC, WHR - SKF: subscapular, suprailiac - BMI - Central fat mass equations: 1. Females: a) 18.53 + 0.20 WC + 0.51BMI	DXA	Pearson correlation coefficient (r), Linear regression	Correlation of body composition measures vs %BF: Females: - HC, r= 0.78 (p< 0.001) - WC, r= 0.82 (p< 0.001) - WHR, r= 0.21 (p>0.05) - BMI, r= 0.82 (p< 0.001)	All measures (except for WHR) showed high validity to assess central fat mass, compared with DXA. All equations (based on WC + BMI + SKF) showed high validity to assess central fat mass.

b) $12.90 + 0.13 \text{ WC} + 0.36 \text{ BMI} + 0.23 \text{ subscapular SKF}$

c) $17.67 + 0.13 \text{ WC} + 0.16 \text{ BMI} + 0.11 \text{ HC} + 0.20 \text{ subscapular SKF}$

2. Males:

a) $33.00 + 0.34 \text{ WC} + 0.48 \text{ BMI}$

b) $25.40 + 0.25 \text{ WC} + 0.32 \text{ BMI} + 0.31 \text{ subscapular SKF}$

c) $22.80 + 0.22 \text{ WC} + 0.31 \text{ BMI} + 0.16 (\text{subscapular} + \text{suprailiac SKF})$

- Subscapular+suprailiac, $r = 0.82$ ($p < 0.001$)

- Subscapular, $r = 0.82$ ($p < 0.001$)

- Suprailiac, $r = 0.72$ ($p < 0.001$)

Males:

- HC, $r = 0.81$ ($p < 0.001$)

- WC, $r = 0.85$ ($p < 0.001$)

- WHR, $r = 0.27$ ($p < 0.001$)

- BMI, $r = 0.81$ ($p < 0.001$)

- Subscapular+suprailiac, $r = 0.84$ ($p < 0.001$)

- Subscapular, $r = 0.80$ ($p < 0.001$)

- Suprailiac, $r = 0.78$ ($p < 0.001$)

Regression analysis of equations vs %BF:

Females (all, $p < 0.05$):

						a) $R^2 = 0.75$	
						b) $R^2 = 0.79$	
						c) $R^2 = 0.82$	
						Males (all, $p < 0.05$):	
						a) $R^2 = 0.77$	
						b) $R^2 = 0.82$	
						c) $R^2 = 0.83$	
Lahav et al. 2018	Females= 113	(20-62)	- AbC, NC, WC	DXA	Linear regression, SEE, Bland-Altman method	Regression analysis of %BFcal vs %BF:	Both equations (based on AbC + NC) showed high validity to assess %BF.
	Males= 111	38.3 $\pm$ 10.5	- SKF: biceps, triceps, subscapular, suprailiac			- Females: $R^2 = 0.65$, SEE=	DW Sum 4SKF equation
	Healthy	(21-60)				4.85 ($p < 0.05$)	was less accurate for
	BMI= 27.2	38.7 $\pm$ 9.8	- %BF equations:			- Males: $R^2 = 0.78$, SEE= 4.66	prediction of %BF than
	<i>Race-ethnicity not described</i>		1. Females, %BFcal= 19.197($\pm$ 7.439)- 0.239($\pm$ 0.035) x height +0.808($\pm$ 0.29) x AbC -0.518($\pm$ 0.128) x NC			($p < 0.05$)	%BFcal.
			2. Males, %BFcal= 10.111($\pm$ 6.541)- 0.239($\pm$ 0.035) x			Agreement:	
						- Females, MD ($\pm$ 95% LoA)= 0.1 (-8.4 to 8.4)	
						- Males, MD ($\pm$ 95% LoA)= 0.0 (-7.8 to 7.8)	

			height +0.808(± 0.029) x AbC-0.518(± 0.128) x NC - SKF, %BFSKF by Durnin/ Womersley (DW Sum 4SKF)			Accuracy of %BFcal vs %BFSKF: - Females, %BFcal (MD)= 0.24; %BFSKF (MD)= -3.01 - Males, %BFcal (MD)= - 0.93; %BFSKF (MD)= -2.48	
Lanham et al. 2001	Females= 40 Healthy BMI= 21.7 Chinese Australian	(18-45) 32.5 ± 8.0	- SKF: biceps, subscapular, suprailiac, triceps - %BF equations: 1. Durnin/Womersley (DW Sum 4SKF) 2. BIA	D ₂ O	Pearson correlation coefficient (r), ANOVA, Bland- Altman method	Correlation of equations vs %BF: - DW Sum 4SKF: r= 0.71, t= - 9.53 (p< 0.001) - BIA equation: r= 0.76, t= 10.58 (p< 0.001) Agreement: - DW, MD (SD): -7.6 (4.5) - BIA, MD (SD): -6.7 (4.1)	Both equations, DW Sum 4SKF and BIA, showed high validity to assess %BF. However, both equations suggest an overestimation of body fat levels for leaner individuals and under estimation for overfat individuals.
Lee et al. 2017	Females= 2015 Males= 2292 Healthy BMI= 26.1	(≥ 18) 45.3 ± 24.2 42.7 ± 22.4	- AC, CC, TC, WC - SKF: subscapular, triceps - BMI - %BF equations:	DXA	Linear regression, SEE	Regression analysis of equations vs %BF: Females: 1) R ² = 0.66, SEE= 3.87 (p< 0.01)	All the equations (based on age, height, weight + WC, AC, CC, TC + subscapular and triceps SKF) showed high validity to assess %BF.

	White, Black, Mexican American, Hispanic, Other		1. Based on: Age, height and weight.			2) R ² = 0.66, SEE= 3.85 (p< 0.01)	
			2. Based on: Age, height, weight and WC			3) R ² = 0.69, SEE= 3.69 (p< 0.01)	
			3. Based on: Age, height, weight, WC, AC, CC and TC			4) R ² = 0.74, SEE= 3.36 (p< 0.01)	
			4. Based on: Age, height, weight, WC, AC, CC, TC, subscapular and triceps SKF			Males: 1) R ² = 0.62, SEE= 3.57 (p=0.16) 2) R ² = 0.74, SEE= 3.01 (p< 0.01) 3) R ² = 0.74, SEE= 3.31 (p< 0.01) 4) R ² = 0.80, SEE= 2.66 (p< 0.01)	
Lukaski et al. 1986	Females= 67 Males= 47 Healthy	(19-50) 27 ± 7	- SKF: biceps, scapula, suprailiac, triceps - BIA - %BF equation (from BD):	UWW	Pearson correlation coefficient (r), regression analysis, SEE	Correlation of BIA vs UWW: - Females, r= 0.95 (p< 0.05) - Males, r= 0.98 (p< 0.05) Regression analysis of %BF:	BIA showed very high validity to assess %BF, compared with UWW. DW Sum 4SFK showed very high to assess %BF, compared with UWW.

	%BF= 25.1 Females; 16.2 males <i>Race-ethnicity not described</i>		1. Brozek - SKF by Durnin/ Womersley (DW Sum 4SKF)			- R ² = 0.98, SEE= 2.20 (p< 0.001) Accuracy between body composition methods and %BF: - DW Sum 4SKF, r= 0.88, SEE= 3.89 (p< 0.05) - BIA, r= 0.93, SEE= 0.66 (p< 0.05)	
Noppa et al. 1979	Females= 153 BMI= 24.6 Swedish	(44-66)	- AC, GLC, WC - SKF: subscapular, triceps - BMI - BF equations: Based on GLC and Sum subscapular and triceps SKF	K ⁴⁰ , Isotope dilution	Pearson correlation coefficient (r), multiple regression analysis, SEE	Correlation of body composition measures vs BF (all, p< 0.05): - AC: r= 0.77 - GLC: r= 0.86 - WC: r= 0.84 - Sum 2SKF: r= 0.77 - Triceps SKF: r= 0.74 - Subscapular SKF: r= 0.71 - BMI: r= 0.88	All body composition measures showed high validity to predict BF, compared with K ⁴⁰ /Isotope dilution. All equations (based on GLC + subscapular and triceps SKF) showed very high validity to predict BF.

						Regression analysis of equations (all, $p < 0.05$):	
						- 44 years, $R^2 = 0.95$, $SEE = 2.77$; 52 years, $R^2 = 0.93$, $SEE = 2.37$; 56 years, $R^2 = 0.95$, $SEE = 2.40$; 60 years, $R^2 = 0.90$, $SEE = 2.92$; 66 years, $R^2 = 0.91$, $SEE = 3.78$; all sample, $R^2 = 0.92$, $SEE = 2.96$	
Pasco et al. 2012	Females= 1076 Males= 1467 Healthy BMI, 3 groups: <25.0, 25.0-29.9, ≥ 30) White	(20-96)	- BMI	DXA	Multiple regression analysis, SEE, CI	Correlation of BMI vs BF (kg): Females: - BMI 25 kg/m ² : BF= 24.0 kg (95%CI 23.7–24.3); BMI 30 kg/m ² : BF= 33.5 kg (95%CI 33.2–33.8) Males: - BMI 25.0 kg/m ² : BF= 17.7 kg (95%CI 17.5–17.9); BMI 30 kg/m ² : BF= 27.2 kg (95%CI 27.0–27.5)	BMI equation showed very high to assess total BF, compared with DXA. Sex and age should be considered when BMI is used to indicate %BF.

Correlation of BMI vs %BF:							
Females:							
- BMI 25 kg/m ² : %BF= 36.7% (95%CI 36.4–37.0); BMI 30 kg/m ² : %BF= 44.2% (95%CI 43.8–44.6)							
Males:							
- BMI 25.0 kg/m ² : %BF= 22.7% (95%CI 22.4–23.0); BMI 30 kg/m ² : %BF= 29.9% (95%CI 29.6–30.2)							
Prediction equation:							
- Sex+age+BMI: R ² = 0.82, SEE= 4.07 (p< 0.001)							
Pollock et al. 1975	Females= 143	(18-50)	- AbC, AC, AkC, CC, ChC, FC, GlC, ShC, TC, WC, WrC	UWW	Pearson correlation coefficient (r), regression analysis, SEE	Correlation of BMI vs %BF (from BD): Young (all, p< 0.05): - Circumferences, r= 0.19-0.58 - Diameters, r= 0.03-0.65	In young females, thigh SKF showed high validity to validity to assess total %BF (from BD), compared with UWW.
	Healthy	20.2 ± 1.2					
	%BF= 24.8	44.7 ± 5.7	- Biacromial D, bi-iliac D, bitrochanter D, ChD, KD, WrD				
	young; 29.8		- Cup size				
	middle-aged						

	<i>Race-ethnicity not described</i>	Two age groups: young (18-22) and middle-aged (33-50)	- SKF: abdomen, axilla, chest, subscapula, suprailium, thigh, triceps - %BF equation (from BD): 1. Brozek 2. Siri			- SKF, r= 0.50-0.73 Middle-aged (all, p< 0.05): - Circumferences, r= 0.40-0.83 - Diameters, r= 0.18-0.71 - SKF, r= 0.49-0.82 Best prediction equations: - Young: SKF+circumferences+diameters, R ² = 0.84, SEE= 0.008 (p< 0.05) - Middle-aged: SKF+circumferences+diameters+cup size, R ² = 0.91, SEE= 0.006 (p< 0.05)	In middle-aged females, all SKF (except for thigh and knee), ChC, AbC, WC, GLC, ChD, bi-iliac D, bitrochanter D, and cup size showed high validity to validity to assess total %BF (from BD), compared with UWW. Equations combining SKF + circumferences + diameters for young females and SKF + circumferences + diameters + cup size for middle-aged females, best predicted %BF.
Pollock et al. 1976	Males= 84 Healthy %BF= 13.4 young; 24.7 middle-aged	(18-55) 19.7 ± 1.5 44.9 ± 4.8	- AbC, AC, AkC, CC, ChC, FC, GLC, ShC, TC, WC, WrC	UWW	Pearson correlation coefficient (r), regression analysis, SEE	Correlation of BMI vs %BF (from BD by UWW): Young (all, p< 0.05): - Circumferences, r= 0.02-0.59	In both, young and middle-aged men, SKF showed high validity to validity to assess total %BF (from

	<i>Race-ethnicity not described</i> Two age groups: young (18-22) and middle-aged (40-55)		- Biacromial D, bi-iliac D, bitrochanter D, ChD, KD, WrD - SKF: abdomen, axilla, chest, subscapula, suprailium, thigh, triceps - %BF equation (from BD): 1. Brozek 2. Siri		- Diameters, r= 0.12-0.47 - SKF, r= 0.73-0.77 Middle-aged (all, p< 0.05): - Circumferences, r= 0.15-0.70 - Diameters, r= 0.18-0.71 - SKF, r= 0.05-0.67 Best prediction equations: - Young: SKF+circumferences+diameters, R²= 0.88, SEE= 0.007 (p< 0.05) - Middle-aged: SKF+circumferences, R²= 0.84, SEE= 0.007 (p< 0.05)	BD), compared with UWW. In middle-aged men, AbC, GLC, TC, WC and bi-iliac D also showed high validity to validity to assess total %BF (from BD), compared with UWW. Equations combining SKF + circumferences + diameters in young men; and SKF + circumferences in middle-aged men, best predicted %BF.	
Pongchaiyakul et al. 2005	Females= 130 Males= 83 Healthy BMI= 23.3	(20-84)	- HC, WC - SKF: biceps, subscapular, suprailiac, triceps	DXA	Linear regression, Bland-Altman method	Regression analysis of body composition measures vs %BF: - Females, R²= 0.68 (p< 0.01)	Both equations showed high validity to assess %BF.

	Rural Thai		- %BF equations:			- Males, R ² = 0.68 (p< 0.01)	In Females, the equation
			1. Females= 0.417 x HC +0.172 x suprailiac SKF +0.46 x BMI - 23.748			Agreement:	was based on HC + suprailiac SKF + BMI.
			2. Males= 0.417 x subscapular SKF+ 0.621 x BMI -0.276 x biceps SKF +0.166 x WC -18.466			- Females, MD (± 95% LoA)= -0.1% (-9.0 to 9.0)	In males, the equation was based on WC + subscapular SKF + BMI.
						- Males, MD (± 95% LoA)= 0.03% (-8.2 to 8.3)	
Simoes et al. 2015	Females= 205 Males= 186 Healthy %BF= 37 Females; 27 males Portuguese	(24-64) 47.4 ± 11.1 48.8 ± 10.8	- AbC, AC, WC, HC, mid-thigh C - SKF: abdominal, biceps, subscapular, suprailiacal, thigh, triceps - %BF equations: 1. Model 1: Females: -23.14+0.25 WC+0.28 HC+0.72 AC -0.27 Weight+0.12 Age. Males: 9.61+0.29 WC+0.19 HC -0.16 Height. 2. Model 2:	DXA	Multiple linear regression, SEE, RRSE	Regression analysis of equations vs %BF: - Model 1: Females: R ² = 0.66, SEE= 4.0 (p< 0.05) Males: R ² = 0.57, SEE= 3.8 (p< 0.05) - Model 2: Females: R ² = 0.69, SEE= 3.8 (p< 0.05) Males: R ² = 0.69, SEE= 3.2 (p< 0.05) - Model 3: Females: R ² = 0.71, SEE= 3.7 (p< 0.05)	All model equations (based on circumferences and SKF) showed high validity to predict %BF.

			Females: 36.12+0.16 thigh SKF+0.14 subscapularSKF+0.28 tricepsSKF -0.18 Height+0.17 Weight+0.10 Age. Males: 32.83+0.11 abdominalSKF+0.17 suprailiacSKF+0.27 tricepsSKF – 0.16 Height+0.11 Weight+0.08 Age.				
			3. Model 3:				
			Females: -8.00+0.20 WC+0.21 HC+0.26 tricepsSKF +0.13 subscapularSKF+0.15 thighSKF –0.16 Weight +0.11 Age. Males: 20.91+0.21 WC+0.22 suprailiacSKF+0.23 tricepsSKF -0.12 Height.				
Smith & Boyce 1977	Females= 61 Healthy %BF= 25.1 <i>Race-ethnicity not described</i>	(25-37) 29.1 ± 3.2	- AbC, AC, AkC, CC, ChC, FC, HC, ShC, TC - Biacromial D, Bi-iliac D, Bi-trochanteric D, ChD, KD, WrD - SKF: abdominal, knee, midaxillary, scapula, suprailiac,	UWW	Pearson correlation coefficient (r), regression analysis, SEE	Correlation of equations vs %BF (from BD): - Circumferences: r= 0.82, SEE= 0.008 (p< 0.05) - SKF+circumferences: r= 0.82, SEE= 0.008 (p< 0.05) - SKF+diameters+circumferenc	All measures, except for height + weight, showed high validity to assess total %BF, compared with UWW. Equation based on weight + SKF + circumferences best predicted %BF.

			thigh, triceps - %BF equation (from BD): 1. Brozek			es: $r = 0.82$, $SEE = 0.008$ ($p < 0.05$) - Heith+weight: $r = 0.65$, $SEE = 0.010$ ($p < 0.05$)	
Steinkamp et al. 1965	Females= 817 Males= 1236 Healthy BF (kg)=19.8 Caucasian/ White, Black (males) Age-groups: 25- 34 and 35-44 Whites, 25-44 Blacks	(25-44)	- AC, AkC, ChC, TC, WC, WrC - Biacromial D, Bi-iliac D, Chest D - SKF: abdomen, arm, chest, scapular	K^{40} , Isotope dilution	Pearson correlation coefficient (r), regression analysis, SEE	Correlation of body composition measures vs BF (kg): - Female, White: $r = 0.38$ -0.88, circumferences; $r = 0.41$ -0.64, diameters; $r = 0.80$ -0.85, SKF in 25-34 years. $r = 0.23$ -0.94, circumferences; $r = 0.35$ -0.74, diameters; $r = 0.71$ -0.88, SKF in 35-44 years (all, $p < 0.05$) - Male, White: $r = 0.61$ -0.91, circumferences; $r = 0.54$ -0.73, diameters; $r = 0.70$ -0.85, SKF in 25-34 years; $r = 0.29$ -0.85, circumferences. $r = 0.23$ -0.50, diameters; $r = 0.34$ -0.60, SKF in 35-44 years (all, $p < 0.05$) - Male, Black (25-44 years): $r = 0.53$ -0.87, circumferences; $r = 0.30$ -0.56, diameters; $r =$ 0.56-0.88, SKF (all, $p < 0.05$) Regression equations: - $R^2 = 0.92$ -0.98, $SEE = 2.044$ - 3.665 ($p < 0.05$)	In White Females aged 25- 44 years and White males aged 25-34 years, SKF showed high validity to assess total BF. In Blacks, chest and abdomen SKF also showed high validity to assess total BF. All prediction equations (based on circumferences + diameters + SKF) showed very high validity to assess total BF, regardless of sex and race.
Thomas et al. 1998	Females= 54 Healthy,	(18-45) 32.5 $\pm$ 2.19	- SKF: biceps, subscapular, suprailiac,	MRI	Pearson correlation coefficient (r), Bland-Altman	Field test correlations vs DXA:	In the group with normal weight (and total sample), DW Sum 4SKF showed

BMI= 22.2
normal weight,
26.6 overweight,
34.6 obese 1, 46.4
obese 2
*Race-ethnicity not
described*

triceps
- BIA
- %BF (from BD):
1. Brozek
2. Siri
- SKF by Durnin/
Womersley (DW Sum
4SKF)

method

Total sample:
- DW Sum 4SKF: $r = 0.88$ ($p < 0.01$)
- BIA: $r = 0.93$ ($p < 0.01$)
Group with normal weight:
- DW Sum 4SKF: $r = 0.79$ ($p < 0.01$)
- BIA: $r = 0.54$ ($p < 0.02$)
Group with overweight:
- DW Sum 4SKF: $r = 0.59$ ($p < 0.02$)
- BIA: $r = 0.84$ ($p < 0.01$)
Group with obesity 1:
- DW Sum 4SKF: $r = 0.28$ ($p < 0.14$)
- BIA: $r = 0.91$ ($p < 0.01$)
Group with obesity 2:
- DW Sum 4SKF: $r = 0.28$ ($p < 0.28$)
- BIA: $r = 0.90$ ($p < 0.01$)
Agreement:
- DW Sum 4SKF vs MRI, MD
($\pm$ 95% LoA)= 13.0 (-2.5 to 28.0)
- BIA vs MRI, MD ($\pm$ 95%
LoA)= 9.8 (2.5 to 19.6)

high validity; in the group
with overweight, it showed
moderate validity; and in
the groups with obesity,
low validity, to assess
%BF, compared with MRI.
BIA showed high validity
in the group with
overweight, and very high
validity in the groups with
obesity, to assess %BF,
compared with MRI.

Tucker et al. 2001	Males= 154 Healthy BMI= 23.8 White	(18-26) 22.2 ± 2.1	- AC, AbC, CC, ChC, FC, HC, NC, TC, WC - Diet questions - Fitness and exercise questions - %BF equation: + 0.00267 * HC2 (cm) + 0.34142 * WC + 1.47496 * Fitness question [possible scores 1-7] - 0.65736 * Weight question [possible scores 1-6] - 0.40478 * Exercise question [possible scores 1-8] - 0.23323 * Cereal question [possible scores 1-9] - 0.10898 * Weight in pounds - 20.88722 (constant)	UWW	Multiple regression, SEE	Regression analysis of proposed equation vs %BF: - R ² = 0.80, SEE= 2.66 (p< 0.05)	This equation (based on HC, WC, and diet + exercise variables) showed high validity to assess %BF.
Vogel et al. 1988	Females= 1194 Males= 319 Healthy	(≥ 17) 24.1 ± 4.5 30.2 ± 8.9	- AbC, AC, AkC, CC, ChC, HC, HdC, NC, ShC, TC, WrC - AkD, Biacromial D,	UWW	Pearson correlation coefficient (r), multiple regression, SEE	Regression equations: Females (all, p< 0.05): - Total sample, R ² = 0.82,	These equations showed very high validity to assess %BF in all groups, regardless of ethnicity.

	BMI <22.9; 22.9-24.9; 25.0-26.9; >26.9		Bi-iliac D, Bi-trochanteric D, ChD, DD, ED, KD, WrD			SEE= 3.598	In Females, the equations were based on HC + FC + NC + WC.
	White, Hispanic, Black		- %BF equations (from BD by Siri):			- Whites, R ² = 0.82, SEE= 3.598	
			1. Females: -35.601 - (0.515 x height)+(0.173 x HC) - (1.574 x FC)-(0.533 x NC)-(0.200 x WC)+(105.328 x Loglo weight)			- Blacks, R ² = 0.73, SEE= 4.076	In males, the equations were based on AbC + NC.
			2. Males: 46.892 - (68.678 x Loglo height)+(76.462 x Loglo (AbC-NC))			- Hispanics, R ² = 0.85, SEE= 3.332	
						Males (all, p< 0.05):	
						- Total sample, R ² = 0.82, SEE= 4.020	
						- Whites, R ² = 0.78, SEE= 3.914	
						- Blacks, R ² = 0.82, SEE= 4.012	
						- Hispanics, R ² = 0.80, SEE= 4.027	
						Accuracy %BF equations:	
						- Females: r= 0.82	
						- Males: r= 0.82	
						Cross-validation of equations:	
						- Females: r= 0.79, SEE= 4.4	
						- Males: r= 0.89, SEE= 3.7	
Wang et al. 1994	Females= 390 Males= 297 Healthy BMI= 15-38	(18-94) 51 ± 19	- AC, ChC, IC, TC, WC - SKF: abdomen, biceps, chest, suprailiac, subscapular, thigh, triceps,	DXA	Regression analysis, SEE	Regression analysis of BMI vs %BF: Females (both, p< 0.05): - Whites, R ² = 0.56, SEE= 5.7	Equations including age, BMI + SKF showed high validity to assess %BF, compared with DXA in all groups. BMI alone is less accurate

White, Asian			umbilicus - BMI - %BF equations: - Females, White: based on BMI, age, abdominal, thigh and triceps SKF - Females, Asian: based on BMI, age, subscapular, suprailiac, thigh and triceps SKF - Males, White: based on BMI, age, abdominal-umbilicus, subscapular and thigh SKF - Males, Asian: based on BMI, age, abdominal, thigh and triceps SKF			- Asian, $R^2 = 0.55$, SEE= 4.4 Males (both, $p < 0.05$): - Whites, $R^2 = 0.35$, SEE= 5.2 - Asian, $R^2 = 0.40$, SEE= 4.9 Equations for predicting %BF: Females (both, $p < 0.05$): - Whites, $R^2 = 0.75$, SEE= 4.4 - Asian, $R^2 = 0.73$, SEE= 3.5 Males (both, $p < 0.05$): - Whites, $R^2 = 0.71$, SEE= 3.5 - Asian, $R^2 = 0.67$, SEE= 3.7	to assess %BF.
Wang et al. 1996	Females= 124 Males= 81 Healthy BMI= 22.4 Chinese	(18-67) 34.2 ± 11.1 32.1 ± 11.6	- SKF: biceps, subscapular, suprailiac, triceps - BIA - BMI - %BF (from BD by Siri): 1. SKF by Durnin/ Womersley (DW Sum 4SKF)	UWW	Pearson correlation coefficient (r), Bland-Altman method	Field test correlations vs UWW: Females: - DW Sum 4SKF: $r = 0.82$ ($p < 0.0001$) Agreement: MD ($\pm$ 95% LoA)= 0.59 (-14.5 to 10.5) - BIA: $r = 0.83$ ($p < 0.0001$) Agreement: MD ($\pm$ 95%	In Females, all field measurements showed high validity to assess %BF, compared with UWW. In males, BIA and BMI showed moderate validity and DW Sum 4SKF high validity to assess %BF, compared with UWW.

						LoA)= 0.63 (-12.0 to 10.0) - BMI: $r = 0.82$ ($p < 0.0001$) Agreement: MD ($\pm$ 95% LoA)= 0.66 (-12.5 to 14.8) Males: - DW Sum 4SKF: $r = 0.70$ ($p < 0.0001$) Agreement: MD ($\pm$ 95% LoA)= 0.36 (-10.0 to 11.3) - BIA: $r = 0.69$ ($p < 0.0001$) Agreement: MD ($\pm$ 95% LoA)= 0.39 (-11.0 to 13.5) - BMI: $r = 0.62$ ($p < 0.0001$) Agreement: MD ($\pm$ 95% LoA)= 0.63 (-10.3 to 14.7)	
Womersley & Durnin 1977	Females= 324 Males= 245 Healthy %BF= 23.2-32.6 Females; 2.8-23.8 males <i>Race-ethnicity not described</i> Age groups: 17- 19, 20-29, 30-39, 40-49, ≥ 50	(17-72)	- SKF: biceps, subscapula, suprailiac, triceps - %BF equations (from BD): 1. Siri	UWW	Spearman correlation, linear regression, SEE	Correlation of SKF vs %BF: - Females: $r = 0.77$ in 17-19 years, 0.83 in 20-29 years, 0.92 in 30-39 years, 0.75 in 40-49 years, 0.85 in ≥ 50 years (all, $p < 0.05$) - Males: $r = 0.74$ in 17-19 years, 0.76 in 20-29 years, 0.70 in 30-39 years, 0.78 in 40-49 years, 0.81 in ≥ 50 years (all, $p < 0.05$) Regression analysis of SKF equations vs %BF: - Females: SEE= 4.70 in 17-	In Females, SKF showed high to very high validity to assess the %BF, compared with UWW, especially in 30-39 years old. In males, SKF showed high validity to assess the %BF, compared with UWW, especially in ≥ 50 years old. SKF equations are especially accurate in 30- 39 years old Females and

						19 years, 5.21 in 20-29 years, 3.67 in 30-39 years, 4.40 in 40-49 years, 4.07 in 50-68 years	40-49 years old males.
						- Males: SEE= 3.89 in 17-19 years, 3.94 in 20-29 years, 4.12 in 30-39 years, 3.62 in 40-49 years, 4.37 in 50-76 years	
Al-Bachir et al. 2016	Males= 213 Healthy %BF= <i>see results</i> Syrian	(18-19)	- SKF: triceps - HC, WC, WHR - %BF equations: 1. %BF= 0.5221 x WC(cm) -18.569 2. %BF= 0.6442 x HC(cm) - 40.071 3. %BF= 39.123 x WHR -10.009 4. %BF= 88.041 x WHtR -17.794	D ₂ O	Linear regression, SEE, Bland-Altman method	%BF values: - WC equation= 13.0 ± 5.6 - WC+SKF equation= 14.5 ± 8.6 - D ₂ O= 21.3 ± 6.4 Regression analysis of equations vs %BF: - WC: R ² = 0.55, SEE= 4.30 (p< 0.01) - HC: R ² = 0.62, SEE= 3.98 (p< 0.01) - WHR: R ² = 0.11, SEE= 6.07 (p< 0.01) - WHtR: R ² = 0.52, SEE= 4.45 (p< 0.01) Agreement between WC equations vs %BF: - MD ($\pm$ 95% LoA)= -5.0 (-12.5 to 1.7)	All the equation but the equation using WHR (low validity) showed high validity to assess the %BF. Equations based on WC, HC and WHtR may accurately predict %BF in Syrian men.

Al-Gindan et al. 2015	Females= 110 Males= 94 Healthy BMI= 25.3 Caucasian, African- American, Hispanic, Asian	(18-86) 44.1 ± 16.4 43.3 ± 15.9	- AC, CC, HC, TC, WC - %BF equations: 1. Females: 0.789 weight (kg) + 0.0786 age (years) – 0.342 height (cm) + 24.5 2. Males: 0.198 weight + 0.478 WC – 0.147 height – 12.8	MRI	Linear regression, SEE, Bland-Altman method	Regression analysis of equations vs %BF: Females: - R²= 0.84, SEE= 3.0 (p< 0.001) Males: - R²= 0.80, SEE= 3.7 (p< 0.001) Agreement: Females: - MD (± 95% LoA)= 0.20 (- 4.9 to 5.2) Males: - MD (± 95% LoA)= -1.3 (- 8.6 to 6.0)	Equation for Females (based on weight, height and age) showed high validity and equation for males (based on weight, height + WC) showed very high validity to assess %BF.
Benito et al. 2019	Females= 48 Males= 36 Healthy, sedentary BMI= 28.2 (overweight) <i>Race-ethnicity not described</i>	(18-50) 37.3 ± 8.1	- SKF: biceps, subscapular, suprailiac, triceps - BIA - %BF (from BD by Siri) 1. SKF by Durnin/ Womersley (DW Sum 4SKF)	DXA	Pearson correlation coefficient (r), Bland-Altman method	Field test correlations vs DXA: Females: - DW Sum 4SKF: r= 0.63 (p< 0.01) - BIA: r= 0.82 (p< 0.01) Males: - DW Sum 4SKF: r= 0.77 (p< 0.01) - BIA: r= 0.81 (p< 0.01) Agreement:	In males, DW Sum 4SKF showed high validity to assess %BF, compared with DXA. In Females, DW Sum 4SKF showed moderate validity to assess %BF, compared with DXA. In both sexes, BIA showed high validity to assess %BF, compared with DXA.

						- SKF vs DXA, MD ($\pm$ 95% LoA)= 0.13 (-9.5 to 9.8) - BIA vs DXA, MD ($\pm$ 95% LoA)= 0.25 (-8.9 to 9.4)	
Bergman et al. 2011	Females= 123 Males= 97 Healthy BMI= 30.0 African-Americans, Mexican-American	(20-50)	- HC, WC - BAI: (HC/Height ^{1.5})-18	DXA	Pearson correlation coefficient (r), LCCC (Pc), Bland-Altman method	Correlation of equations vs %BF: - r= 0.85 (p< 0.001) Concordance (bias correction factor): - Pc= 0.95 (p< 0.001) Agreement: - MD ($\pm$ 95% LoA)= 0.3 (-10.0 to 11.5)	The BAI equation showed high validity to assess %BF.
Bhat et al. 2005	Males= 141 Healthy BMI= 21.9 Indian	(30-50) 39.3 $\pm$ 76.2	- SKF: biceps, subscapular, suprailiac, triceps - BIA - %BF (from BD by Siri) 1. SKF by Durnin/Womersley (DW Sum 4SKF)	D ₂ O	Pearson correlation coefficient (r), Bland-Altman method	Field test correlations vs D ₂ O: - DW Sum 4SKF: r= 0.92 (p< 0.001) - BIA: r= 0.92 (p< 0.001) Agreement: - DW Sum 4SKF vs D ₂ O, MD ($\pm$ 95% LoA)= -1.2 (-8.0 to 5.6) - BIA vs D ₂ O, MD ($\pm$ 95% LoA)= 1.0 (-4.0 to 6.0)	Both measures, DW Sum 4SKF and BIA, showed very high validity to assess %BF, compared with D ₂ O method.
Deurenberg et al. 1991	Females= 708 Males= 521 Healthy	($\geq$ 21)	- BMI - %BF equations (from BD by Siri):	UWW	Pearson correlation coefficient (r), Linear regression, SEE, ANOVA	Correlation of BMI vs %BF: - Females: r= 0.51 in 21-25 years, 0.89 in 26-35 years, 0.81 in 36-45 years, 0.75 in	In Females aged 26-45 years, BMI showed high validity; in Females aged 21-25 years and 56-65 years, moderate validity to

	BMI= 21.0-40.9 Age group 21-25, 26-35, 36-45, 46-55, 56-65 <i>Race-ethnicity not described</i>		- %BF equation: $1.20 \times \text{BMI} + 0.23 \times \text{age} - 10.8 \times \text{sex} - 5.4$			46-55 years, 0.50 in 56-65 years (all, $p < 0.01$) - Males: $r = 0.47$ in 21-25 years, 0.92 in 26-35 years, 0.74 in 36-45 years, 0.80 in 46-55 years, 0.72 in 56-65 years (all, $p < 0.01$) Regression analysis of age- and sex-specific BMI equations vs %BF: - $R^2 = 0.79$, $\text{SEE} = 4.1\%$	assess %BF, compared with UWW. In males ≥ 26 years, BMI showed high to very high validity; in males aged 21-25 years, moderate validity to assess %BF, compared with UWW. The age-sex specific BMI equation showed high validity to assess %BF.
Eston et al. 1995	Females= 9 Males= 47 Healthy %BF= <i>see results</i> Chinese	(27-33) 30.3 ± 2.0 (24-43) 34.9 ± 4.6	- SKF: abdominal, biceps, calf, chest, subscapular, suprailiac, thigh, triceps - %BF equations (from BD by Brozek): 1. Durnin/ Womersley (DW Sum 4SKF) 2. Jackson/Pollock (JP Sum 3SKF), in males 3. Jackson, Pollock and Ward (JPW Sum 3SKF), in Females	UWW	Pearson correlation coefficient (r), SEE, ANOVA	%BF values: - DW Sum 4SKF = 20.1 ± 1.8 Females; 15.9 ± 0.6 males - JP Sum 3SKF = 11.7 ± 0.6 - JPW Sum 3SKF = 18.5 ± 2.0 - UWW = 25.9 ± 1.5 Females; 17.9 ± 0.6 males Correlation of equations vs %BF: - DW Sum 4SKF: $r = 0.89$, $\text{SEE} = 2.5$ ($p < 0.01$) Females; $r = 0.73$, $\text{SEE} = 2.9$ ($p < 0.01$) males - JP/JPW Sum 3SKF: $r = 0.93$, $\text{SEE} = 2.0$ ($p < 0.01$) Females; $r = 0.67$, $\text{SEE} = 3.1$ ($p < 0.01$) males ANOVA between methods:	The DW Sum 4SKF equation showed high validity to assess %BF. However, it seems to overestimate %BF in males. JP Sum 3SKF and JPW Sum 3SKF equations showed very high validity to assess %BF in Females but moderate in males.

						- UWW: mean (SD): 20.0 (1.8), $p < 0.01$ Females; 15.9 (0.61) $p < 0.01$ males - DW Sum 4SKF: mean (SD): 25.9 (1.5), $p < 0.01$ Females; 17.9 (0.58) $p < 0.01$ males - JP/JPW Sum 3SKF: mean (SD): 18.5 (2.0) $p < 0.01$ Females; 11.7 (0.61) $p < 0.01$ males	
Fedewa et al. 2018	Females= 91 Males= 97 Healthy BMI= 24.4 Caucasian, African-American, Hispanic	(18-40) 21.8 $\pm$ 4.8	- WC - BAI - %BF equations: 1. BMI 2. WC 3. BAI	DXA	Pearson correlation coefficient (r), linear regression, SEE	Correlation of equations vs %BF: - BMI: $r = 0.60$ ($p < 0.001$) - WC: $r = 0.72$ ($p < 0.001$) - BAI: $r = 0.64$ ($p < 0.001$) Regression analysis of equations: Females: - BMI: $R^2 = 0.30$, SEE= 5.08 ($p < 0.001$) - WC: $R^2 = 0.41$, SEE= 4.64 ($p < 0.001$) - BAI: $R^2 = 0.32$, SEE= 5.00 ($p < 0.001$) Males: - BMI: $R^2 = 0.37$, SEE= 4.87 ($p < 0.001$)	In Females, all equations showed moderate validity to assess %BF. In males, only WC equation showed high validity to assess %BF. BMI and BAI showed moderate validity to assess %BF. BAI and WC should be used over BMI alone.

Garcia et al. 2005	Females= 71	(20-66)	- ChC, HC, TC, WC	DXA	Pearson correlation coefficient (r), multiple linear regression, RMSE, Bland-Altman method	- WC: R ² = 0.53, SEE= 4.21 (p< 0.001)	Correlation of equations vs %BF:	Both equations showed very high validity to assess %BF.
	Males= 46	46.7 ± 14.1	- SKF: abdominal, biceps, calf, chin, chest, hip, knee, subscapular, thigh, triceps			- BAI: R ² = 0.41, SEE= 4.70 (p< 0.001)		
	Healthy	49.1 ± 11.9	- Bone breadths: ankle, chest, elbow, knee, wrist					
	BMI= 27.8		- Chest depth					
	White (German)		- %BF equations:					
			1. SKF by Durnin/ Womersley (DW Sum 4SKF)					
			2. Females, BFMNew (kg)= -77.538 + {(0.424					
			x HC) + [8.777 x (log chin + log triceps SKF + log subscapular SKF)] + (3.128 x knee breadth)}					
			3. Males, BFMNew (kg)= -40.750 + (0.397 x WC) + [6.568 x (log triceps SKF + log subscapular SKF + log					
	</							

abdominal SKF)]

Hassager et al. 1986	Females= 130 Males= 98 Healthy <i>Fat status and race-ethnicity not described</i>	(20-72)	- Height, weight - SKF: triceps, subscapular - %BF: fat mass/body weight x 100	ADP	Multiple linear regression, SEE, ANOVA	Regression analysis of equations vs %BF: - Height+weight+age: Females: $r = 0.95$, SEE= 3.4% ($p < 0.05$) Males: $r = 0.87$, SEE= 3.7% ($p < 0.05$) - Height+weight+age+Sum 2SKF: Females: $r = 0.80$, SEE= 3.4%8 ($p < 0.05$) Males: $r = 0.83$, SEE= 3.2% ($p < 0.05$)	Both equations showed high to very high validity to assess %BF. Including Sum 2SKF has not advantage in predicting %BF.
Hicks et al. 2000	Females= 150 Healthy BMI= 25.5 American Indian	(18-60) 34.3 ± 10.4	- HC, WC - SKF: abdominal, axillar, chest, subscapular, suprailium, thigh, triceps - %BF equations (from BD by Siri): 1. Jackson, Pollock and Ward (JPW Sum 7SKF) 2. Jackson, Pollock and Ward (JPW Sum 3SKF)	UWW	Linear regression, SEE, Bland-Altman method	Cross-validation of existing equations: - JPW Sum 7SKF: $R^2 = 0.69$, SEE= 0.008 ($p < 0.05$); MD: 4.0 - JPW Sum 3SKF: $R^2 = 0.68$, SEE= 0.008 ($p < 0.05$); MD: - 2.1 - Heyward's NIR: $R^2 = 0.62$, SEE= 0.009 ($p < 0.05$); MD: - 7.7 - Futrex-5000 NIR: $R^2 = 0.66$, SEE= 5.50 ($p < 0.05$); MD: 9.7 Ethnic-specific equations,	Ethnic-specific SKF and NIR (including HC) equations showed high validity to assess %BF. These equations should be chosen over the classical JPW SKF and NIR equations for American Indian female population.

3. Heyward's NIR equations
 4. Futrex-5000 NIR equation
 5. Ethnic-specific equations:
 a) SKF, BD= 1.06198316– 0.00038496(Sum 3SKF)- 0.00020362(age)
 b) NIR, BD= Prediction Equation:
 1.08275966- 0.00090439 (HC) 0.0369861(Sum2ΔOD2)+ 0.0004167(height)+ 0.0000866(FIT index) –0.0001894(age)

validation:
 - SKF, $R^2= 0.67$, SEE= 0.0084 (p< 0.05); MD: 0.6
 - NIR, $R^2= 0.74$, SEE= 0.00798 (p< 0.05); MD: 2.5
 Ethnic-specific equations, cross-validation:
 - SKF, $R^2= 0.77$, SEE= 0.0068 (p< 0.05)
 - NIR, $R^2= 0.72$, SEE= 0.0076 (p< 0.05)

Jackson et al. 2002	Females= 359 Males= 296 Healthy BMI= 27.1 Black, White	(17-65) 35.1 ± 12.6	- BMI - %BF equations (from BD by Siri): 1. Gallagher et al. 2. Deurenberg et al. 3. Jackson/Pollock (JP Sum 3SKF), in males 4. Jackson, Pollock and Ward (JPW Sum 3SKF), in Females	UWW	Multiple regression analysis, SEE, ANOVA	Regression analysis of BMI vs %BF: - Females: $R^2= 0.78$, SEE= 4.63% (p< 0.05) - Males: $R^2= 0.68$, SEE= 4.90% (p< 0.05) Regression equations: - Model I (log transformed BMI and %BF): Females: $R^2= 0.78$, SEE= 4.7; males: $R^2= 0.67$, SEE= 4.9	BMI showed high validity to assess body fat, compared with UWW. BMI is dependent of age and sex, and of race in Females.

				- Model II (+age): Females: $R^2 = 0.80$, SEE= 4.4; males: $R^2 = 0.72$, SEE= 4.6 - Model III (+race): Females: $R^2 = 0.81$, SEE= 4.3; males: $R^2 = 0.72$, SEE= 4.6 Mean ($\pm$ SD) between measured and estimated %BF, by BMI and interaction terms: - Females, interaction BMI <25kg/m² and race: Black, 2.0 $\pm$ 4.1; White -0.8 $\pm$ 4.8 - Males, interaction BMI >30kg/m² and age: <30 years, 3.3 $\pm$ 4.3; 30-45 years, -0.1 $\pm$ 3.9; $\geq$ 45 years, -1.3 $\pm$ 3.7 Cross-validation of published equations: - Females, Black: $r = 0.89$ (all), SEE= 4.2-4.8 ($p < 0.05$) - Females, White: $r = 0.88$ (all), SEE= 4.3-5.4 ($p < 0.05$) - Males, Black: $r = 0.79$-0.80, SEE= 4.5-4.9 ($p < 0.05$) - Males, White: $r = 0.85$-0.86, SEE= 4.7-4.8 ($p < 0.05$)			
Jackson et al. 2008	Females= 706 Males= 423	(17-35) 21.8 $\pm$ 3.1	- SKF: chest, abdominal and thigh (males), triceps, supra- ilium and thigh	DXA	GLM, regression coefficients (r), Bland-Altman	Regression analysis of BF%- GEN vs BF%-DXA: - Females: $r = 0.85$, GLM=	The equation BF%-GEN showed high validity in Females and very high validity in males to assess

	Healthy BMI= 26.8 White, Hispanic, African-American		(females) - %BF equation: 1. BF%-GEN (%BF from Siri's two- component percentage fat equation)	method	0.88 (p< 0.001) - Males: r= 0.93, GLM= 0.78 (p< 0.001) Agreement: - Females, MD (± 95% LoA)= -0.17 (-4.5 to 11.4) - Males, MD (± 95% LoA)= 0.03 (-5.3 to 8.5)	%BF.
Kagawa, Byrne & Hills 2008	Females= 121 Males= 95 Healthy BMI= 27.3 Australian	(20-64) 44.2 ± 10.9 47.1 ± 10.5	- AbC, HC, WC, WHtR - BMI	DXA Spearman's correlation coefficients, multiple linear regression, SEE	Correlation of body composition measures vs %BF: Females: - WHtR (WC): r= 0.78 (p< 0.01) - WHtR (AbC): r= 0.77 (p< 0.01) - BMI: r= 0.87 (p< 0.01) Males: - WHtR (WC): r= 0.74 (p< 0.01) - WHtR (AbC): r= 0.82 (p< 0.01) - BMI: r= 0.78 (p< 0.01) Regression equations: - WHtR (WC): R ² = 0.69, SEE= 5.39 - WHtR (AbC): R ² = 0.72,	WHtR and BMI equations showed high validity to assess %BF, compared with DXA. WHtR calculated with AbC may be better than using WC.

						SEE= 5.12 - BMI: R ² = 0.75, SEE= 4.82	
Kanellakis et al. 2010	Females= 196 Healthy BMI= 30.7 (overweight/ obese) Caucasian	(50-76) 62.6 ± 5.9	- WC, WHR - SKF: biceps, subscapular, suprailiac, triceps - Fat mass (FM) equations: 1. BIA= 38.475 + 0.207×weight-0.092×r esistance/height ² + 0.291×reactance/height 2 2. SKF= -31.913 + 0.333 × WC + 0.840 × BMI + 0.064 × SumSKF	DXA	Multiple linear regression, SEE, Bland-Altman method	Regression analysis of BIA equation vs %BF: - R ² = 0.83, SEE= 2.65 (p< 0.0001) Regression analysis of SKF equation vs %BF: - R ² = 0.80, SEE= 1.62 (p< 0.0001) Agreement: - BIA: MD (± 95% LoA)= 0.26 (-2.86 to 3.38) - SKF: MD (± 95% LoA)= 0.09 (-5.32 to 5.14)	The SKF equation (including WC + BMI) showed high validity and the BIA equation very high validity to assess FM, in postmenopausal Females.
Kanellakis et al. 2012	Females= 277 Healthy BMI= 29.2 Caucasian	(50-75) 62 ± 6.2	- HC, WC - SKF: biceps, triceps - %BF equations (from BD by Siri and Brozek): 1. %BF Martarelli et al. (BMI) 2. %BF Visser et al. (SKF, Siri equation) 3. %BF Visser et al. (BMI, Siri equation) 4. %BF Visser et al.	DXA	Pearson correlation coefficient (r), ICC, SEE, Bland-Altman method	Correlation of equations vs %BF: - Model 1: r= 0.71 (p< 0.001), SEE= 4.86, ICC= 0.823 - Model 2: r= 0.52 (p< 0.001), SEE= 4.81, ICC= 0.651 - Model 3: r= 0.79 (p< 0.001), SEE= 3.42, ICC= 0.883 - Model 4: r= 0.52 (p< 0.001), SEE= 4.77, ICC= 0.636 - Model 5: r= 0.79 (p< 0.001), SEE= 3.38, ICC= 0.878	Model 1, 3 and 5 (all based on BMI) showed high validity to assess %BF, in postmenopausal Females. Models based on SKF and WC showed moderate validity to assess %BF, in postmenopausal Females.

			(SKF, Brozek equation)			- Model 6: r= 0.56 (p< 0.001), SEE= 5.76, ICC= 0.718	
			5. %BF Visser et al. (BMI, Brozek equation)			- Model 7: r= 0.66 (p< 0.001), SEE= 2.60, ICC= 0.789	
			6. %BF Lean et al. (WC)			Agreement: - Model 1: MD (± 95% LoA)= 1.96 (-11.68 to 7.76)	
			7. %BF Lean et al. (BMI)			- Model 2: MD (± 95% LoA)= 8.49 (-1.13 to 18.11)	
						- Model 3: MD (± 95% LoA)= 3.56 (-3.29 to 10.41)	
						- Model 4: MD (± 95% LoA)= 5.77 (-3.77 to 15.34)	
						- Model 5: MD (± 95% LoA)= 1.21 (-5.56 to 7.98)	
						- Model 6: MD (± 95% LoA)= 4.29 (-7.23 to 15.81)	
						- Model 7: MD (± 95% LoA)= 4.43 (-5.96 to 14.82)	
Kanellakis et al. 2017	Females= 408 Males= 234 Healthy BMI= 25.8 Caucasian	(18-80) 41.3 ± 15.3	- CC, FC, HC, NC, WC - SKF: abdominal, biceps, calf, subscapular, suprailiac, thigh, triceps - BMI - %BF equations: 1. %BF= -0.615-10.948 x sex + 0.321 x WC x 0.502 x HC x -0.39 x	DXA	Multiple linear regression, ICC, LCCC (Pc), Bland-Altman method	Regression analysis of models vs %BF: - Model 1: R ² = 0.85, SEE= 0.185 (p< 0.0001), ICC= 0.95, Pc= 0.91 - Model 2: R ² = 0.92, SEE= 0.208 (p< 0.0001), ICC= 0.97, Pc= 0.95 Agreement: - Model 1: MD (± 95%	Model 1 (based on WC + HC + FC) showed high validity and Model 2 (based on WC + HC + FC + SKF) very high validity to assess %BF, compared with DXA.

			$\text{FC} -19.768 \times \text{height(m)}$ $2. \%BF = -27.787 - 5.515 \times \text{sex} -8.419 \times \text{height} +0.145 \times \text{WC} +0.270 \times \text{HC} +7.509 \times \text{log of thigh SKF} +20.090 \times \text{log of sum of SKF (bicep + tricep + suprailiac + subscapular)} -0.445 \times \text{FC}$			$\text{LoA} = 0.052 (-8.10 \text{ to } 8.04)$ $\text{- Model 2: MD (} \pm 95\% \text{ LoA)} = 0.148 (-6.20 \text{ to } 5.90)$	
Kholi et al. 2009	Females= 103 Males= 105 Healthy BMI equally distributed between normal weight, overweight and obesity South Asians	(30-65) 45.0 ± 0.8	- FC, HC, WC - SKF: biceps, calf, subscapular, suprailiac, triceps - %BF equations: 1. Females, NEW= - 25.26 -17.36 x (height) + 0.44 x (mass) + 0.34 x (HC) + 8.87 x (log-tricepsSKF) + 0.14 x (age) 2. Males, NEW= - 15.57 + 0.30 x (weight) + 0.23 x (WC) – 3.38 x (FC) + 8.83 x (log-tricepsSKF) + 6.70 x (log-bicepsSKF) + 0.13 x (age) 3. Durnin/ Womersley (DW Sum 4SKF, BD by Siri)	DXA	Multiple linear regression (r), SEE, PE, RSME, Bland-Altman method	Correlation of new equations vs %BF: - Females: r= 0.97, SEE= 2.30 (p< 0.001), PE= 2.30, RSME= 3.27 - Males: r= 0.94, SEE= 2.80 (p< 0.001), PE= 2.80, RSME= 4.25 Agreement: - New, Females, MD (± 95% LoA)= -0.26 (-7.42 to 7.36) - New, males, MD (± 95% LoA)= -0.36 (6.88 to 6.15) - DW Sum 4SKF, Females, MD (± 95% LoA)= -5.13 (-14.96 to 4.69) - DW Sum 4SKF, males, MD (± 95% LoA)= -4.05 (-11.75 to 3.65)	Both ethnic-specific equations (based on circumferences + SKF) showed very high validity to assess %BF. These equations should be chosen over the classical DW Sum 4SKF equations for South Asian population.

Lam et al. 2013	Females= 52 Males= 53 Healthy BMI= 28.1 Chinese	(>21) 39.3 ± 11.6	- HC, WC - BMI - %BF equation: BAI= (hip/heightx1.5) - 18	DXA	Pearson correlation coefficient (r), multiple linear regression, Bland-Altman method	Correlation of BMI vs %BF: - Females: r= 0.87 (p< 0.01) - Males: r= 0.81 (p< 0.01) Correlations of BAI vs DXA: - Females: r= 0.82 (p< 0.01) - Males: r= 0.74 (p< 0.01) Regression models of BMI vs DXA-derived adiposity: - Adjusted for gender: R²= 0.76 (p< 0.05) - Adjusted for age and gender: R²= 0.75 (p< 0.05) Regression models of BAI vs DXA-derived adiposity: - Adjusted for gender: R²= 0.69 - Adjusted for age and gender: R²= 0.69 Agreement: - BAI: MD (± 95% LoA)= 5.77 (-14.15 to 2.61)	Both, the BAI and BMI equations showed high validity to assess %BF, compared with DXA. However, the BMI equation showed higher validity than the BAI equation to assess %BF.
Leahy et al. 2012	Females= 618 Males= 518 Healthy BMI= 25.1 Caucasian	(18-81) 39.5 ± 17.0 31.7 ± 14.6	- AbC, AC, CC, FC, HC, TC - SKF: abdominal, biceps, chest, forearm, iliac crest, medial calf, midaxilla, subscapular, supraspinal, thigh,	DXA	Spearman's Rank correlations, SEE, Bland-Altman method	Correlation of equations vs %BF: - New, Females: rs= 0.92, SEE= 3.0 (p< 0.05) - New, males: rs= 0.95, SEE= 2.5 (p< 0.05)	These new equations (based on SKF and adding also AbC in Females) showed very high validity to assess %BF, compared with DXA. The DW Sum 4SKF

triceps
- %BF equations:
1. Females= (age) x 0.1
+ (log AbC x 41.8) +
(log bicepsSKF x 11.9)
+ (log subscapularSKF
x 5.6) + (log
medialcalfSKF x 8.3) –
79.0
2. Males= (age) x 0.3 +
(log abdominalSKF x
5.6) - (log
subscapularSKF x 6.5)
+ (log medialcalfSKF x
3.9) + 6.6
3. Durnin/ Womersley
(DW Sum 4SKF, BD
by Siri)

- DW Sum 4SKF, Females: rs
= 0.86, SEE= 4.1 (p< 0.05)
- DW Sum 4SKF, males: rs =
0.92, SEE= 3.4 (p< 0.05)
Agreement:
- New, Females, MD (± 95%
LoA)= 0.1 (-5.7 to 5.9)
- New, males, MD (± 95%
LoA)= 0.0 (-4.8 to 4.9)
- DW Sum 4SKF, Females,
MD (± 95% LoA)= 0.8 (-7.1
to 8.5)
- DW Sum 4SKF, males, MD
(± 95% LoA)= 1.0 (-5.3 to
7.4)

equation showed high to
very high validity, but
appear to become less
accurate as %BF increases.
Therefore, the new
equations should be
chosen.

Lean, Han &
Deurenberg 1996

Females= 84
Males= 63
Healthy
BMI= 24.9
White

(17-65)
40 ± 13.5

- AC, HC, TC, WC,
WHR
- Arm span, leg length
- SKF: biceps,
subscapular, suprailiac,
triceps
- BMI
- %BF:
12 equations
combining
anthropometrics + age
+ sex (including DW

UWW

Multiple linear
regression, SEE,
Bland-Altman
method

Correlation of body
composition measures vs %BF
- Females: r= -0.54 to -0.79
(all, p< 0.001)
- Males= r= -0.32 to -0.88 (all,
p< 0.001)
Best regression equations:
- Best, Females: BMI +
triceps-SKF + age: R²= 0.79,
SEE= 4.0
- Best, males: WC + triceps-
SKF + age: R²= 0.86, SEE=

Both new equations (based
on BMI + triceps SKF and
age in Females; WC +
triceps SKF and age in
males) showed very high
validity to assess %BF.
The classical DW Sum
4SKF, JPW Sum 7SKF,
and JP Sum 7SKF showed
also very high validity to
assess %BF.
The new equations require
less time/measures to

			Sum 4SKF, BD by Siri; JPW Sum 7SKF; JP Sum 7SKF)			3.2 Other published equations: - DW Sum 4SKF, Females: R ² = 0.81, SEE= 4.6 - DW Sum 4SKF, males: R ² = 0.81 SEE= 3.8 - JPW Sum 7SKF, Females: R ² = 0.75, SEE= 3.6 - JP Sum 7SKF, males: R ² = 0.84, SEE= 3.2	assess %BF.
Lee et al. 2021	Females= 10009 Males= 7599 Healthy BF (kg)=19.0 Females; 15.4 males Korean	(>19) 49.6 ± 16.0 49.1 ± 15.6	- WC - BMI - %BF equations: 1. age, height, weight, and WC 2. age, height, weight, WC and serum creatinine 3. age, height, weight, WC, serum creatinine level and physical activity 4. age, height, weight, WC, serum creatinine level, physical activity, smoking habit and alcohol use	DXA	Linear regression, SEE, ICC, Bland- Altman method	Regression analysis of equation 4 vs %BF: Females: - R ² = 0.74, SEE= 2.69, ICC= 0.95 (p< 0.001) Males: - R ² = 0.83, SEE= 2.23, ICC= 0.91 (p< 0.001) Agreement: - Females, MD (± 95% LoA)= -0.00 (-4.42 to 4.42) - Males, MD (± 95% LoA)= 0.01 (-5.44 to 5.46)	Equation 4 (based on age, height, weight + WC and other variables) showed very high validity in males and high validity in Females to assess %BF.
Macias et al. 2007	Females= 82	(21-48)	- HC, WC, WHR	ADP	Multiple linear regression, SEE,	New equation: R ² = 0.97, PE= 2.96 (p< 0.05)	This BIA equation showed very high validity to assess

	Males= 73 Healthy BMI= 25.9 Mexican	34.3 ± 7.6 33.9 ± 7.3	- BIA - %BF equation: 0.7374 * (Ht2 /R) + 0.1763 * (body weight) - 0.1773 * (Age) + 0.1198 * (Xc) - 2.4658		PE, Bland-Altman method	Mean ± SD values BIA vs %BF: - Females, BIA= 36.1 ± 6.1; ADP= 36.2 ± 6.5 - Males, BIA= 25.8 ± 5.9; ADP= 25.9 ± 6.9 Agreement: - MD (± 95% LoA)= -0.87 (- 6.56 to 4.82)	%BF.
Manios et al. 2012	Females= 276 Healthy BMI= 20.5-42.0 White	(47-79) 62 ± 6.2	- WC, GIC - SKF: biceps, suprailiac, subscapular, triceps - %BF equation: FM= 0.069 x bicepsSKF +0.553 x weight -14.655 x height + 0.218 x GIC - 9.830	DXA	Multiple linear regression, ICC, SEE, Bland-Altman method	Regression analysis of equation vs %BF: - R ² = 0.93, SEE= 2.14, ICC= 0.983 (p< 0.0001) Agreement: - MD (± 95% LoA)= 0.16 (- 4.05 to 4.36)	This equation (based on weight, height + GIC + biceps SKF) showed very high validity to assess %BF, in postmenopausal Females.
Nickerson et al. 2016	Females= 63 Males= 67 Healthy %BF= <i>see results</i> Caucasian	(18-37) 22 ± 5	- SKF: abdominal, chest, mid-axilla, suprailium, subscapular, thigh, triceps - BMI-%BF equations: 1. Jackson et al. (BMIJA) 2. Deurenberg et al. (BMIDE)	4C model: - UWW - DXA - BIA	Pearson correlation coefficient (r), CE, SEE, TE, Bland- Altman method	%BF values: - BMIJA=23.9 ± 6.9 - BMIDE=23.3 ± 6.2 - BMIGA=23.8 ± 7.5 - BMIZA=21.4 ± 8.9 - BMIWO=24.8 ± 5.9 - JP and JPW Sum 7SKF =16.8 ± 7.1	JP and JPW Sum 7SKF equations showed high validity to assess %BF and it is recommended over BMI-based equations (which showed moderate validity).

3. Gallagher et al.
(BMIGA)

4. Zanovec at al.
(BMIZA)

5. Womersley and
Durnin (BMIWO)

- SKF-%BF equation
(from BD by Brozek):

6. Jackson/Pollock (JP
Sum 7SKF), in males;
Jackson, Pollock and
Ward (JPW Sum
7SKF), in Females

$$- 4C=21.6 \pm 7.8$$

Equation correlations vs %BF:

- BMIJA: $r= 0.56$, $SEE= 6.4$,
 $TE= 7.3$, $CE= 2.4$ ($p< 0.001$)

- BMIDE: $r= 0.67$, $SEE= 5.8$,
 $TE= 6.1$, $CE= 1.8$ ($p< 0.001$)

- BMIGA: $r= 0.64$, $SEE= 5.9$,
 $TE= 6.8$, $CE= 2.2$ ($p< 0.001$)

- BMIZA: $r= 0.73$, $SEE= 5.3$,
 $TE= 6.2$, $CE= -0.2$ ($p=0.745$)

- BMIWO: $r= 0.66$, $SEE= 4.4$,
 $TE= 6.7$, $CE= 3.2$ ($p< 0.001$)

- JP and JPW Sum 7SKF: $r=$
 0.88 , $SEE= 3.4$, $TE= 6.0$, $CE=$
 -4.8 ($p< 0.001$)

Agreement:

- BMIJA: MD ($\pm$ 95%
LoA)= 1.5 (-12.0 to 17.0)

- BMIDE: MD ($\pm$ 95%
LoA)= 1.5 (-10.0 to 12.0)

- BMIGA: MD ($\pm$ 95%
LoA)= 1.5 (-10.0 to 17.0)

- BMIZA: MD ($\pm$ 95%
LoA)= 0.0 (-12.0 to 11.0)

- BMIWO: MD ($\pm$ 95%
LoA)= 1.5 (-8.0 to 15.0)

- JP and JPW Sum 7SKF: MD
($\pm$ 95% LoA)= -5.0 (-11.0 to
2.0)

O'Connor et al. 2010	Females= 705 Males= 428 Healthy BMI= 26.6 Non-Hispanic White (NHW), Hispanic, African-American (AA)	(17-35) 21.4 ± 3.1	- SKF: 1. iliac crest, thigh, triceps, for Females 2. abdominal, chest, thigh, for males -BMI - SKF by: Jackson, Pollock and Ward (JPW Sum 3SKF), in Females; Jackson/Pollock (JP Sum 3SKF), in males - %BF equations: 1. Model Sum 3SKF 2. Model Sum 3SKF+BMI	DXA	Regression coefficients estimates, SE, Bland-Altman method	Regression analysis of equations vs %BF: - Sum 3SKF: Females, EST= 0.169, SE= 4.01% (95% CI= 3.62%-4.36%) (p< 0.001); males, EST= 0.195, SE= 3.07% (95% CI= 2.77%-3.40%) (p< 0.001) - Sum 3SKF+BMI: Females, EST= -0.0007, SE= 3.64% (95% CI= 3.41%-3.89%) (p< 0.001); males, EST= -0.0005, SE= 3.12% (95% CI= 2.82%-3.44%) (p< 0.001) Agreement: - Females, MD (± 95% LoA)= 1.2 (-1.3 to 1.1) - Males, MD (± 95% LoA)= 1.4 (-1.5 to 1.2)	These ethnic-specific generalized equations (based on Sum 3SKF + BMI) are valid and provide accurate prediction to assess %BF.
Pascale et al. 1956	Males= 88 Healthy BD (g/cc)=1.1 Caucasian	(17-25) 22.1 ± 1.95	- SKF: abdominal, chest, subscapular, triceps - %BF equations (from BD): 1. Brozek	UWW	Pearson correlation coefficient (r), multiple linear regression, SEE	Correlation of SKF vs %BF (from BD): - Chest: r= -0.79 to -0.83 (p< 0.05) - Abdominal: r= -0.72 to -0.77 (p< 0.05) - Triceps: r= -0.77 (p< 0.05) - Subscapular: r= -0.74 (p< 0.05)	All SKF showed high validity, to assess %BF, compared with UWW. Equation including chest and triceps SKF showed high validity, to assess %BF. This equation showed a slightly improvement, compared with those of Brozek (although the SKF included were not the

						Regression equation: - $R^2 = 0.85$, $SEE = 0.006$ Brozek equation: - $R^2 = 0.87$, $SEE = 0.007$	same).
Ramirez-Zea et al. 2006	Females= 123 Males= 114 Healthy BMI= 23.5 Guatemalan	(18-56) 26.5 ± 6	- AbC, AC, CC, HC, TC, WC - SKF: abdominal, subscapular, suprailiac, triceps - %BF equations (from BD by Siri): 1. Females: 19.420 - (weight x 0.385) - (height x 0.215) - (AbC x 0.265) 2. Males: -48.472 -(weight x 0.257) - (AbC x 0.989)	UWW	Multiple linear regression, RMSE, PE	Regression equations (model- building): - Females: $R^2 = 0.78$, RMSE= 3.77 - Males: $R^2 = 0.74$, RMSE= 3.46 Regression equations (validation): - Females: $R^2 = 0.79$, PE= 3.53 - Males: $R^2 = 0.77$, PE= 3.70	Equations including AbC showed high validity, to assess %BF in Guatemalan population.
Ramos-Jiménez et al. 2018	Females= 142 Males= 142 Healthy BMI= 24.9 Mexican	(18-35) 21.5 ± 3.9 22.1 ± 3.4	- WC - SKF: abdominal, biceps, iliac crest, subscapular, medial calf, midaxilla, supraspinal, thigh, triceps - %BF equations (from	DXA	Multiple linear regression, SEE, Bland-Altman method	Regression analysis of equations vs %BF: - Females: $R^2 = 0.98$, $SEE = 0.04$ ($p < 0.001$) - Males: $R^2 = 0.78$, $SEE = 0.03$ ($p < 0.001$) Agreement:	Equations (based on WC + SKF) showed high to very high validity to assess %BF.

			BD by Siri): 1. Females= $0.20 \times \text{WC} + 0.25 \times \text{supraspina SKF} + 0.21 \times \text{thigh SKF} + 0.25 \times \text{midaxila SKF} - 0.13 \times \text{biceps SKF}$ 2. Males= $26.02 + 0.26 \times \text{WC} + 0.17 \times \text{abdominal SKF} + 0.31 \times \text{thigh} + 0.13 \times \text{subscapular SKF} - 23.59 \times \text{height}$			- Females, MD ($\pm$ 95% LoA)= 0.1 (-8.0 to 10.5) - Males, MD ($\pm$ 95% LoA)= 0.1 (-9.8 to 10.3)	
Rush et al. 1997	Females= 82 Healthy BMI= 30 White (New Zealand European), Polynesian	(18-27) 22 $\pm$ 2	- AbC, HC, WC, WHR - SKF: biceps, subscapular, suprailiac, triceps - BMI - %BF equations (from BD by Brozek): 1. SKF by Durnin/Womersley (DW Sum 4SKF) 2. White= $0.313 \text{ AbC} + 0.246 \text{ HC} - 0.313 \text{ height} + 37.62$ 3. Polynesians= $0.188 \text{ AbC} + 0.265 \text{ HC} - 7.916$	Isotope dilution	Pearson correlation coefficient (r), multiple linear regression, SEE, Bland-Altman method	Correlation of DW Sum 4SKF vs %BF: - Whites: $r = -0.47$ ($p = 0.002$) - Polynesian: $r = -0.39$ ($p = 0.01$) Regression analysis of body composition measures vs %BF: - Circumferences: $R^2 = 0.87$, SEE= 3.65%, for Whites; $R^2 = 0.83$, SEE= 3.16%, for Polynesian ($p < 0.05$) - DW Sum 4SKF: $R^2 = 0.84$, SEE= 3.98%, for Whites; $R^2 = 0.72$, SEE= 4.01%, for Polynesian ($p < 0.05$) Best equations:	DW Sum 4SKF showed high validity, to assess %BF, compared with Isotope dilution, in Polynesian. Both, new race-specific (based on AbC + HC) and DW Sum 4SKF equations showed very high validity, to assess %BF, compared with Isotope dilution, in Whites and Polynesian.

						- White: $R^2 = 0.86$, SEE = 3.83% ($p < 0.05$) - Polynesian: $R^2 = 0.81$, SEE = 3.40% ($p < 0.05$)	
Segheto et al. 2016	Females= 395 Males= 311 Healthy BMI= 25.3 Brazilian	(20-59) 36.0 ± 12.8	- WC, WHR - SKF: abdomen, iliac crest, triceps, in Females; chest, subscapular, triceps, in males - BAI - BIA - BMI - % BF (from BD by Siri): 1. Jackson, Pollock & Ward, in Females (JPW Sum 3SKF) 2. Jackson/ Pollock, in males (JP Sum 3SKF)	DXA	Pearson correlation coefficient (r), Bland-Altman method	Field test correlations vs DXA: Females: - WC: $r = 0.82$ ($p < 0.001$) - WHR: $r = 0.55$ ($p < 0.001$) - JPW Sum 3SKF: $r = 0.86$ ($p < 0.001$) - BAI: $r = 0.78$ ($p < 0.001$) - BIA: $r = 0.83$ ($p < 0.001$) - BMI: $r = 0.82$ ($p < 0.001$) Males: - WC: $r = 0.83$ ($p < 0.001$) - WHR: $r = 0.65$ ($p < 0.001$) - JP Sum 3SKF: $r = 0.89$ ($p < 0.001$) - BAI: $r = 0.72$ ($p < 0.001$) - BIA: $r = 0.75$ ($p < 0.001$) - BMI: $r = 0.76$ ($p < 0.001$) Agreement: - Females: BAI vs DXA, MD (± 95% LoA) = 4.1 (-12.4 to 11.6)	All the measurements (except for WHR) showed high validity to assess %BF, compared with DXA.

Shafer et al. 2010							- Males: BAI vs DXA, MD ($\pm$ 95% LoA)= -1.0 (-10.4 to 12.5)
	Females= 68	(≥ 18)	- SKF: biceps, subscapular, suprailiac, triceps	ADP	Linear regression, RMSE, ANOVA, Bland-Altman method	Regression analysis:	DW Sum 4SKF showed very low to low validity to assess %BF, compared with ADP.
	Males= 63	35.6 $\pm$ 13.3				- Group with normal weight: R ² = 0.07 (p< 0.06), RMSE= 0.01	
	Healthy	44.9 $\pm$ 15.2	- %BF (from BD by Siri)			- Group with overweight: R ² = 0.15 (p< 0.009), RMSE= 0.00	
	BMI= 22.6 normal weight, 27.3 overweight, 33.7 obese	47.2 $\pm$ 12.4	1. Durnin/ Womersley (DW Sum 4SKF)			- Group with obesity: R ² = 0.00 (p= 0.93), RSME= 0.01	
	Race-ethnicity not described					ANOVA:	
						- Group with normal weight: mean $\pm$ SD= -2.57 $\pm$ 5.11 (p< 0.05)	
						- Group with overweight: mean $\pm$ SD= 0.85 $\pm$ 4.28 (p< 0.05)	
						- Group with obesity: mean $\pm$ SD= 6.72 $\pm$ 5.39 (p< 0.05)	
						Agreement:	
						- Group with normal weight, MD ($\pm$ 95% LoA)= 0.00 (-0.01 to 0.02)	
						- Group with overweight, MD ($\pm$ 95% LoA)= 0.00 (-0.02 to 0.02)	
						- Group with obesity, MD ($\pm$ 95% LoA)= -0.01 (-0.03 to 0.01)	

						0.00)	
Skoufas et al. 2018	Females= 437 Males= 272 Healthy BMI= 25.7 Caucasian	(18-80) 40.4 ± 15.3	- AbC, FC, NC, WC - SKF: abdominal, biceps, subscapular, suprailiac, triceps 1. Females equation: %Abdominal FM=-121.368 - (17.322 x Height(m)) + (76,452 x logAbC(cm)) + (18.341 x logSuprailiac SKF(mm)) + (8802 x logAbdominal SKF(mm)) 2. Males equation: %Abdominal FM=-174.117 - (15.247 x Height (m)) + (104.503 x logAbC(cm)) + (9907 x logSuprailiac SKF(mm)) + (7971 x logAbdominal SKF(mm))	DXA	Correlation coefficient, ICC, Bland-Altman method	Regression analysis of abdominal fat mass equations vs %BF: - Females: R²= 0.89 (p< 0.001), ICC= 0.940 - Males: R²= 0.94 (p< 0.001), ICC= 0.957 Agreement: - Females, MD (± 95% LoA)= 0.20 (-9.15 to 9.56) - Males, MD (± 95% LoA)= 0.41 (-8.37 to 9.19)	Equations (based on AbC + SKF) showed high to very high validity, to assess %BF.
Slaughter et al. 1988	Females= 32 Males= 36 Healthy %BF= 15.8 Females; 27.0 males	22.6 ± 2.7	- SKF: abdominal, anterior suprailiac, biceps, subscapula, medial calf, midaxillary, mid-thigh, suprailiac, triceps - %BF equations (from	D ₂ O, photon absorptiometry, UWW	Regression analysis, SEE	Regression analysis of SKF vs %BF: - Sum SKF triceps+calf: R²= 0.82, SEE= 3.6 (p< 0.05) - Sum SKF triceps +subscapular: R²= 0.82, SEE=	These equations, based on Sum 2SKF showed very high validity to assess %BF.

	Black and White		BD): 1. Siri			3.5 (p< 0.05) Prediction equations: - Sum SKF triceps+calf: B= 0.61, SEE= 3.8 in Females; B= 0.73, SEE= 3.8 in males (p< 0.05) - Sum SKF triceps +subscapular: B= 1.33, SEE= 3.9 in Females; B= 1.21, SEE= 3.6 in males (p< 0.05)	
Stout et al. 1994	Males= 57 Healthy %BF= <i>see results</i> Caucasian	(18-33) 22 ± 3	- SKF: abdominal, chest, thigh - %BF equation (from BD, by Brozek) 1. Jackson/Pollock (JP Sum 3SKF) 2. BIA	UWW	Correlation coefficient (r), CE, TE, SEE	%BD values: - JP Sum 3SKF=12.6 ± 5.2 - BIA=17.8 ± 4.6 - UWW=15.1 ± 6.2 Correlation of equations vs %BF: - JP Sum 3SKF, r= 0.90 (p=0.05), CE= -2.5, TE= 3.6, SEE= 2.7 - BIA, r= 0.74 (p=0.05), CE= 2.8, TE= 5.0, SEE= 4.2	The JP Sum 3SKF equation and BIA showed very high validity to assess %BF.
Sun et al. 2005	Females= 491 Males= 100 Healthy BMI= 26.3 Canadian	(19-60) 42.1 ± 10.3	- WHtR - BMI - BIA	DXA	Pearson correlation coefficient (r), ANOVA, Bland-Altman method	Correlation of BIA vs %BF: - Whole group: r= 0.88 (p< 0.05) - Females: r= 0.85 (p< 0.05) - Males: r= 0.78 (p< 0.05) Agreement BIA vs %BF:	BIA showed high validity to assess %BF, compared with DXA. BIA overestimates %BF in lean adults and underestimates %BF in obese adults.

	obese, obese Anglo-Celtic Australian		2. Lukaski 1986 3. Segal et al. 1988 - %BF: 1. SKF by Durnin/ Womersley (DW Sum 4SKF)			ANOVA: Females: - DW Sum 4SKF= 37.4 vs 34.2% (p< 0.0001) - BIA= 35.1 vs 37.4% (p< 0.01) Males: - DW Sum 4SKF= 24.6 vs 24.7% (p>0.05) - BIA= 28.2 vs 24.6 % (p< 0.001) Agreement: Females: - DW Sum 4SKF vs DXA, MD (± 95% LoA)= 3.0 (-6.0 to 12.0) - BIA vs DXA, MD (± 95% LoA)= 3.9 (-5.0 to 9.6) Males: - DW Sum 4SKF vs DXA, MD (± 95% LoA)= -0.6 (-9.8 to 7.3) - BIA vs DXA, MD (± 95% LoA)= -0.6 (-15.0 to 14.8)	low validity to assess %BF, compared with DXA.
Yao et al. 2002	Females= 38 Males= 33 Healthy	(35-49) 42.8 ± 0.6 43.1 ± 0.7	- SKF: abdominal, biceps, mid-thigh, subscapular, suprailiac, triceps	D ₂ O	Pearson correlation coefficient (r), regression analysis, SEE, Bland-Altman	Regression analysis of equations vs %BF: - DW Sum 4SKF, Siri: R ² = 0.88, SEE= 2.71 (p< 0.001)	All equations showed very high validity to assess %BF.

	BMI= 24.9 Chinese		- BMI - %BF equations: 1. Durnin/ Womersley (DW Sum 4SKF), BD by Siri and by Brozek 2. Wang et al.		method	- DW Sum 4SKF, Brozek et al.: $R^2= 0.88$, $SEE= 2.71$ ($p< 0.001$) - Wang et al.: $R^2= 0.87$, $SEE= 2.88$ ($p< 0.001$) Agreement: - DW Sum 4SKF, Siri: MD ($\pm$ 95% LoA)= 0.23 (-4.4 to 6.6) - DW Sum 4SKF, Brozek et al.: MD ($\pm$ 95% LoA)= 0.00 (-5.5 to 5.3) - Wang et al.: MD ($\pm$ 95% LoA)= 0.70 (-8.1 to 7.9)	
Zanovec et al. 2009	Females= 145 Males= 133 Healthy BMI= 17-37 Black, White	(18-24) 20.4 $\pm$ 1.5	- BMI - Physical activity levels	DXA	Pearson correlation coefficients (r), multiple linear regression analysis, RMSE, Bland-Altman method	Correlation of BMI vs %BF: - Females, White: $r= 0.77$ ($p< 0.001$) - Females, Black: $r= 0.90$ ($p< 0.001$) - Males, White: $r= 0.71$ ($p< 0.001$) - Males, Black: $r= 0.87$ ($p< 0.001$) Regression analysis: - Sex+BMI+race: $R^2= 0.81$, RMSE= 4.07 - Sex+BMI+race+physical activity: $R^2= 0.83$, RMSE= 3.87 Agreement (sex+BMI+race): MD ($\pm$ 95% LoA)= 0.0 (-	BMI equations showed very high validity to assess %BF. Including physical activity implies greater accuracy.

AbC, Abdominal Circumference; AC, Arm Circumference; AD, Arm Diameter; ADP, Air-Displacement Plethysmography; AkC, Ankle Circumference; AkD, Ankle Diameter; BAI, Body Adiposity Index; BD, Body Density; BIA, Bioelectrical Impedance Analysis; BMI, Body Mass Index; C, Circumference; CC, Calf Circumference; ChC, Chest Circumference; ChD, Chest Diameter; D, Diameter; DD, Deltoid Diameter; DXA, Dual energy X-ray Absorptiometry; D₂O, Deuterium Oxide; ED, Elbow Diameter; FC, Forearm Circumference; FFM, Fat Free Mass; GlC, Gluteal Circumference; HC, Hip Circumference; HdC, Head Circumference; ³H₂O, Tritium Water; IC, Iliac Circumference; KC, Knee Circumference; KD, Knee Diameter; K⁴⁰, Potassium-40; MRI, Magnetic Resonance Imaging; NC, Neck Circumference; ShC, Shoulder Circumference; SKF, Skinfolks; TC, Thigh Circumference; UWW, Under Water Weighing; WC, Waist Circumference; WHR, Waist to Hip Ratio; WHtR, Waist to Height Ratio; WrC, Wrist Circumference; WrD, Wrist Diameter; %BF, Body Fat percentage.

ANOVA, indicates analysis of variance; CE, Constant Error; CI, Coefficient Intervals; CV_{RSME}, Coefficient of Variation of the Root-Mean-Square Error; EST, Regression Coefficients Estimates; GLM, General Linear Models; ICC, Intra-class Correlation Coefficient; LCCC, Lin's Concordance Coefficient; LoA, Limits of Agreement; ME, Mean Error; MD, Mean Difference; MSD, Mean Signed Difference; PE, Pure Error; RMSE, Root Mean Square Error; RRSE, Root Relative Square Error; SD, Standard Deviation; SE, Standard Error; SEE, Standard Error Estimate; TE, Total Error.

Table 29. Levels of evidence of estimation equations studies in adults.

EQUATION/ INDEX	WHOLE GROUP	FEMALES	MALES	19-33 YRS	34-49 YRS	50-64 YRS	GROUP WITH NORMAL WEIGHT	GROUP WITH OVERWEIGHT	GROUP WITH OBESITY	Caucasian/ White	African- American/ Black	Hispanic	Asian
D/W	●	●	●	●	●	□	●	●	○	●	●	●	●
JP	N/A	N/A	●	●	●	■	N/A	■	N/A	◐	N/A	■	■
JPW	N/A	●	N/A	●	●	N/A	●	◐	N/A	◐	■	●	●
AbC+SKF	●	●	■	◐	●	N/A	■	◐	N/A	◐	N/A	■	N/A
WC	●	●	●	●	●	■	●	●	Δ	●	●	●	■
WC+SKF	●	◐	●	■	●	■	●	●	N/A	●	■	◐	◐
WC+SKF+BMI	●	◐	◐	■	◐	◐	◐	■	■	◐	N/A	N/A	■
CIRCUM	●	●	◐	◐	◐	N/A	■	●	■	●	◐	◐	N/A
C+SKF	●	●	●	N/A	●	◐	●	●	◐	●	■	■	N/A
C+D+SKF	●	◐	◐	◐	◐	N/A	N/A	N/A	N/A	■	N/A	N/A	■
SKF	●	●	●	●	●	●	◐	●	□	●	◐	◐	◐
SKF+BMI	●	●	◐	■	■	■	◐	◐	■	●	■	■	■
BAI	●	●	●	N/A	●	N/A	■	●	□	●	◐	●	■
BIA	●	●	●	●	●	●	●	●	●	●	N/A	●	●
BMI	●	●	●	●	●	●	●	●	●	●	●	●	●
WHtR	●	●	●	◐	●	N/A	□	□	□	□	■	●	N/A
WHR	◐	◐	□	■	◐	N/A	■	■	■	□	N/A	■	N/A
AbC+NC	◐	■	◐	■	■	N/A	■	◐	N/A	■	■	■	N/A

● Indicates strong evidence of validity; ◐ moderate evidence of validity; ○ strong evidence of low/null validity; ■ limited evidence of validity*;

□ inconclusive evidence of validity**.

*Limited evidence of validity derived from one single study; **Inconclusive evidence of validity derived from inconsistent results.

D/W, Durnin/Womersley equation; JP, Jackson/Pollock equation; JPW, Jackson, Pollock and Ward equation; AbC+NC, equations combining abdominal circumference + neck circumference; AbC+SKF, equations combining abdominal circumference + skinfolds; WC, equations based on waist circumference; WC+SKF, equations combining waist circumference + skinfolds; WC+SKF+BMI, equations combining waist circumference + skinfolds + body mass index; Circum, equations combining various circumferences; C+SKF, equations combining circumferences + skinfolds; C+D+SKF, equations combining circumferences + diameters + skinfolds; SKF, equations implying sum of skinfolds;

SKF+BMI, equations combining skinfolds + body mass index; BAI, body adiposity index equations; BIA, bioelectrical impedance analysis equations; BMI, body mass index equations; WHR, waist to hip ratio equations; WHtR, waist to height ratio equations; N/A, not applicable.

Most common methods for body composition assessment

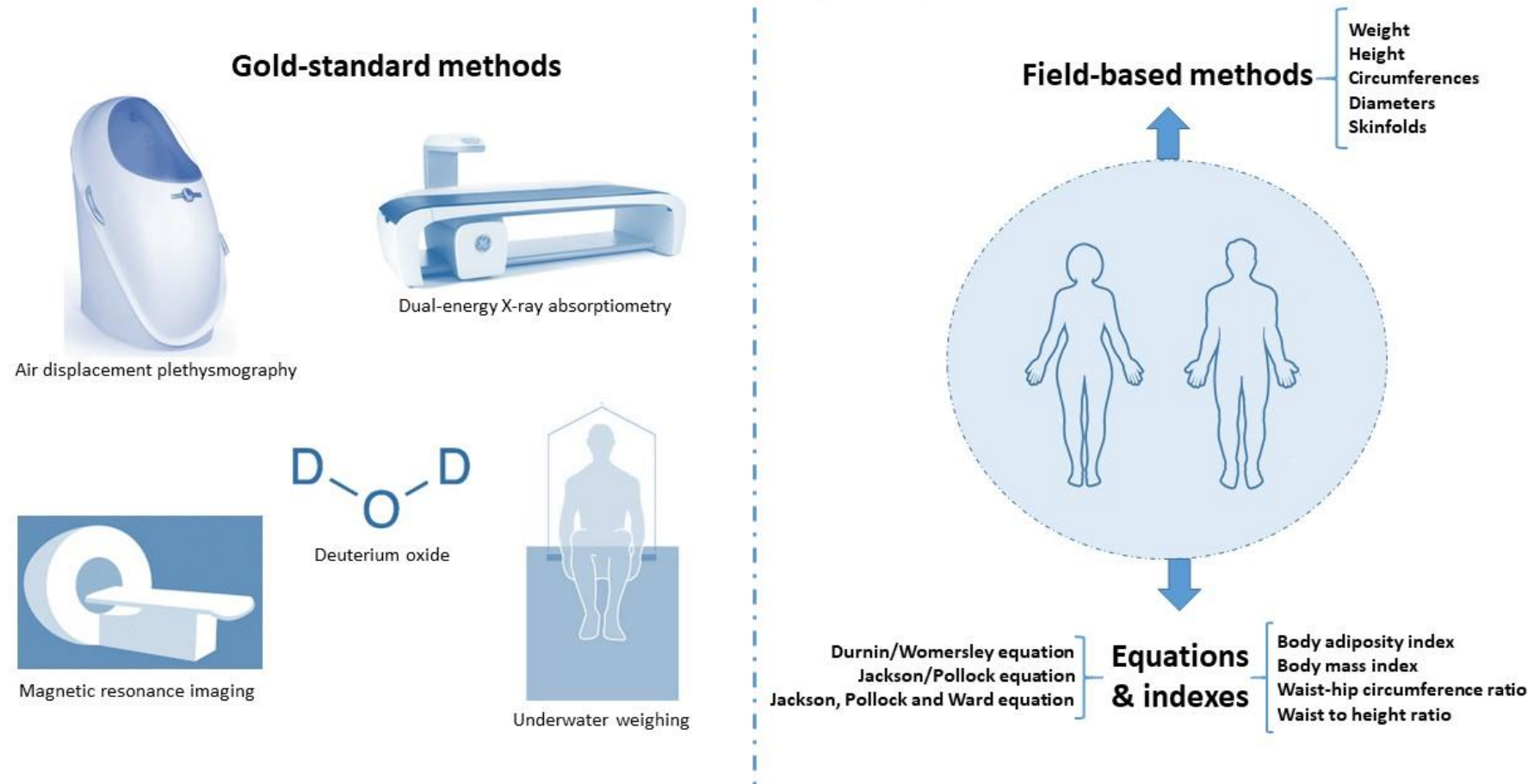


Figure 9. Most common methods for body composition assessment.

Table 30. Advantages and limitations of the “gold-standard” methods, and valid field-based methods and equations used for the assessment of body composition in adults.

GOLD-STANDARD METHODS²⁸²⁻²⁸⁴		
Advantages		Limitations
Air-displacement plethysmography	High accuracy, relatively short time of measurement, non-invasive	High cost, large space, non-portable, specialized technicians, some contraindications (e.g. claustrophobia, anxiety), usually available only in research environments
Deuterium oxide	High accuracy, easy administration	High cost, specialized technicians, usually available only in research environments
Dual-energy X-ray absorptiometry	High accuracy, relatively short time of measurement, assessment of regional body composition, bone mineral density	High cost, large space, long time of calibration, non-portable, specialized technicians, usually available only in research/clinical environments
Magnetic resonance imaging	High accuracy, non-invasive	High cost, large space, long time of measurement, non-portable, specialized technicians, some contraindications (e.g. metallic objects or pacemakers, claustrophobia), usually available only in research environments
Underwater weighing	High accuracy, non-invasive	High cost, large space, long time of measurement, non-portable, specialized technicians, some contraindications (e.g. claustrophobia, anxiety), usually available only in research environments
FIELD-BASED METHODS^{285*}		
Waist Circumference	Low cost, easy of calculation, short time of measurement, non-invasive, no specialized technicians required, adaptable to different settings, implies central adiposity measurement	Superficial measurement, may be influenced by the timing of measurement (e.g., preprandial and postprandial period), require the

		individual to remove clothing to obtain a more accurate measurement, specific populations' cutoffs
Body Adiposity Index	Low cost, easy of calculation, short time of measurement, non-invasive, no specialized technicians required, adaptable to different settings, implies central adiposity measurement	Superficial measurement, may be influenced by the timing of measurement (e.g., preprandial and postprandial period), require the individual to remove clothing to obtain a more accurate measurement
Body Mass Index	Low cost, easy of calculation, short time of measurement, non-invasive, no specialized technicians required, adaptable to different settings	Superficial measurement, may be influenced by the timing of measurement (e.g., preprandial and postprandial period), overall index of adiposity, misleading information about fat distribution in specific populations (e.g., athletes), specific populations' cutoffs
EQUATION/ INDEXES^{283, 284*}		
Durnin/Womersley equation	Low cost, easy of calculation, relatively short time of measurement (depending on the number of SKF measured, four SKF may involve more time), SKF implies a subcutaneous adipose tissue measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, SKF may require the individual to remove clothing to obtain a more accurate measurement
Jackson/Pollock equation	Low cost, easy of calculation, relatively short time of measurement (depending on the number of SKF measured, seven SKF may involve more time), SKF implies a subcutaneous adipose tissue measurement, non-invasive, adaptable to different settings	Specific for the male population, medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, SKF may require the individual to remove clothing to obtain a more accurate measurement
Jackson, Pollock and Ward equation	Low cost, easy of calculation, relatively short time of measurement (depending on the number of SKF measured, seven SKF may involve more time), SKF implies a subcutaneous adipose tissue measurement, non-invasive, adaptable to different settings	Specific for the female population, medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, SKF may require the individual to remove clothing to obtain a more accurate measurement

Equations combining abdominal circumference + skinfolds (SKF)	Low cost, easy of calculation, SKF implies a subcutaneous adipose tissue measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, relatively long time of measurement (depending on the number of SKF measured), require the individual to remove clothing to obtain a more accurate measurement
Equations based on waist circumference	Low cost, easy of calculation, short time of measurement, non-invasive, can be extrapolated to the general population, no specialized technicians required, adaptable to different settings	Superficial measurements, require the individual to remove clothing to obtain a more accurate measurement
Equations combining waist circumference + SKF	Low cost, easy of calculation, SKF implies a subcutaneous adipose tissue measurement, non-invasive, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, relatively long time of measurement (depending on the number of SKF measured), require the individual to remove clothing to obtain a more accurate measurement
Equations combining waist circumference + SKF + Body Mass Index	Easy of calculation, SKF implies a subcutaneous adipose tissue measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, longer time of measurement, SKF may require the individual to remove clothing to obtain a more accurate measurement, BMI implies specific populations' cutoffs
Equations combining various circumferences	Low cost, easy of calculation, non-invasive, no specialized technicians required, adaptable to different settings	Superficial measurements, longer time of measurement (depending on the number of circumferences measured), may require the individual to remove clothing to obtain a more accurate measurement
Equations combining circumferences + SKF	Low cost, easy of calculation, SKF implies a subcutaneous adipose tissue measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, longer time of measurement, may require the individual to remove clothing to obtain a more accurate measurement

Equations implying sum of SKF	Low cost, easy of calculation, SKF implies a subcutaneous adipose tissue measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, relatively long time of measurement (depending on the number of SKF measured), SKF may require the individual to remove clothing to obtain a more accurate measurement
SKF + Body Mass Index	Low cost, easy of calculation, SKF implies a subcutaneous adipose tissue measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Medium-high cost depending on the type of caliper (plastic calipers are the most affordable), SKF requires tester experience, longer time of measurement, SKF may require the individual to remove clothing to obtain a more accurate measurement, BMI implies specific populations' cutoffs
Body Adiposity Index equations	Low cost, easy of calculation, short time of measurement, non-invasive, no specialized technicians required, adaptable to different settings	Different population equations, may require the individual to remove clothing to obtain a more accurate measurement
Bioelectrical Impedance Analysis equations	Easy of calculation, short time of measurement, non-invasive, can be extrapolated to the general population, adaptable to different settings	Different population equations, overestimation in body fat in individuals with severe obesity
Body Mass Index	Low cost, easy of calculation, short time of measurement, non-invasive, can be extrapolated to the general population, no specialized technicians required, adaptable to different settings	Specific populations' cutoffs
Waist to Height Ratio equations	Low cost, easy of calculation, short time of measurement, non-invasive, can be extrapolated to the general population, no specialized technicians required, adaptable to different settings	Superficial measurements, may require the individual to remove clothing to obtain a more accurate measurement

*Both the field-based methods and the equations/indexes summarized in this table have presented validity for estimating body composition in adults.

Chapter III. Reliability of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study IV*).

In our electronic search, 17010 records were retrieved and 42 additional records identified through other sources (i.e., reference checking). After removing duplicates, 15275 articles were identified, of which 15105 were excluded after reading the title and abstract, and 41 articles after reading the full text. Reasons for exclusion were: (i) age range (n= 8); (ii) not a field-based test (n= 18); (iii) not a reliability study (n= 7); (iv) not an original study (n= 3); (v) not a full test (n= 1); (vi) special populations (n= 3); and (vii) inappropriate protocol (n= 1). Finally, 129 studies were considered eligible for our systematic review (see PRISMA flow-chart in **Figure 10**). Of these 129 studies, 114 (88%) were classified as ‘high quality’ and 16 (12%) as ‘low quality’; no article was classified as ‘very low quality’ (see **Table 31**). The sample sizes of the individual studies ranged from 13 to 260. The time interval between tests ranged from 1 to 54 days, but in most it was 1 or 2 weeks, except for those examining intra-session test–retest data.

Cardiorespiratory fitness was analyzed in 33 studies, of which 30 were classified as high quality. The most common field-based physical fitness tests for assessing cardiorespiratory fitness were the 20-m SRT (ten studies, nine of them classified as high quality), the 6-minute walk test (four studies, all of high quality) and the 6-minute step test (four studies, all of high quality).

The reliability of field-based physical fitness tests assessing **musculoskeletal fitness** was investigated in 92 studies, 78 of which were classified as high quality. The most common field-based physical fitness tests for musculoskeletal fitness were the handgrip test (23 studies, 22 of high quality) to assess isometric muscular fitness; the Biering-Sørensen test (11 studies, all of high quality), the partial curl-up test (six studies, five of high quality), the trunk flexion sustained test (11 studies, all of high quality), the push-ups test (five studies, four of high quality), and the sit-to-stand test (eight studies, seven of high quality) to assess endurance muscular fitness; and the sit-and-reach test (19 studies, 13 of high quality) and toe-to-touch test (ten studies, eight of high quality) to measure flexibility.

The reliability of field-based physical fitness tests assessing **motor fitness** was investigated in 22 studies, all classified as high quality. The most common field-based tests for motor fitness were the TUG (three studies), and the T-test (three studies) (see **Table 32**).

ICC was the statistical method most frequently used to assess relative reliability. To assess absolute reliability, SEM, CV, analysis of variance, and the Bland–Altman method were employed (see **Table 32**).

Table 33 shows a summary of the different levels of evidence found for the reliability of cardiorespiratory fitness tests.

Table 34 shows a summary of the different levels of evidence found for the reliability of musculoskeletal and motor fitness tests.

Figure 11 summarizes the evidence-based proposal of field-based physical fitness test reliability in adults. Our findings indicate: (i) a strong level of evidence exists for the high reliability of the **cardiorespiratory fitness** tests: the 20-m SRT, 6-min step and 6-min walk test. The 20-m SRT is a maximal test that offers high reliability in young people while the step tests and the 6-min walk test emerged as possible alternative submaximal tests to assess cardiorespiratory fitness in adults who are not very physically fit and/or find it difficult to run (e.g., obese individuals); (ii) among the **musculoskeletal fitness** tests, the handgrip test using a JAMAR dynamometer offers high test-retest reliability yet this dynamometer does not allow for adapting grip span to the individual's hand size with the same precision as the TKK dynamometer; (iii) the back-leg test, Biering-Sørensen and its modified versions, trunk flexion sustained, back-leg strength and 5-reps sit-to-stand tests show high reliability; (iv) the bilateral side and prone bridge tests provide results with moderate reliability; (v) the sit-and-reach test and its modified versions, and the toe-to-touch test show high reliability; and finally, among the motor fitness tests (vi) the single-leg stand test and its modified versions demonstrate low reliability; and (vii) the T-test and its modified version provide results demonstrating moderate reliability. There is inconclusive evidence regarding the reliability of balance and gait speed tests. Due to the low number of studies, evidence for the reliability of a large number of other field-based physical fitness tests was limited.

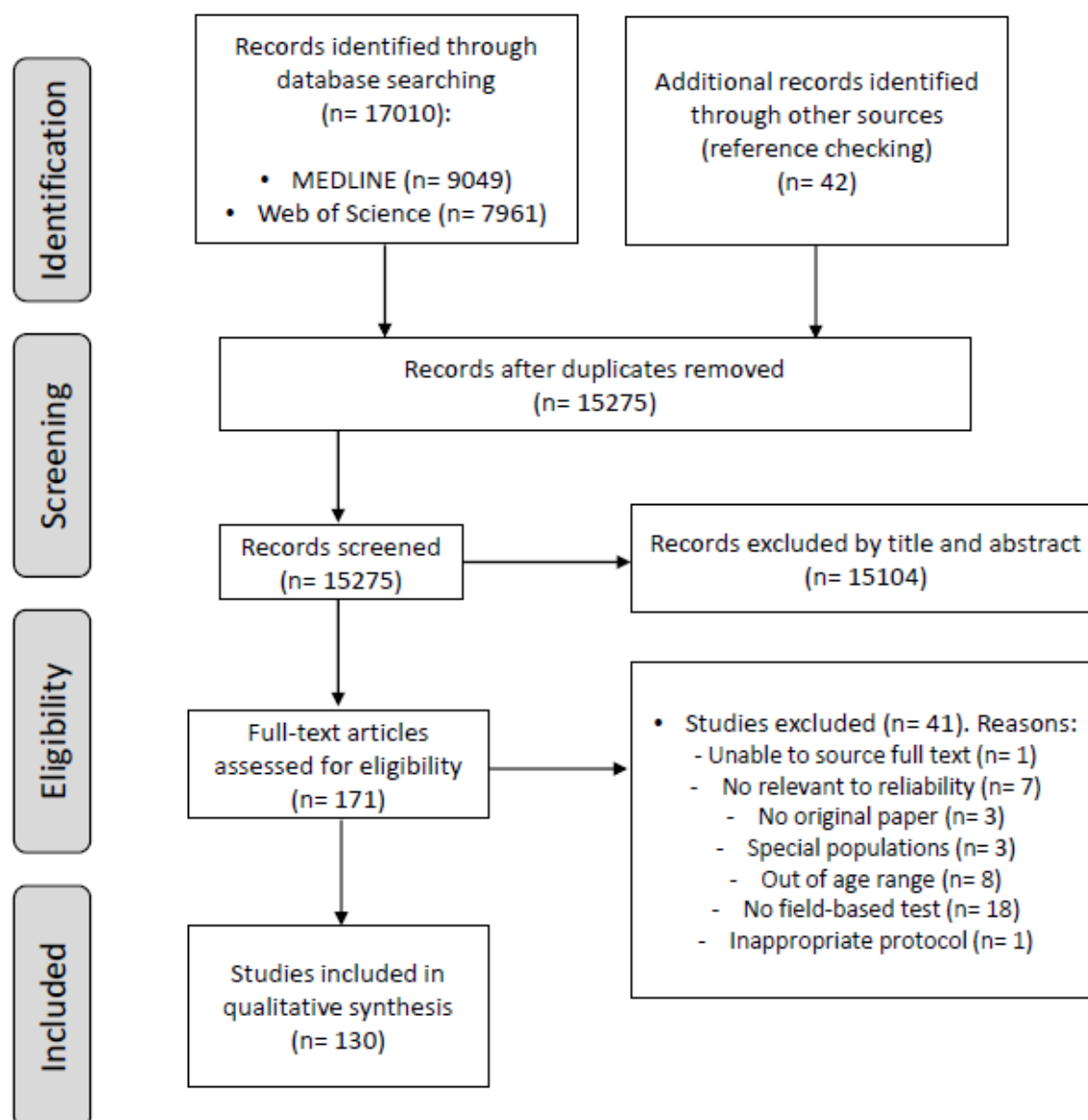


Figure 10. PRISMA flow-chart showing the selection of original studies.

Table 31. Quality assessment of fitness reliability studies.

Study	Fitness Component	Description of the participants	Time interval	Description of results	Statistical analysis	Total score
<i>Low quality studies (2-5)</i>						
DiNucci et al. 1990	Cardiorespiratory fitness	1	0	2	2	5
Chatterjee 2008	Cardiorespiratory fitness	2	0	1	2	5
Sartor et al. 2016	Cardiorespiratory fitness	2	0	1	2	5
Broer & Galles 1958	Musculoskeletal fitness	0	0	1	1	2
Cotten DJ 1972	Musculoskeletal fitness	1	0	1	1	3
Harvey & Scott 1967 _a	Musculoskeletal fitness	0	2	1	1	4
Wells & Dillon 1952	Musculoskeletal fitness	1	1	1	2	5
Harvey & Scott 1967 _b	Musculoskeletal fitness	0	2	2	1	5
Noble 1975	Musculoskeletal fitness	1	2	1	1	5
Jackson & Langford 1989	Musculoskeletal fitness	1	2	1	1	5
DiNucci et al. 1990	Musculoskeletal fitness	1	0	2	2	5
Hoeger & Hopkins 1992	Musculoskeletal fitness	1	2	1	1	5
Liemohn et al. 1994 _b	Musculoskeletal fitness	0	2	1	2	5
Knudson & Johnston 1995	Musculoskeletal fitness	1	1	1	2	5
Stephens et al. 1996	Musculoskeletal fitness	0	2	1	2	5
Wood & Baumgartner 2004	Musculoskeletal fitness	0	2	1	2	5
Agarwal & Kiely 2006	Musculoskeletal fitness	1	2	1	1	5
<i>High quality studies (6-8)</i>						
Leger & Boucher 1980	Cardiorespiratory fitness	1	2	2	1	6
Leger et al. 1988	Cardiorespiratory fitness	1	2	1	2	6
Sproule et al. 1993	Cardiorespiratory fitness	2	2	1	1	6
Curb et al. 2006	Cardiorespiratory fitness	2	2	1	1	6

Penry et al. 2011	Cardiorespiratory fitness	2	1	1	2	6
Santa Maria et al. 1976	Cardiorespiratory fitness	2	2	2	1	7
Berriault et al. 2009	Cardiorespiratory fitness	2	2	1	2	7
Taylor et al. 2001	Cardiorespiratory fitness	1	2	2	2	7
Gabriel et al. 2010	Cardiorespiratory fitness	2	2	1	2	7
Reuter et al. 2011	Cardiorespiratory fitness	2	2	1	2	7
Carvalho et al. 2015	Cardiorespiratory fitness	2	2	1	2	7
Simmonds et al. 1998	Cardiorespiratory fitness	2	2	2	2	8
Laukkanen et al. 2000	Cardiorespiratory fitness	2	2	2	2	8
Larsen et al. 2002	Cardiorespiratory fitness	2	2	2	2	8
Tsigilis et al. 2002	Cardiorespiratory fitness	2	2	2	2	8
Buckley et al. 2004	Cardiorespiratory fitness	2	2	2	2	8
Sykes & Roberts 2004	Cardiorespiratory fitness	2	2	2	2	8
Cooper et al. 2005	Cardiorespiratory fitness	2	2	2	2	8
Ritchie et al. 2005	Cardiorespiratory fitness	2	2	2	2	8
Lamb & Rogers 2007	Cardiorespiratory fitness	2	2	2	2	8
Larsson & Reynisdottir 2008	Cardiorespiratory fitness	2	2	2	2	8
Metsios et al. 2008	Cardiorespiratory fitness	2	2	2	2	8
Flouris et al. 2010	Cardiorespiratory fitness	2	2	2	2	8
Aandstad et al. 2011	Cardiorespiratory fitness	2	2	2	2	8
Kim et al. 2011	Cardiorespiratory fitness	2	2	2	2	8
Mikawa et al. 2012	Cardiorespiratory fitness	2	2	2	2	8
Wilken et al. 2012	Cardiorespiratory fitness	2	2	2	2	8
Jurgensen et al. 2015	Cardiorespiratory fitness	2	2	2	2	8
Arcuri et al. 2016	Cardiorespiratory fitness	2	2	2	2	8
Giacomantonio et al. 2018	Cardiorespiratory fitness	2	2	2	2	8

Cadenas-Sanchez et al. 2016*	Musculoskeletal fitness	N/A	2	1	2	5
Hyytiäinen et al. 1991	Musculoskeletal fitness	1	2	2	1	6
Dijkstra et al. 1994	Musculoskeletal fitness	0	2	2	2	6
Liemohn et al. 1994a	Musculoskeletal fitness	2	2	1	1	6
Hui et al. 1999	Musculoskeletal fitness	2	1	1	2	6
Hui et al. 2000	Musculoskeletal fitness	2	1	1	2	6
Curb et al. 2006	Musculoskeletal fitness	2	2	1	1	6
López-Miñarro et al. 2007	Musculoskeletal fitness	2	1	1	2	6
Van Meeteren et al. 2007	Musculoskeletal fitness	2	1	1	2	6
Clemons et al. 2010	Musculoskeletal fitness	1	2	1	2	6
Ten Hoor et al. 2016	Musculoskeletal fitness	2	1	1	2	6
Reijnierse et al. 2017	Musculoskeletal fitness	2	1	1	2	6
De Blaiser et al. 2018	Musculoskeletal fitness	2	1	1	2	6
Wear 1963	Musculoskeletal fitness	2	1	2	2	7
Gauvin et al. 1990	Musculoskeletal fitness	2	2	2	1	7
Shephard et al. 1990	Musculoskeletal fitness	2	2	1	2	7
Hamilton et al. 1992	Musculoskeletal fitness	2	2	2	1	7
Coldwells et al. 1994	Musculoskeletal fitness	1	2	2	2	7
Hamilton et al. 1994	Musculoskeletal fitness	2	1	2	2	7
Minkler & Patterson 1994	Musculoskeletal fitness	2	2	1	2	7
Diener et al. 1995	Musculoskeletal fitness	1	2	2	2	7
Ito et al. 1996	Musculoskeletal fitness	2	2	1	2	7
Suni et al. 1996	Musculoskeletal fitness	1	2	2	2	7
Boissy et al. 1999	Musculoskeletal fitness	1	2	2	2	7
Latimer et al. 1999	Musculoskeletal fitness	1	2	2	2	7
Ljungquist et al. 1999	Musculoskeletal fitness	2	2	1	2	7

Peolsson et al. 2001	Musculoskeletal fitness	2	2	1	2	7
Perret 2001	Musculoskeletal fitness	2	2	1	2	7
Shechtman et al. 2001	Musculoskeletal fitness	2	2	2	1	7
Shechtman et al. 2003	Musculoskeletal fitness	2	2	1	2	7
Ortiz et al. 2005	Musculoskeletal fitness	2	2	1	2	7
Shechtman et al. 2005	Musculoskeletal fitness	2	2	1	2	7
Wolinsky et al. 2005	Musculoskeletal fitness	2	1	2	2	7
Coldham et al. 2006	Musculoskeletal fitness	2	2	1	2	7
LeBrasseur et al. 2008	Musculoskeletal fitness	2	2	1	2	7
Gruther et al. 2009	Musculoskeletal fitness	1	2	2	2	7
Leard et al. 2009	Musculoskeletal fitness	2	1	2	2	7
Svantesson et al. 2009	Musculoskeletal fitness	2	2	1	2	7
Negrete et al. 2010	Musculoskeletal fitness	2	2	1	2	7
Reiman et al. 2010	Musculoskeletal fitness	2	2	1	2	7
Atamaz et al. 2011	Musculoskeletal fitness	2	2	1	2	7
Reuter et al. 2011	Musculoskeletal fitness	2	2	1	2	7
Chkeir et al. 2012	Musculoskeletal fitness	2	2	1	2	7
Durall et al 2012	Musculoskeletal fitness	2	2	1	2	7
Haitz et al. 2014	Musculoskeletal fitness	1	2	2	2	7
Fielitz et al. 2016	Musculoskeletal fitness	1	2	2	2	7
Kahraman et al. 2016 _a	Musculoskeletal fitness	1	2	2	2	7
Plant et al. 2016	Musculoskeletal fitness	2	1	2	2	7
Vincent & Britten 1980	Musculoskeletal fitness	2	2	2	2	8
Kippers & Parker 1987	Musculoskeletal fitness	2	2	2	2	8
Robertson & Magnusdottir 1987	Musculoskeletal fitness	2	2	2	2	8
Lusardi et al. 1991	Musculoskeletal fitness	2	2	2	2	8

Jackson et al. 1996	Musculoskeletal fitness	2	2	2	2	8
Moreland et al. 1997	Musculoskeletal fitness	2	2	2	2	8
Simmonds et al. 1998	Musculoskeletal fitness	2	2	2	2	8
Keller et al. 2001	Musculoskeletal fitness	2	2	2	2	8
Mathiowetz 2002	Musculoskeletal fitness	2	2	2	2	8
Tsigilis et al. 2002	Musculoskeletal fitness	2	2	2	2	8
Ritchie et al. 2005	Musculoskeletal fitness	2	2	2	2	8
Bohannon et al. 2006	Musculoskeletal fitness	2	2	2	2	8
Evans et al. 2007	Musculoskeletal fitness	2	2	2	2	8
Cowley & Swensen 2008	Musculoskeletal fitness	2	2	2	2	8
Moir et al. 2008	Musculoskeletal fitness	2	2	2	2	8
Bozic et al. 2010	Musculoskeletal fitness	2	2	2	2	8
Bohannon et al. 2011	Musculoskeletal fitness	2	2	2	2	8
Ayala et al. 2012	Musculoskeletal fitness	2	2	2	2	8
Gerodimos 2012	Musculoskeletal fitness	2	2	2	2	8
Greene et al. 2012	Musculoskeletal fitness	2	2	2	2	8
Savva et al. 2013	Musculoskeletal fitness	2	2	2	2	8
Whitehead et al. 2012	Musculoskeletal fitness	2	2	2	2	8
Wilken et al. 2012	Musculoskeletal fitness	2	2	2	2	8
Brotons-Gil et al. 2013	Musculoskeletal fitness	2	2	2	2	8
Clemons 2014	Musculoskeletal fitness	2	2	2	2	8
del Pozo-Cruz et al. 2014	Musculoskeletal fitness	2	2	2	2	8
van den Dolder et al. 2014	Musculoskeletal fitness	2	2	2	2	8
Kahraman et al. 2016 _b	Musculoskeletal fitness	2	2	2	2	8
Larsson et al. 2015	Musculoskeletal fitness	2	2	2	2	8
Juan-Recio et al. 2018	Musculoskeletal fitness	2	2	2	2	8

Curb et al. 2006	Motor fitness	2	2	1	1	6
Suni et al. 1996	Motor fitness	1	2	2	2	7
Kinzey et al. 1998	Motor fitness	2	2	1	2	7
Ljungquist et al. 1999	Motor fitness	2	2	1	2	7
Piva et al. 2004	Motor fitness	2	2	1	2	7
Ortiz et al. 2005	Motor fitness	2	2	1	2	7
Wolinsky et al. 2005	Motor fitness	2	1	2	2	7
LeBrasseur et al. 2008	Motor fitness	2	2	1	2	7
Gribble et al. 2013	Motor fitness	2	2	1	2	7
Haitz et al. 2014	Motor fitness	1	2	2	2	7
Lopez-Plaza et al. 2018	Motor fitness	1	2	2	2	7
Simmonds et al. 1998	Motor fitness	2	2	2	2	8
Rinne et al. 2001	Motor fitness	2	2	2	2	8
Tsigilis et al. 2002	Motor fitness	2	2	2	2	8
Punakallio 2004	Motor fitness	2	2	2	2	8
Ritchie et al. 2005	Motor fitness	2	2	2	2	8
Beekhuizen et al. 2009	Motor fitness	2	2	2	2	8
Sassi et al. 2009	Motor fitness	2	2	2	2	8
Spagnuolo et al. 2010	Motor fitness	2	2	2	2	8
Munro & Herrington 2011	Motor fitness	2	2	2	2	8
Whitehead et al. 2012	Motor fitness	2	2	2	2	8
Wilken et al. 2012	Motor fitness	2	2	2	2	8
Raya et al. 2013	Motor fitness	2	2	2	2	8
Kahraman et al. 2016 _b	Motor fitness	2	2	2	2	8
Manderoos et al. 2016	Motor fitness	2	2	2	2	8

N/A, Not Available.

Table 32. Overview of field-based physical fitness tests reliability studies in adults.

Author	Participants	Age (y)	Test	Design/Time Interval	Statistical Methods	Main Outcome And Conclusion
Low quality studies						
Cardiorespiratory fitness (n= 3)						
Chatterjee et al. 2008	Females= 32	20.4-24.8	20-m shuttle run	Test-retest (N/A)	Bland-Altman method, paired t-test	MD (SD)= -0.13 (± 1.9 ml·kg ⁻¹ ·min ⁻¹); t= 0.32 p= 0.74 with 95% LoA. The reliability of the 20-m SRT is high.
DiNucci et al. 1990	Females= 57 Males= 143	N/A	1-mile run	Test-retest (N/A) and inter-rater	ICC	Test-retest ICC= 0.96-0.98. Inter-rater ICC= 0.93-0.97. The reliability of the 1-mile run test is high.
Sartor et al. 2016	Females= 18 Males= 63	18-67	Ruffier-Dickson	Test-retest (≥ 1 wk)	ICC, Pearson correlation coefficient (r)	Squat HR peak: ICC= 0.86; r= 0.76, p< 0.001. Start HR rec: ICC= 0.79 ; r= 0.66, p< 0.01. Ruffier-Dickson Index (RDI): ICC= 0.82 ; r= 0.70, p< 0.001. Model 1: the same as the RDI. Model 2 (sex, age, height, and RDI): ICC= 0.91 ; r= 0.83, p< 0.001 Model 3 (sex, start HR rec/Age ² , start HR rec/H ²): ICC= 0.93; r= 0.88, p< 0.001. The Ruffier-Dickson test shows high repeatability when HR recovery features is included.
Musculoskeletal fitness (n= 15)						
Isometric muscular fitness (n= 1)						

Stephens et al. 1996	n= 48 N/A (females/ males)	20-60	Handgrip test (Tekdyne and Jamar dynamometers)	Test-retest (1-5 d and intra-session (60 s)	ICC, SEM	Test-retest intra-session / intra-days (dominant - nondominant hands): Tekdyne dynamometer: ICC= 0.954 - 0.984 / 0.971 - 0.976. SEM= 0.19-0.29 / 0.22 (both hands). Standard Jamar dynamometer: ICC= 0.957 / 0.975; SEM= 5.24 / 4.09. Modified Jamar dynamometer: ICC= 0.981 / 0.991; SEM= 3.35 / 2.4. The reliability of Tekdyne and Jamar dynamometers is high.
<i>Endurance muscular fitness (n= 6)</i>						
Wood & Baumgartner 2004	Females= 32	N/A (College age)	Push-up (Bent- knee)	Test-retest (1 d)	ICC	ICC (95% CI)= 0.83 (0.89-0.95). The reliability of the push-up test is moderate.
Harvey & Scott 1967a	Females= 60	N/A (College age)	Sit-up (30 s., feet held, hands behind the head, knee joint 65°)	Test-retest (≈ 10 min)	Pearson correlation coefficient (r), paired t-test	r= 0.78; t= 2.91 (p≤ 0.01). The reliability of the modified sit-up test is insufficient.
Noble 1975	Females= 48 Males= 48	N/A (College age)	Modified flexion- rotation trunk (60 s, elbows touch a board on the opposite side) Modified flexion- rotation trunk (no time limit)	Test-retest (more than 48h)	Pearson correlation coefficient (r)	Timed (60 s time limit): r= 0.80 (males); r= 0.91 (females). Untimed (no time limit): r= 0.84 males); r= 0.91 females). The reliability of the modified flexion-rotation trunk test is acceptable for Females in both timed and untimed administrations; however, a vague statistical approach is used.

Knudson & Johnston 1995	n= 103 N/A (females /males)	17-23	Modified partial curl-up (2 min, legs elevated on a bench 90°, hands interlocked on the abdomen)	Test-retest (72h)	ICC	ICC= 0.94 (males). ICC= 0.88 (females). The reliability of the modified partial curl-up test is high for males and moderate for Females.
DiNucci et al. 1990	Females= 57 Males= 143	N/A	Flexed-arm hang Sit-up (1 min, full range, held feet, arms folded across the chest)	Test-retest (N/A) and inter-rater	ICC	Test-retest reliability: Flexed-arm hang test: ICC= 0.96-0.97. Sit-up test: ICC= 0.91-0.91. Inter-rater reliability: Flexed-arm hang test: ICC= 0.92-0.94. Sit-up test: ICC= 0.83-0.85. Test-retest reliability is high for both test. The inter-rater reliability is high for the flexed-arm hang and moderate for the modified sit-up test.
Agarwal & Kiely 2006	Females= 16 Males= 14	20-74	1 kg arm lift 30-s sit-to-stand	Test-retest (1 h and 14 d for short and long interval test-retest)	Pearson correlation coefficient (r), Bland- Altman method	r> 0.8 for both tests. 1 kg arm lift test: Bias plot p> 0.05. The 30-s chair stand test shows heteroscedasticity. The reliability of both tests is insufficient.
<i>Flexibility (n= 8)</i>						
Wells & Dillon 1952	Females= 100	N/A (College age)	Sit-and-reach	Test-retest (0 min)	Pearson correlation coefficient (r)	r= 0.98. The reliability of the sit-and-reach test is acceptable; however, a vague statistical approach is used.
Jackson & Langford 1989	Females= 52 Males= 52	20-45	Sit-and-reach	Test-retest (same d)	Pearson correlation coefficient (r)	r= 0.99 (males); r= 0.99 (females).

						The reliability of the sit-and-reach test is high; however, a vague statistical approach is used.
DiNucci et al. 1990	Females= 57 Males= 143	N/A	Sit-and-reach	Test-retest (N/A) and inter-rater	ICC	Inter-rater ICC= 0.92-0.95. Test-retest ICC= 0.96-0.97. The reliability of the sit-and-reach test is high.
Hoeger & Hopkins 1992	Short-term reliability: Females= 257 Long term reliability: Females= 31	19-89	Modified sit-and-reach	Test-retest (0 min for short-term and 24-48 h for long-term reliability)	Pearson correlation coefficient (r)	r= 0.94 / 0.89 (short-term /long-term reliability). The reliability of the modified sit-and-reach test is acceptable; however, a vague statistical approach is used.
Broer & Galles 1958	Females= 50	18-31	Toe-to-touch	Test-retest (N/A)	Pearson correlation coefficient (r)	r= 0.973. The reliability of the toe-to-touch test is acceptable; however, a vague statistical approach is used.
Harvey & Scott 1967b	Females= 100	N/A (college age)	Toe-to-touch	Test-retest (0 min)	Pearson correlation coefficient (r), paired t-test	r= 0.86-0.98; $p \leq 0.05$ between the mean of the best trial and trails 1 and 2; $p > 0.05$ between of best trial and trials 3 and 4. The reliability of the toe-to-touch test is acceptable
Liemohn et al. 1994b	Females= 20 Males= 20	24.5 $\pm$ 5.47	Sit-and-reach Back-saver sit-and-reach	Test-retest (0 min)	Pearson correlation coefficient (r)	Sit-and-reach test: r= 0.98. Back-saver sit-and-reach test: r= 0.98. The reliability of these tests is acceptable; however, a vague statistical approach is used.

Cotten DJ 1972	Females= 26 Males= 31	N/A (college age)	Sit-and-reach Modified sit-and-reach Cureton Scott-French bobbing	Test-retest (N/A)	Pearson correlation coefficient (r)	Sit-and-reach test: $r = 0.95 / 0.98$ (females/males). Modified sit-and-reach test: $r = 0.98 / 0.98$ (females/males). Cureton test: $r = 0.98 / 0.96$ (females/males). Scott-French bobbing test: $r = 0.99 / 0.98$ (females/males). The reliability of these tests is acceptable; however, a vague statistical approach is used.

High quality studies

Cardiorespiratory fitness (n= 30)

Sproule et al. 1993	Females= 4 Males= 16	20-35	20-m shuttle run	Test-retest (3 d - 4 wk)	Pearson correlation coefficient (r)	$r = 0.91$. The 20-m SRT is reliable; however, a vague statistical approach is used.
Leger et al. 1988	n= 81	20-45	20-m shuttle run	Test-retest (1 wk)	Pearson correlation coefficient (r), paired t-test	$r = 0.95$; t-test $p > 0.05$. The 20-m SRT is high reliable.
Tsigilis et al. 2002	Females= 66 Males= 29	Females= 19.4 ± 2.7 Males= 19.5 ± 2.7	20-m shuttle run	Test-retest (1 wk)	ICC	ICC (95% CI)= $0.93 (0.89-0.95)$. The 20-m SRT is a high reliable test.
Cooper et al. 2005	Males= 21	21.8 ± 3.6	20-m shuttle run	Test-retest (1-2 wk)	Bland-Altman method	MD $\pm$ 95% LOA= -0.4 ± 2.7 ml kg/min, $t = -1.37$, $p = 0.190$ and homoscedasticity $0.223 (p = 0.330)$. Log transformation of these data reduced heteroscedasticity to $0.056 (p = 0.808)$ with bias -0.007 ± 0.049 ml kg/min, $t = -1.35$, $p = 0.190$. The 20-m SRT is a reliable test.

Lamb & Rogers 2007	Females= 13 Males= 22	Females= 19.6 ± 1.0 Males= 20.9 ± 1.5	20-m shuttle run	Test-retest (1 wk)	ICC, Pearson correlation coefficient (r), Bland-Altman method	MD ± 95% LOA= -5.3 ± 16.3 for trial 1–trial 2; -4.3 ± 19.2 for trial 1–trial 3 and 0.8 ± 18.2 for trial 2–trial 3 for the total number of shuttles. Table method for estimate VO2max (by Léger et al. (1988)), r / ICC / MD ± 95% LOA= 0.96 / 0.95 / -1.4 ± 5.0 ml kg-1 min-1 for trial 1–trial 2; 0.95 / 0.95 / 0.0 ± 5.5 ml kg-1 min-1 for trial 2–trial 3. Equation method for estimate VO2max (by Brewer et al (1988)), r / ICC / MD ± 95% LOA= 0.96 / 0.95 / -1.1 ± 4.7 for trial 1–trial 2; 0.95 / 0.95 / 0.0 ± 5.0 ml kg-1 min-1 for trial 2–trial 3. The reliability of 20-m SRT to measure and predict VO2max (by different methods) is high. But it does not appear to be reliable enough for the purpose of monitoring changes in VO2max due to non-random reasons.
Aandstad et al. 2011	Males= 41	34.8 ± 4.0	20-m shuttle run	Test-retest (2 d)	ICC, Pearson correlation coefficient (r), Bland-Altman method	ICC= 0.96; r= 0.96; MD ± 95% LOA= -2.7 ± 10.1 number of shuttles. ICC= 0.95; r= 0.56; MD ± 95% LOA= -0.8 ± 3.1 mil kg -1 min -1. The reliability of the 20-m SRT is high.
Kim et al. 2011	Females= 29 Males= 90	Females= 20.8 ± 2.2 Males= 22.1 ± 2.2	20-m shuttle run	Test-retest (1 wk)	Pearson correlation coefficient (r), paired t-test, Bland-Altman method	r= 0.87 (males) / r= 0.97 (females). Paired T-test p= 0.003 (males); T-test p= 0.05 (females). MD ± 95% LOA= -2.72 ± 8.33 (male); -2.24 ± 5.94 (female) for number of shuttles. The reliability of the 20-m SRT is insufficient.

Metsios et al. 2008	Males= 40	21.6 ± 2	20-m shuttle run 20-m squared shuttle run	Test-retest (4 d)	Pearson correlation coefficient (r), Bland-Altman method, CV	The 20-m SRT: $r=0.72$ ($p<0.001$); Mean ± 95% CI= 52.3 ± 2.28 ml·kg⁻¹·min⁻¹; 95% LOA= 0.6 ± 6.8 ml·kg⁻¹·min⁻¹; CV= ± 6.8%. $p>0.05$ mean test-retest predicts VO2 max. 20-m squared shuttle run test: $r=0.85$ ($p<0.001$); Mean ± 95% CI= 47.8 ± 2.19 ml·kg⁻¹·min⁻¹; 95% LOA= 0.3 ± 4.8 ml·kg⁻¹·min⁻¹; CV= $5.2 \pm \%$. $p>0.05$ mean test-retest predicts VO2 max. The reliability of both test is high.
Penry et al. 2011	Females= 33 Males= 28	18-33	20-m shuttle run Cooper's 12 min run	Test-retest (≥ 1 d)	G study analysis (φ)	The 20-m SRT: $\varphi=0.98$; Cooper's 12 min run: $\varphi=0.96$ (after adjustment for a single test instrument and 2 test occasions). The reliability coefficients of both tests are high.
Flouris et al. 2010	Males= 45	18-29	15-m squared shuttle run	Test-retest (≥ 4 d)	Pearson correlation coefficient (r), ANOVA, Bland-Altman method, CV	ANOVA $p>0.05$ between test–retest measured and predicted VO2max. $r=0.78$, $p<0.001$ (measured VO2max); $r=0.92$, $p<0.001$ (predicted VO2max). MD ± 95% LoA (ml·kg⁻¹·min⁻¹) / % CV= - 0.4 ± 1.9 / 3.0 for trial 1; - 0.01 ± 1.9 / 1.9 for trial 2. The 15-m squared shuttle run test is a high reliable test for measuring and predicting VO2max.
Mikawa et al. 2012	Males= 14	40-59	15-m shuttle walk and run (speed start in 2.7 km/h, increasing 15 m/min)	Test-retest (1 wk)	ICC, ANOVA	ICC= 0.98 (VO2max); ICC= 0.92 (HR max). ANOVA $p>0.05$ between test-retest. The reliability of the 15-m shuttle walk and run test is high.

Taylor et al. 2001	LBP (with or without sciatica): Females= 26 Males= 18	48.2 (14.2)	10-m shuttle walk (speed range from 1.8 to 8.53 km/h)	Test-retest reliability (1 w)	ICC, Bland-Altman method	ICC _{1,1} = 0.99 MD (LoA 95%)= 2.5 m (52 m, -47m). The reliability of the 10-m shuttle walk test is high.
Jurgensen et al. 2015	Obese females= 46	18-46	10-m shuttle walk (speed increase. 0.17 m/s ⁻¹ every minute)	Test-retest (2-7 d)	ICC, Bland-Altman method	ICC (95% IC)= 0.93 (0.87-0.96). p> 0.05 between test-retest (m). MD ± 95% LoA (m)= -5.9 (70.2 to -82.2). The reliability of 10-m shuttle walking test is high.
Curb et al. 2006	n= 203 N/A (females/males)	35-71	6-minute walk	Test-retest (< 1 wk)	Pearson correlation coefficient (r)	r= 0.90. The 6-minute walk test is reliable; however, a vague statistical approach is used.
Reuter et al. 2011	Females= 9 Males= 12	29.8 ± 6.9	6-minute walk	Test-retest (≥ 1 d)	ICC	ICC= 0.819. The reliability of the 6-minute walk test is moderate.
Larsson & Reynisdottir 2008	Obese group (BMI: 39.6 kg-m ⁻²): Females= 27 Males= 16	21-62	6-minute walk	Test-retest (1 wk)	ICC, Pearson correlation coefficient (r), S _w , SD _{diff} , Bland-Altman method, ANOVA, CV	ICC= 0.96; r= 0.94; S _w (m)= 25.3; SD _{diff} (m)= 31.6; ± 95% LoA (m)= -45.6 - 79.5. P< 0.001; CV= 4.7%. The reliability of the 6-minute walk test is high and can be recommended for obese participants.
Wilken et al. 2012	n= 20 N/A (females /males)	18 - 43	6-minute walk	Test-retest (5-10 d)	ICC, MDC 95%	ICC (95% CI)= 0.93 (0.79-0.97). p< 0.05 between-session. MDC 95% (m)= 81.25. The reliability of the 6-minute walk test is high.

Simmonds et al. 1998	Low back pain group: Females= 28 Males= 16 Control group: Females= 27 Males= 21	Low back pain= 21-63 Control= 22-54	5-minute walk	Test-retest (2 wk)	ICC, SEM	Low back pain group: ICC (day-to-day)= 0.87; SEM (m)= 36.2. Control group was ICC (day-to-day)= 0.60; SEM (m)= 59.7. The reliability of the test is moderate for the low-back pain group and insufficient for the control group.
Gabriel et al. 2010	Females= 56	45-65	400-meter walk	Test-retest (1 wk)	ICC	ICC (95% CI)= 0.95 (0.92 - 0.97); p< .0001. The reliability of the 400-meter walk test is high.
Laukkanen et al. 2000	Females= 29 Males= 26	30-55	2-km walk	Test-retest (15 wk)	ICC, Bland-Altman method	ICC= 0.88 (males); ICC= 0.91 (females). 95% LoA (ml/kg/min)= -2.8 to 6.6 (males); -2.7 to 4.5 (females). The reliability of the 2-km walk test is high.
Larsen et al. 2002	Females= 55 Males= 57	18-26	1.5-mile	Test-retest (1 d)	ICC	ICC= 0.93. The reliability of the 1.5-mile test is high.
Leger & Boucher 1980	Females= 11 Males= 49	19-40	The Universite de Montreal track	Test-retest (within 10 d, at least 2 d separated)	Pearson correlation coefficient (r)	r= 0.97 (all); r= 0.77 (under 30 years old); r= 0.97 (older than 30 years old); r= 0.97 (males); r= 0.88 (females). This test is reliable for middle-age adults, males and females, and insufficient for young adults. A vague statistical approach is used.
Beriault et al. 2009	Females= 15 Males= 6	55.7 ± 11.2	6-minute step	Test-retest (≥ 2h)	ICC, paired t-test	ICC (95% CI)= 0.948 (0.877 – 0.978; p< 0.001. t-test p> 0.05. The reliability of the 6-minute step test is high.

Carvalho et al. 2015	Females= 31	20-45	6-minute step	Test-retest (2 d)	ICC	ICC (95% CI)= 0.97 (0.898-0.993); $p < 0.001$. The reliability of the 6-minute step test is high.
Arcuri et al. 2016	Males= 42	Females= 36 $\pm$ 15 Males= 41 $\pm$ 19	6-minute step	Test-retest ($\approx$ 1 wk)	ICC, paired t-test	ICC (95% CI)= 0.90 (0.85 – 0.93), $p < 0.001$. t-test $p = 0.65$. The reliability of the 6-minute step test is high.
Giacomantonio et al. 2018	At risk of CVD: Females= 16 Males= 12	39-74	6-minute step	Test-retest (1 wk)	ICC	ICC (95% CI)= 0.92 (0.83–0.97) (HR peak); 0.93 (0.84–0.97) (VO ₂ peak); 0.93 (0.83–0.97) (min). The reliability of the 6-minute test is high in people at risk of CVD.
Buckley et al. 2004	Females= 6 Males= 7	24.4 $\pm$ 4.6	Chester step (30cm step height)	Test-retest (5-7 d) and inter-trial	Bland-Altman method	Test-retest reliability: Bias $\pm$ 95% LOA= 0.8 $\pm$ 3.7 ml O ₂ /kg/min ($p > 0.05$) Inter-trial reliability: The correlation between the inter-trial average and the inter-trial difference in VO ₂ max showed that the 95% LoA in both trials was not heteroscedastic (T1: $r = 20.139$, $p = 0.650$; T2: $r = 20.254$, $p = 0.426$). Stages III and IV showed lower inter-trial 95% LoA for RPE and HR than stages I and II, and the opposite was found for actual VO ₂ . Inter-trial biases $p > 0.05$. The Chester step test is reliable on a test-retest and inter-trial basis.

Sykes et al. 2004	Females= 47 Males= 21	18–52 (30.6 ± 9.7)	Chester step (30cm step height)	Test-retest (1 wk)	Bland-Altman method	Bias ± 95% LOA= -0.7 ± 4.5 ml O ₂ /kg/min The Chester step test is a reliable test based on a small measurement error.
Ritchie et al. 2005	Females= 9 Males= 10	55-70	2-level step	Test-retest (1 wk)	ICC	ICC (95% CI)= 0.92 (0.85-0.96) (HR). The reliability of the 2-level step test is high.
Santa Maria et al. 1976	Males= 27	22.54 ± 4.12	OSU Step	Test-retest (within 4 d)	Pearson correlation coefficient (r)	r= 0.69. The reliability of the OSU Step test is insufficient.

Musculoskeletal fitness (n= 78)

Maximal isometric muscular fitness (n= 25)

Reijnierse et al. 2017	MyoAge cohort: Females= 115 Males= 67 Grey Power cohort: Females= 118 Males= 55	18-30 20-70	Handgrip test (Jamar hand-held hydraulic dynamometer)	Test-retest (0 min)	ICC, Bland-Altman method	ICC= 0.94 - 0.98 for the agreement between attempts 1 vs 2; 1 vs 3; 2 vs 3; 1, 2 vs 3; 1 vs 1, 2, 3. MD (kg)= -1.75 to 1.52. The reliability of the handgrip test is high.
Hamilton et al. 1994	Females= 17 Males= 16	20-55	Handgrip test (Jamar dynamometer)	Test-retest (0 min)	ICC	ICC> 0.93 for the four methods: (1) the score of one trial, (2) the mean score of two trials, (3) the mean score of three trials, and (4) the highest score of three trials. The reliability of the handgrip test is high.
Shechtman et al. 2001	Females= 4 Males= 9	18-55	Handgrip test (Jamar and BTE Primus dynamometers)	Test-retest (10 min)	Pearson correlation coefficient (r)	Jamar dynamometer: r= 0.92 (static grip). BTE Primus dynamometer: r= 0.97 (static grip); r= 0.85 (dynamic grip).

						The reliability of Jamar and BTE Primus dynamometers is acceptable; however, a vague statistical approach is used.
Shechtman et al. 2003	Females= 100 Males= 80	18-49	Handgrip test (Jamar and BTE Primus dynamometers)	Test-retest (15 min)	ICC, ANCOVA	Jamar dynamometer: ICC= 0.979 - 0.983 for the right and left hands and for the sitting and standing positions. BTE-Primus dynamometer: ICC= 0.973 - 0.982 for the right and left hands and for the sitting and standing positions. Grip scores in the first session were significantly greater than in the second session (p= 0.003). The reliability of the handgrip test is inconclusive for both instruments and both positions.
Shechtman et al. 2005	Females= 50 Males= 50	20-40	Handgrip test (Jamar and DynEx dynamometers)	Test-retest (10 min)	ICC, ANCOVA	DynEx dynamometer: ICC= 0.9864. Jamar dynamometer: ICC= 0.9856. ANOVA p= 0.271 between grip scores in the 1 and 2 trials. The reliability of the handgrip test is high for both instruments.
Coldham et al. 2006	Females= 11 Males= 11	23-72	Handgrip test (Jamar dynamometer)	Test-retest (30 s)	ICC, SEM	One trial: ICC (95% CI)= 0.95 (0.89 – 0.98); SEM= 3.04. Mean of three trials: ICC (95% CI)= 0.85 (0.67 – 0.94); SEM= 4.84. Highest of three trials: ICC (95% CI)= 0.95 (0.89 – 0.98); SEM= 2.93. The handgrip test reliability for one trial and the highest of three trials is high. For the mean of three trials is moderate.

Svantesson et al. 2009	Females= 34	20-29	Handgrip test (Jamar and Gripping dynamometers)	Test-retest (4-11 d)	ICC, Pearson correlation coefficient (r), SRD%, Cronbach α	Jamar dynamometer: Maximal voluntary contraction (N) for the left / right hands: ICC= 0.93 / 0.87; r= 0.93 / 0.87; SRD%= 11.1 / 15.4; Cronbach α= 0.963 / 0.932. Gripping dynamometer: Maximal voluntary contraction (N) for the left /right hands: ICC= 0.93 / 0.92; r= 0.95 / 0.93; SRD%= 11.7 / 13.7; Cronbach α= 0.972 / 0.960. Gripping dynamometer: Sustained maximal voluntary contraction (N) for the left / right hands: ICC= 0.93 / 0.95; r= 0.95 / 0.93; Cronbach α= 0.975 / 0.965. The reliability of the Jamar dynamometer is moderate and high for the left and right hands, respectively. The reliability of the Gripping dynamometer is high for both hands and for the maximal voluntary contraction and sustained maximal voluntary contraction.
Mathiowetz et al. 2002	Females= 30 Males= 30	20-50	Handgrip test (Rolyan and Jamar dynamometers)	Inter-instrument (5 min)	ICC	ICC= 0.90 to 0.97. The inter-instrument reliability is high.
Gerodimos 2012	Males= 30	26.1 $\pm$ 5.6	Handgrip test (Jamar dynamometer)	Test-retest (1 d)	ICC, Bland-Altman method, SEM, ANOVA	ICC= 0.94-0.98 for preferred hand; ICC= 0.96-0.98 for non-preferred hand. ANOVA p> 0.05. MD ($\pm$ 95% LoA): 0.35 kg (-3.32 to 4.02) for the preferred hand; 0.32 kg (-2.97 to 3.61) for the non-preferred hand. The reliability of handgrip test is high for both hands.

Savva et al. 2013	Females= 9 Males= 10	Females= 21-23 Males= 21-26	Handgrip test (Jamar dynamometer)	Test-retest (1 wk)	ICC, Bland-Altman method, SEM	ICC (95% CI)= 0.96 (0.91-0.98). MD (95% LoA)= -0.17 (-6.8 to 5.5). SEM= 2.79 1 st testing session; SEM= 2.94 2 nd testing session. The reliability of handgrip test is high.
Bohannon et al. 2011	Females= 14 Males= 14	14-85	Handgrip test (Jamar dynamometer)	Test-retest (4-10 d)	ICC, paired t-test, CV _{SEM} , CV _{ME} , SEM, ME	Scores from first trial, best trial, and mean scores (left and right hands) were: ICC= 0.937 - 0.965; t-tests p> 0.05; CV _{SEM} = 5.2 – 7.1; ME= 16.7 – 24.9; CV _{ME} = 4.9 – 7.1.; SEM= 17.7 - 25.6. The reliability of handgrip test is high.
Van Meeteren et al. 2007	Females= 11 Males= 137	21.8 ± 4.2	Handgrip test (Lode handgrip dynamometer)	Test-retest (≥ 1 min)	ICC	ICC= 0.93 / 0.91 (dominant /non-dominant hand). The reliability of handgrip test is high for both hands.
Boissy et al. 1999	Females= 5 Males= 5	25-67	Handgrip test (modified prehension. Lafayette Model 78001)	Test-retest (1 wk)	ICC, SEM	ICC= 0.98 / 0.95; SEM= 16 / 66 (left / right hand). The reliability of handgrip test is high for both hands.
Reuter et al. 2011	Females= 9 Males= 12	29.8 ± 6.9	Handgrip test (Smedley hand dynamometer)	Test-retest (≥ 1 d)	ICC	ICC dominant / non-dominant hand: Initial max: ICC= 0.924 / 0.967. Final max: ICC= 0.924 /0.947. Duration at 50% of initial max.: ICC= 0.582 / 0.572. Work performed: ICC= 0.593 /0.753. Strength decrement index: ICC= 0.465 / 0.408. The initial max. and final max. scores show a high reliability, but insufficient duration at 50% of initial max., work performed and strength decrement index.

Chkeir et al. 2012	Females= 6 Males= 10	Females= 25.8 ± 1.0 Males= 33.4 ± 10.2	Handgrip test (Grip-Ball dynamometer)	Test-retest (2 d)	ICC	Pressure 100 kPa: ICC (95% CI)= 0.953 (0.889 - 0.983). Pressure 125 kPa: ICC (95% CI)= 0.981 (0.954 - 0.993). Pressure 125 kPa: ICC (95% CI)= 0.991 (0.979 - 0.997). The reliability of the handgrip test is high.
Lusardi et al. 1991	Females= 30	19-64	Handgrip test (Jamar dynamometer)	Test-retest (2min)	ICC, Pearson correlation coefficient (r), ANOVA	Jamar dynamometer: ICC= 0.822; r= 0.823; ANOVA p= 0.086. The reliability of the handgrip test is moderate.
Hamilton et al. 1992	Females= 29	20-39	Handgrip test (Jamar dynamometer)	Test-retest (≈1 wk)	Pearson correlation coefficient (r)	r= 0.784 - 0.876; p< 0.001. The handgrip test reliability is insufficient.
Cadenas-Sánchez et al. 2016	N/A	N/A	Calibrated known weight: Handgrip test (analog and digital TKK dynamometers)	Test-retest (50-60 s)	Bland-Altman method	Analog TKK dynamometer: MD= 0.09 - 0.33. Digital TKK dynamometer: MD= 0.04 - 0.25. The reliability for both instruments are high.
Tsigilis et al. 2002	Females= 66 Males= 29	Females= 19.4 ± 2.7 Males= 19.5 ± 2.7	Handgrip test (model N/A)	Test-retest (1 wk)	ICC	ICC (95% CI)= 0.94 (0.92-0.96). The reliability of the handgrip test is high.

Bohannon et al. 2006	Females= 16 Males= 14	20-76	Handgrip test (MicroFET 4 hand-grip dynamometer)	Test-retest (1 d)	ICC, paired t-test, TEM	Left hand: First trial: ICC (95% CI)= 0.974 (0.947-0.987); t-test p= 0.323; TEM= 22.7. Best trial: ICC (95% CI)= 0.981 (0.960-0.991); t- test p= 0.487; TEM= 19.6. Mean score: ICC (95% CI)= 0.980 (0.958- 0.990); t-test p= 0.700; TEM= 19.6. Right hand: First trial: ICC (95% CI)= 0.979 (0.957-0.990); t-test p= 0.505; TEM= 22.7. Best trial: ICC (95% CI)= 0.985 (0.969-0.993); t- test p= 0.700; TEM= 17.8. Mean score: ICC (95% CI)= 0.984 (0.966- 0.992); t-test p= 0.783; TEM= 19.1. The reliability of the handgrip test is high.
Peolsson et al. 2001	Females= 24 Males= 8	20-64	Handgrip test (Jamar dynamometer)	Test-retest (1 d - 1 wk) and inter-rater	ICC	Test-retest reliability: ICC= 0.94 - 0.98 (both hands). Inter-rater reliability: ICC= 0.98 (both hands). Test-retest and inter-rater reliability is high for right and left handgrip test.
Plant et al. 2016	Females= 15 Males= 10	23-67	Handgrip test Grip fatigue (10s) (J-Tech Tracker Freedom wireless dynamometer (electronic) and Baseline® Hydraulic Hand Dynamometer (manual))	Test-retest (1 wk) and inter-rater	ICC	Test-retest reliability ICC (95% IC) (manual/electronic dynamometers): Handgrip test = 0.96 (0.91-0.98) / 0.96 (0.91- 0.98). Grip fatigue test = 0.86 (0.68-0.94) / 0.80 (0.53- 0.91). The inter-rater reliability ICC (95% IC) (manual/electronic dynamometers): Handgrip test = 0.99 (0.97-1.0) / 0.99 (0.97-1.0). Grip fatigue test = 0.76 (0.29-0.92) / 0.6 (-0.3- 0.87).

						The test-retest and inter-rater reliability is high for the handgrip test using both the electronic and manual dynamometers. The test-retest reliability for the grip fatigue test is moderate with both instruments, and the inter-rater reliability is insufficient with both instruments.
Ten Hoor et al. 2016	Females= 23 Males= 22	18-35	Back-leg-chest isometric muscular fitness (Baseline Back-leg-chest dynamometer)	Test-retest (2-5 d)	ICC, Pearson correlation coefficient (r), Bland-Altman method, SRD	Females: ICC (95% CI)= 0.93 (0.90-0.98); r= 0.92; SRD= 15.0; SRD%= 26.5. Males: ICC (95% CI)= 0.92 (0.88-0.98); r= 0.93; SRD= 26.2; SRD%= 19.9. Bland-Altman-plots showed high agreement. The back-leg test shows high reliability.
Whitehead et al. 2012	Females= 42 Males= 137	Females= 37.5 ± 8.4 Males= 38.2 ± 7.3	Back-leg-chest muscular fitness (Baseline Back-leg-chest dynamometer)	Test-retest (10 d)	ICC, CV	ICC (90% CI)= 0.89 (0.66-0.97); CV%= 7.8. The back-leg test shows high reliability.
Coldwells et al. 1994	Females= 13 Males= 23	19-30	Back-leg-chest muscular fitness (Takei dynamometer)	Test-retest (1 d)	Pearson correlation coefficient (r), paired t-test	Position 1 for measure leg muscular fitness: r= 0.80; p< 0.001. Position 2 for measure back muscular fitness: r= 0.91; p< 0.001. p> 0.05 for leg muscular fitness and p< 0.05 for back muscular fitness. The Takei dynamometer shows moderate and high reliability for leg and back muscular fitness, respectively. But significant differences are found for test-retest of back muscular fitness.

Endurance muscular fitness (dynamic or isometric) (n= 31)

Curb et al. 2006	n= 203 N/A (female/ males)	35-71	5-repetition sit-to-stand 10-repetition sit-to-stand	Test-retest (< 1 wk)	Pearson product-moment correlation (r)	5-repetition sit-to-stand test: r= 0.80. 10-repetition sit-to-stand test: r= 0.84. Both tests show an acceptable reliability, however, a vague statistical approach is used.
Wolinsky et al. 2005	Females= 49 Males= 31	50-64	5-repetition sit-to-stand	Test-retest (5-45 d)	ICC	ICC= 0.72. The reliability of 5-repetition sit-to-stand test is insufficient.
Bohannon et al. 2011	Females= 14 Males= 14	14-85	5-repetition sit-to-stand	Test-retest (4-10 d)	ICC, paired t-test, CV_{SEM} , CV_{ME} , SEM, ME	Scores from first trial, best trial, and mean scores were: ICC= 0.954 - 0.975; t-tests were $p < 0.05$; $CV_{SEM} = 4.5 - 6.2$; $ME = 0.3 - 0.5$; $CV_{ME} = 4.5 - 6.2$; SEM= 0.3 - 0.5. The reliability of the 5-repetition sit-to-stand test is high.
Wilken et al. 2012	Females= 50 Males= 130 Interrater reliability (n= 25) Test-retest reliability (n= 20)	18-43	5-repetition sit-to-stand Timed stair ascent (12 steps)	Test-retest (5-10 d) and inter-rater	ICC	Test-retest reliability: 5-repetition sit-to-stand test: ICC (95% CI)= 0.92 (0.75-0.97). Timed stair ascent: ICC (95% CI)= 0.90 (0.76-0.96). Inter-rater reliability: 5-repetitions sit-to-stand test: ICC (95% CI)= 0.99 (0.97-0.99). Timed stair ascent: ICC (95% CI)= 0.97 (0.92-0.98). The test-retest and inter-rater reliability for these tests are high.

Simmonds et al. 1998	Low-back pain group (n= 44): Females= 28 Males= 16 Control group (n= 48): Females= 27 Males= 21	Low back pain= 21-63 Control= 22-54	5-repetition sit-to-stand Biering-Sørensen	Test-retest (2 wk), inter-rater and day-to-day	ICC, SEM	Low back pain group: ICC/SEM for test-retest; day-to-day; inter-rater reliability: 5-repetition sit-to-stand test (s): 0.45 / 6.66; 0.89 / 2.3; 0.99 / 0.41. Biering-Sørensen test (s): 0.91 / 10.3; 0.88 / 8.7; 0.99 / 1.4. Control group: ICC/SEM for test-retest; day-to-day; interrater reliability: 5-repetition sit-to-stand test (s): 0.83 / 0.59; 0.89 / 0.45; 0.91 / 0.21. Biering-Sørensen test (s): 0.73 / 14.3; 0.68 / 17.6; 0.99 / 1.2. The test-retest, day-to-day, and interrater reliability range from moderate to high in both tests and both groups. Except to test-retest reliability for 5-repetition sit-to-stand test in low back pain group and the Biering-Sørensen test in control group.
Kahraman et al. 2016a	Low back pain Females= 14 Males= 24	18–55	30-s sit-to-stand	Intra-rater reliability (48-72 h)	ICC, SEM, MDC, CV	ICC 2,1 (95% CI)= 0.94 (0.89–0.97); SEM (repeats)= 1.17; MDC= 3.24; CV= 32.44 The reliability of the 30-s sit-to-stand test is high.
Ritchie et al. 2005	Females= 9 Males= 10	55-70	60-s sit-to-stand Lift and reach task (1 min, weighted basket: 5 kg Females/ 8 kg males)	Test-retest (1 wk)	ICC	60-s sit-to-stand test: ICC (95% CI)= 0.80 (0.65-0.89). Lift and reach task test: ICC (95% CI)= 0.66 (0.43-0.80). The reliability is insufficient for the lift and reach task test, and moderate for the 60-s sit-to-stand test.

LeBrasseur et al. 2008	Males= 30	38.7 ± 1.4	Timed stair ascent (12 steps) Loaded stair ascent (12 steps) Lift and lower task (1min, weighted basket equivalent to 15% of body weight)	Test-retest (2-7 d)	ICC, Bland-Altman method	Timed stair ascent test: ICC (95% CI)= 0.791 (0.486-0.915); Bias (95% CI) LOA= 0.175 (-0.250; 0.375) 0.688–1.038. Loaded stair ascent test: ICC (95% CI)= 0.940 (0.847-0.976); Bias (95% CI) LOA= 0.205 (0.095; 0.315) 0.254–0.664. Lift and lower test: ICC (95% CI)= 0.925 (0.775-0.975); Bias (95% CI) LOA= -2.53 (-4.10; -0.960) 8.91–3.85. The test-retest reliability is high for all test and with minimal bias.
Ljungquist et al. 1999	Back/neck pain group: Females= 10 Males= 14 Control group: Females= 7 Males= 4	Patients= 30-58 Control= 27-59	Modified timed stair ascent (18-20 steps) Biering-Sørensen (up to 180 s) Trunk flexion sustained (hands on thigh, up to 90 s)	Test-retest (8 d) and inter-rater	ANOVA, Bland-Altman method	Test-retest reliability: ANOVA $p > 0.05$ for all tests. Inter-rater reliability: Modified timed stair ascent test: MD (LoA): 0.30 (-3.44; 4.04). Biering-Sørensen test: MD (LoA)= 0.15 (-1.61; 1.93). Trunk flexion sustained test: MD (LoA)= -0.1 (-3.4;3.2). The test-retest and inter-rater reliability is acceptable for all test.

Suni et al. 1996	Females= 20 Males= 20	25-59	Modified push-up (40 s) Single-leg squat (dynamic squat knee flex 90°) Trunk flexion sustained (trunk flex 90° (thigh-trunk) up to 240 s))	Test-retest (6-8 d) and inter-rater	MD, CV, ICC, SEM	Test-retest reliability: Modified push-up test (reps): MD (95% CI)= 3.0 (2.1 to 3.9); CV= 0.6 %. Single-leg squat test (points): MD (95% CI)= 0.1 (-0.7 to 0.5); CV= 12.1 %. Trunk flexion sustained test (s): MD (95% CI)= 7.9 (-5.7 to 21.5); CV= 3.7 % Inter-rater reliability: Modified push-up test (reps): ICC= 0.88; SEM= 2.6. Single-leg squat test (points): ICC= 0.86; SEM= 0.9. Trunk flexion sustained test (s): ICC= 0.76; SEM= 20.1. The test-retest reliability shows small mean differences. The inter-rater reliability is moderate for modified push-up and single-leg squat tests, but insufficient for trunk flexion sustained test.
Larsson et al. 2015	Intra-rater reliability (test-retest): Males= 20 Inter-rater reliability: Females= 4 Males= 33	19-46	Bilateral side bridge Pull-ups (maximal reps) Modified step-up/ascent (step 40 cm, 20kg, 75 reps, rate of 25 step/min) Modified Biering-Sørensen	Test-retest (1 wk) and inter-rater	ICC, ANOVA, Bland-Altman method, SEM, SRD	Test-retest reliability: Side bridge left /right (s): ICC (95% IC)= 0.82 (0.59–0.93)/0.91 (0.78–0.96) ; SEM= 18/15; SEM%= 16.8/13.2; SRD= 50/42. Pull-ups (maximal reps): ICC (95% IC)= 0.94 (0.86–0.98); SEM= 1; SEM%= 14.2; SRD= 4. Modified step-up test 20 kg left /right (reps): ICC (95% IC)= 0.97 (0.92–0.99)/0.88 (0.38–0.95); SEM= 1/3; SEM%= 2/3; SRD= 4/7. Modified Biering-Sørensen test (s): ICC (95% IC)= 0.85 (0.66–0.94); SEM= 11; SEM%= 11; SRD= 32. Inter-rater reliability: ICC (95% IC)= 0.99 (0.99-1) for all tests; SEM= 0.3 - 1 (ranging between <1 repetition to 4

repetitions and 1s to 3 s); SEM %= 0.9 - 5.5.

The test-retest reliability ranges from moderate to high. The inter-rater reliability is high for all tests.

Kahraman et al. 2016 _b	non-specific low back pain: Females= 14 Males= 24	18-55	Single-leg squat (reps in 30 s each leg (knee flex 60°) Lateral step (20cm, reps in 30 s each leg) Partial curl-up (max rep 1min) Sit-up (full range, max rep 1 min, held feet) Trunk flexion sustained (arms folded across the chest, trunk flex 60°, held feet) Bilateral side bridge Biering-Sørensen	Test-retest (48-72 h)	ICC, CV, SEM, MDC	Single-leg squat test: ICC (95% IC)= 0.87(0.77-0.93); CV= 32.78; SEM= 2.22; MDC= 6.15. Right lateral step test: ICC (95% IC)= 0.93 (0.87-0.96); CV= 33.60; SEM= 1.87; MDC= 5.18. Left lateral step test: ICC (95% IC)= 0.92 (0.86-0.96); SEM= 2.00; CV= 34.19; MDC= 5.54. Partial curl-up test: ICC (95% IC)= 0.90 (0.82-0.95); CV= 45.71; SEM= 3.40; MDC= 9.42. Sit-up (1min) test: ICC (95% IC)= 0.83 (0.71-0.91); CV= 37.16; SEM= 3.20; MDC= 8.87. Trunk flexion sustained test: ICC (95% IC)=0.97 (0.94-0.98); CV= 112.98; SEM= 5.96; MDC= 16.52. Bilateral side bridge test (Right/left): ICC (95% IC)= 0.95 (0.96-0.97)/ 0.96 (0.92-0.98); CV= 87.2/84.04; SEM= 6.37/5.42; MDC= 17.66/15.02. Biering-Sørensen test: ICC (95% IC)= 0.88 (0.79-0.93); CV= 78.09; SEM= 8.71; MDC= 24.14. The reliability for these test are high for the lateral step down, partial curl-up, trunk flexion sustained, and bilateral side bridge tests, and moderate for the sit-up, single-leg squat, and Biering-Sørensen tests.
Robertson & Magnusdottir 1987	Females= 12 Males= 12	Females= 20-31 Males= 22-39	Partial curl-up	Test-retest (15 min)	ANOVA, Pearson correlation coefficient (r)	ANOVA= F (5,70)= 0.27, p> 0.05; r= 0.93 (females). ANOVA= F (5,70)= 1.96, p> 0.05; r= 0.97 (males).

The partial curl-up test shows high levels of consistency between trials.

Vincent & Britten 1980	Males= 19	N/A (College age)	Trunk flexion sustained (hands interlocked behind head)	Test-retest (1 d)	Pearson correlation coefficient (r)	r= 0.71. The reliability of the trunk flexion sustained test is insufficient.
Brotons-Gil et al. 2013	Females= 16 Males= 35	23.2 ± 4.07	Flexion-rotation trunk	Test-retest (1 wk)	ICC, ANOVA, SEM	ICCs= ≥ 0.83; SEMs= ≤ 7.54 reps. The ICCs between trials increased, and SEMs decreased with test repetition, reaching an ICC of 0.94 and an SEM of 6.46 reps between trail 3 and trail 4. Males: ICC= 0.87 to 0.93; SEM= 6.70 to 7.27. ANOVA $p \leq 0.05$ (between trials 2 vs. 1, and 3 vs 2). Females: ICC= 0.63 to 0.93; SEM= 4.24 to 7.44. ANOVA $p > 0.05$ (between trials 2 vs. 1, and 3 vs 2). The reliability of the flexion-rotation trunk test range between moderate (trials 2 vs. 1) and high (trials 3 vs 2).
Juan-Recio et al. 2018	Males= 27	24.1 ± 2.9	Flexion-rotation trunk	Test-retest (1 mo)	ICC, TE, MDC, Change in the mean	ICC= 0.86; TE (%) (mean – 90% confidence limits)= 11.11 (8.99–14.68); MDC (%)= 16.66; Change in the mean (%) (mean – 90% confidence limits)= 12.16 (6.81–17.78). The reliability for the flexion-rotation trunk test is high.

Clemons 2014 _a	Females= 31	19-36	90° Modified flexed-arm hang (s)	Test-retest (3-7 d) and inter-rater	ICC, ANOVA	Test-retest (average scores of two days): ICC= 0.98; $F_{1,30} = 3.809$; $p = 0.60$. Inter-rater: ICC= 0.99; $F_{1,30} = 1.356$; $p = 0.253$ (day 1). ICC= 0.99; $F_{,30} = 0.675$; $p = 0.418$ (day 2). The test-retest and inter-rater reliability of this test is high.
Tsigilis et al. 2002	Females= 66 Males= 29	Females= 19.4 ± 2.7 Males= 19.5 ± 2.7	Bent-arm hang (s) Sit-up (30 s, full range, hands interlocked behind the neck, held feet)	Test-retest (1 wk)	ICC	Bent-arm hang test (s): ICC (95% CI)= 0.89 (0.84-0.93). Modified sit-up test (reps): ICC (95% CI)= 0.83 (0.65-0.90). The reliability of both tests is moderate.
Negrete et al. 2010	Females= 34 Males= 12	18-45	Push-up (push-ups completed in each of the 3 15-second bouts. 45 s rec) Modified pull-up (horizontal bar on the rack)	Test-retest (3-5 d)	ICC, SEM, MDC	Push-up test (rep): ICC= 0.958; SEM= 1; MDC= 2. Modified pull-up test (rep): ICC= 0.989; SEM= 1; MDC= 2. The reliability of both test is high.
Fielitz et al. 2016	Females= 3 Males= 12	21.2 ± 1.0	Push-up (2-minute, Army Physical Fitness Test (APFT))	Test-retest (≥ 1 wk) and inter-rater	ICC, Pearson correlation coefficient (r), Kappa statistic, modified Kappa, Attribute Agreement Analysis	Test-retest reliability: ICC= 0.25–0.52. $r = 0.48 - 0.99$. Kappa = 0.14 - 0.52. Modified kappa = 0.14 - 0.52. Intra-rater agreement for individual push-up repetitions ranges from 41.8% to 84.8%. Inter-rater reliability: The overall scores of the Raters varied between 3.0 and 35.7 push-ups. Post hoc comparisons showed significant increase in the grand mean of push-ups

from trials 1–3 to trial 4 ($p < 0.05$), and over the 4 trials ($p < 0.05$). $r = 0.10 - 0.97$.

The test-retest and intra-rater reliability of push-up to evaluate upper-body muscular fitness is insufficient.

Whitehead et al. 2012	Females= 42 Males= 137	Females= 37.5 ± 8.4 Males= 38.2 ± 7.3	Push-up (cadence 2 min) Single-leg squat (isometric, wall support) Single-leg prone bridge	Test-retest (10 d)	ICC, CV	Cadence push-up test (reps): ICC (90%CI)= 0.87 (0.82-0.91); CV%= 13.4. Single-leg squat test (s): ICC (90%CI)= 0.85 (0.80-0.90); CV%= 46.6. Single-leg prone bridge test (s): ICC (90%CI)= 0.86 (0.81-0.9); CV%= 21.3. The reliability of these tests is moderate.
Diener et al. 1995	Test-retest reliability: Females= 21 Males= 11 Inter-tester reliability: Females= 10 Males= 10	Test-retest: 18-43 Inter-tester: 18-49	Partial curl-up (reach 8.89 cm with fingertip, maximum reps up to 60s)	Test-retest (10-15 min) and inter-tester (1 wk)	ICC, Pearson correlation coefficient (r), paired t-test	Test-retest reliability: ICC= 0.980; $r = 0.978$; t-test higher performance in the 2 nd trial $p \leq 0.01$. Inter-tester reliability: ICC= 0.853; $r = 0.762$; t-test higher performance in the 2 nd session regardless of the tester $p \leq 0.01$. The test-retest reliability is high and the inter-tester reliability is moderate.

Hyytiäinen et al. 1991	Males= 30	35-44	Trunk flexion sustained (hands on thigh, feet without support, up to 240 s) Modified partial curl-up (feet without support) Biering-Sørensen	Test-retest (1 wk) and inter-rater	Pearson correlation coefficient (r), Wilcoxon signed rank test, Cohen's Kappa	Test-retest reliability: Trunk flexion sustained test: $r = 0.93$, $p > 0.05$. Biering-Sørensen test: $r = 0.74$; $p > 0.05$. Modified partial curl-up test: $\text{Kappa} = 0.57$. Inter-rater reliability: Trunk flexion sustained test: $r = 0.90$, $p > 0.05$. Biering-Sørensen test: $r = 0.80$; $p < 0.05$. Modified partial curl-up: $\text{Kappa} = 0.78$. Test-retest and inter-rater reliability for trunk flexor sustained test is high and for Biering-Sørensen insufficient. Test-retest and inter-rater reliability for modified partial curl-up test is insufficient and good, respectively.
Moreland et al. 1997	Females= 24 Males= 15	22-57	Partial curl-up (reach 8-12 cm with fingertip, maximum reps paced 25 rep/min) Trunk flexion sustained (hands at head, feet without supports)	Inter-rater	ICC, SEM	Partial curl-up test (rep): $\text{ICC} = 0.89$, $\text{SEM} = 8$. Trunk flexion sustained test (s): $\text{ICC} = 0.51$, $\text{SEM} = 35$. The inter-rater reliability for partial curl-up test is high, but insufficient for the trunk flexion sustained test.
Durall et al. 2012	Females= 43 Males= 17	Females= 20.9 ± 1.9 Males= 21.4 ± 3.0	Trunk flexion sustained (arms folded across the chest, trunk flex 60° , held feet) Prone bridge	Test-retest (30 s)	ICC	Trunk flexion sustained test: $\text{ICC} (95\% \text{ CI}) = 0.71 (0.38-0.91)$. Prone bridge test: $\text{ICC} (95\% \text{ CI}) = 0.95 (0.87-0.99)$. The reliability for the trunk flexion sustained test is insufficient, and for the prone bridge test high.

De Blaiser et al. 2018	Females= 15 Males= 14 (Healthy-active)	25.5 ± 2.1	Prone bridge	Test-retest (2 dy) and inter-rater	ICC (2,1), SEM	Intra-tester reliability: ICC (95% CI)= 0.89 (0.78 - 0.95), SEM= 20.9 s. Inter-tester reliability: ICC (95% CI)= 0.87 (0.73 -0.93), SEM= 20.75 s. Test-retest and inter-rater reliability for the prone bridge test is moderate.
Evans et al. 2007	Females= 8 Males= 16	16-57	Trunk flexion sustained (arms folded across the chest, trunk flex 60°, held feet) Bilateral side bridge	Test-retest (2 wk) and inter-rater	ICC, paired t-test, SEM	Intra-rater reliability: Trunk flexion sustained: Rater A: t-test p> 0.05; ICC (95% CI)= 0.95 (0.89–0.98); SEM= 59.1. Rater B: t-test p> 0.05; ICC (95% CI)= 0.95 (0.89– 0.98); SEM= 77.6. Right side bridge test: Rater A: t-test p> 0.05; ICC (95% CI)= 0.82 (0.63–0.92); SEM= 14.9. Rater B: t-test p> 0.05; ICC (95% CI)= 0.81 (0.61–0.91); SEM= 15.2. Right side bridge test: Rater A: t-test p> 0.05; ICC (95% CI)= 0.82 (0.63–0.92); SEM= 14.9. Rater B: t-test p> 0.05; ICC (95% CI)= 0.81 (0.61–0.91); SEM= 15.2. Inter-rater reliability: Trunk flexion sustained: Session 1: t-test p> 0.05. ICC (95% CI)= 0.97 (0.92–0.999); SEM= 60.2. Session 2: t-test p> 0.05. ICC (95% CI)= 0.98 (0.94–0.99); SEM= 50.2. Right side bridge test: Session 1: t-test p= 0.04. ICC (95% CI)= 0.91 (0.78–0.96); SEM= 10.5. Session 2: Differences between raters (p= 0.04). ICC (95% CI)= 0.89 (0.77–0.95); SEM= 11.1. Left side bridge test: Session 1: t-test p> 0.05. ICC (95% CI)= 0.82 (0.63–0.92); SEM= 18.9. Session 2: No differences between raters (p> 0.05). ICC

						(95% CI)= 0.91 (0.81–0.96); SEM= 12.4.
						Test-retest and intra-rater reliability for the trunk flexion sustained test is high, and for bilateral side bridge test is moderate.
Ito et al. 1996	Chronic low-back pain group: Females= 53 Males= 37 Control: Females= 60 Males= 40	CLBP= 33-48 Control= 35-49	Trunk flexion sustained (up to 5 min, lied on the floor, elevated legs 90° hip- knee and arms folded across the chest) Modified Biering-Sørensen (lied on the floor)	Test-retest (72 h)	Pearson correlation coefficient (r)	Trunk flexion sustained test for the control / chronic low-back pain groups: r= 0.95 - 0.89 / 0.91 - 0.85. Modified Biering-Sørensen test for the control / chronic low-back pain groups: r= 0.97 - 0.94 / 0.93 - 0.95. Both tests show high reliability in healthy and chronic low-back pain people; however, a vague statistical approach is used.
del Pozo-Cruz et al. 2014	Low-back pain group: Females= 12 Males= 19	Females= 46.01 ± 8.15 Males= 45.95 ± 9.17	Trunk flexion sustained (lied on the floor, elevated legs 90° hip- knee and arms folded across the chest) Modified Biering-Sørensen (lied on the floor)	Test-retest (1 wk)	ICC, SEM, SRD, Kappa statistic, Bland-Altman method	Trunk flexion sustained test: Females: p> 0.05 between the testing sessions. ICC (95% CI)= 0.96 (0.92–0.99); SEM= 2.67 (3.40%); SRD= 7.41 (9.50%); Kappa= 1. Males: p> 0.05 between the testing sessions. ICC (95% CI)= 0.97 (0.96–0.99); SEM= 3.53 (4.70%); SRD= 9.78 (12.95%); Kappa=1. Modified Biering-Sørensen test: Females: p> 0.05 between the testing sessions. ICC (95% CI)= 0.96 (0.94–0.98); SEM= 6.92 (13.00%); SRD= 19.17 (36.20%); Kappa=1 Males: p> 0.05 between the testing sessions. ICC (95% CI)= 0.97 (0.94–0.98); SEM= 6.54 (6.70%); SRD= 19.33 (18.75%); Kappa= 1. Bias for trunk flexion sustained test values were higher on day 7 than on day 1; and for modified Biering-Sørensen test negative.

The reliability for both tests is high.

Reiman et al. 2010	Females= 34 Males= 16	22-38	Trunk flexion sustained (arms folded across the chest, trunk flex 60°, held feet) Modified Biering-Sørensen (held by a partner)	Test-retest (1 wk)	ICC, Pearson correlation coefficient (r)	Trunk flexion sustained test: ICC= 0.93; r= 0.84. Modified Biering-Sørensen test: ICC= 0.97; r= 0.90. The reliability for both test is high.
Latimer et al. 1999	Current/Previous nonspecific low back pain and asymptomatic Females= 13 Males= 10	35.9 (15.7) 36.6 (13.7) 28.8 (8.2)	Biering-Sørensen	Inter-rater (15 min)	ICC; SEM	Total: ICC (95% IC)= 0.85 (0.76–0.90); SEM (s)= 15.6. Current nonspecific low back pain group: ICC (95% IC)= 0.88 (0.73–0.95); SEM (s)= 11.6. Previous nonspecific low back pain group: ICC (95% IC)= 0.77 (0.52–0.90); SEM (s)= 17.5. Asymptomatic group: ICC (95% IC)= 0.83 (0.62–0.93); SEM (s)= 17.4. The reliability of the Biering-Sørensen test is moderate for current nonspecific low back pain and asymptomatic participants; but insufficient for previous nonspecific low-back participants.

Keller et al. 2001	Low-back pain group: Females= 24 Males= 7 Control group: Females= 24 Males= 7	23-61	Biering-Sørensen	Test-retest (5-10 d)	ICC, CD	Low-back pain group: ICC= 0.93-0.98; CD= 57%. Control group: ICC= 0.80-0.98; CD= 54%. The reliability of the Biering-Sørensen test for low-back pain participants is high and moderate for healthy participants.
Gruther et al. 2009	Chronic low back pain group= 22 Healthy control group= 19 Chronic headache group= 15	18–60	Biering-Sørensen	Test-retest (3 ws)	ICC, AUC (ROC curves), Bland-Altman method	ICC= 0.59; AUC= 0.93; 54.96; MD (SD) (s)= 7.24 (9.8) between day 1 and day 2; p= 0.46; 95% prediction interval= – 88.69, 103.1. The reliability of the Biering-Sørensen test for low-back pain, headache and healthy participants is acceptable and shows an excellent diagnostic accuracy.
Greene et al. 2012	Females= 43 Males= 17	21.1 ± 2.2	Bilateral side bridge Modified side bridge (feet-elevated side support test)	Test-retest (1 wk)	ICC	Side bridge: ICC= 0.78 / 0.91 (right/left side bridge). Modified side bridge: ICC= 0.87 / 0.86 (right/left side bridge). The reliability of the modified side bridge test is moderate, high for the left side bridge test but insufficient for the right-side bridge test.
<i>Explosive muscular fitness (n= 11)</i>						
Clemons 2010 _b	Females= 24 Males= 19	Females= 22.1 ± 1.8 Males= 23.1 ± 2.9	Modified medicine ball put (45° incline bench, 6kg females, 9 kg males)	Test-retest (≥ 2 min)	ICC	ICC= 0.94 (female); ICC= 0.92 (male). The reliability of the medicine ball put test is high.

Cowley & Swensen 2008	Females= 24	20.9 ± 1.1	Front and side abdominal power (2 kg medicine ball)	Test-retest (1 d)	ICC, Bland-Altman method, CV, SEM	Front abdominal power test: p> 0.05 between the testing sessions. ICC (95% CI)= 0.95 (0.89–0.98). Absence of heteroscedasticity. LOA= 36.8%. CV= 9.8%. Side abdominal power test: p> 0.05 between the testing sessions. ICC (95% CI)= 0.93 (0.85–0.97). There was heteroscedasticity (random error (cm)= 67.5; Bias (cm)= 6.4; SEM (cm)= 24). The reliability of front and side abdominal power tests is high There is heterocedasticity in the side abdominal power test.
Tsigilis et al. 2002	Females= 66 Males= 29	Females= 19.4 ± 2.7 Males= 19.5 ± 2.7	Standing long jump	Test-retest (1 wk)	ICC	ICC (95% CI)= 0.89 (0.84-0.93). The reliability of the standing long jump test is moderate.
Whitehead et al. 2012	Females= 42 Males= 137	Females= 37.5 ± 8.4 Males= 38.2 ± 7.3	Standing long jump	Test-retest (10 d)	ICC, CV	ICC (90%CI)= 0.97 (0.96-0.98); CV%= 3.1. The reliability of the standing long jump test is high.
Moir et al. 2008	Females= 35 Males= 35	Females= 21.0 ± 2.06 Males= 21.5 ± 2.30	Countermovement vertical jump	Test-retest (1 wk)	ICC, ANOVA, CV	Females: ICC= 0.87 - 0.95; ANOVA p> 0.05 between sessions; CV= 4.4 - 6.6%. Males: ICC= 0.87 - 0.96; ANOVA p< 0.05 between session 1 and 2 for the mean of the 3 trials; CV= 3.6 - 5.6%. The reliability of the countermovement jump test ranges from moderate to high for the highest attempt and the mean of three attempts.

Suni et al. 1996	Females= 20 Males= 20	25-59	Jump and reach	Test-retest (6-8 d) and inter-rater	MD, CV, ICC, SEM	Test-retest reliability: MD (95% CI)= 1.7 (-0.2 to 3.6); CV (%)= 2.4. Inter-rater reliability: ICC= 0.98; SEM= 3.0. The test-retest and inter-rater reliability for the jump and reach test is high.
Kahraman et al. 2016 _b	non-specific low back pain: Females= 14 Males= 24	18-55	Single-leg hop	Test-retest (48-72 h)	ICC, CV, SEM, MDC	Right single-leg hop test: ICC (95% IC)= 0.98 (0.96-0.99); CV= 34.39; SEM= 5.97; MDC= 16.55. Left single-leg hop test: ICC (95% IC)= 0.97 (0.94-0.98); CV= 35.94; SEM= 7.93; MDC= 21.98. The reliability of the single-leg hop test is high.
Munro & Herrington 2011	Females= 11 Males= 11	Females= 22.3 ± 3.7 Males= 22.8 ± 3.1	Hop Test Sequence: Single leg hop Timed hop (6m) Triple hop Crossover hop	Test-retest (1 wk)	ICC	Single leg hop test: ICC= 0.80 (females); ICC= 0.80 (males). Timed hop test: ICC= 0.84 (females); ICC= 0.60 (males). Triple hop test: ICC= 0.80 (females); ICC= 0.92 (males). Crossover hop test: ICC= 0.87 (females); ICC= 0.86 (males). The reliability of these test is moderate for females and range from moderate to high in males, except the timed hop test that shows insufficient reliability.
Haitz et al. 2014	Females= 18 Males= 22	18-24	Hop Test Sequence: Single leg hop Timed hop (6m)	Test-retest (1 wk) and inter-rater	ICC, SEM, SRD	Test-retest reliability: Single leg hop test: ICC= 0.91; SEM (cm)= 7.7; SRD (cm)= 21.3. Timed hop test: ICC= 0.92; SEM (s)= 0.1; SRD (s)= 0.3. Triple hop test: ICC= 0.95; SEM (cm)= 25.1;

			Triple hop Crossover hop			SRD (cm)= 69.7. Crossover hop test: ICC= 0.94; SEM (cm)= 26.1; SRD (cm)= 72.5. Inter-rater reliability: Single leg hop test: ICC= 1.00; SEM (cm)= 1.7; SRD (cm)= 4.8. Timed hop test: ICC= 0.93; SEM (s)= 0.1; SRD (s)= 0.3. Triple hop test: ICC= 1.00; SEM (cm)= 5.6; SRD (cm)= 15.5. Crossover hop test: ICC= 1.00; SEM (cm)= 4.7; SRD (cm)= 13.0. Test-retest and inter-rater reliability of these test is high.
Ortiz et al. 2005	Females= 25 nonathletes	22 ± 2	Figure-eight hop Up-and-down hop Side-to-side hop	Test-retest (2 wk), inter-rater and day-to-day	ICC, SEM	Test-retest reliability: Figure-eight hop test: ICC= 0.92; SEM (s)= 0.13. Up-and-down hop test: ICC= 0.93; SEM (s)= 1.04. Side-to-side hop test: ICC= 0.89; SEM (s)= 0.87. Inter-rater reliability: Figure-eight hop test: ICC= 0.99; SEM (s)= 0.23. Up-and-down hop test: ICC= 0.99; SEM (s)= 0.36. Side-to-side hop test: ICC= 0.99; SEM (s)= 0.27. Day-to-day reliability: Figure-eight hop test: ICC= 0.92; SEM (s)= 0.82. Up-and-down hop test: ICC= 0.92; SEM (s)= 0.96. Side-to-side hop test: ICC= 0.88; SEM (s)= 0.95. Test-retest, inter-rater and day-to-day reliability for the figure-eight hop and for the up-and-down hop test is high. For the side-to-side hop test inter-rater reliability is high, but test-retest and day-to-day reliability is moderate.

Flexibility (n= 21)

Liemohn et al. 1994 ^a	Females= 20 Males= 20	24.5 ± 5.47	Sit-and-reach	Test-retest (0 min)	Pearson correlation coefficient (r)	r= 0.98. The reliability of sit-and-reach test is high; however, a vague statistical approach is used.
Shephard et al. 1990	Females= 41 Males= 39	45-75	Sit-and-reach	Test-retest (8 mo)	ICC	ICC= 0.83. The reliability of sit-and-reach test is moderate.
Tsigilis et al. 2002	Females= 66 Males= 29	Females= 19.4 ± 2.7 Males= 19.5 ± 2.7	Sit-and-reach	Test-retest (1 wk)	ICC	ICC (95% CI)= 0.94 (0.91-0.96). The reliability of sit-and-reach test is high.
Bozic et al. 2010	Males= 84	21.3 ± 2.6	Sit-and-reach	Test-retest (1 wk) and intra-trial (2 min)	ICC, paired t-test, CV	Test-retest reliability: ICC (95% CI)= 0.94 (0.91-0.96); t-test p> 0.05; CV (95% CI)= 6.7 (5.8–8.0). Intra-trial reliability within the first testing session: ICC (95% CI)= 0.98 (0.97–0.99); t-test p> 0.05; CV (95% CI)= 3.5 (3.2–3.8). The test-retest and intra-trial reliability of sit-and-reach test is high.
Kahraman et al. 2016 _b	non-specific low back pain: Females= 14 Males= 24	18-55	Sit-and-reach	Test-retest (48-72 h)	ICC, CV, SEM, MDC	ICC (95% IC)= 0.98 (0.96-0.99); CV% = -176.78; SEM (cm)= 1.93; MDC (cm)= 5.35. The reliability of sit-and-reach test is high.

Hui et al. 1999	Females= 99 Males= 62	17-41	Sit-and-reach Back-saver sit-and-reach V sit-and-reach	Test-retest (≥ 1 d)	ICC	Sit-and-reach test: ICC (end score / difference score)= 0.98 / 0.97 (males); 0.96 / 0.91 (females). Back-saver sit-and-reach test (left-right legs): ICC (end score / difference score)= 0.93-0.98 / 0.92-0.95 (males); 0.97-0.97; 0.95-0.93 (females). V sit-and-reach test: ICC (end score / difference score)= 0.96 / 0.96 (males); 0.89 / 0.88 (females). The reliability of these tests is high.
Leard et al. 2009	Females= 12 Males= 6	23.7 $\pm$ 3.6	Back-saver sit-and-reach	Test-retest (0 min) and inter-rater	ICC	Test-retest reliability: Tester A: ICC= 0.985; Tester B: ICC= 0.996. Inter-rater reliability: Test 1: ICC= 0.844; Test 2: ICC= 0.849. The test-retest reliability is high, and the inter-rater reliability moderate.
Ayala et al. 2012	Females= 87 Males= 156	Females= 20.7 $\pm$ 1.6 Males= 21.3 $\pm$ 2.5	Sit-and-reach Toe-to-touch	Test-retest (4 wk)	ICC, TE, paired t-test	Sit-and-reach test: ICC (90% CI)= 0.92 (0.88-0.95); TE (90% CI)= 8.74 (7.55-10.43). T-test $p > 0.05$. Toe-to-touch test: ICC (90% CI)= 0.89 (0.84-0.93); TE (90% CI)= 9.86 (8.56-11.67). T-test $p > 0.05$. The reliability for both tests is high.
Atamaz et al. 2011	Females= 10 Males= 36	18-28	Chair Sit-and-reach Modified back-saver sit-and-reach Toe-to-touch	Test-retest (1 wk)	ICC, CV, CD	Active subject: Test-retest reliability: ICC= 0.89 - 0.94; CV _{pooled} = 7.7 - 16.5. CV _{mean_individual} = 5.2 to 12.7. CD= 21.3 - 45.6. Inter-rater reliability: ICC= 0.72 - 0.90; CV _{pooled} = 11.7 - 28.0. CV _{mean_individual} = 9.4 to 19.0. CD= 32.3 - 77.4. Controls:

						Test-retest reliability: ICC= 0.89 - 0.97; CV_{pooled}= 7.2 - 26.6; CV_{mean_individual}= 5.7 - 20.2; CD = 19.9 - 70.8. Inter-rater reliability: ICC= 0.71 - 0.94; CV_{pooled}= 14.5 - 23.2; CV_{mean_individual}= 11.4 - 18.7; CD= 40.1 - 64.1. For both the active and control subjects, the test- retest reliability ranges from moderate to high and the inter-rater reliability range from insufficient to high.
Hui et al. 2000	Females= 96 Males= 62	17-41	Sit-and-reach Modified back- saver sit-and- reach Back-saver sit- and-reach V sit-and-reach	Test-retest (≥ 1 d)	ICC	Sit-and-reach test: ICC= 0.98 (males); 0.96 (females). Modified back-saver sit-and-reach test (left-right legs): ICC= 0.96 -0.97 (males); 0.97 - 0.97 (females). Back-saver sit-and-reach test (left-right legs): ICC= 0.93-0.98 (males); 0.97-0.97 (females). V sit-and-reach test: ICC= 0.96 (males); 0.89 (females). The reliability of this test is high, except for V sit- and-reach test in female that is moderate.
López-Miñarro et al. 2007	Females= 47 Males= 58	Females= 24.1 $\pm$ 5.53 Males= 23.7 $\pm$ 4.09	Sit-and-reach Back-saver sit- and-reach Modified back saver sit-and- reach V sit-and-reach Toe-to-touch	Test-retest (N/A)	ICC	ICC= 0.93-0.97. The reliability of these test is high.

Minkler & Patterson 1994	Females= 51 Males= 48	18-35	Modified sit-and-reach	Test-retest (0 min)	ICC	ICC= 0.99. The reliability of modified sit-and-reach is high.
Wear 1963	Males= 116	19-24	Sit-and-reach Trunk lift Supine back extension	Test-retest (0 min)	Pearson correlation coefficient (r)	Sit-and-reach test: r= 0.94. Trunk lift test: r= 0.96. Supine back extension test: r= 0.92. The reliability of these tests is acceptable; however, a vague statistical approach is used.
Kippers & Parker 1987	Females= 17 Males= 16	21.6 ± 3.4	Toe-to-touch	Test-retest (0 min)	Pearson product-moment correlation (r), paired t-test	r= 0.88-0.97; t-test p> 0.05. The reliability of toe-to-touch test is high.
Gauvin et al. 1990	Females= 26 Males= 47	18-73	Toe-to-touch	Test-retest (10-30 s) and inter-rater	ICC	Test-retest reliability: ICC= 0.98. Inter-tester reliability: ICC= 0.95. The test-retest and inter-tester reliability of toe-to-touch test is high.
Dijkstra et al. 1994	Females= 15 Males= 15	26.5 ± 4.3	Toe-to-touch	Test-retest (0 min) and inter-rater	Residual standard deviations as a measure, t-test, ANOVA	Test-retest reliability: 1.65-1.73 cm degrees in the 1 st and 2 nd sessions. ANOVA p< 0.001 between subject. Inter-rater reliability: 1.37-1.51cm between observers (a-b-c). T-test p< 0.05 between the 1 st measurement of the 1 st session and the last measurement of the 2 nd session The reliability of toe-to-touch test is insufficient.

Perret et al. 2001	Females= 16 Males= 16	21-80	Toe-to-touch	Test-retest (≈ 15 min) and inter-rater	ICC, Bland-Altman method	Test-retest reliability: ICC= 0.99. Inter-rater reliability: MD $\pm$ SD = -0.6 $\pm$ 2.8. Test-retest and inter-rater reliability for toe-to-touch is high.
Hyytiäinen et al. 1991	Males= 30	35-44	Toe-to-touch Side-bending of the trunk	Test-retest (1 wk) and inter-rater	Pearson correlation coefficient (r), Wilcoxon signed rank test.	Test-retest reliability: Toe-to-touch test: $r = 0.93$, $p < 0.001$. Side-bending of the trunk test: $r = 0.82-0.87$, $p > 0.05$. Inter-rater reliability: Toe-to-touch test: $r = 0.96$, $p > 0.05$. Side-bending of the trunk test: $r = 0.84-0.88$, $p > 0.05$. Test-retest and inter-rater reliability for side-bending of the trunk test and inter-rater reliability for toe-to-touch test is acceptable, but test-retest reliability for toe-to-touch test is insufficient.
Jackson et al. 1996	Females= 142 Males= 118	Females= 22 ± 4.7 Males= 23 ± 4.8	The trunk lift (FITNESSGRAM battery)	Test-retest (4 d - 4 wk)	ICC	Best trial: ICC (95% CI)= 0.96 (0.94-0.99). Single trial: ICC (95% CI)= 0.86 (0.81-0.90). The reliability for the best trial is high and for a single trial moderate.
Van den Dolder et al. 2014	Females= 40 Males= 20	45.8 $\pm$ 12.1	Hand-behind-back	Test-retest (2 wk) and inter-rater	ICC, SEM, SRD	Test-retest reliability: ICC (95% CI)= 0.95 (0.91-0.97); SEM [cm (%)] = 0.43 (1.27%); SRD (cm)= 1.28. Inter-rater reliability: CC (95% CI)= 0.96 (0.93-0.98); SEM [cm (%)] = 0.26 (0.76%); SRD (cm)= 0.77. The test-retest and inter-rater reliability for hand-behind-back test is high.

Suni et al. 1996	Females= 20 Males= 20	25-59	Side-bending of the trunk Shoulder-neck mobility	Test-retest (6-8 d) and inter-rater	MD, CV, ICC, SEM	Test- retest reliability: Shoulder-neck mobility test (point scale): Right: Ratio of discordant pairs= 2.3 (0.7 to 7.1); CV (%)= N/A / Left: Ratio of discordant pairs= 2.3 (0.7 to 7.1); CV (%)= N/A. Side-bending of the trunk test (average of both sides; cm): MD= -0.5 (-1.3 to 0.3); CV (%)= 4.7. Inter-rater reliability: Shoulder-neck mobility (point scale): Right: ICC= 0.61; SEM= N/A / Left: ICC= 0.62; SEM= N/A. Side-bending of the trunk test (average of both sides; cm): ICC= 0.92; SEM= 1.4. Only the side-bending of the trunk test appears to provide good reliability.
Motor fitness (n= 23)						
Curb et al. 2006	n= 203 Females= N/A Males= N/A	35-71	Static balance: Single-leg stand test (up to 30 s) 10-foot fast walk test	Test-retest (< 1 wk)	Pearson product-moment correlation	Single-leg stand test: r= 0.69. 10-foot fast walk test: r= 0.57. The reliability of these test is insufficient.
Suni et al. 1996	Females= 20 Males= 20	25-59	Static balance: Single-leg stand with eyes open (up to 1 min), closed, and with head turns (up to 30 s)	Test-retest (6-8 d) and inter-rater	MD, CV, ICC, SEM	Test-retest reliability: Eyes open (s): MD= 3.7 (-2.2 to 9.6); CV (%)= 5.0. Eyes closed (s): MD= 0.6 (-1.6 to 2.8); CV (%)= 10.9. Head turns (s): MD= 2.6 (-0.7 to 5.9); CV (%)= 4.0. Inter-rater reliability: Eyes open (s): ICC= 0.76; SEM= 13.3. Eyes closed (s): ICC= 0.18; SEM= 4.6. Head turns (s): ICC= 0.28; SEM= 7.5.

						Test-retest and inter-rater reliability of this tests is insufficient.
Wolinsky et al. 2005	Females= 49 Males= 31	50-64	Static balance: Single-leg stand (up to 30 s) Semi-tandem stand eyes open Tandem stand with eyes open/ closed Gait speed	Test-retest (5-45 d)	ICC	Single-leg stand test (s): ICC= 0.60. Semi-tandem stand test (s): ICC= 0.52. Tandem stand with eyes open test (s): ICC= 0.27. Tandem stand with eyes closed test (s): ICC= 0.58. Gait speed test (s): ICC= 0.56. The reliability for all these tests is insufficient.
Tsigilis et al. 2002	Females= 66 Males= 29	Females= 19.4 ± 2.7 Males= 19.5 ± 2.7	Flamingo balance Plate tapping Shuttle run 10x5m	Test-retest (1 wk)	ICC	Flamingo balance test (number): ICC (95% CI)=0.73 (0.54-0.83). Plate tapping test (s): ICC (95% CI)= 0.57 (0.12-0.78). Shuttle run 10x5m test (s): ICC (95% CI)= 0.86 (0.80-0.91). The reliability for the shuttle run 10x5m test is moderate and for the flamingo balance and plate tapping tests insufficient.
Ritchie et al. 2005	Females= 9 Males= 10	55-70	Static balance: The Romberg, eyes open (parallel,	Test-retest (1 wk)	Percent agreement statistic	Percent agreement statistic (parallel /semi-tandem /tandem positions): 94.7% / 73.7% /52.6% (point scale).

			semitandem and tandem stance)			The reliability for the tandem balance is moderate, and for the parallel and semitandem balance the agreement is high.
Rinne et al. 2001	Females= 11 Males= 14	36-72	Static balance: Single-leg stand on a narrow bar Dynamic balance: Tandem walking forwards Tandem walking backwards	Test-retest (1 wk) and inter-rater	ICC, Bland-Altman method	Test-retest reliability ICC / MD (95% CI): Tandem walking forwards (s): 0.91 / 0.3 (0.1 to 0.6). Tandem walking backwards (s): 0.85 / 0.2 (-0.4 to 0.8). Standing on a narrow bar (touches): 0.92 / 1.0 (0.3 to 1.6). Inter-rater reliability ICC / MD (95% CI): Tandem walking forwards (s): 0.88 / -0.1 (-0.5 to 0.3). Tandem walking backwards (s): 0.96 / 0.2 (-0.1 to 0.6). Standing on a narrow bar (touches): 0.96 / -1.7 (-2.4 to -1.0). The test-retest reliability is high for the tandem walking forwards and the standing on a bar; and it is moderate for the tandem walking backwards. The inter-rater reliability of the tandem walking forwards is moderate, and for the tandem walking backwards and the standing on a bar is high.
Kinzey et al. 1998	Females= 11 Males= 9	18-35	Dynamic balance: Star excursion balance	Test-retest (1 wk)	ICC, SEM, Spearman brown prophecy ($r_{cc'}$)	ICC= 0.67-0.87; SEM= 3.43-4.78; and $r_{cc'} = 0.86 - 0.98$ for the different directions (left- and right- anterior, and left- and right- posterior). The reliability of star excursion balance test is moderate. A learning period before being evaluated with this test is suggested.

López-Plaza et al. 2018	Males= 27	24.54 ± 3.05	Dynamic balance: Star excursion balance	Test-retest (1 mo)	ICC, SEM	Dominant leg: ICC (95% CI)= 0.87 (0.73-0.94), SEM= 2.12 (1.66 -2.93). Non-dominant leg: ICC (95% CI)= 0.74 (0.50-0.87), SEM= 3.23 (2.54 - 4.45). The reliability of star excursion balance test is moderate for dominant leg and insufficient for non-dominant leg.
Gribble et al. 2013	Females= 19 Males= 10	18-50	Dynamic balance: Star excursion balance	Test-retest (5 min)	ICC	Normalized maximum excursion distances: ICC= 0.86 - 0.92. Non-normalized excursion distances: ICC= 0.89 - 0.94. The reliability of star excursion balance test ranges from moderate to high for both the normalized and non-normalized maximal excursion distances.
Punakallio 2004	Males= 29	33-56	Dynamic balance: The functional balance (walk forwards and backwards across a wooden plank that was long, 9/5 cm wide/thick)	Test-retest (2 mo)	ICC	ICCs (95% CI) (s): from 0.84 (0.69-0.92) to 0.96 (0.92-0.98). ICCs (95% CI) (s + errors): from 0.78 (0.58-0.89) to 0.93 (0.87-0.97). The reliability of the test ranged from moderate to high when it is scored in seconds, and from insufficient to high when recorded in seconds + errors.
Spagnuolo et al. 2010	Females= 53 Males= 11	57 ± 10	Timed up-and-go (3 m)	Test-retest (30 s)	ICC	ICC (95% CI)= 0.936 (0.895-0.961). The reliability of the timed up-and-go test is high.

Simmonds et al. 1998	Low back pain group: Females= 28 Males= 16 Control group: Females= 27 Males= 21	Low back pain group= 21-63 Control group= 22-54	Timed up-and-go (3 m) 50-foot fast walk 50-foot preferred speed walk	Test-retest (2 wk), inter-rater and day-to-day	ICC, SEM	Low-back pain group (ICC/SEM for test-retest; day-to-day; inter-rater reliability): 50-foot fast walk test (s): 0.99 / 0.43; 0.80 / 1.74; 0.99 / 0.22. 50-foot preferred speed walk (s): 0.95 / 0.99; 0.64 / 2.56; 0.98 / 0.55. Timed up-and-go test (s): 0.98 / 0.42; 0.98 / 0.99; 0.99 / 0.29. Control group (ICC/SEM for test-retest; day-to-day; inter-rater reliability): 50-foot fast walk test (s): 0.91 / 0.26; 0.65 / 0.51; 0.91 / 0.25. 50-foot preferred speed walk (s): 0.86 / 0.67; 0.48 / 1.17; 0.98 / 0.28. Timed up-and-go test (s): 0.89 / 0.21; 0.84 / 0.25; 0.88 / 0.22. The test-retest, day-to-day, and interrater reliability for these tests range from moderate to high.
Piva et al. 2004	Knee osteoarthritis group: Females= 43 Males= 7	test-retest = 57 ± 8 inter-rater= 54 ± 9	Expanded timed up-and-go (15.2 m)	Test-retest (2 min) and inter-rater	ICC	Test-retest reliability: ICC (95% C)= 0.95 (0.72-0.98). Inter-rater reliability: ICC (95% C)= 0.98 (0.94-0.99). Test-retest and inter-rater reliability for the expanded timed up-and-go test is high in adults with knee osteoarthritis.
Ljungquist et al. 1999	Back/neck pain group: Females= 10 Males= 14 Control group:	Back/neck pain= 30-58 Control= 27-59	2x20 m Loaded 2x20 m (4 and 8 kg each hand for females	Test-retest (8 d) and inter-rater	ANOVA, Bland-Altman method	Test-retest reliability= ANOVA p> 0.05 between test occasions. Inter-rater reliability:

	Females= 7 Males= 4		and males, respectively) *Self-selected			2x20 m: MD (LoA)= 0.25 (-0.63; 1.13). 2x20 m with burden: MD (LoA)= 0.05 (-1.33; 1.43).
						Test-retest and inter-rater reliability of self- selected walking speed tests is high.
LeBrasseur et al. 2008	Males= 30	38.7 ± 1.4	50-m fast walk (no run) Loaded 50-m fast walk (no run)	Test-retest (2-7 d)	ICC, Bland-Altman method	50-m walk test: ICC (95% CI)= 0.987 (0.967- 0.995); Bias (95% CI) LOA= 0.935 (0.509; 1.361) 0.798–2.667. Loaded 50m walk: ICC (95% CI)= 0.977 (0.941- 0.991); Bias (95% CI) LOA= 0.472 (-0.071; 1.016) 1.804–2.748.
						The reliability for both tests is high.
Munro & Herrington 2011	Females= 11 Males= 11	Females= 22.3 ± 3.7 Males= 22.8 ± 3.1	T-test	Test-retest (1 wk)	ICC	Females: ICC= 0.96. Males: ICC= 0.82.
						The reliability of the T-test is high for females and moderate for males.
Raya et al. 2013	Males= 97	18-39	T-test Edgren side step (10 sec) Illinois agility	Test-retest (1-2 d) and inter-rater	ICC	Test-retest reliability: T-test (s): ICC (95% CI)= 0.83 (0.75-0.88). Edgren side step test (m): ICC (95% CI)=0.62 (0.48-0.73). Illinois agility test (s): ICC (95% CI)= 0.68 (0.55-0.78). Inter-rater reliability: T-test (s): ICC (95% CI)= 0.98 (0.97-0.99). Edgren side step test (m): ICC (95% CI)= 0.92 (0.88-0.94). Illinois agility test (s): ICC (95% CI)= 0.99 (0.98-0.99).

Test-retest reliability of the T-test is moderate; and for the Edgren side step test and the Illinois agility test is insufficient. Inter-rater reliability for these three tests is high.

Sassi et al. 2009	Females= 34 Males= 52	Females= 22.6 ± 1.4 Males= 22.4 ± 1.5	Modified T-test	Test-retest (≥ 3 min)	ICC, CV, Bland-Altman method	ICC (95% CI)= 0.92 (0.84-0.96) /0.95 (0.91-0.97) (females/males). 95% LOA= 0.33 /0.37 (females/males). CV%= 2.6 /2.7 (females/males).
						The reliability of modified T-test is high in females and males.
Beekhuizen et al. 2009	Females= 9 Males= 17	18-30	Hexagon agility (double-leg hopping)	Test-retest (1 h between 1st and 2nd sessions, and 48 h between 1st and 3rd)	ICC, paired t-test, Bland-Altman method, MDC 95% CI	ICC (same day / between-day)= 0.938 (p> 0.001) / 0.924 (p< 0.001). T-test p< 0.001 between sessions 1 and 2. T-test p= 0.18 between sessions 2 and 3. Bland-Altman (same day / between-day)= scattered below the zero-reference line / spread evenly above and below the zero point. MDC 95% IC (s)=1.015.
						The reliability of the hexagon agility test is high.

Ortiz et al. 2005	Females= 25	22 ± 2	Modified hexagon agility (hopping single-legged) Zig-zag run	Test-retest (2 wk), inter-rater and day-to-day	ICC, SEM	Test-retest reliability: Modified hexagon hop test: ICC= 0.84; SEM= 0.70. Zig-zag run test: ICC= 0.97; SEM= 0.26. Inter-rater reliability: Modified hexagon hop test: ICC= 0.99; SEM= 0.20. Zig-zag run test: ICC= 0.97; SEM= 0.21. Day-to-day reliability: Modified hexagon hop test: ICC= 0.76; SEM= 0.72. Zig-zag run test: ICC= 0.94; SEM= 0.30. Test-retest and inter-rater reliability for modified hexagon hop and the zig-zag run tests is high. Day-to-day reliability for the modified hexagon hop test is moderate and high for Zig-zag run test.
Wilken et al. 2012	Females= 50 Males= 130 Test-retest (n= 20) / Inter-rater (n= 25)	18-43	4-square step	Test-retest (5-10 d) and inter-rater	ICC	Test-retest reliability: ICC (95% CI)= 0.86 (0.75-0.96). Inter-rater reliability: ICC (95% CI)= 0.99 (0.99-0.99). Test-retest and inter-rater reliability of the four-square step test is moderate and high, respectively.
Whitehead et al. 2012	Females= 42 Males= 137	Females= 37.5 ± 8.4 Males= 38.2 ± 7.3	Pro-agility	Test-retest (10 d)	ICC, CV	ICC (90% CI)= 0.90 (0.86-0.93), CV (%)= 3.1. The reliability of the pro-agility test is high.
Manderoos et al. 2016	Females= 25 Males= 27	28-55	Agility test track	Test-retest (1 wk between 1st and 2nd sessions and 1.5 h between 2nd and 3rd)	ICC, SEM, MDC 95% CI	ICC (one week / same day)= 0.94 / 0.91 (females); 0.94/0.95 (males). SEM (s) (one week / same day)= 0.20 / 0.27 (females); 0.19 / 0.13 (males). MDC 95% CI (s) (one week / same day)= 0.56 / 0.76 (females); 0.51 / 0.37(males).

The test reliability of the agility test track is high.

ANOVA, Analysis of Variance; AUC, Area Under the Curve; CVD, Cardiovascular Disease; CD, Critical Difference, CI, Coefficient of Interval; CV, Coefficient of Variation; CV_{ME}, Coefficient of Variation of Method Error; CV_{SEM}, Coefficient of Variation of Standard Error of Measurement; ICC, Intraclass Correlation Coefficient; LoA, Limits of Agreement; MD, Mean Difference; MDC, Minimal Detectable Change; MDC 95% CI, Minimal Detectable Change at 95% of Coefficient of Interval; ME, Method Error; N/A: Not Available; SD, Standard Deviations; SD_{diff}, Standard Deviation of the Differences; SEE, Standard Error Estimate; SEM, Standard Error of Measurement; SRD, Smallest Real Difference; S_w, Measurement Error; PCA, Percentage Close Agreement; R, Pearson Correlation Coefficient; TEM, Technical Error of Measurement; TE, Typical Error; VO_{2max}, Maximum Oxygen Uptake.

Table 33. Levels of evidence of cardiorespiratory fitness tests.

Field-based physical fitness test	Strong	Moderate	Limited
Cardiorespiratory fitness tests			
<i>Shuttle run tests</i>			
20-m shuttle run	●		
20-m square shuttle run			○
15-m square shuttle run			○
15-m shuttle walk/run			○
10-m shuttle walk			○
<i>Distance and time-based run/walk tests</i>			
6-minute walk	●		
5-minute walk			○
1.5-mile run/walk			●
2-km walk			●
400-m walk			●
University Montreal			●
Cooper's 12-minute run			●
<i>Step tests</i>			
6-minute step	●		
Chester step		●	
2-level step			●
OSU step			○

● = Indicates high reliability. ● = Indicates moderate reliability. ○ = Indicates low/null reliability.
 ■ = Indicates inconclusive reliability.

Table 34. Levels of evidence of musculoskeletal and motor fitness tests.

Field-based physical fitness test	Strong	Moderate	Limited
Musculoskeletal fitness test			
<i>Maximal isometric muscular fitness</i>			
Handgrip test (Jamar)	●		
Handgrip test (TKK)			○
Handgrip test (DynEx)			○
Handgrip test (BTE Primus)		◐	
Handgrip test (Grippit)			○
Handgrip test (Royal)			○
Handgrip test (Smedley)			○
Handgrip test (Lode)			○
Handgrip test (Grip-Ball)			○
Handgrip test (Lafayette)			○
Back-leg-chest test (Baseline)		◐	
Back-leg-chest test (Takei)			○
<i>Endurance muscular fitness (dynamic or isometric)</i>			
<i>Trunk endurance muscular fitness</i>			
Original/Modifications	●		
Biering-Sørensen			
Original/Modified side bridge		◐	
Original/ Modified prone bridge		◐	
Sit-up		◐	
Original/Modified partial curl-up		◐	
Trunk flexion sustained	●		
Flexion-rotation trunk		◐	
<i>Lower-body endurance muscular fitness</i>			
Sit-to-stand (5 reps)	●		
Sit-to-stand (10 reps)			○
Sit-to-stand (30 sec)			○
Sit-to-stand (60 sec)			○
Original/Modified timed stair ascent		◐	
Step-up (40 cm, 20 kg load)			○
Single-leg squats (isometric/dynamic)			○
<i>Upper-body endurance muscular fitness</i>			
Original/Modifications push-up		◐	
Original/Modified pull-up			○
Bent-arm hang			○
Modified flexed-arm hang			○
Lift and reach task			■
<i>Explosive muscular fitness</i>			
<i>Lower-body explosive muscular fitness</i>			
Standing long jump		◐	
Contramovement vertical jump			○
Jump and reach			○
Hop sequence		◐	
Figure-eight hop			○

Up-and-down hop				○
Side-to-side hop				○
<i>Trunk /upper-body explosive muscular fitness</i>				
Modified medicine ball put				○
Front and side abdominal power				○
Flexibility				
<i>Hamstring-low back flexibility</i>				
Original/Modifications sit-and-reach	●			
Toe-to-touch	●			
Trunk lift			◐	
Side-bending			◐	
Shoulder-neck mobility				○
Hand behind back				○
Motor fitness test				
<i>Static Balance</i>				
Original/Modifications single-leg stand (open/closed eyes)	■			
Parallel, semi-tandem and tandem stand (open/closed eyes)				■
<i>Dynamic Balance</i>				
Star excursion balance				■
Tandem walking				○
Functional balance				○
<i>Gait speed - Agility</i>				
Original/Modified timed up-and-go	●			
Original/Modified t-test	●			
Original/Modified hexagon agility			◐	
Original/Modifications gait speed				○
4-square step				○
10x5 shutter run				○
Zig-zag run				○
Edgren side step				■
Illinois agility				■
Pro-agility				○
Agility track				○
Plate tapping				■

● = Indicates high reliability. ◐ = Indicates moderate reliability. ○ = Indicates low/null reliability.
 ■ = Indicates inconclusive reliability.

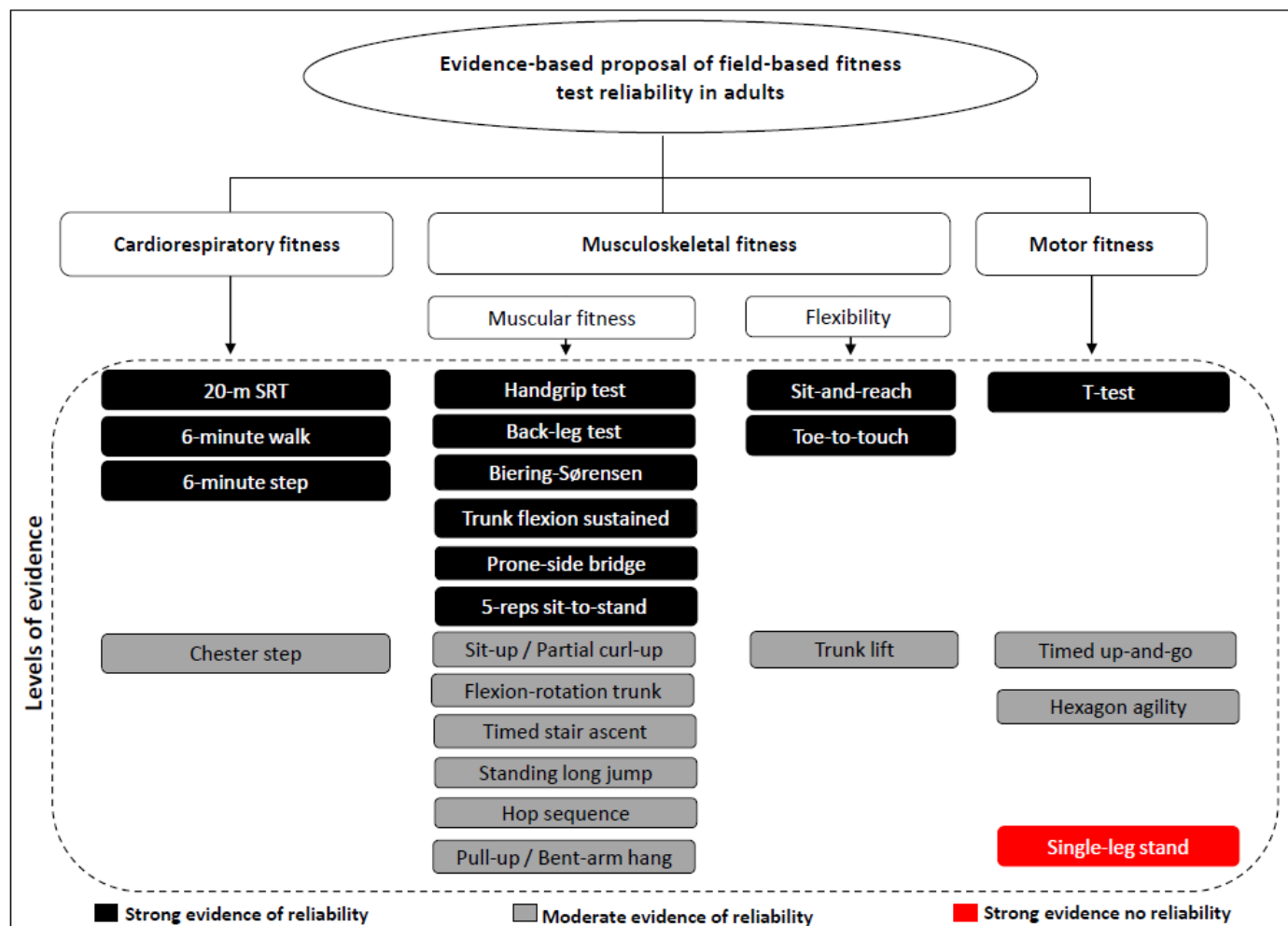


Figure 11. Evidence-based proposal of field-based physical fitness test reliability in adults.

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

The final sample size was composed of 410 adults aged 18-64 years (49.5% females). The descriptive characteristics of the participants, distributed by sex, age and, physical activity level are shown in **Table 35**. The mean age of the sample was 42 ($\pm$ 13.06) years old. Overall, significant differences were found according to sex, age, and physical activity level. Regarding sex, males presented higher anthropometric values than females (all, $p < 0.001$), except for %BF and tricipital skinfold, being higher in females (both, $p < 0.001$). Males performed faster the 2-km walk test and completed more stages in the 20-m SRT than females ($p < 0.001$); measured and estimated VO_{2max} was also higher in males than females ($p < 0.001$). Regarding age, the 50-64 years old group presented higher BMI, %BF, waist and neck circumference and subscapular skinfold than the 18-34 years old group (all, $p < 0.01$), and higher WC than the 35-49 years old group ($p < 0.01$); the 35-49 years old group presented higher BMI, %BF and WC than the 18-34 years old group (all, $p < 0.01$); the 50-64 years old group performed slower the 2-km walk test, completed less stages in the 20-m SRT and had measured and estimated VO_{2max} than their younger counterparts (all, $p < 0.001$). Regarding physical activity level, active participants shower lower anthropometric values (all, $p < 0.01$), expect for neck circumference ($p > 0.05$), and better performance in cardiorespiratory variables than non-active ones (all, $p < 0.001$).

Criterion-related validity

Bivariate correlation analysis between measured VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) with estimated VO_{2max} by 2-km walk test and 20-m SRT, and anthropometric variables, distributed by whole sample, sex, age groups, and physical activity level are displayed in **Table 36**. In the whole sample, measured VO_{2max} was associated with estimated VO_{2max} by 2-km walk test ($r = 0.784$, $p < 0.01$), and 20-m SRT ($r = 0.875$, $p < 0.01$). Sex, age, physical activity level, and all anthropometric variables were associated with measured and estimated VO_{2max} ($r = -0.148$ to -0.751 , all $p < 0.001$). %BF, tricipital skinfold and sum triceps+subscapular skinfold had the strongest association with measured VO_{2max} ($r = -0.750$, $r = -0.648$, $r = -0.656$, respectively; all, $p < 0.001$), with estimated VO_{2max} by 2-km walk test ($r = -0.730$, $r = -0.699$, $r = -0.723$, respectively; all, $p < 0.001$), and with estimated VO_{2max} by 20-m SRT ($r = -0.751$, $r = -0.650$, $r = -0.636$, respectively; all, $p < 0.001$). Similar

results were found when the sample was distributed by sex, age, and physical activity level.

Figure 12 shows the scatterplot of relationship between measured with estimated $\text{VO}_{2\text{max}}$ by Oja's equation and time in the 2-km walk test, and relationship between measured with estimated $\text{VO}_{2\text{max}}$ by Leger's equation, and final stage in the 20-m SRT. The measured $\text{VO}_{2\text{max}}$ was associated with the estimated $\text{VO}_{2\text{max}}$ by Oja's equation ($R^2=0.614$, $p<0.01$; $\text{SEE}=5.027$) and time in the 2-km walk test ($R^2=0.429$, $p<0.01$; $\text{SEE}=6.117$), and with estimated $\text{VO}_{2\text{max}}$ by Leger's equation ($R^2=0.766$, $p<0.01$; $\text{SEE}=3.917$) and final stage in the 20-m SRT ($R^2=0.774$, $p<0.01$; $\text{SEE}=3.865$). The association between measured $\text{VO}_{2\text{max}}$ with estimated $\text{VO}_{2\text{max}}$ by Oja's equation remained the same when the sample was distributed by sex, age, and physical activity level. The association between measured $\text{VO}_{2\text{max}}$ with time in the 2-km walk test was slightly higher in females ($R^2=0.448$, $p<0.01$) than in males ($R^2=0.300$, $p<0.01$); in the 50-64 years group ($R^2=0.517$, all, $p<0.01$) than in the younger ones ($R^2=0.346$ and $R^2=0.397$, respectively; both, $p<0.01$); and in non-active ($R^2=0.449$, $p<0.01$) than in active participants ($R^2=0.286$, $p<0.01$) (see **Figure 13**). The association between measured $\text{VO}_{2\text{max}}$ with estimated $\text{VO}_{2\text{max}}$ by Leger's equation and final stage in the 20-m SRT remained the same when the sample was distributed by sex, age, and physical activity level (see **Figure 14**).

Figure 15 shows the Bland–Altman difference plot between the measured and estimated $\text{VO}_{2\text{max}}$, by Oja's and Leger's equations. It can be observed that the difference was nearly 0 for both equations. The difference between measured and estimated $\text{VO}_{2\text{max}}$ by Oja's equation was $-0.30 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -12.86 to 12.26 , $p<0.001$), and between measured and estimated $\text{VO}_{2\text{max}}$ by Leger's equation was $0.86 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -7.30 to 9.02 , $p=0.051$). Heteroscedasticity was observed between the measured and estimated $\text{VO}_{2\text{max}}$ difference with the measured and estimated $\text{VO}_{2\text{max}}$ mean by Oja's equation ($R^2=0.117$). The effect size (Cohen d) of the mean differences between measured and estimated $\text{VO}_{2\text{max}}$ by Oja's equation was -0.141 . The differences between the measured and estimated $\text{VO}_{2\text{max}}$, by Oja's equation remained the same when the sample was distributed by sex and age groups. When the sample was distributed by physical activity level, the difference was lower in the non-active [$-0.19 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -13.45 to 13.06 , $p<0.001$; $R^2=0.249$)] than in the active participants [$-3.36 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -15.31 to 8.59 , $p<0.001$; $R^2=0.064$)] (**Figure 16**).

The difference between the measured and estimated $\text{VO}_{2\text{max}}$, by Leger's equation was lower in males [$0.08 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -8.92 to 9.08 , $p=0.525$; $R^2=$

0.002)] than in females [$1.65 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -5.24 to 8.54, $p = 0.617$; $R^2 = 0.001$)]. It was also lower in the 18-34 years group [$0.47 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -7.49 to 8.43, $p = 0.116$; $R^2 = 0.020$)] and the 35-49 years group [$0.16 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -8.88 to 9.21, $p = 0.375$; $R^2 = 0.007$)] than in the 50-64 years group [$1.79 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -5.44 to 9.03, $p = 0.137$; $R^2 = 0.017$)] (**Figure 16**).

Table 37 shows the stepwise lineal regression analysis predicting the $\text{VO}_{2\text{max}}$. The 2-km walk test showed that total time represented the 42% of explained variance for measured $\text{VO}_{2\text{max}}$ ($\text{SEE} = 6.136 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p < 0.001$). When heart rate was added, the explained variance increased until 47%. Finally, when age, physical activity level and %BF were included, the explained variance increased until 70% ($\text{SEE} = 4.408 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p < 0.001$).

The same results were observed when replacing %BF by BMI, or sum tricripital+subscapular skinfold, WC, HC and lean mass in the prediction equations (*data not shown*).

The 20-m SRT showed that the final stage represented the 78% of explained variance for measured $\text{VO}_{2\text{max}}$ ($\text{SEE} = 3.841 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p < 0.001$). When physical activity and %BF mass was included, the explained variance only increased an additional 2% ($\text{SEE} = 3.603 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p < 0.001$).

The same results were observed when replacing %BF by BMI, or sum tricripital+subscapular skinfold, WC, HC and lean mass in the prediction equations (*data not shown*).

Reliability

Test-retest reliability of 2-km walk test and 20-m SRT is shown in **Table 38**. Significant mean differences between T1 and T2 were found in the time to complete the 2-km walk test (-1.48 ± 0.51 seconds, $p = 0.004$) and in the final stage reached in the 20-m SRT (0.04 ± 0.01 , $p = 0.002$). The effect size (Cohen d) of the mean differences was -0.014 and 0.002, respectively. Non-significant differences were found between T1 and T2 in the estimated $\text{VO}_{2\text{max}}$ by Oja's ($-0.29 \pm 0.20 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p > 0.05$) and Leger's equations ($0.03 \pm 0.04 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p > 0.05$). The ICCs reported a high reproducibility, ranging from 0.95 to 0.99 (all, $p < 0.001$) in both tests. All the analyzed error measurements showed low values ($\text{RMSE} = 0.38\text{-}10.09$; $\%CV = 0.71\text{-}7.73$; $\text{SEE} = 0.25\text{-}9.88$). **Table 39** shows test-retest reliability of 2-km walk test and 20-m SRT,

distributed by sex, age, and physical activity levels. Overall, similar results to those of the whole sample were found.

Figure 17 shows the Bland–Altman difference plot between the 2-km walk test and 20-m SRT. The systematic error was nearly 0 for all the cases: estimated $\text{VO}_{2\text{max}}$ by Oja's equation [$-0.29 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -8.13 to 7.54 , $p= 0.034$)], time to complete the 2-km walk test [-1.48 seconds (95% LOA= -21.07 to 18.10 , $p= 0.041$)], estimated $\text{VO}_{2\text{max}}$ by Leger's equation [$0.03 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LOA= -1.51 to 1.58 , $p= 0.086$)], and final stage reached in the 20-m SRT [0.04 (95% LOA= -0.45 to 0.53 , $p= 0.172$)]. Heteroscedasticity (T1-T2 variability) was observed in the estimated $\text{VO}_{2\text{max}}$ by Oja's equation ($R^2= 0.012$), and time to complete the 2-km walk test ($R^2= 0.011$).

Figure 18 displayed the results analyzed by sex, age, and physical activity level. Overall, similar results to those of the whole sample were found.

Table 35. Descriptive characteristics of the participants, stratified by sex, age and, physical activity level.

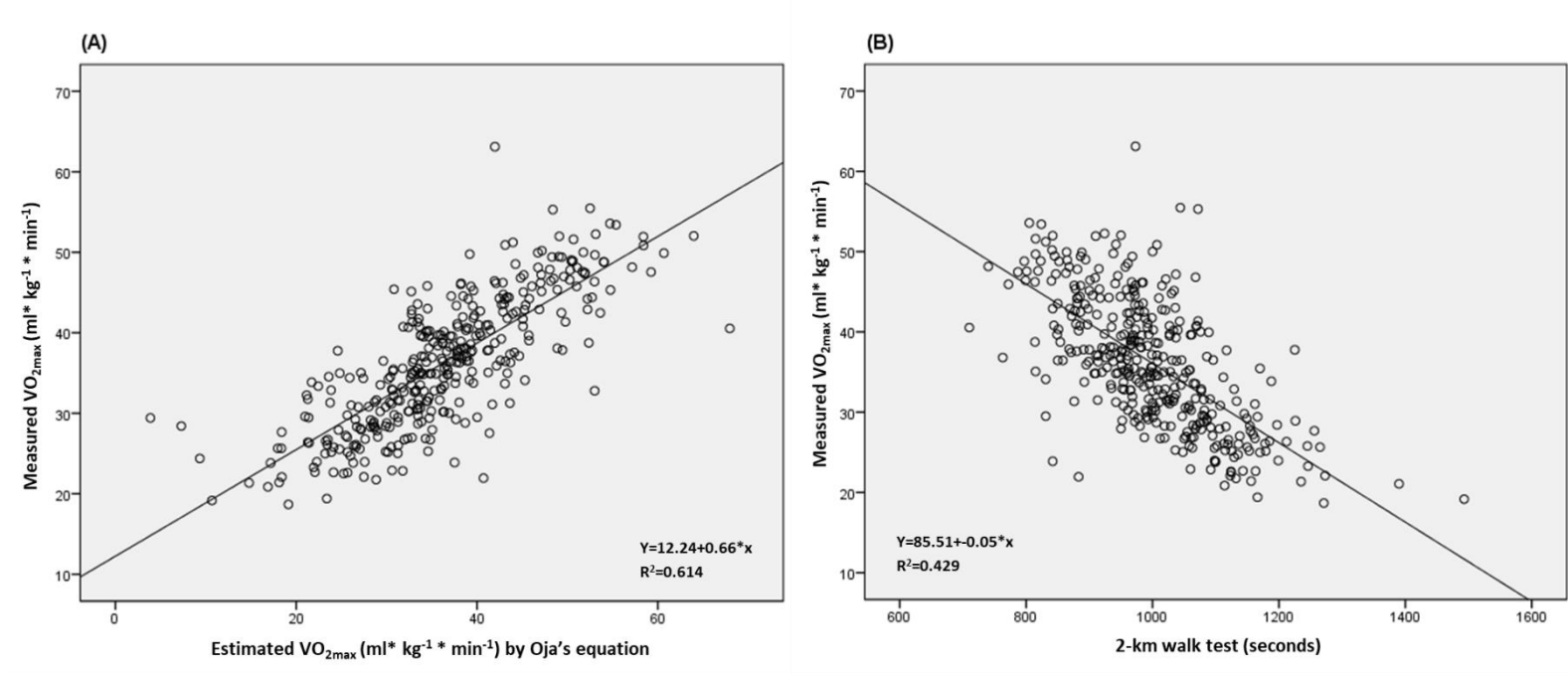
	All (n= 410)	Sex		Age groups			Physical activity levels	
		Females (n= 203)	Males (n= 207)	18-34 yr (n= 136)	35-49 yr (n= 131)	50-64 yr (n= 143)	Non-Active (n= 195)	Active (n= 215)
Age (years)	41.86 (13.06)	41.48 (12.84)	42.24 (13.28)	26.22 (0.36)	42.49 (0.36)	56.17 (0.35)	43.18 (13.10)*	40.62 (12.95)
Basal heart rate (bpm)	69.51 (12.09)	72.25 (12.20)***	66.82 (11.39)	71.83 (13.46) ^{a^}	67.39 (11.11)	69.24 (11.25)	73.51 (12.23)***	65.88 (10.80)
Maximal heart rate (bpm)	173.29 (13.77)	171.8 (13.0)*	174.66 (14.37)	182.47 (12.20) ^{a^}	173.68 (9.17) ^{b^}	164.41 (12.91) ^{c^}	172.79 (15.25)	173.73 (12.42)
Weight (kg)	71.16 (15.49)	62.06 (9.36)***	80.13 (15.12)	68.80 (16.24)	72.19 (16.29)	72.48 (13.77)	71.99 (16.82)	70.32 (14.14)
Height (cm)	168.25 (9.19)	161.38 (5.98)***	175.00 (6.36)	168.46 (0.79)	168.42 (0.80)	167.89 (0.77)	167.33 (9.17)	169.02 (9.12)
Body Mass Index	24.99 (4.10)	23.85 (3.45)***	26.12 (4.38)	24.07 (4.27) ^{a^}	25.32 (4.48)	25.56 (3.38) ^{c^}	25.55 (4.62)**	24.47 (3.50)
Body fat (%)	24.85 (7.47)	28.64 (6.30)***	21.11 (6.62)	22.77 (7.69) ^{a^}	24.99 (7.08)	26.70 (7.15) ^{c^}	27.38 (7.31)***	22.57 (6.89)
Lean mass (kg)	50.37 (10.58)	41.56 (3.92)***	59.06 (7.41)	49.85 (10.13)	51.09 (10.81)	50.22 (10.83)	49.10 (10.72)*	51.45 (10.31)
Waist circumference (cm)	83.14 (12.66)	76.90 (9.25)***	89.26 (12.58)	78.11 (11.68) ^{a^}	83.28 (13.11) ^{b^}	87.80 (11.34) ^{c^}	85.54 (14.21)***	80.91 (10.64)
Hip circumference (cm)	99.38 (8.69)	98.80 (7.27)	99.95 (9.87)	98.69 (9.75)	100.30 (9.19)	99.19 (6.96)	101.13 (9.10)***	97.77 (8.00)
Neck circumference (cm)	35.30 (3.83)	32.30 (2.00)***	38.24 (2.79)	34.40 (3.65)	35.44 (3.91)	36.01 (3.80) ^{c^}	35.31 (4.12)	35.27 (3.57)
Tricipital skinfold (mm)	16.92 (7.48)	19.69 (6.42)***	14.23 (7.46)	16.13 (0.64)	17.03 (0.65)	17.58 (0.63)	19.39 (7.67)***	14.72 (6.56)
Subscapular skinfold (mm)	19.50 (10.27)	20.33 (9.19)	18.68 (11.19)	17.37 (11.72)	19.46 (9.83)	21.57 (8.72) ^{c^}	22.74 (10.99)***	16.57 (8.62)
2-km walk test								
Total time (min:seg)	16.63 (1.80)	17.26 (1.66)***	16.00 (1.72)	16.50 (1.66)	16.14 (1.71) ^{b^}	17.16 (1.88) ^{c^}	17.32 (1.83)***	16.01 (1.54)
Final heart rate (bpm)	149.79 (21.36)	151.55 (20.05)***	148.05 (22.48)	151.89 (22.99)	151.38 (21.81)	146.46 (18.99)	152.54 (21.59)*	147.36 (20.90)
VO _{2max} (ml* kg ⁻¹ * min ⁻¹)	36.58 (10.14)	33.09 (6.43)***	40.06 (11.84)	40.38 (10.67)	37.97 (9.24) ^{b^}	31.89 (8.47) ^{c^}	32.65 (9.79)***	40.14 (9.12)
20-meter shuttle run test								
Final stage	5.10 (2.82)	3.71 (2.17)***	6.48 (2.72)	6.29 (2.71)	5.65 (2.69) ^{b^}	3.51 (2.26) ^{c^}	3.87 (2.31)***	6.20 (2.79)
VO _{2max} (ml* kg ⁻¹ * min ⁻¹)	35.42 (8.51)	31.30 (6.57)***	39.50 (8.25)	39.04 (8.26)	37.11 (8.04) ^{b^}	30.60 (6.76) ^{c^}	31.74 (6.98)***	38.70 (8.42)
Maximal treadmill test								
VO _{2max} absolute (ml* min ⁻¹)	2557.15 (720.28)	1999.87 (367.17)***	3097.71 (545.33)	2703.81 (744.90)	2655.09 (698.22) ^{b^}	2329.57 (720.28) ^{c^}	2299.89 (633.97)***	2784.21 (716.77)
VO _{2max} relative (ml* kg ⁻¹ * min ⁻¹)	36.13 (8.14)	32.78 (6.59)***	39.39 (8.20)	39.48 (7.86) ^{a^}	36.94 (7.69) ^{b^}	32.23 (7.17) ^{c^}	32.17 (6.74)***	39.67 (7.67)
RER final	1.20 (0.08)	1.18 (0.08)***	1.21 (0.08)	1.22 (0.07)	1.20 (0.07) ^{b^}	1.17 (0.08) ^{c^}	1.20 (0.08)	1.20 (0.08)

Differences between sex, and between physical activity level: *p< 0.05, **p< 0.01, ***p< 0.001; differences between age groups: ^a = 18-34 years and 35-49 years, ^b = 35-49 years and 50-64 years, ^c = 18-35 years and 50-64 years ([^]p< 0.05; [^]p< 0.01, ^{^^}p<0.001).

Results are expressed as mean (standard deviation).

VO_{2max} estimated in the 2-km walk test by Oja's equation; VO_{2max} estimated in the 20-meter shuttle run test by Leger's equation.

Difference between sex, and physical activity levels measured with an independent t-test, difference between age groups measured with one-way repeated measures analysis of Variance (ANOVA).



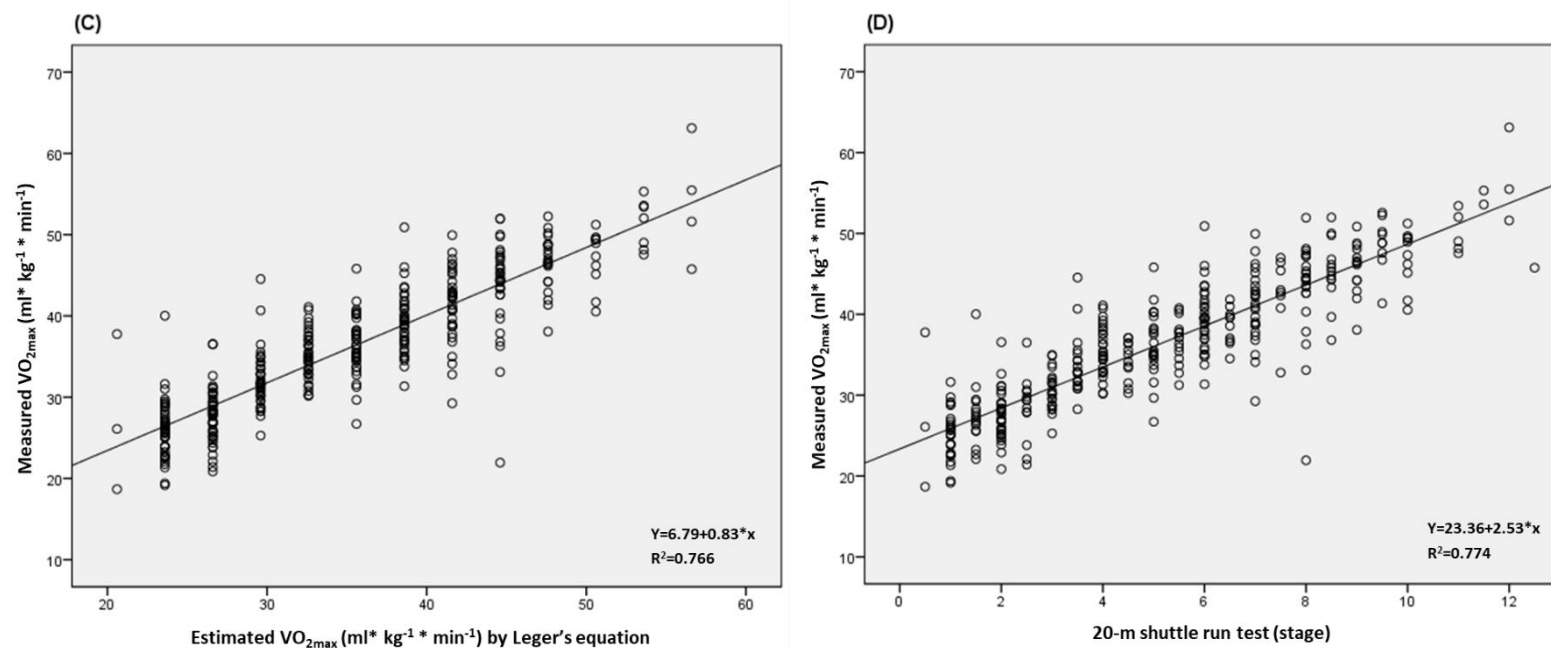
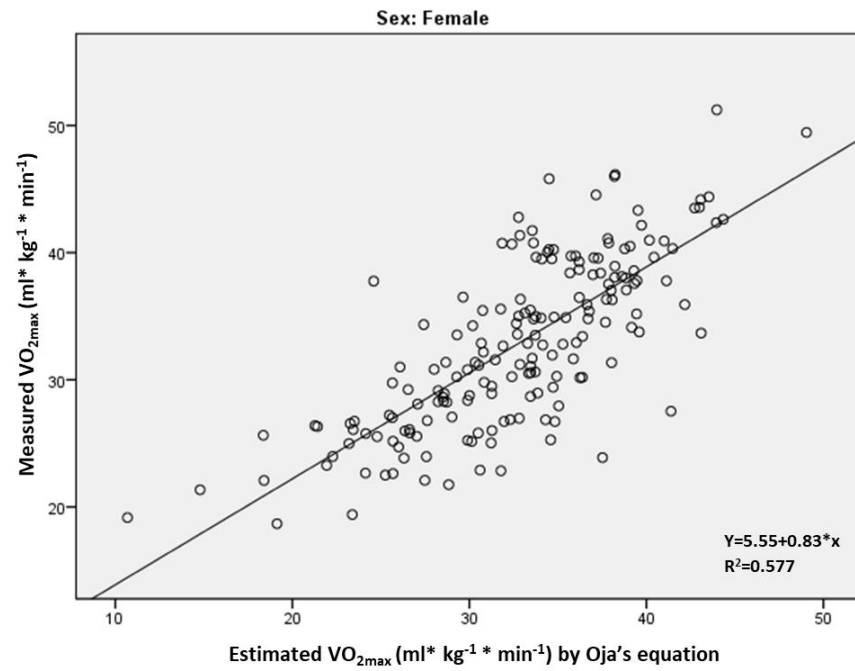


Figure 12. Scatterplot of relationship between measured with estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Oja's equation (estimated from the 2-km walk test) and Leger's equation (estimated from the 20-m shuttle run test); and relationship between measured with total time in the 2-km walk test and final stage in the 20-m shuttle run test.

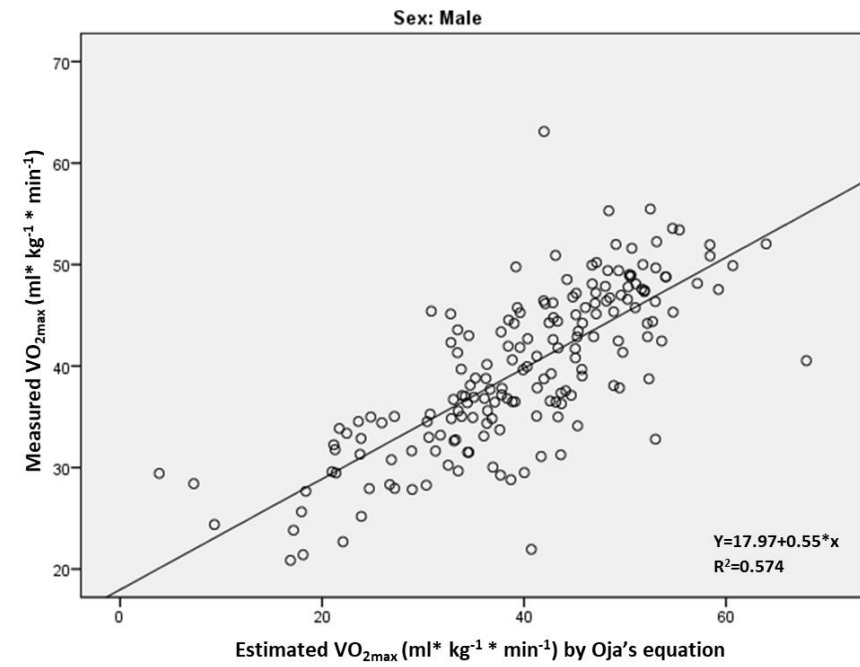
(A), relationship between measured and estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Oja's equation; (B), relationship between measured VO_{2max} (ml* kg⁻¹ * min⁻¹) and total time (sec) in the 2-km walk test; (C), relationship between measured and estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Leger's equation; (D), relationship between measured VO_{2max} (ml* kg⁻¹ * min⁻¹) and final stage in the 20-m shuttle run test.

R^2 indicates lineal coefficient of determination.

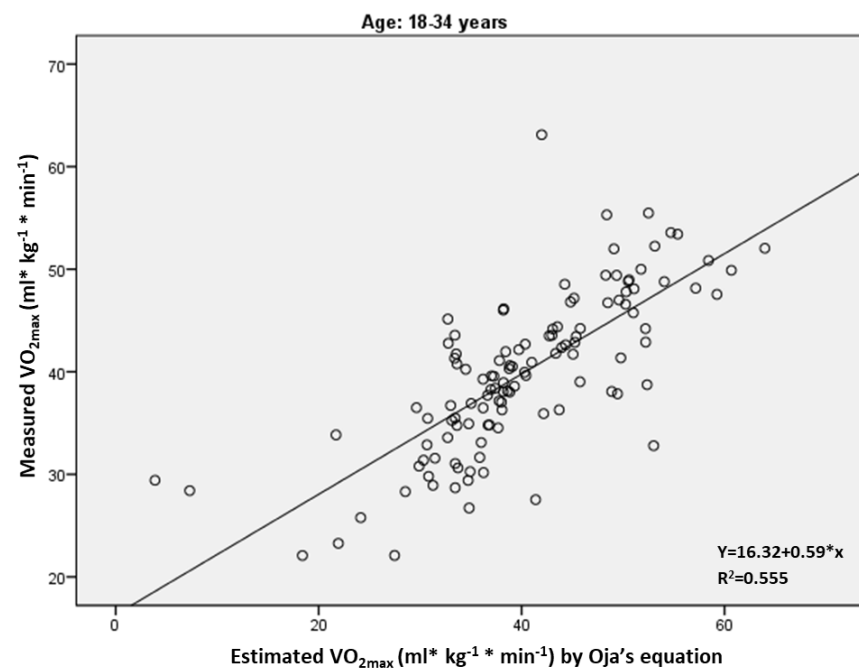
(A.1)



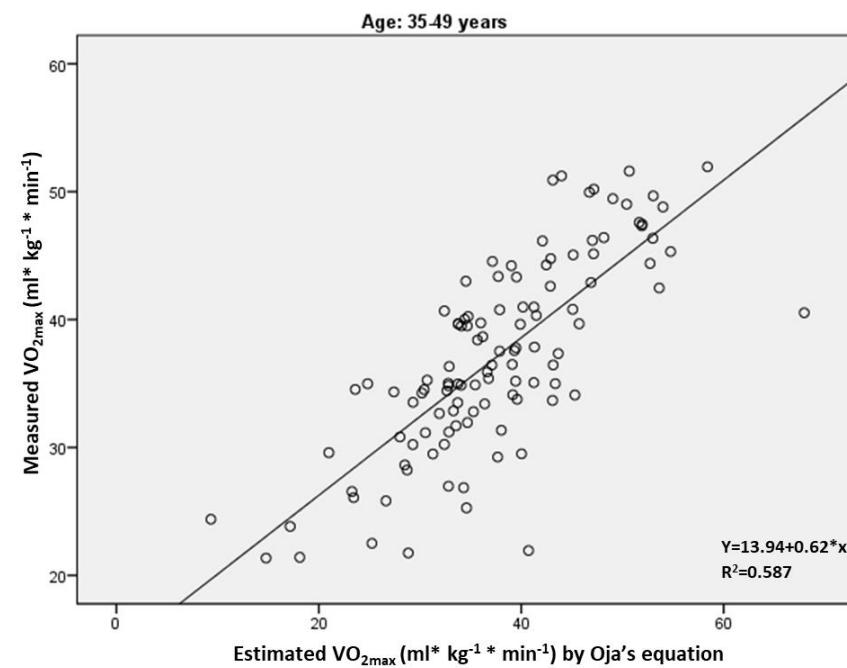
(A.2)



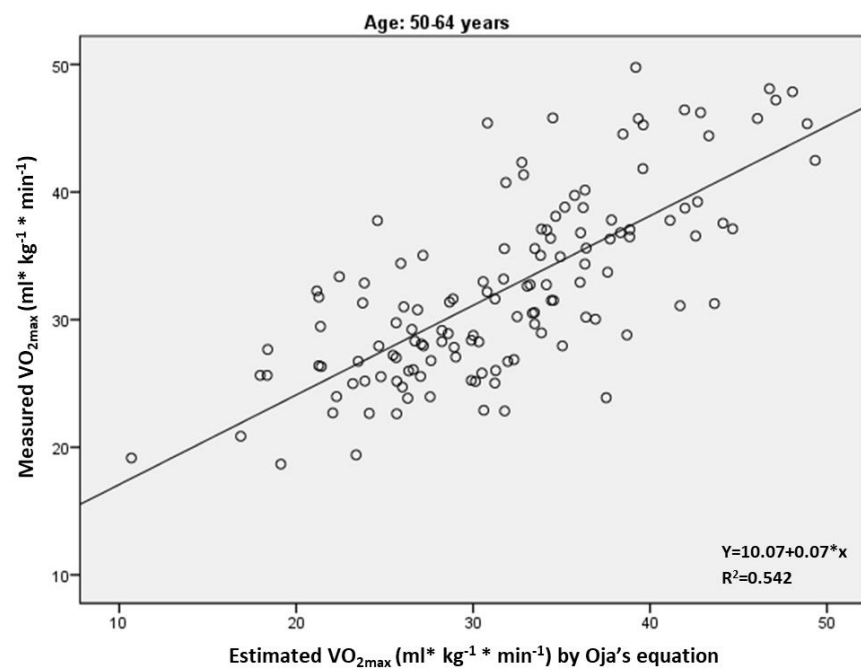
(A.3)



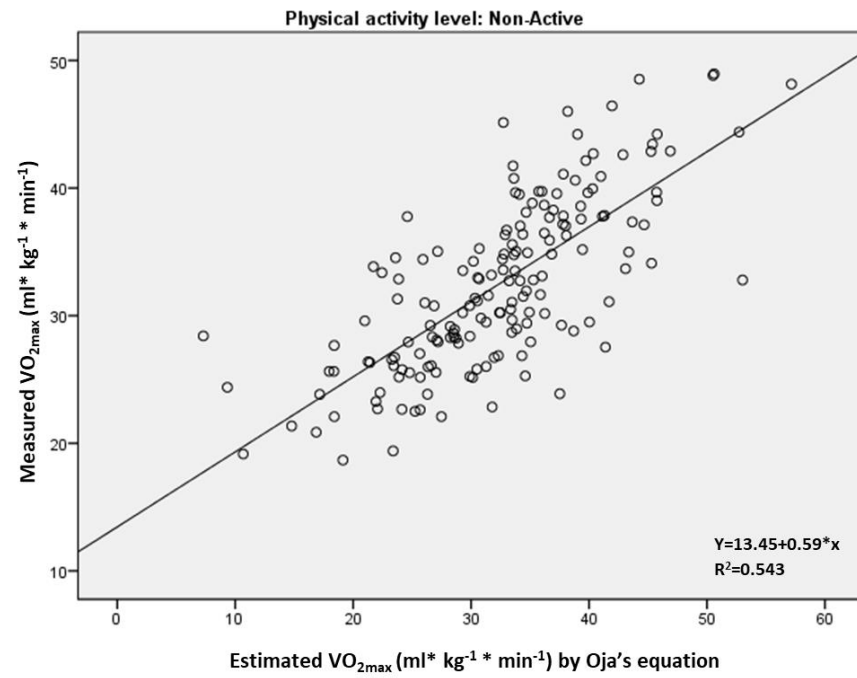
(A.4)



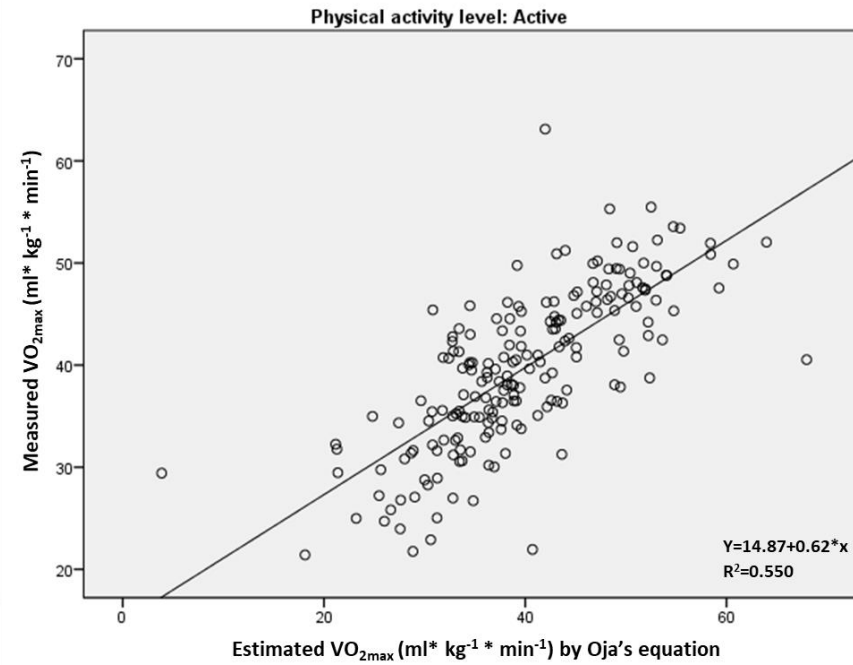
(A.5)



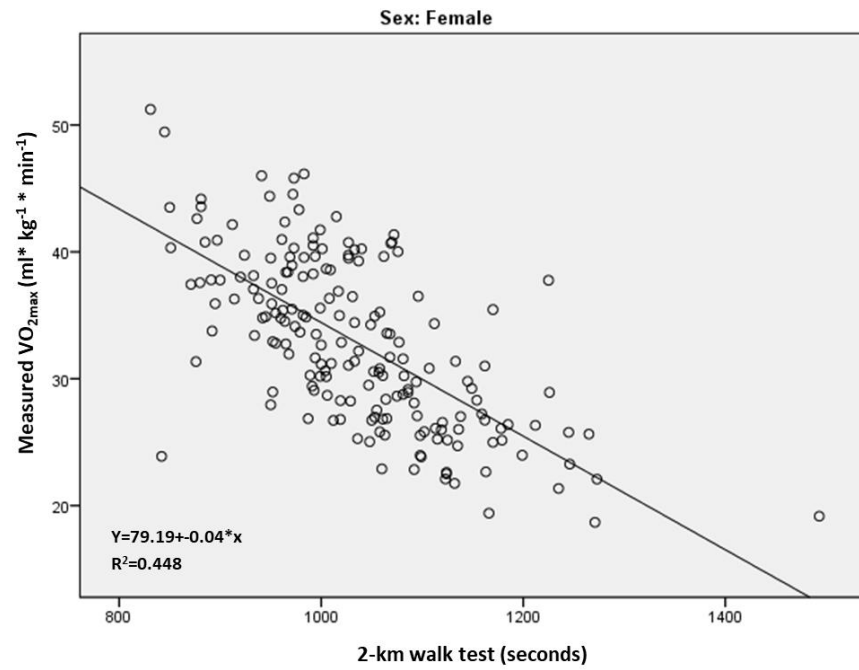
(A.6)



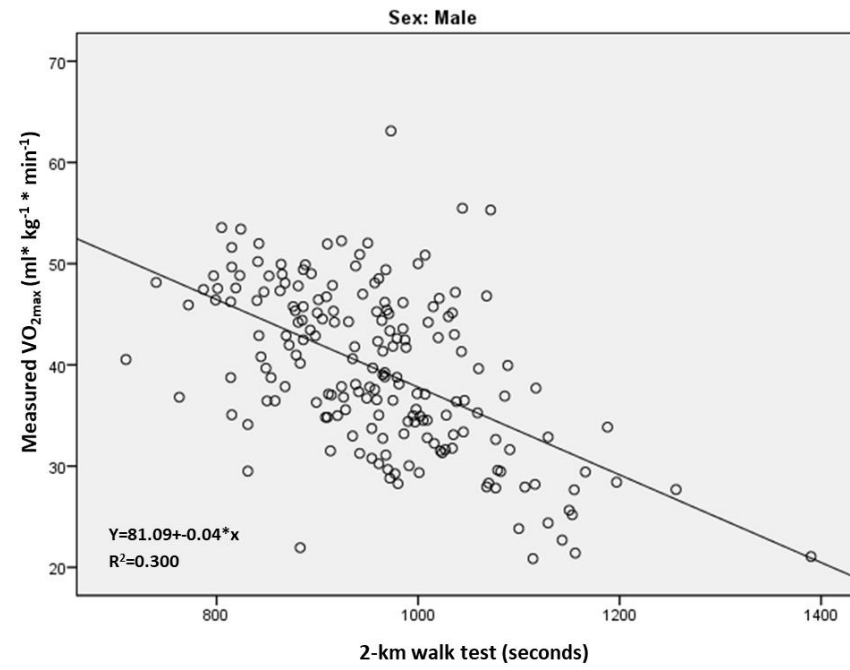
(A.7)



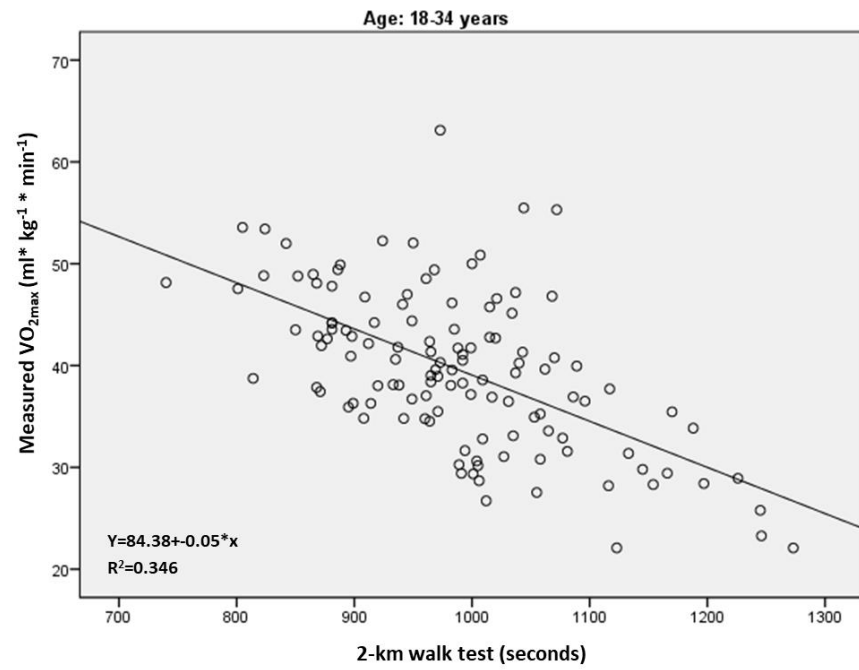
(B.1)



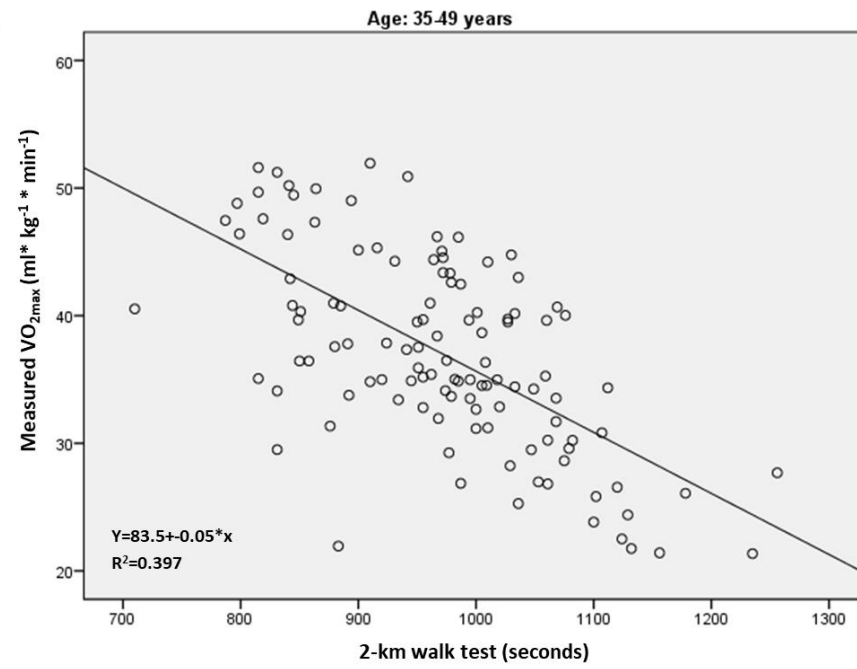
(B.2)



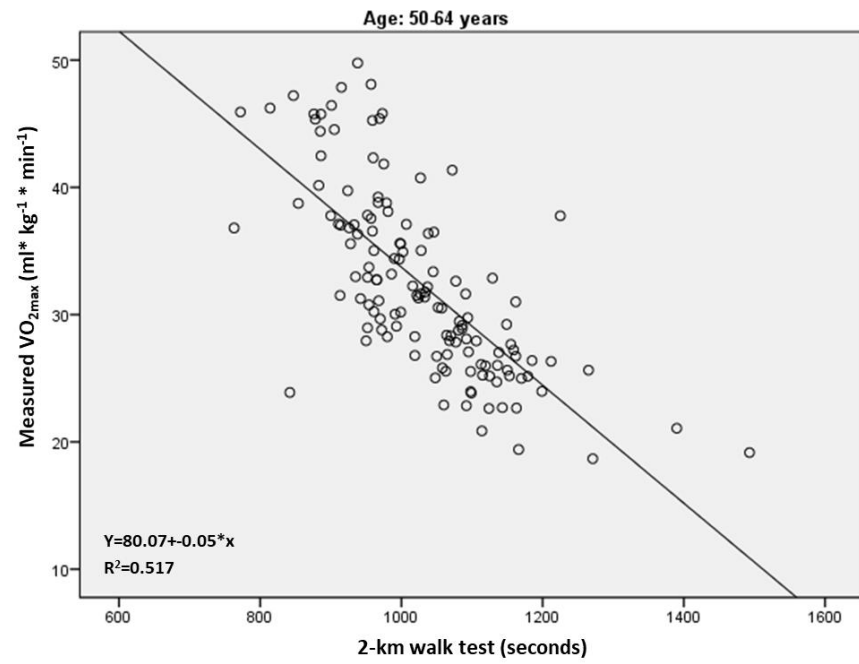
(B.3)



(B.4)



(B.5)



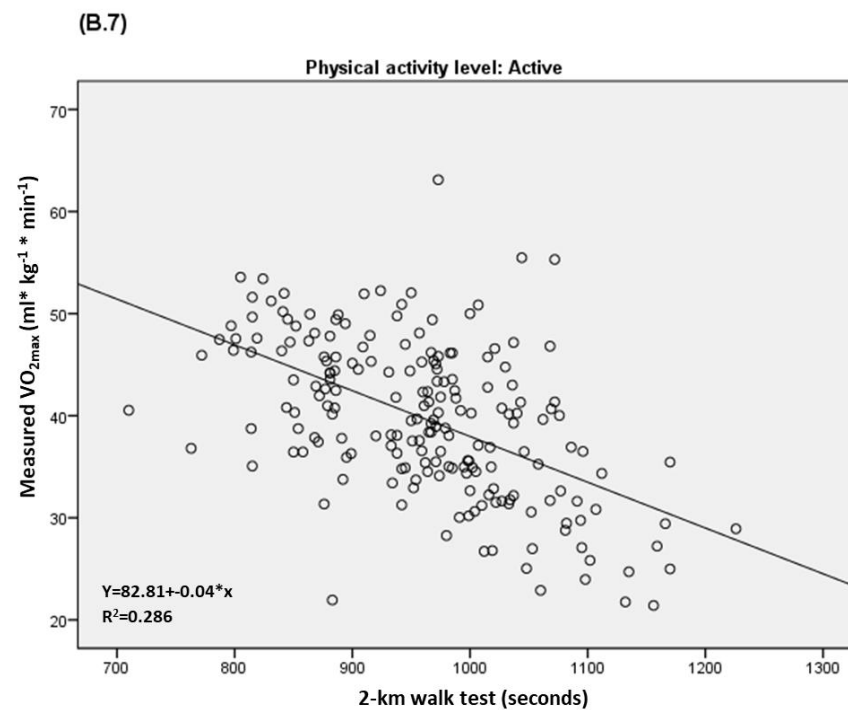
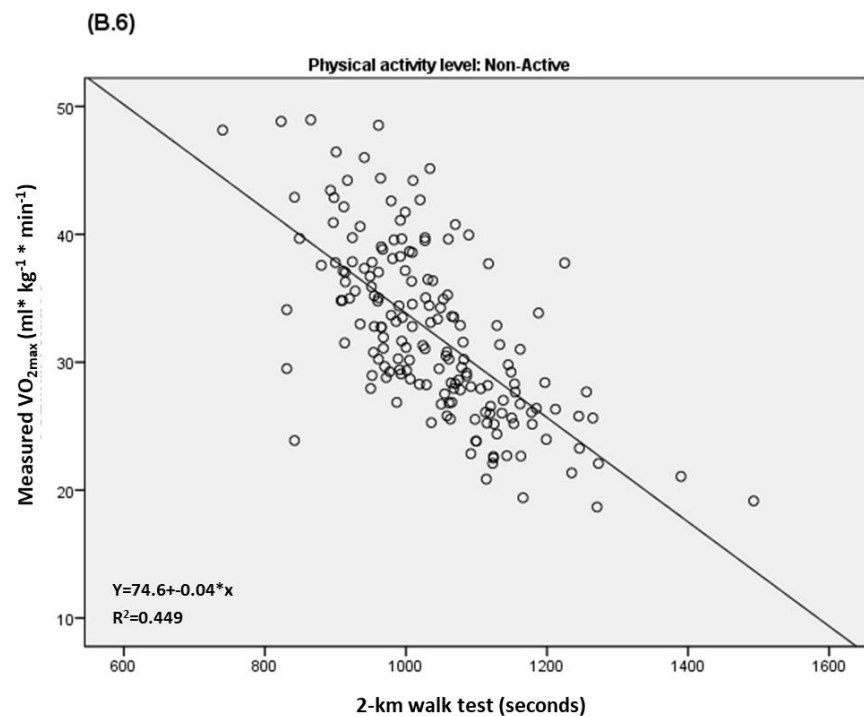
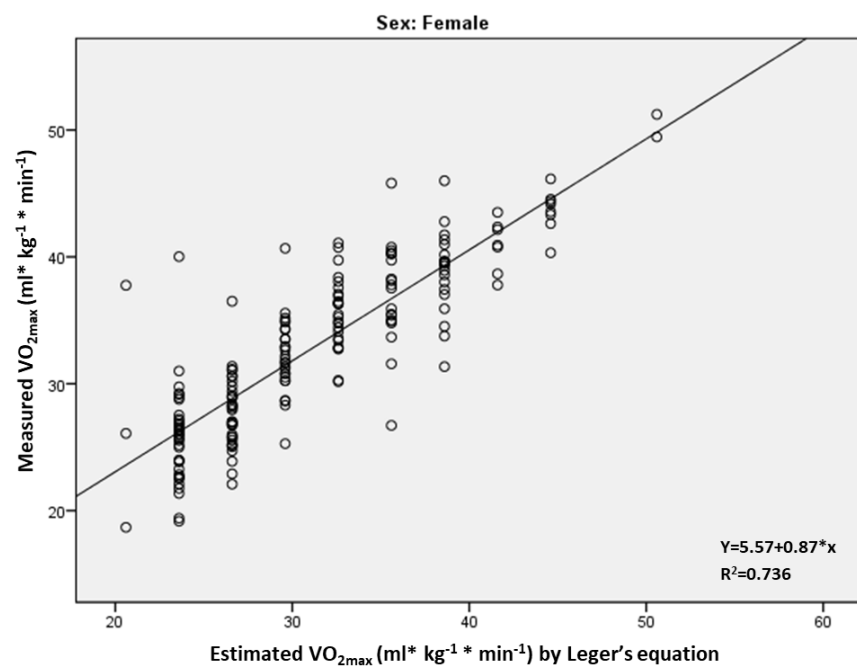


Figure 13. Scatterplot of relationship between measured with estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Oja's equation (estimated from the 2-km walk test), and relationship between measured with total time in the 2-km walk test; distributed by sex, age and physical activity levels.

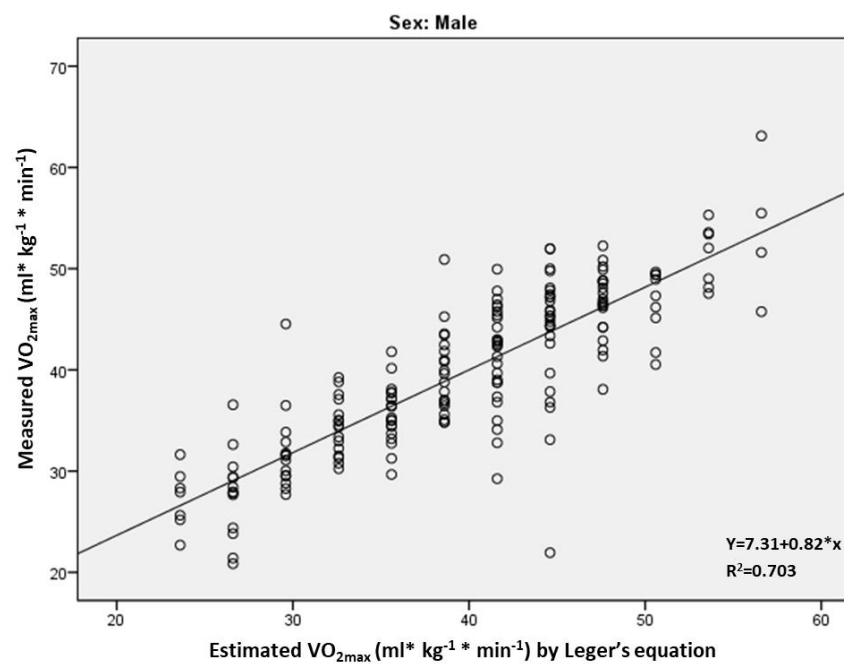
(A.1-A.2), relationship between measured and estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Oja's equation, distributed by sex, (A.3-A.5), age (A.6-A.7) and, physical activity level; (B.1-B.2), relationship between measured VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) and total time (sec) in the 2-km walk test, distributed by sex, (B.3-B.5), age (B.6-B.7), and physical activity level.

R^2 indicates lineal coefficient of determination.

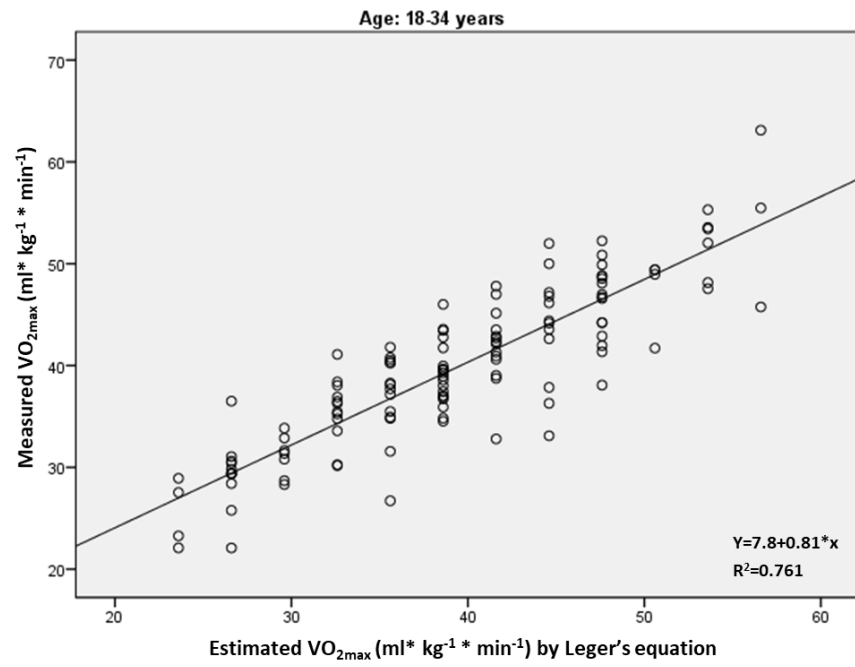
(A.1)



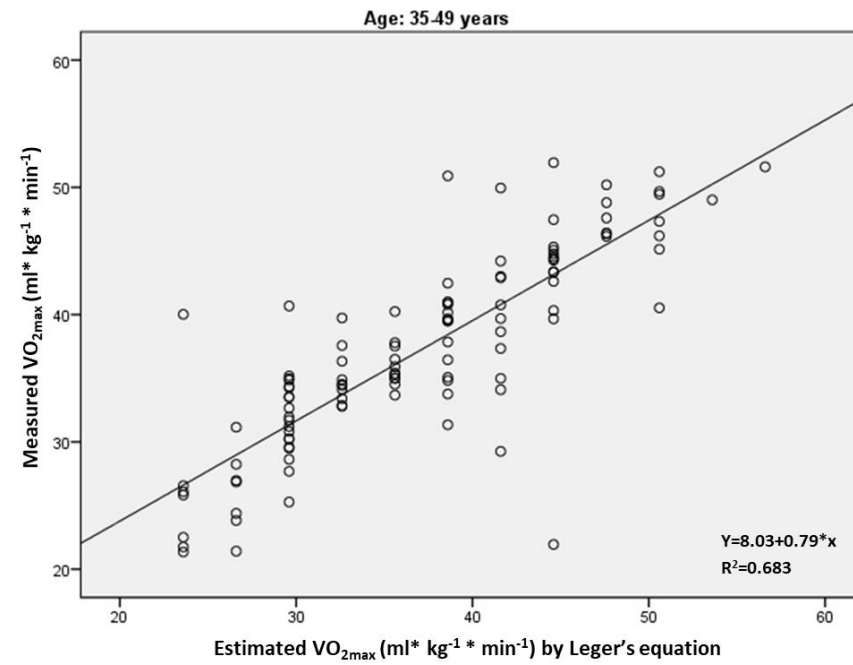
(A.2)



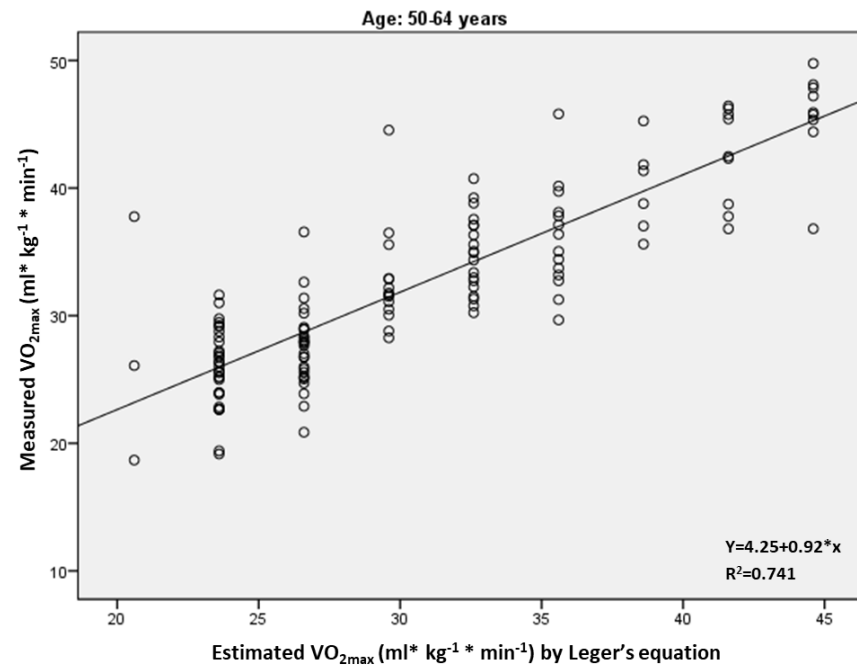
(A.3)



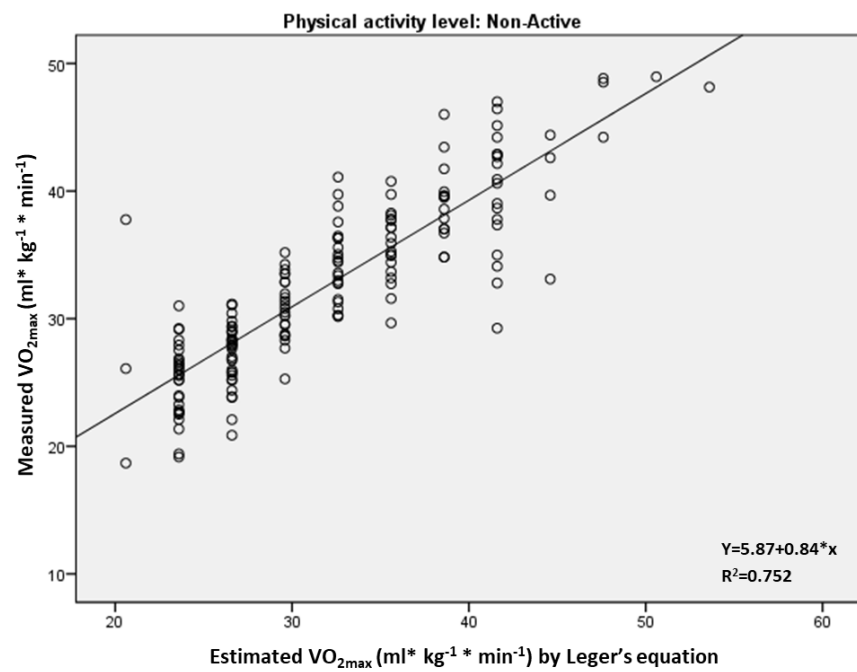
(A.4)



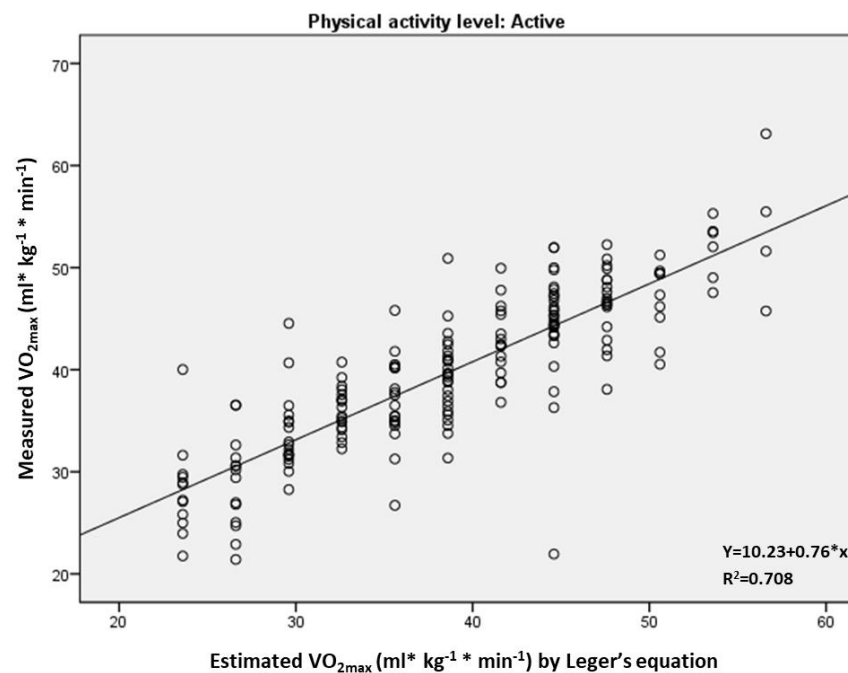
(A.5)



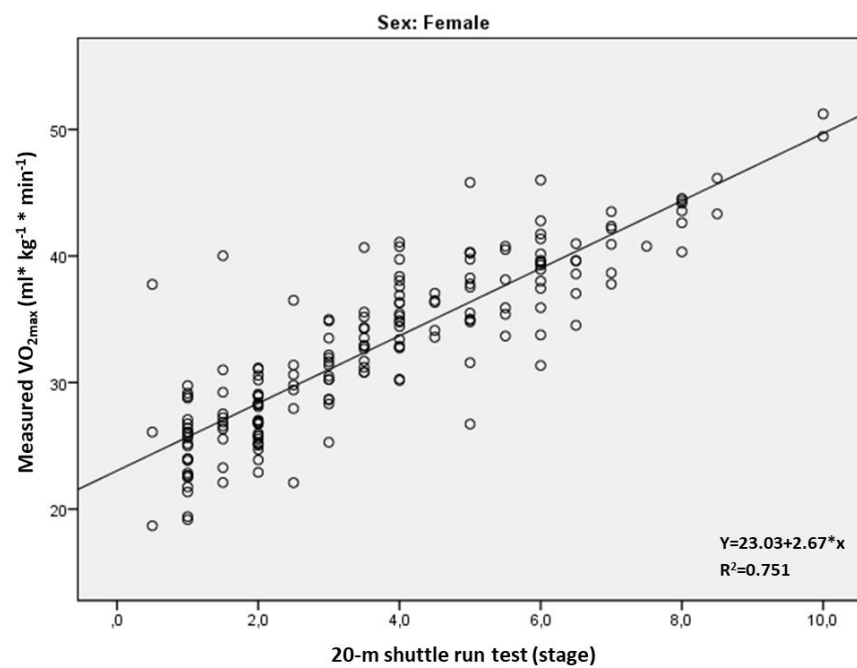
(A.6)



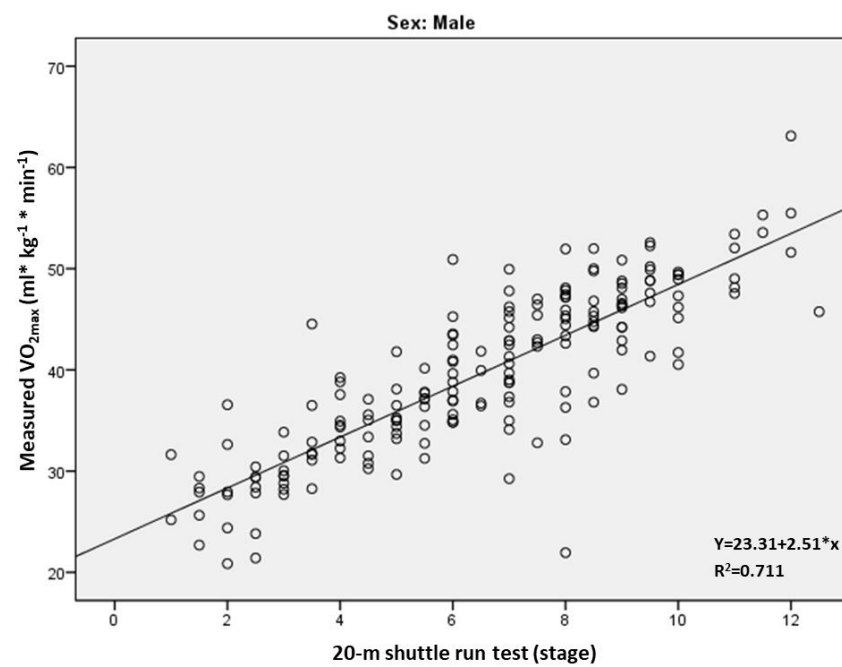
(A.7)



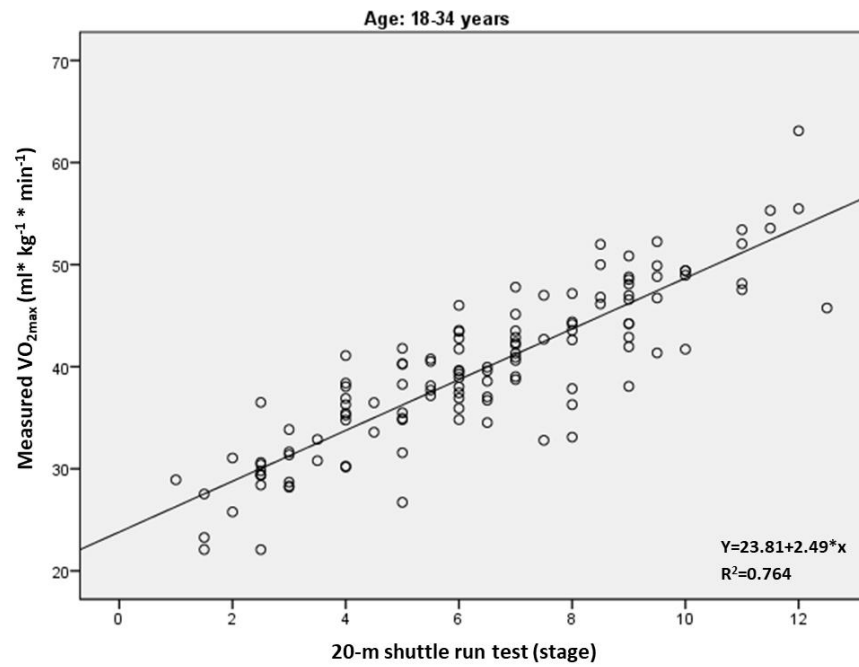
(B.1)



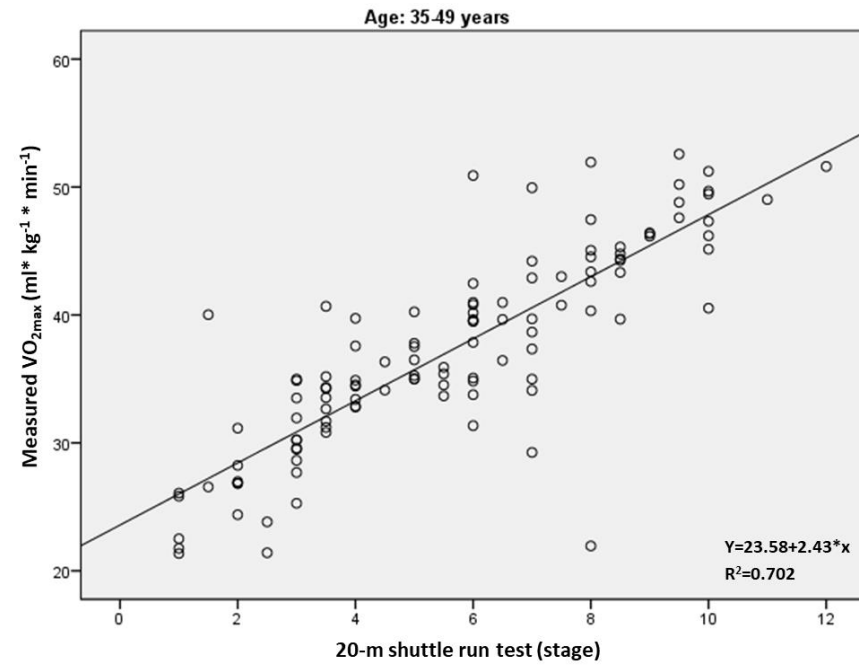
(B.2)



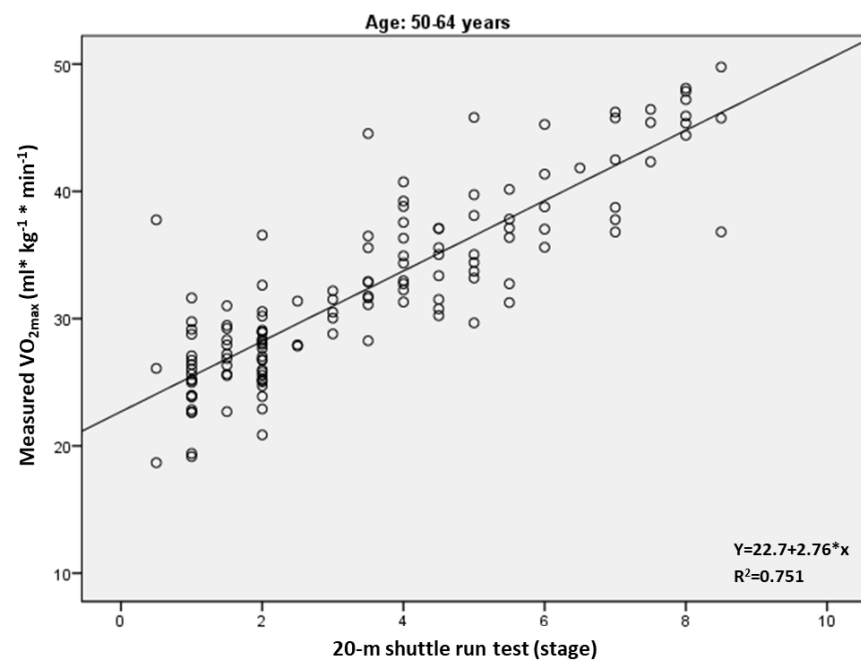
(B.3)



(B.4)



(B.5)



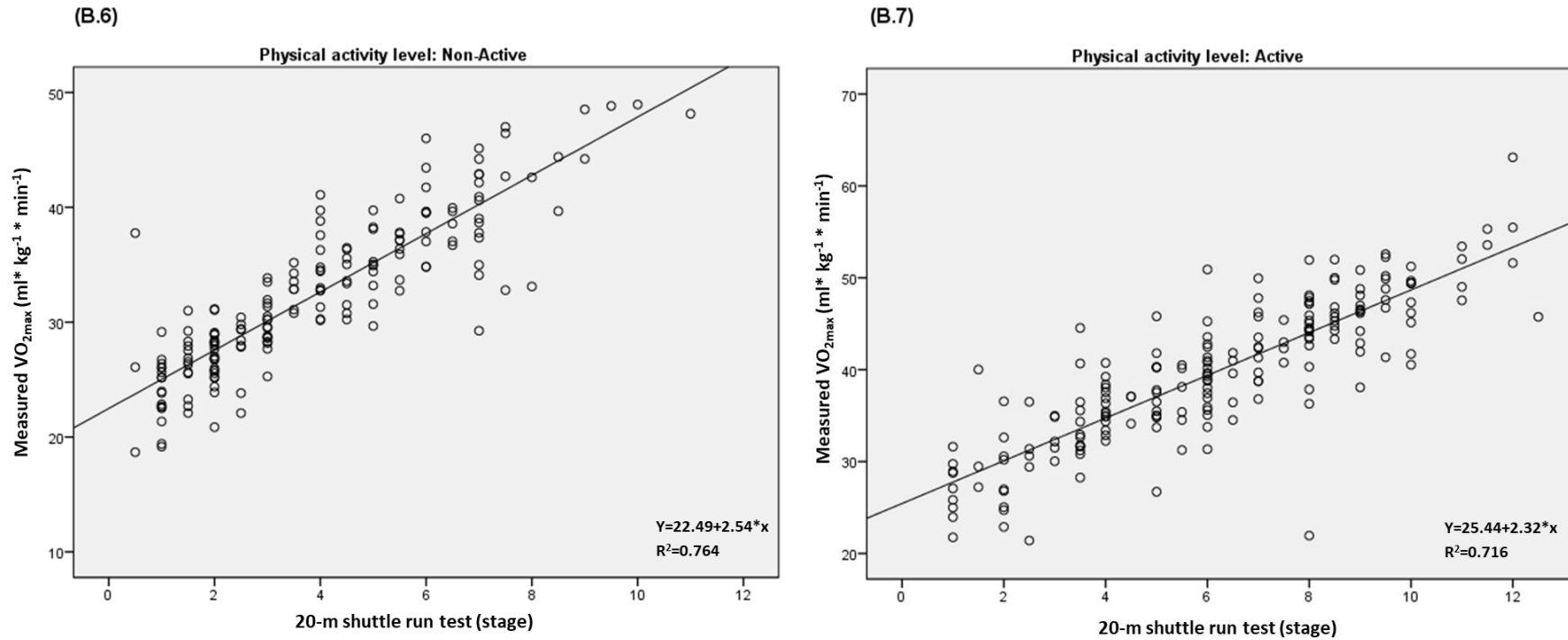


Figure 14. Scatterplot of relationship between measured with estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Leger's equation (estimated from the 20-m shuttle run test), and relationship between measured with final stage in the 20-m shuttle run test; distributed by sex, age, and physical activity levels.

(A.1-A.2), relationship between measured and estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Leger's equation, distributed by sex; (A.3-A.5 age; (A.6-A.7), and physical activity level; (B.1-B.2), relationship between measured VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) and final stage in the 20-m shuttle run test, distributed by sex; (B.3-B.5), age; (B.6-B.7), and physical activity level.

R^2 indicates lineal coefficient of determination.

Table 36. Bivariate correlation analysis between measured VO_{2max} (ml* kg⁻¹ * min⁻¹) with estimated VO_{2max} by 2-km walk test, by 20-meter shuttle run test, and anthropometric variables, distributed by whole sample, sex, age, and physical activity levels.

	Measured VO _{2max}	Sex	Age	PA level	Weight	Height	BMI	Body fat (%)	Lean mass	Waist circ.	Hip circ.	Neck circ.	Tricipital SKF	Subscapular SKF	SUM triceps + subscapular SKF
All															
Measured VO _{2max}	1	0.406	-0.380	0.460	-0.148	0.309	-0.391	-0.750	0.241	-0.338	-0.421	0.047	-0.648	-0.594	-0.656
Estimated VO _{2max} by 2-km walk test	0.784	0.380	-0.384	0.385	-0.202	0.376	-0.513	-0.730	0.222	-0.377	-0.449	0.022	-0.699	-0.668	-0.723
Estimated VO _{2max} by 20-m SRT	0.875	0.483	-0.446	0.409	-0.009	0.416	-0.298	-0.751	0.355	-0.251	-0.372	0.149	-0.650	-0.563	-0.636
Female															
Measured VO _{2max}	1	-0.440	0.415	-0.446	0.069	-0.509	-0.695	-0.008	-0.583	-0.439	-0.426	-0.539	-0.534	-0.569	
Estimated VO _{2max} by 2-km walk test	0.760	-0.499	0.351	-0.463	0.274	-0.623	-0.681	-0.113	-0.685	-0.482	-0.512	-0.613	-0.620	-0.659	
Estimated VO _{2max} by 20-m SRT	0.858	-0.513	0.370	-0.374	0.119	-0.450	-0.679	0.010	-0.578	-0.424	-0.389	-0.545	-0.508	-0.553	
Male															
Measured VO _{2max}	1	-0.402	0.477	-0.560	-0.028	-0.617	-0.689	-0.281	-0.728	-0.518	-0.486	-0.620	-0.666	-0.681	

Estimated VO _{2max} by 2-km walk test	0.757	-0.422	0.383	-0.612	0.090	-0.740	-0.705	-0.225	-0.699	-0.515	-0.461	-0.687	-0.725	-0.744
Estimated VO _{2max} by 20-m SRT	0.839	-0.556	0.420	-0.447	0.097	-0.560	-0.667	-0.144	-0.662	-0.465	-0.418	-0.600	-0.649	-0.659
	Measured VO _{2max}	Sex	PA level	Weight	Height	BMI	Body fat (%)	Lean mass	Waist circ.	Hip circ.	Neck circ.	Tricipital SKF	Subscapular SKF	SUM triceps + subscapular SKF
18-34 years														
Measured VO _{2max}	1	0.502	0.447	-0.130	0.347	-0.354	-0.756	0.249	-0.220	-0.429	0.172	-0.733	-0.571	-0.662
Estimated VO _{2max} by 2-km walk test	0.745	0.498	0.370	-0.260	0.408	-0.540	-0.777	0.169	-0.320	-0.575	0.129	-0.800	-0.740	-0.797
Estimated VO _{2max} by 20-m SRT	0.873	0.596	0.418	-0.032	0.503	-0.314	-0.766	0.375	-0.119	-0.441	0.332	-0.742	-0.574	-0.670
35-49 years														
Measured VO _{2max}	1	0.369	0.429	-0.223	0.329	-0.477	-0.778	0.212	-0.362	-0.540	0.015	-0.746	-0.697	-0.753
Estimated VO _{2max} by 2-km walk test	0.766	0.437	0.308	-0.132	0.459	-0.522	-0.648	0.336	-0.226	-0.361	0.140	-0.734	-0.616	-0.699
Estimated V VO _{2max} by 20-m SRT	0.826	0.536	0.392	0.028	0.459	-0.313	-0.784	0.394	-0.205	-0.439	0.197	-0.736	-0.623	-0.702
50-64 years														
Measured VO _{2max}	1	0.486	0.496	-0.018	0.289	-0.272	-0.687	0.309	-0.170	-0.382	0.156	-0.511	-0.482	-0.551

Estimated VO _{2max} by 2-km walk test	0.736	0.380	0.450	-0.094	0.349	-0.425	-0.732	0.256	-0.337	-0.444	0.028	-0.601	-0.596	-0.659
Estimated VO _{2max} by 20-m SRT	0.861	0.576	0.409	0.163	0.424	-0.125	-0.710	0.456	-0.051	-0.305	0.239	-0.554	-0.448	-0.546
	Measured VO _{2max}	Sex	Age	Weight	Height	BMI	Body fat (%)	Lean mass	Waist circ.	Hip circ.	Neck circ.	Tricipital SKF	Subscapular SKF	SUM triceps + subscapular SKF
Non-Active														
Measured VO _{2max}	1	0.323	-0.445	-0.178	0.269	-0.386	-0.701	0.169	-0.322	-0.341	0.039	-0.514	-0.487	-0.535
Estimated VO _{2max} by 2-km walk test	0.737	0.234	-0.402	-0.304	0.294	-0.570	-0.655	0.105	-0.389	-0.379	-0.061	-0.626	-0.598	-0.653
Estimated VO _{2max} by 20-m SRT	0.867	0.441	-0.485	-0.012	0.368	-0.263	-0.731	-0.304	-0.227	-0.317	0.156	-0.535	-0.433	-0.508
Active														
Measured VO _{2max}	1	0.445	-0.332	-0.100	0.337	-0.377	-0.735	0.254	-0.273	-0.421	0.070	-0.691	-0.609	-0.686
Estimated VO _{2max} by 2-km walk test	0.742	0.449	-0.358	-0.117	0.439	-0.466	-0.733	0.263	-0.302	-0.450	0.100	-0.704	-0.675	-0.731
Estimated VO _{2max} by 20-m SRT	0.841	0.495	-0.433	0.027	0.454	-0.291	-0.719	0.368	-0.186	-0.341	0.180	-0.674	-0.603	-0.674

Abrev.: BMI, Body Mass Index; Circ., circumferences; SKF, Skinfolds; 20-m SRT, 20-meters shuttle run test.
 Bold values were significant at p< 0.01.

Table 37. Stepwise linear regression model predicting VO_{2max} by 2-km walk test and, by 20-meter shuttle run test.

Model	Independent variables	β	p value	r	R ²	R ² change	SEE
2-km walk test							
1	2-km total time	-0.652	<0.001	0.652	0.423	0.425	6.136
2	2-km total time	-0.700	<0.001	0.685	0.466	0.044	5.903
	2-km heart rate	-0.216	<0.001				
3	2-km total time	-0.625	<0.001	0.703	0.489	0.025	5.772
	2-km heart rate	-0.176	<0.001				
	Sex	0.175	<0.001				
4	2-km total time	-0.559	<0.001	0.777	0.600	0.110	5.111
	2-km heart rate	-0.224	<0.001				
	Sex	0.200	<0.001				
	Age	-0.344	<0.001				
5	2-km total time	-0.486	<0.001	0.795	0.626	0.027	4.940
	2-km heart rate	-0.175	<0.001				
	Sex	0.205	<0.001				
	Age	-0.331	<0.001				
	Physical activity level	0.182	<0.001				
6	2-km total time	-0.313	<0.001	0.841	0.702	0.076	4.408
	2-km heart rate	-0.113	0.001				
	Sex	0.065	0.065				
	Age	-0.239	<0.001				
	Physical activity level	0.158	<0.001				
	% Body fat	-0.403	<0.001				
20-m shuttle run test							
1	Final stage	0.881	<0.001	0.881	0.776	0.777	3.841
2	Final stage	0.893	<0.001	0.881	0.776	0.000	3.843
	Sex	-0.023	0.423				
3	Final stage	0.911	<0.001	0.882	0.776	0.001	3.843

	Sex	-0.032	0.285				
	Age	-0.028	0.333				
4	Final stage	0.848	<0.001	0.888	0.786	0.0011	3.753
	Sex	-0.017	0.570				
	Age	-0.009	0.746				
	Physical activity level	0.116	<0.001				
5	Final stage	0.709	<0.001	0.897	0.803	0.017	3.603
	Sex	-0.053	0.069				
	Age	0.001	0.985				
	Physical activity level	0.110	<0.001				
	% Body fat	-0.206	<0.001				

β, Standardized regression coefficient; r, correlation coefficients; R², adjusted coefficients of determination; SEE, Standard Error of Estimate.

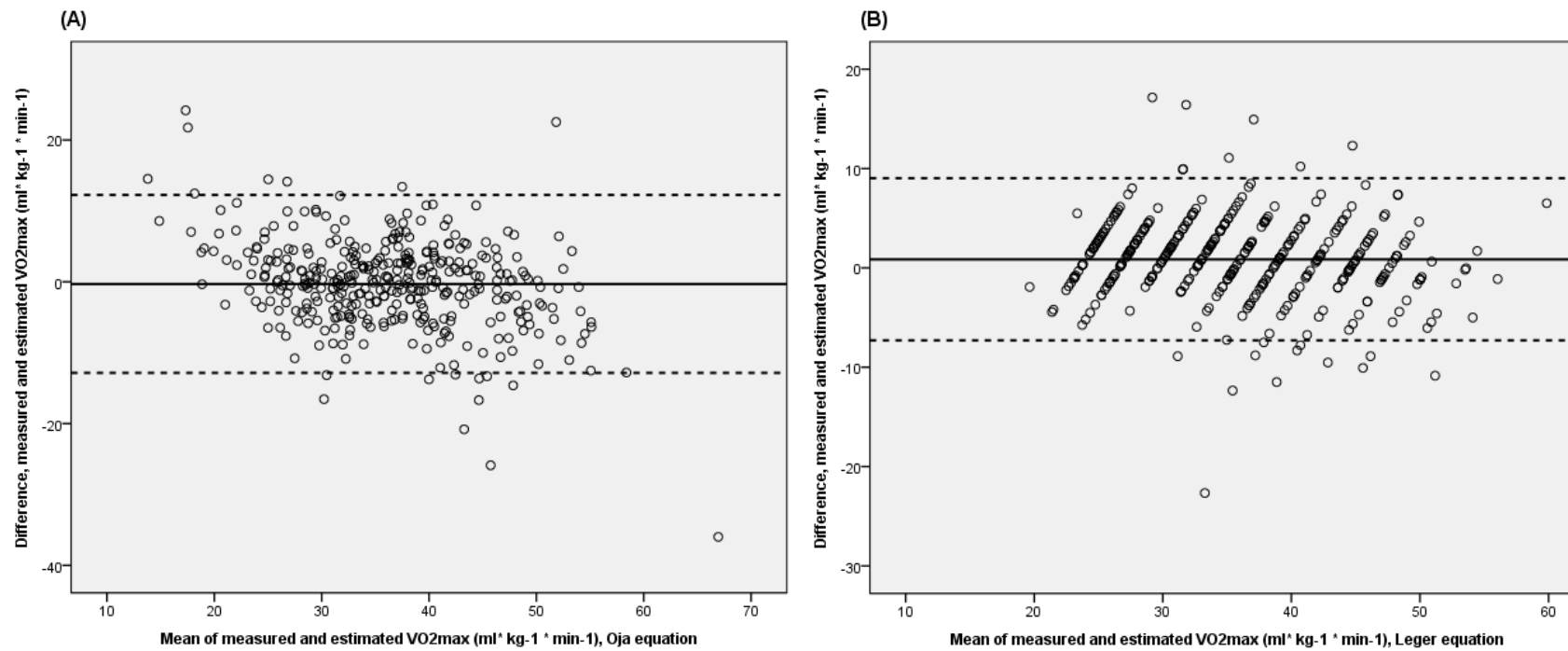
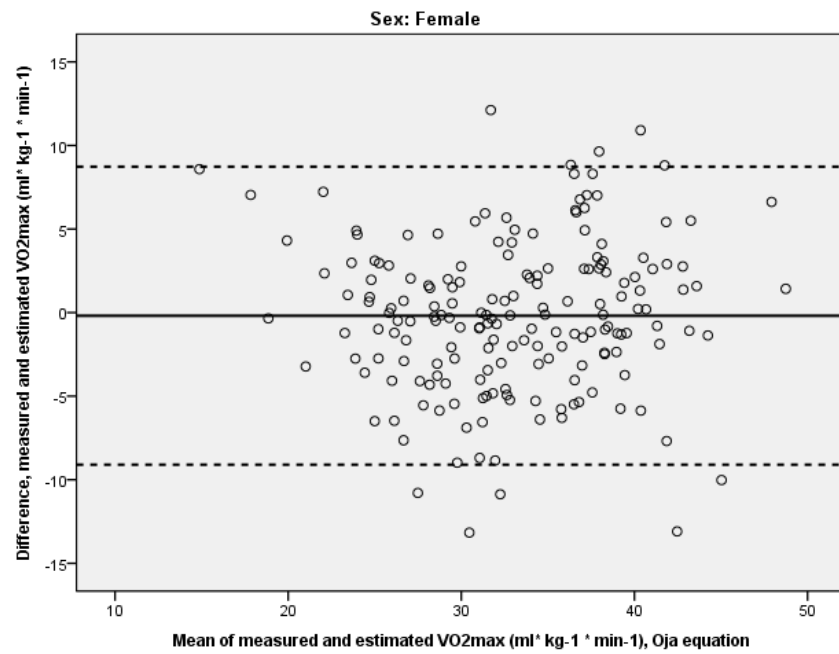


Figure 15. Bland-Altman plot of agreement between measured and estimated $\text{VO}_{2\text{max}}$ (ml*kg⁻¹*min⁻¹) by Oja's equation (estimated from the 2-km walk test), and Leger's equation (estimated from the 20-m shuttle run test).

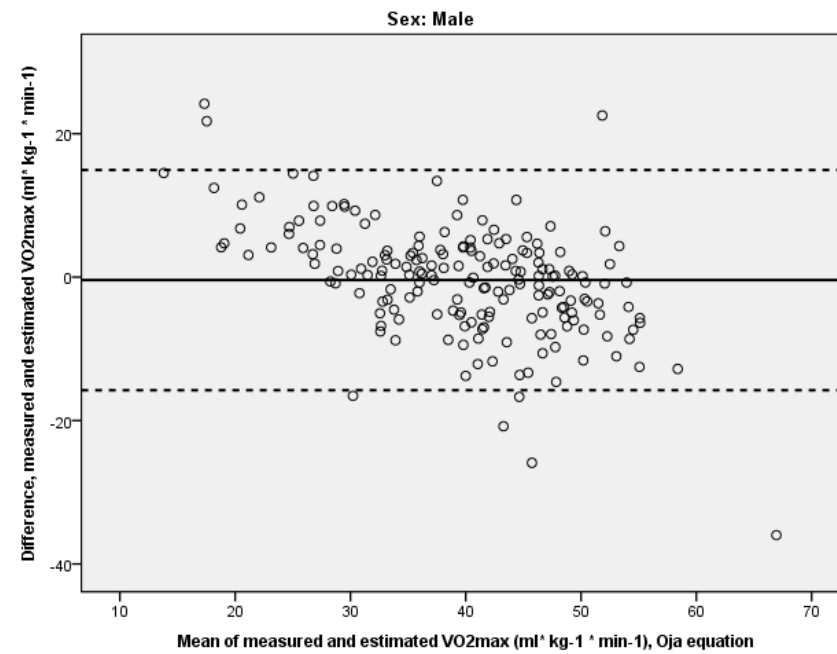
(A), agreement between measured and estimated $\text{VO}_{2\text{max}}$ (ml*kg⁻¹*min⁻¹) by Oja's equation; (B), agreement between measured and estimated $\text{VO}_{2\text{max}}$ (ml*kg⁻¹*min⁻¹) by Leger's equation.

The central line represents the mean differences between measured and estimated $\text{VO}_{2\text{max}}$. The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean difference $\pm$ 1.96 SD of the differences), respectively.

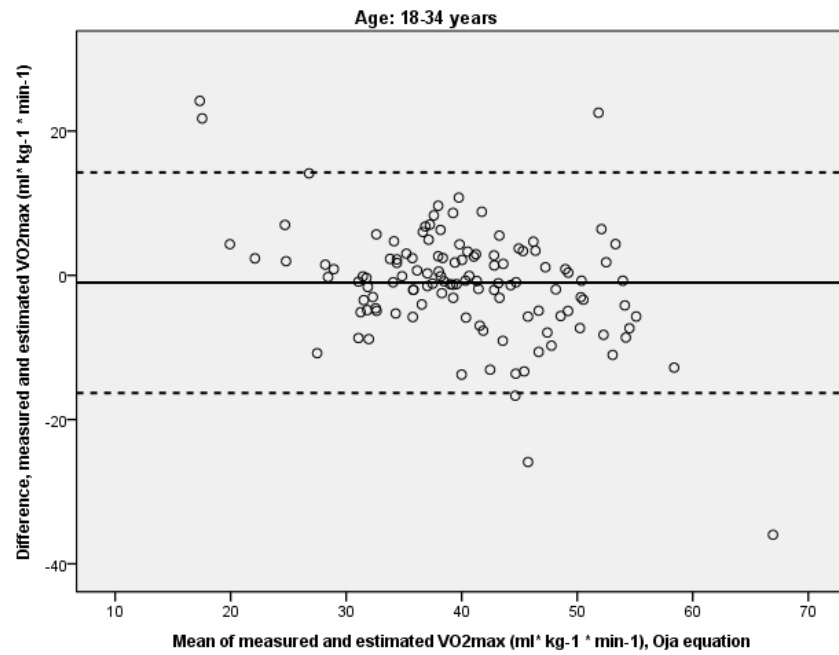
(A.1)



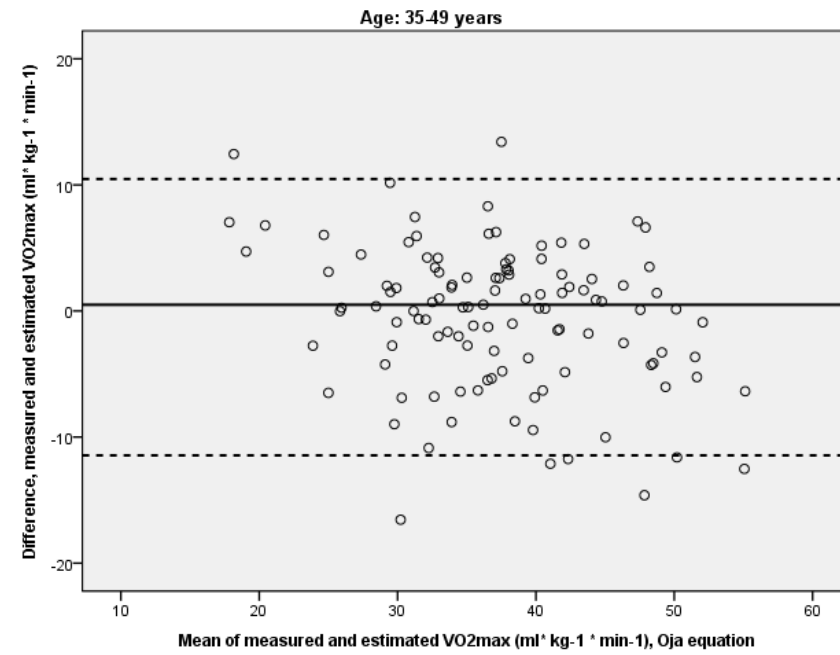
(A.2)



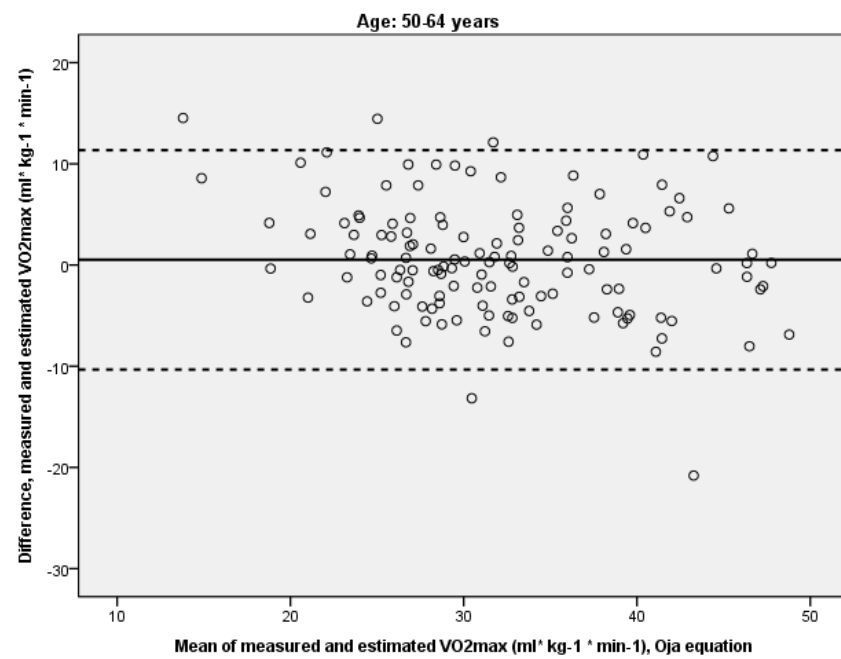
(A.3)



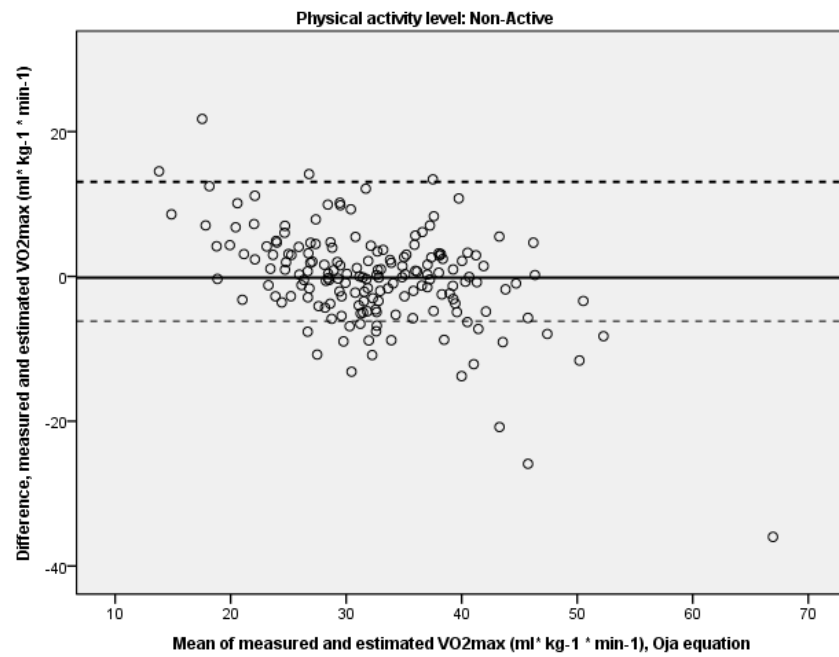
(A.4)



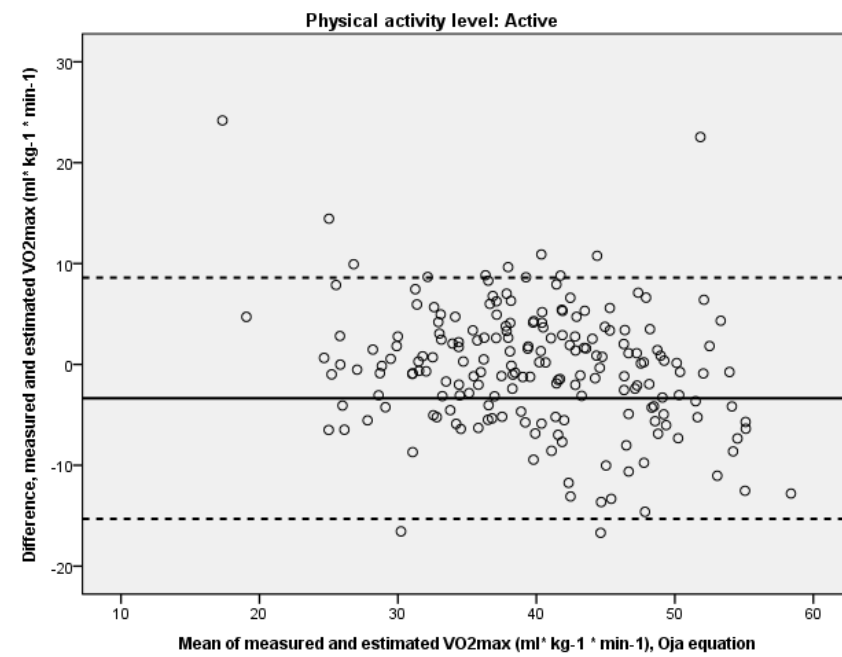
(A.5)



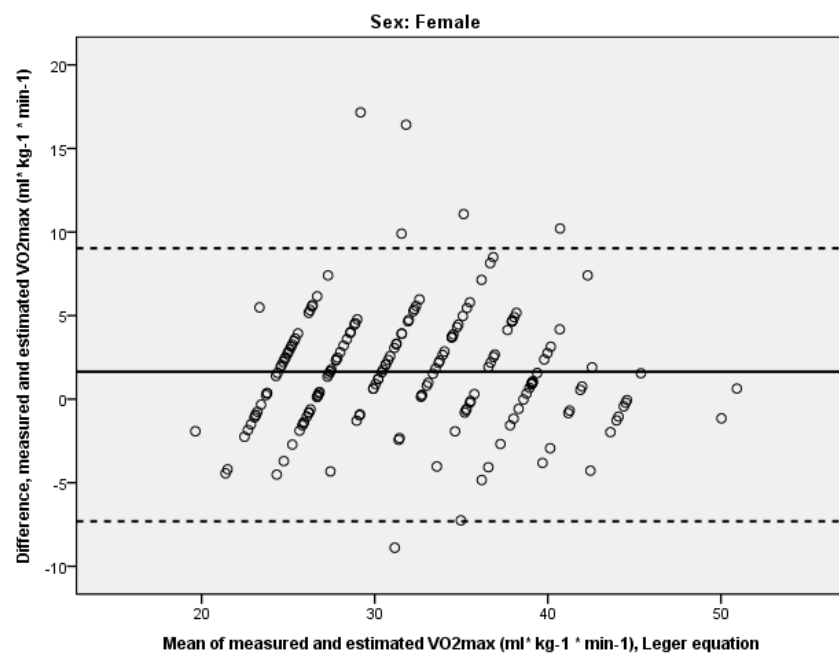
(A.6)



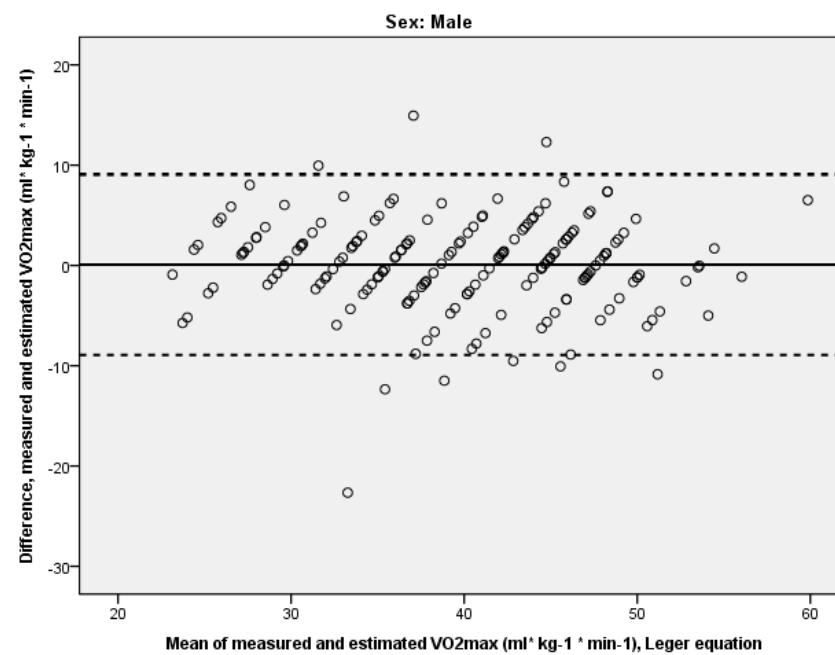
(A.7)



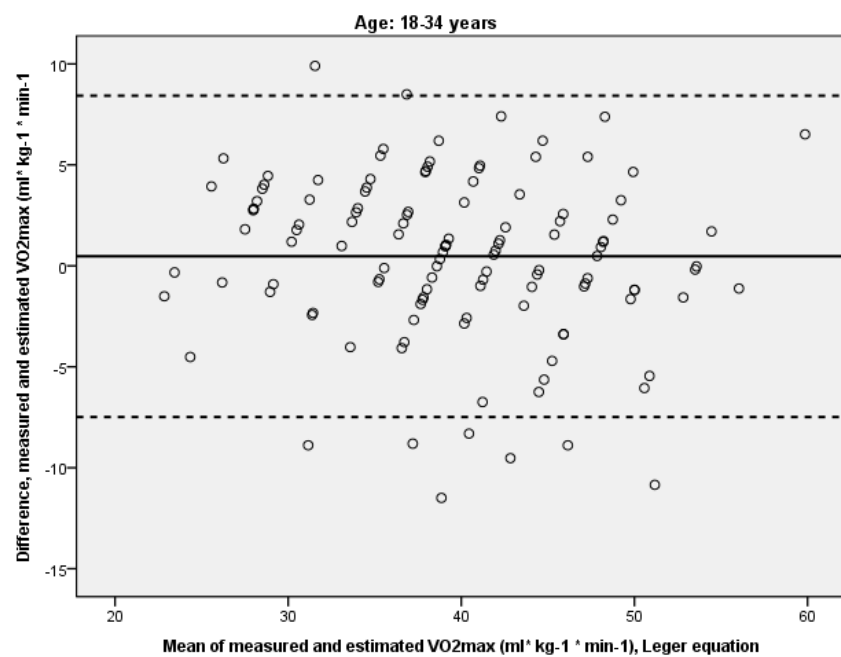
(B.1)



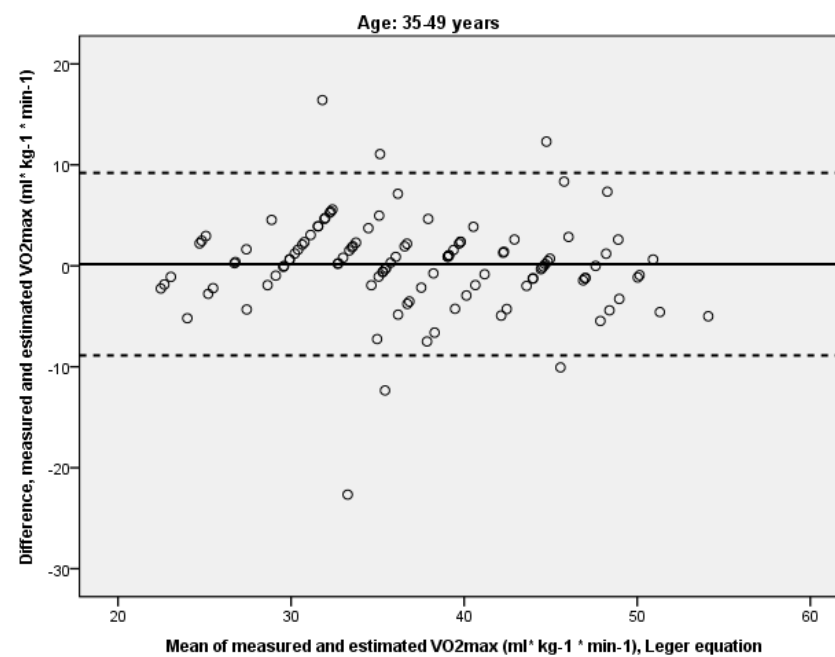
(B.2)



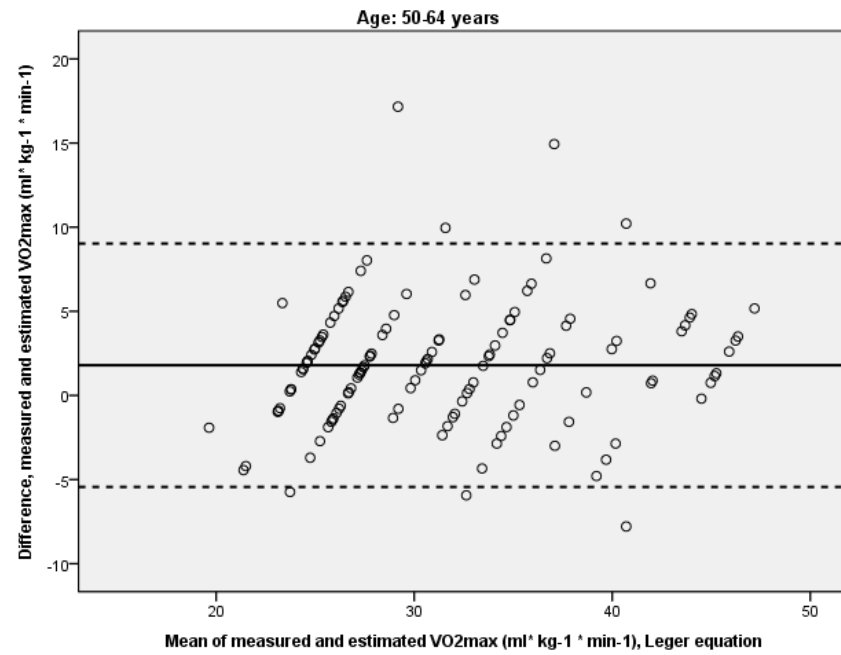
(B.3)



(B.4)



(B.5)



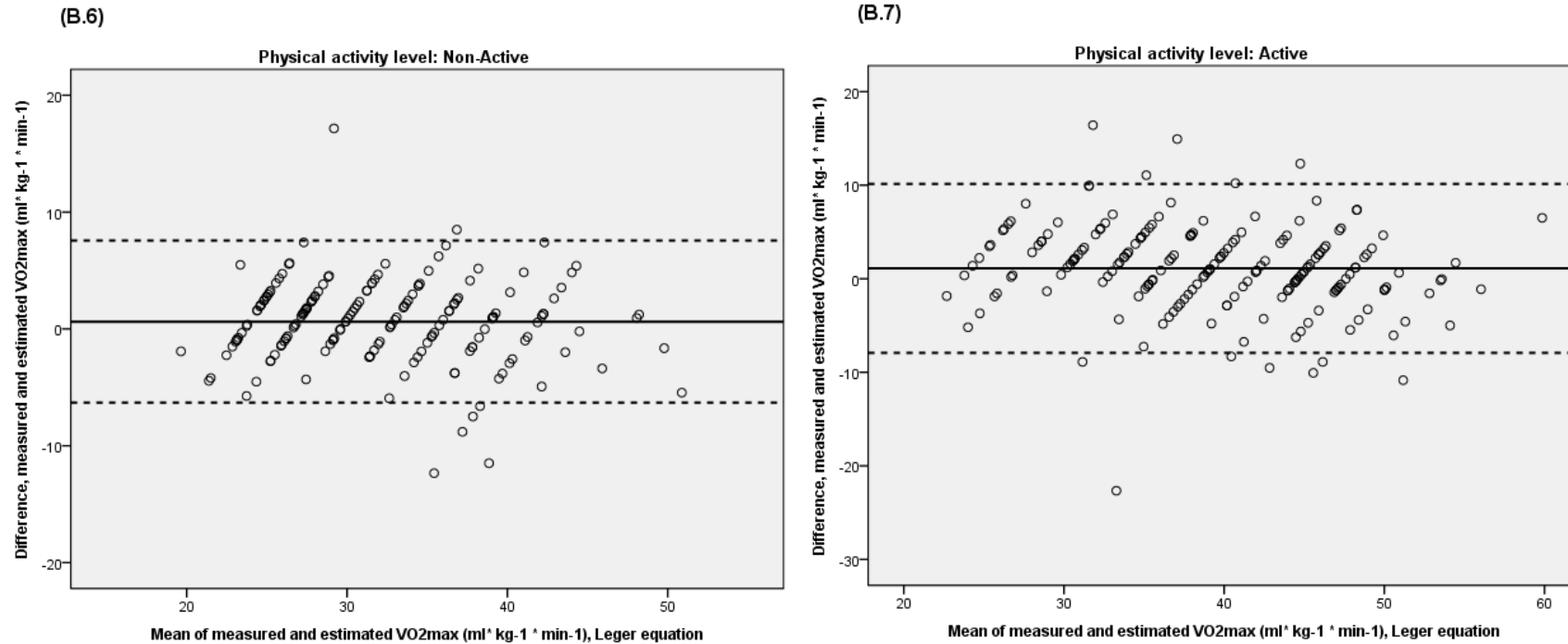


Figure 16. Bland-Altman plot of agreement between measured and estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation (estimated from the 2-km walk test) and Leger's equation (estimated from the 20-m shuttle run test), distributed by sex, age, and physical activity levels. (A.1-A.2), relationship between measured and estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation, distributed by sex; (A.3-A.5), distributed by age; (A.6-A.7), distributed by physical activity level. (B.1-B.2), distributed by sex; (B.3-B.5), distributed by age; (B.6-B.7), distributed by physical activity level. The central line represents the mean differences between measured and estimated $\text{VO}_{2\text{max}}$. The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean difference ± 1.96 SD of the differences), respectively.

Table 38. Test-retest reliability of 2-km walk test and 20-m shuttle run test in the whole sample.

	T1*	T2*	Intertrial difference (T2-T1)	p value	Cohen's <i>d</i>	ICC (95% CI)**	SSE	MSE	RMSE	% Error	% CV	% SEM	SEE
All													
2-km walk test (sec)	997.70 ± 108.22	994.81 ± 104.80	-1.48 ± 0.51	0.004	-0.014	0.99 (0.99-0.99)	38479.00	101.80	10.09	1.29	0.71	0.15	9.88
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	36.58 ± 10.14	36.43 ± 9.62	-0.29 ± 0.20	0.206		0.95 (0.94-0.96)	6042.61	16.02	4.00	5.02	7.73	0.79	3.80
20-m SRT (stage)	5.10 ± 2.82	5.15 ± 2.85	0.04 ± 0.01	0.002	0.015	0.99 (0.99-0.99)	54.50	0.14	0.38	3.16	3.45	0.98	0.25
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	35.42 ± 8.51	35.42 ± 8.49	0.03 ± 0.04	0.433		0.99 (0.99-0.99)	234.00	0.62	0.78	2.18	1.57	0.09	0.78

2-km walk test is expressed as total time to complete the test (sec), and as estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Oja's equation; 20-m shuttle run test is expressed as final stage reached, and as estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Leger's equation.

*T1 refers to test (trial 1) and T2 to retest (trial 2).

T2-T1 refers to retest (trial 2) minus test (trial 1). Values are displayed as mean ± SD.

**All the ICCs were significant at p< 0.001.

ICC, intraclass correlation coefficients; CI, confident interval; SSE, sum of squared errors; MSE, mean sum of squared errors; RMSE, root mean sum of squared errors; %Error, percentage error; %SEM, standard error of measurement; %CV, percentage coefficient of variation; SEE, standard error of estimate.

Table 39. Test-retest reliability of 2-km walk test and 20-m shuttle run test, distributed by sex, age, and physical activity levels.

	T1*	T2*	Intertrial difference (T2-T1)	p value	Cohen’s <i>d</i>	ICC (95% CI)*	SSE	MSE	RMSE	% Error	% CV	% SEM	SEE
Female													
2-km walk test (sec)	1035.55 ± 99.81	1034.28 ± 99.96	-2.34 ± 0.66	0.001	-0.023	0.99 (0.99-0.99)	16963.00	89.75	9.47	1.42	0.63	0.23	9.17
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	33.09 ± 6.43	32.78 ± 6.19	-0.21 ± 0.17	0.214		0.96 (0.95-0.97)	1059.95	5.63	2.37	6.01	5.10	0.65	2.27
20-m SRT (stage)	3.71 ± 2.17	3.72 ± 2.16	0.03 ± 0.01	0.051		0.99 (0.99-0.99)	9.50	0.05	0.22	2.37	4.26	0.87	0.22
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	31.30 ± 6.57	31.22 ± 6.53	0.06 ± 0.06	0.286		0.99 (0.99-0.99)	126.16	0.67	0.82	2.73	1.86	0.20	0.81
Male													
2-km walk test (sec)	959.86 ± 103.16	955.34 ± 94.35	-0.62 ± 0.77	0.423		0.99 (0.99-0.99)	21516.00	113.84	10.67	2.19	0.79	0.07	10.52
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	40.06 ± 11.84	40.08 ± 10.97	-0.36 ± 0.37	0.328		0.94 (0.92-0.95)	4982.56	26.36	5.13	6.44	9.02	0.91	4.84
20-m SRT (stage)	6.48 ± 2.72	6.55 ± 2.75	0.05 ± 0.02	0.013	0.018	0.99 (0.99-0.99)	14.75	0.08	0.28	2.42	2.97	0.76	0.27
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	39.50 ± 8.25	39.55 ± 8.18	0.00 ± 0.05	1.000		0.99 (0.99-0.99)	108.00	0.56	0.75	2.28	1.35	0.00	0.75
18-34 years													
2-km walk test (sec)	990.19 ± 99.646	987.56 ± 98.91	-2.26 ± 0.67	0.001	-0.023	0.99 (0.99-0.99)	7541.00	60.81	7.80	1.46	0.54	0.23	7.37
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	40.38 ± 10.67	39.90 ± 9.71	-0.72 ± 0.39	0.068		0.95 (0.92-0.96)	2441.67	19.69	4.43	5.56	7.72	1.79	4.00
20-m SRT (stage)	6.29 ± 2.71	6.39 ± 2.71	0.03 ± 0.02	0.117		0.99 (0.99-0.99)	6.50	0.05	0.23	1.98	2.51	0.50	0.23
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	39.04 ± 8.26	39.18 ± 8.15	0.00 ± 0.07	1.000		0.99 (0.99-0.99)	90.00	0.72	0.84	2.57	1.54	0.00	0.83

35-49 years													
2-km walk test (sec)	968.65 ± 102.44	963.97 ± 100.13	-0.91 ± 0.83	0.275		0.99 (0.99-0.99)	9730.00	82.46	9.08	1.73	0.67	0.09	9.07
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	37.97 ± 9.24	38.20 ± 9.55	0.07 ± 0.33	0.827		0.96 (0.94-0.97)	1555.01	13.29	3.64	7.38	6.78	0.19	3.64
20-m SRT (stage)	5.65 ± 2.69	5.71 ± 2.73	0.06 ± 0.02	0.011	0.023	0.99 (0.99-0.99)	39.00	0.33	0.57	5.23	3.31	1.93	0.26
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	37.11 ± 8.04	36.97 ± 8.11	0.03 ± 0.06	0.657		0.99 (0.99-0.99)	45.00	0.38	0.62	1.88	1.19	0.07	0.61
50-64 yr													
2-km walk test (sec)	1029.91 ± 112.88	1028.17 ± 105.21	-1.26 ± 1.06	0.239		0.99 (0.99-0.99)	21208.00	155.94	12.49	1.71	0.86	0.12	12.10
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	31.89 ± 8.47	31.74 ± 7.58	-0.20 ± 0.33	0.535		0.93 (0.90-0.95)	2045.82	15.04	3.87	9.00	8.63	0.65	3.59
20-m SRT (stage)	3.51 ± 2.26	3.51 ± 2.27	0.02 ± 0.02	0.183		0.99 (0.99-0.99)	9.00	0.07	0.26	3.22	5.18	0.84	0.26
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	30.60 ± 6.76	30.59 ± 6.74	0.06 ± 0.07	0.368		0.99 (0.99-0.99)	99.00	0.72	0.85	3.55	1.98	0.22	0.85
Non-Active													
2-km walk test (sec)	1039.06 ± 109.66	1034.12 ± 106.53	-1.66 ± 0.62	0.009	-0.016	0.99 (0.99-0.99)	12922.00	72.60	8.52	1.13	0.57	0.16	8.36
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	32.65 ± 9.79	32.51 ± 8.58	-0.40 ± 0.32	0.210		0.94 (0.92-0.95)	3261.30	18.32	4.28	5.46	9.24	1.22	3.79
20-m SRT (stage)	3.87 ± 2.31	3.91 ± 2.32	0.03 ± 0.01	0.052		0.99 (0.99-0.99)	11.25	0.06	0.25	2.38	4.50	0.93	0.25
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	31.74 ± 6.98	31.83 ± 6.96	0.11 ± 0.06	0.052		0.99 (0.99-0.99)	117.00	0.65	0.80	2.45	1.78	0.37	0.80
Active													
2-km walk test (sec)	960.57 ± 92.59	960.05 ± 90.18	-1.27 ± 0.79	0.111		0.99 (0.99-0.99)	25436.00	127.82	11.31	2.19	0.83	0.13	10.96
2-km walk test (ml* kg ⁻¹ * min ⁻¹)	40.14 ± 9.12	39.93 ± 9.18	-0.17 ± 0.26	0.515		0.95 (0.94-0.96)	2767.02	13.97	3.73	6.27	6.61	0.43	3.66

20-m SRT (stage)	3.87 ± 2.31	3.91 ± 2.32	0.04 ± 0.01	0.012	0.016	0.99 (0.99-0.99)	13.00	0.07	0.26	2.23	2.87	0.73	0.25
20-m SRT (ml* kg ⁻¹ * min ⁻¹)	31.74 ± 6.98	31.83 ± 6.96	-0.04 ± 0.05	0.407		0.99 (0.99-0.99)	117.00	0.59	0.77	2.33	1.41	0.12	0.76

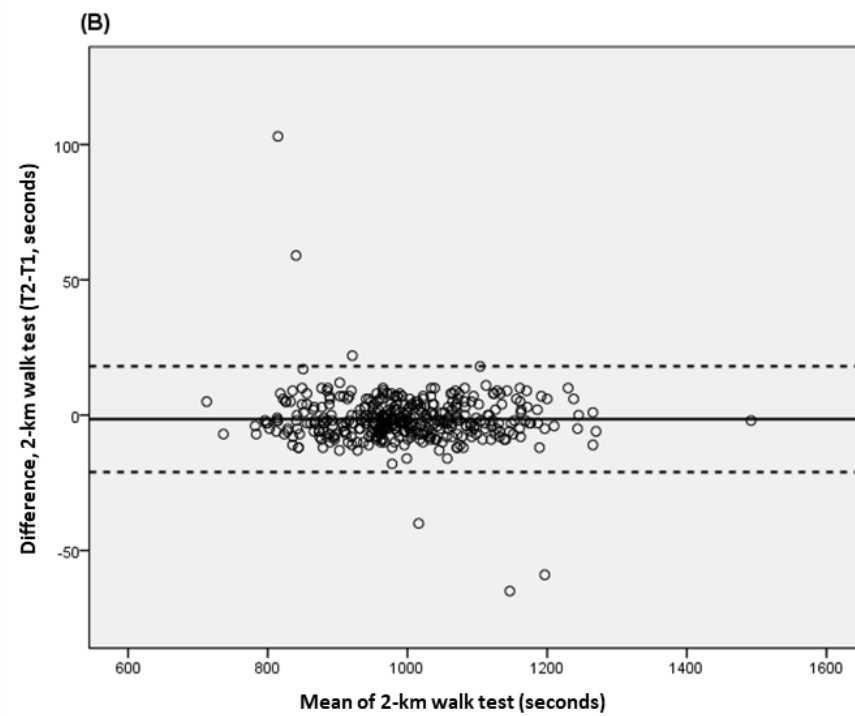
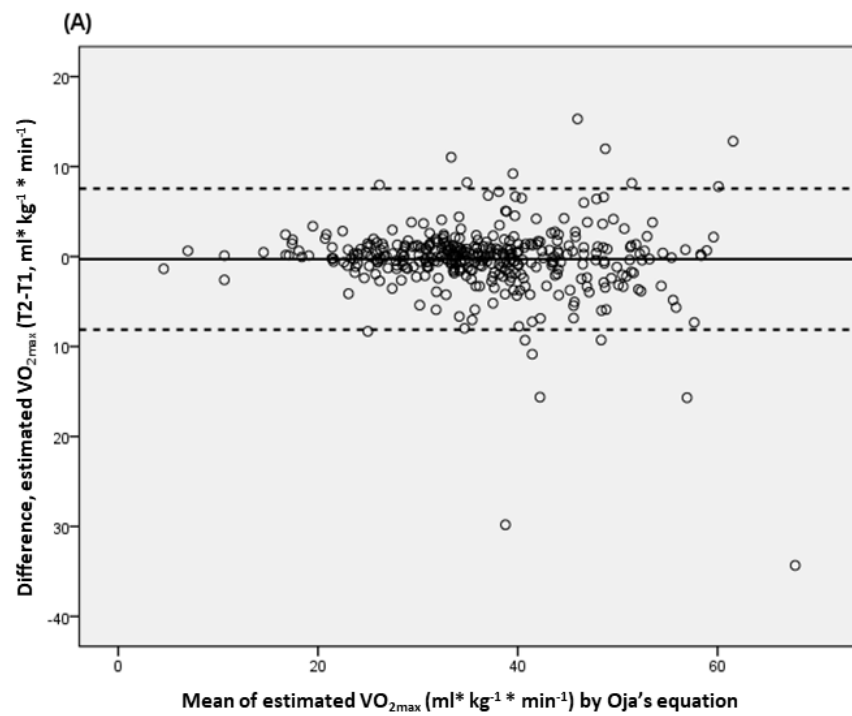
2-km walk test is expressed as total time to complete the test (sec), and as estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Oja's equation; 20-m SRT is expressed as final stage reached, and as estimated VO_{2max} (ml* kg⁻¹ * min⁻¹) by Leger's equation.

*T1 refers to test (trial 1) and T2 to retest (trial 2).

T2-T1 refers to retest (trial 2) minus test (trial 1). Values are displayed as mean ± SD.

*All the ICCs were significant at p< 0.001.

ICC, intraclass correlation coefficients; CI, confident interval; SSE, sum of squared errors; MSE, mean sum of squared errors; RMSE, root mean sum of squared errors; %Error, percentage error; %SEM, standard error of measurement; %CV, percentage coefficient of variation; SEE, standard error of estimate.



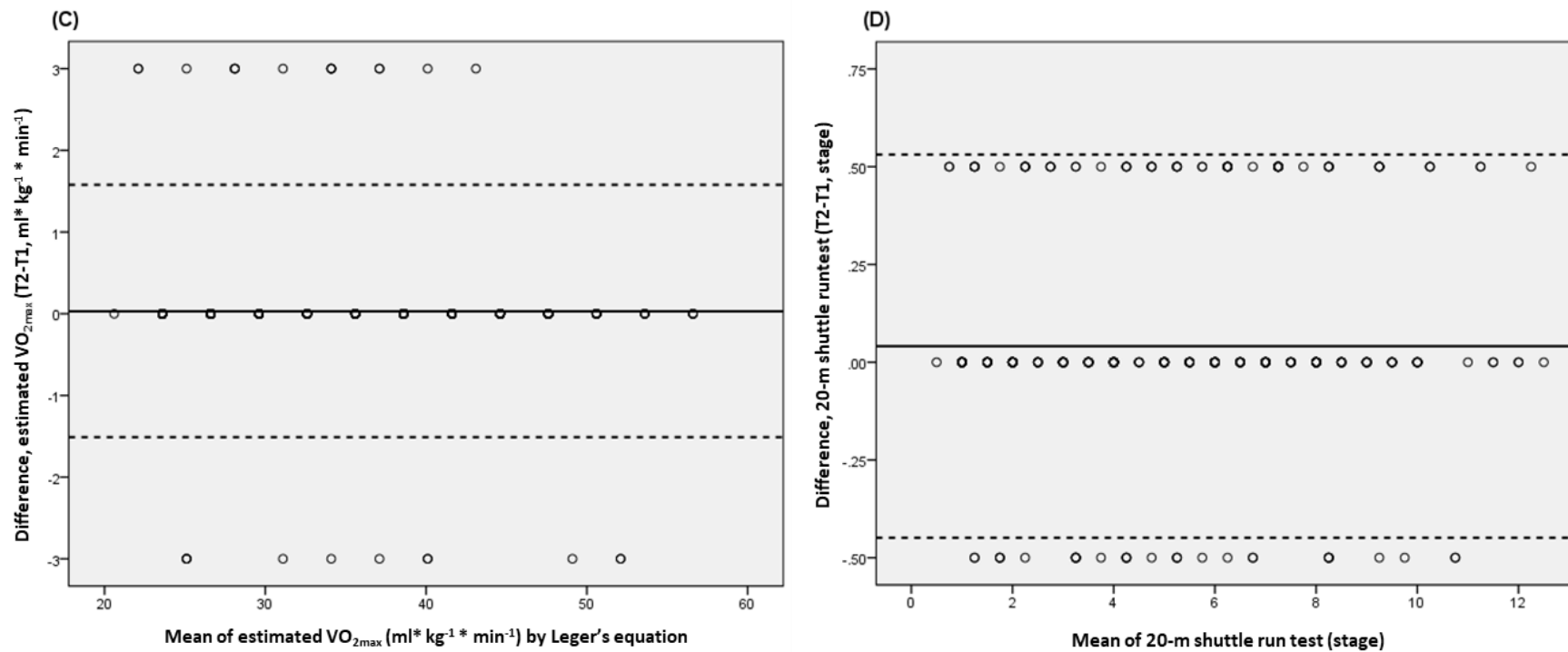
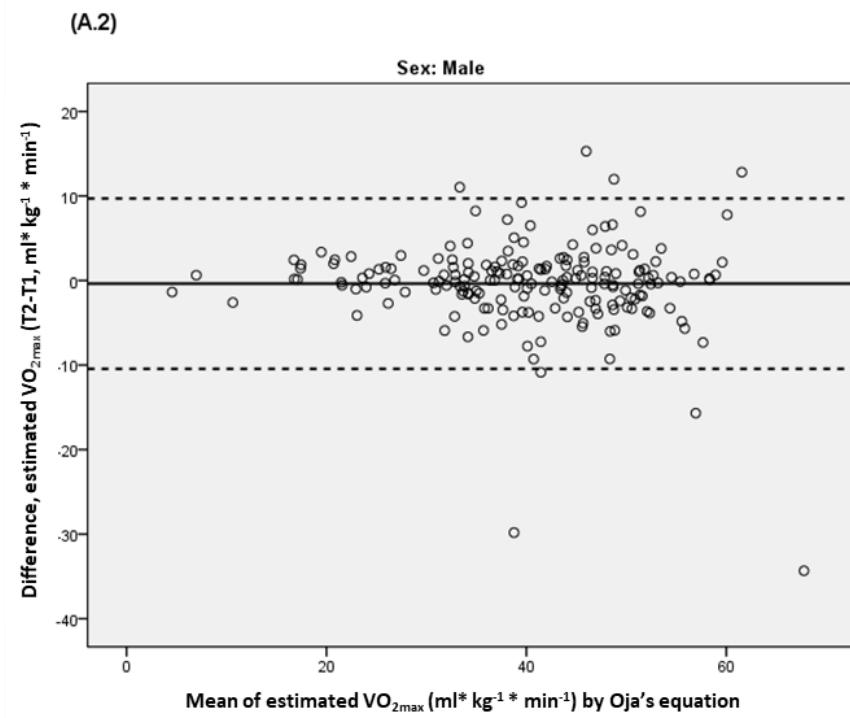
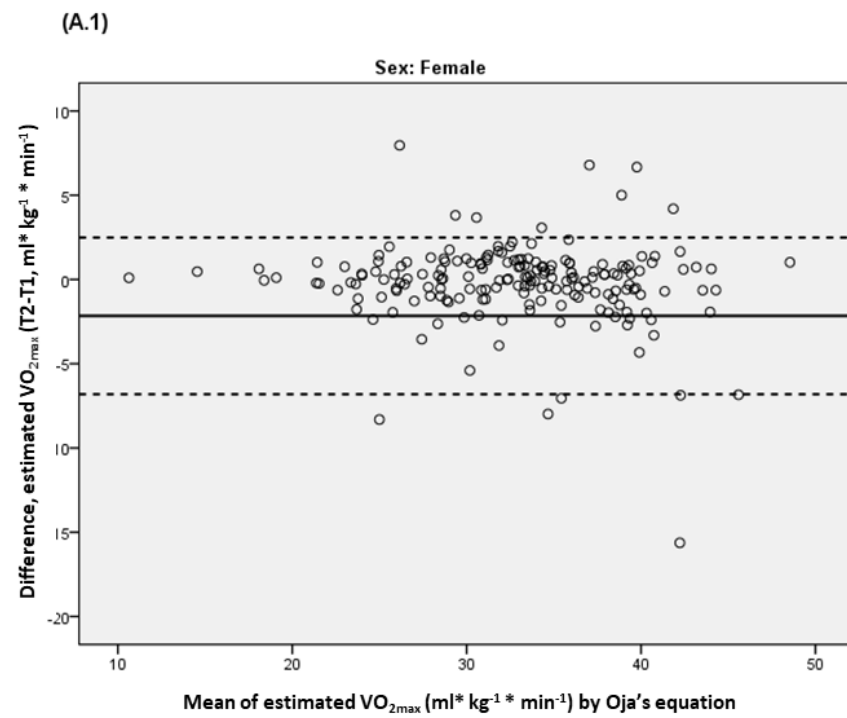
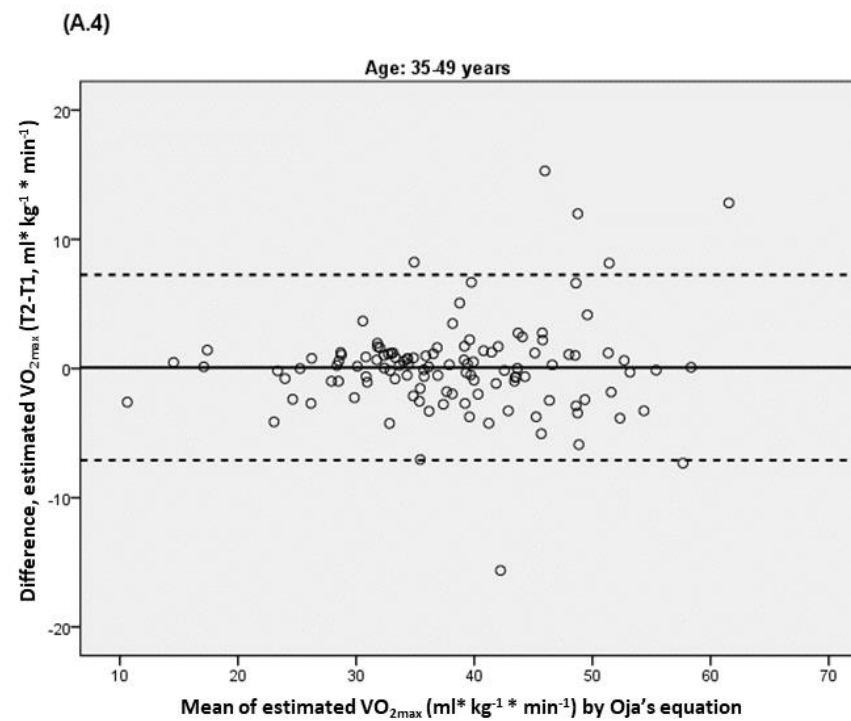
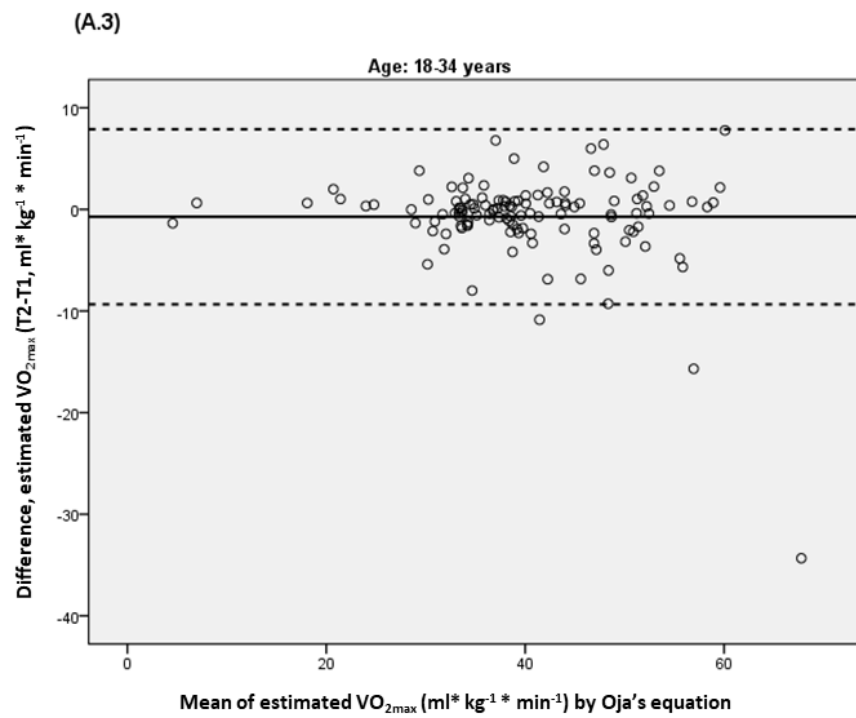


Figure 17. Bland-Altman plot of the estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation, and Leger's equation, the 2-km walk test (seconds), and the 20-m shuttle run test (stage).

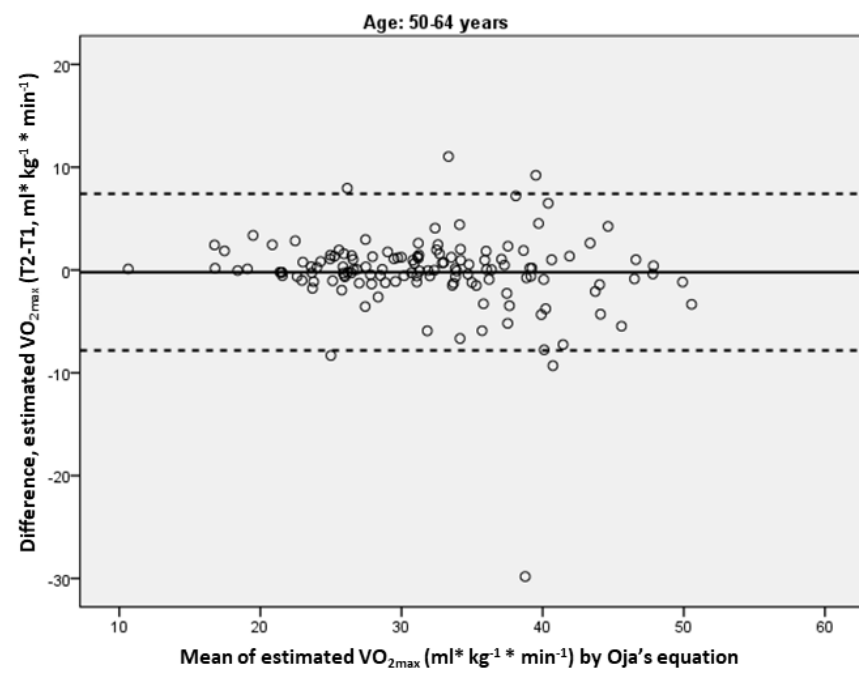
(A), agreement between mean and difference in estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Oja's equation; (B), agreement between mean and difference in total time (sec) in the 2-km walk test; (C), agreement between mean and difference in estimated $\text{VO}_{2\text{max}}$ ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) by Leger's equation; (D), agreement between mean and difference in final stage in the 20-m shuttle run test.

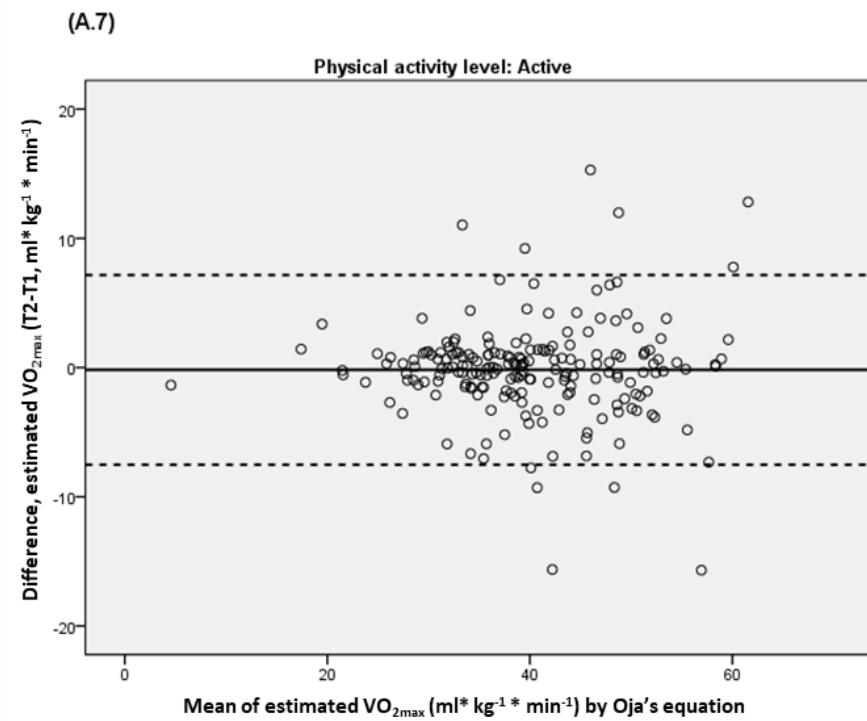
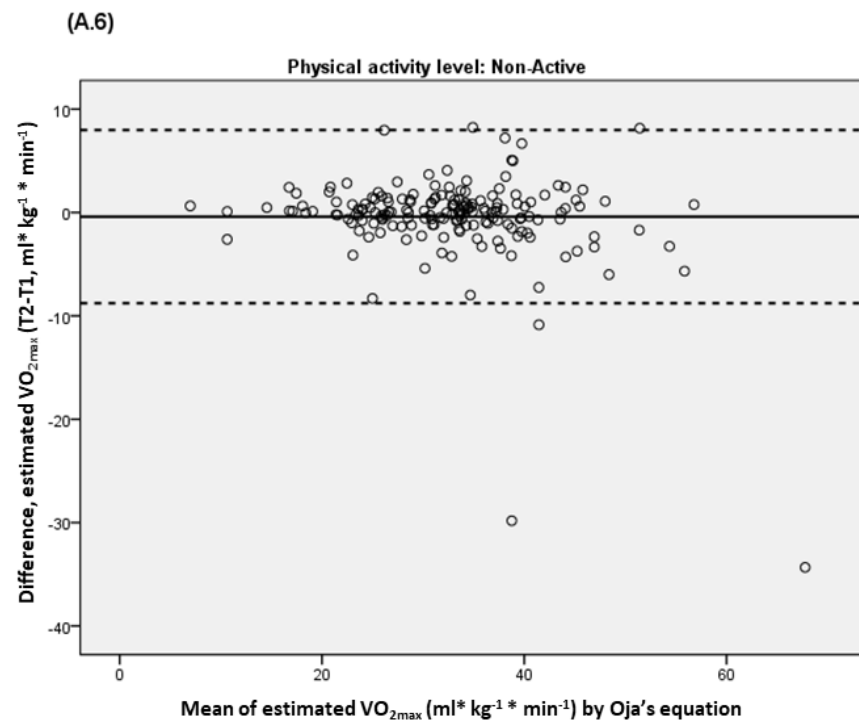
The central line represents the mean differences between retest (T2) and test (T1). The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

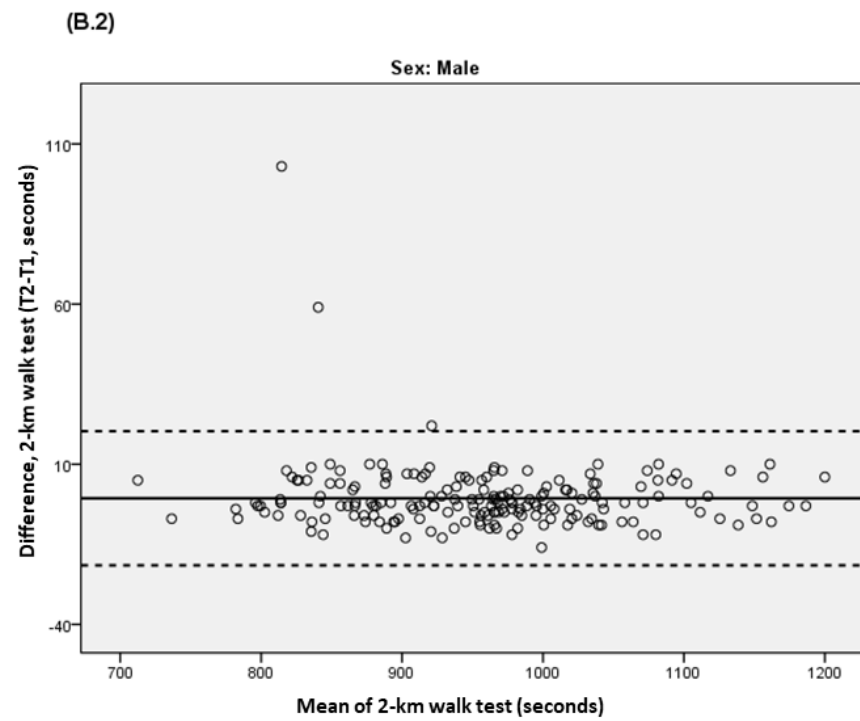
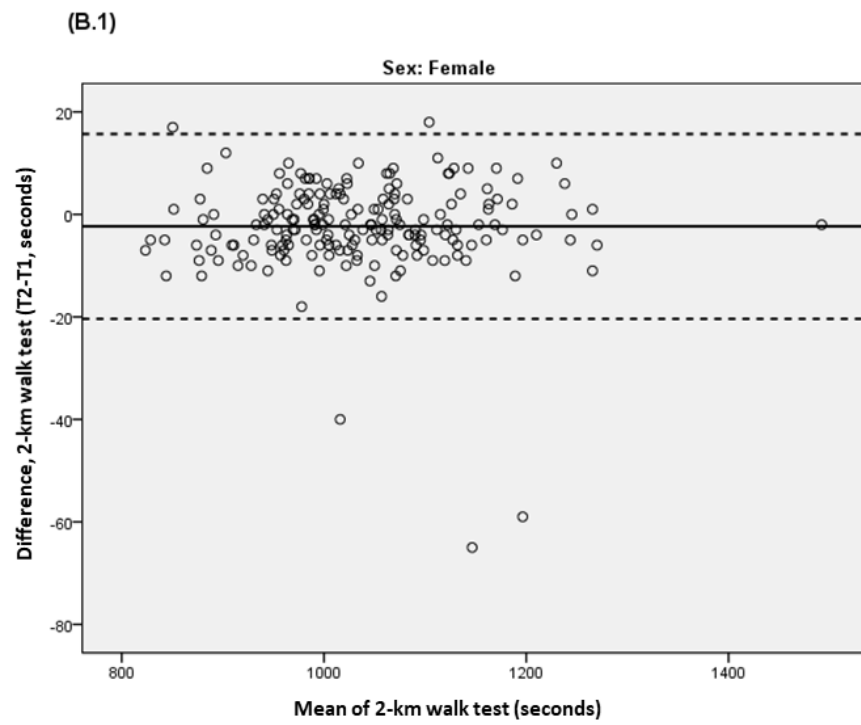




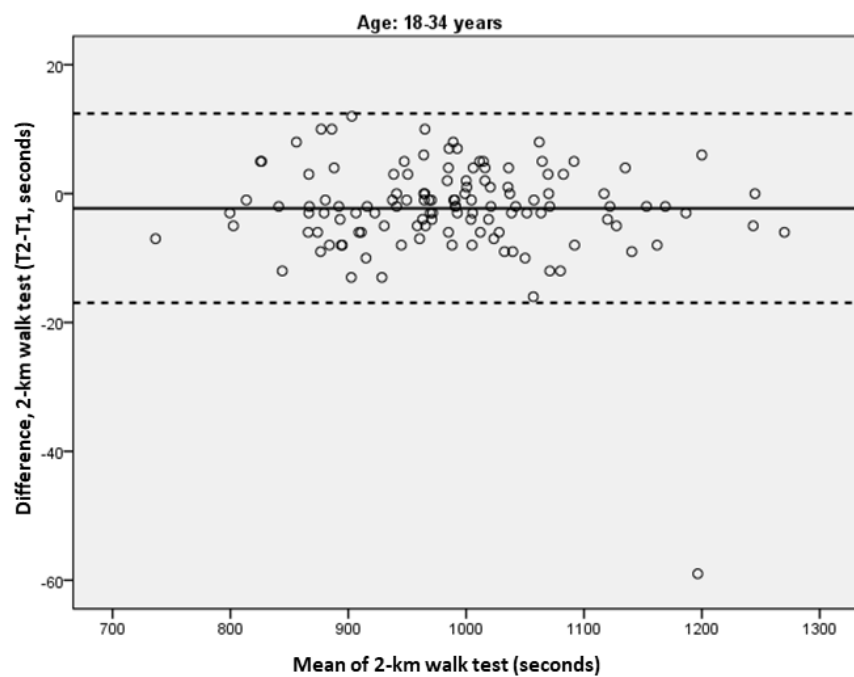
(A.5)



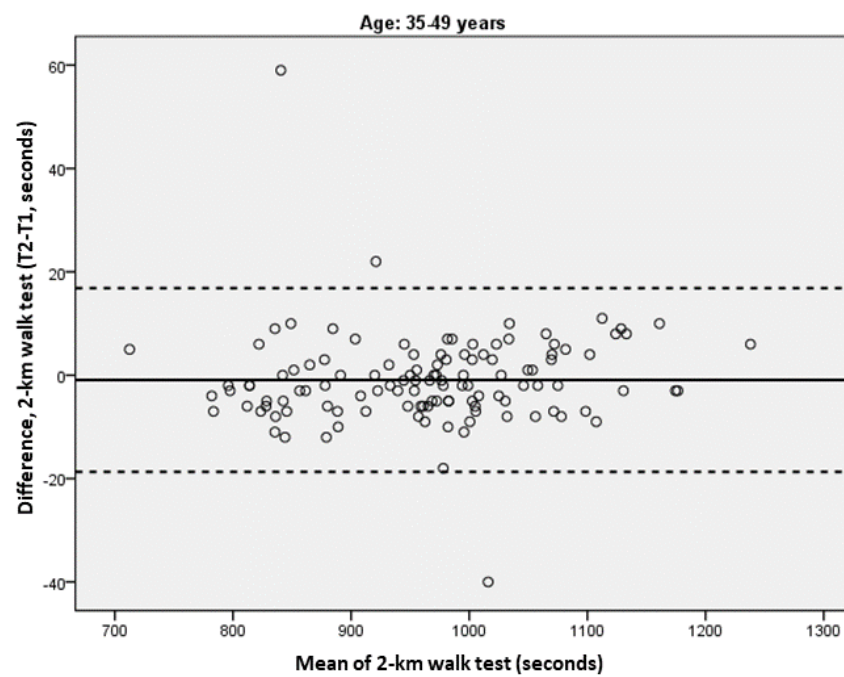




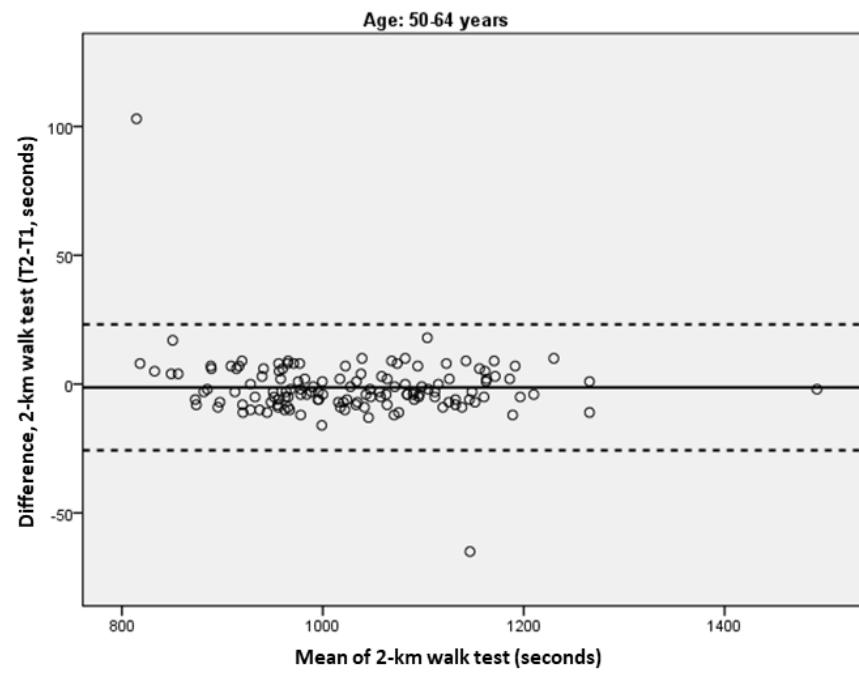
(B.3)



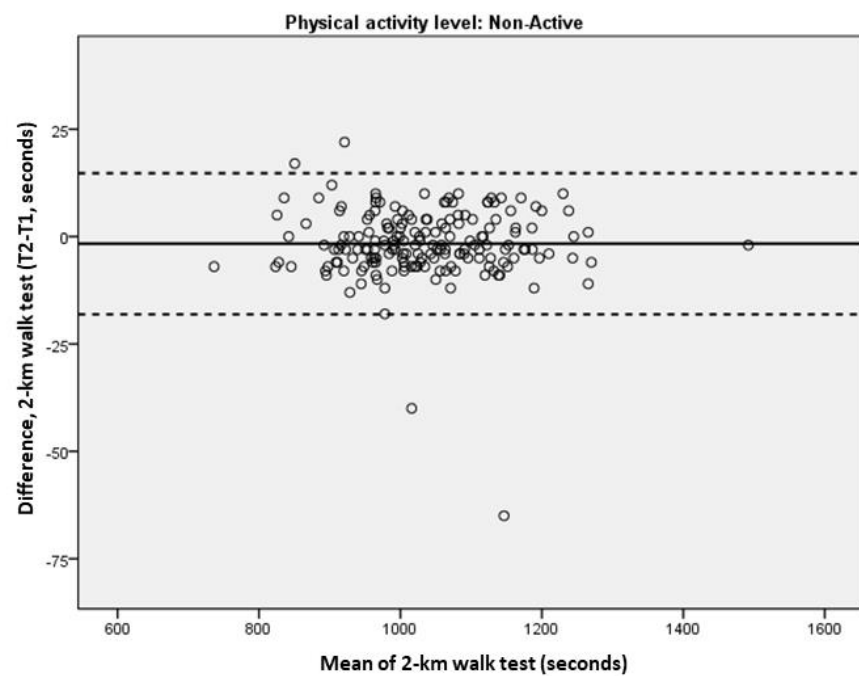
(B.4)



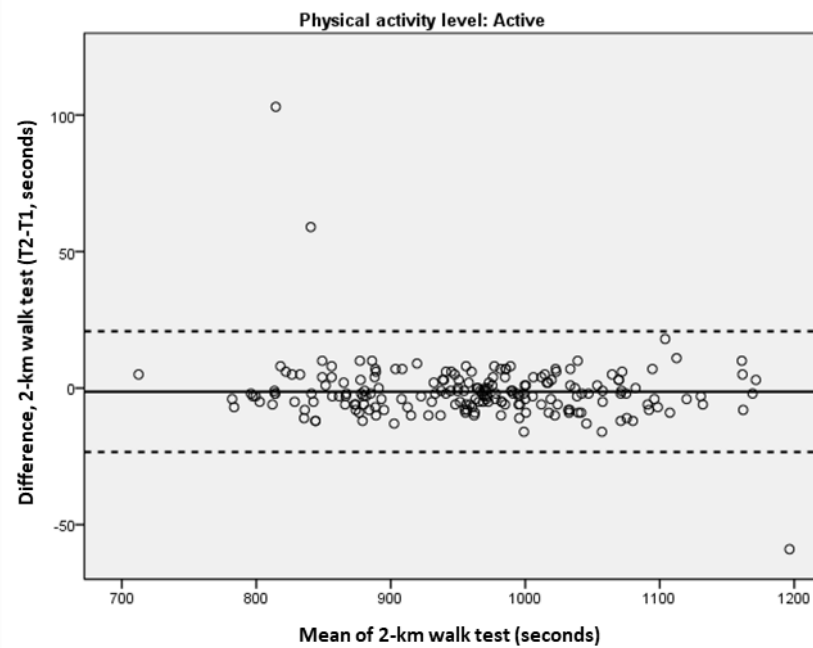
(B.5)



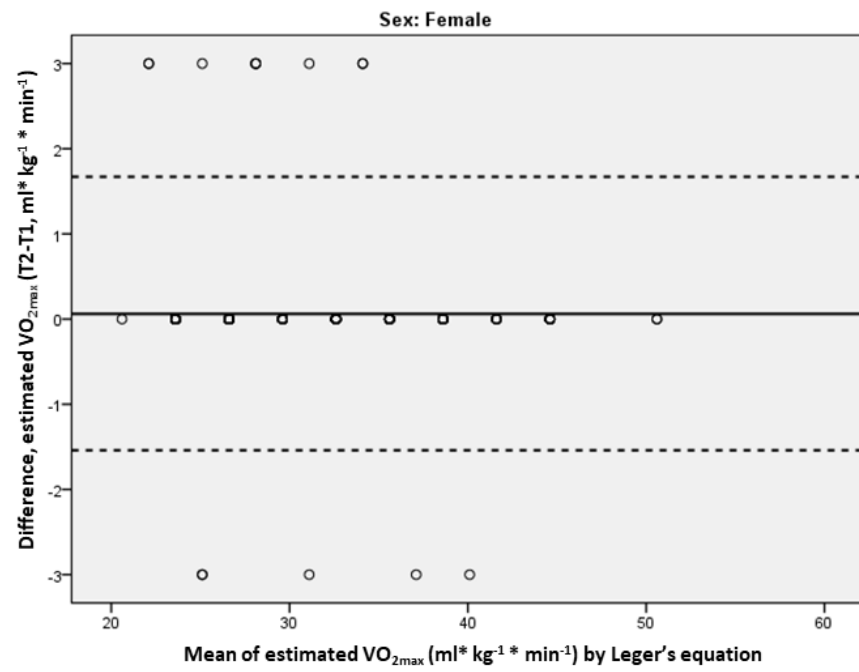
(B.6)



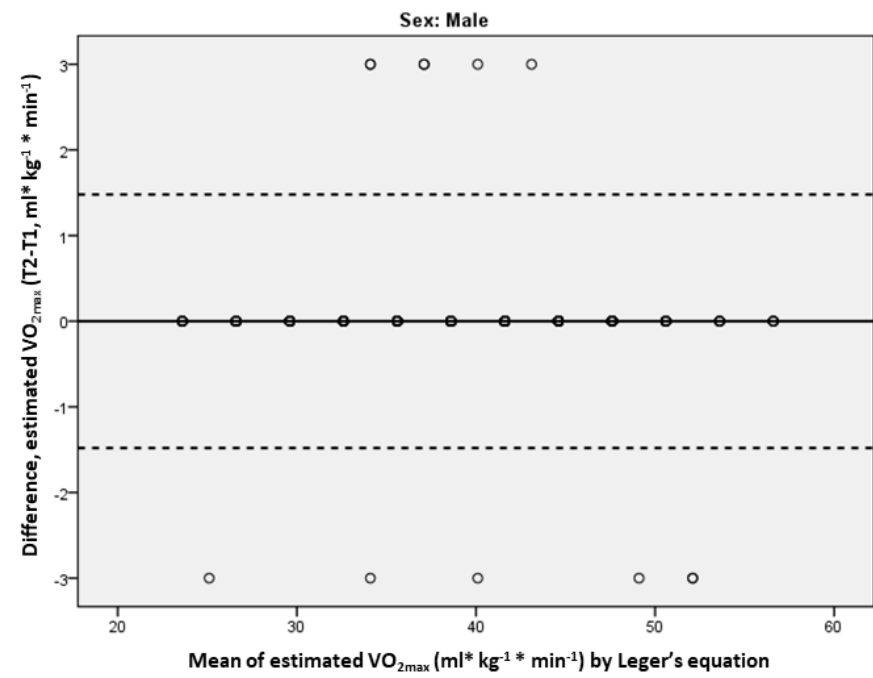
(B.7)



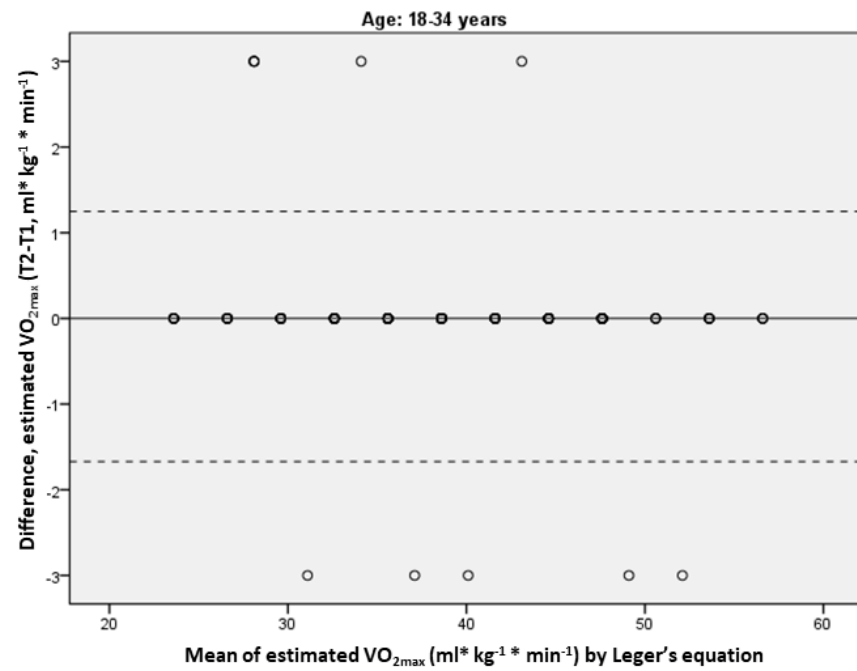
(C.1)



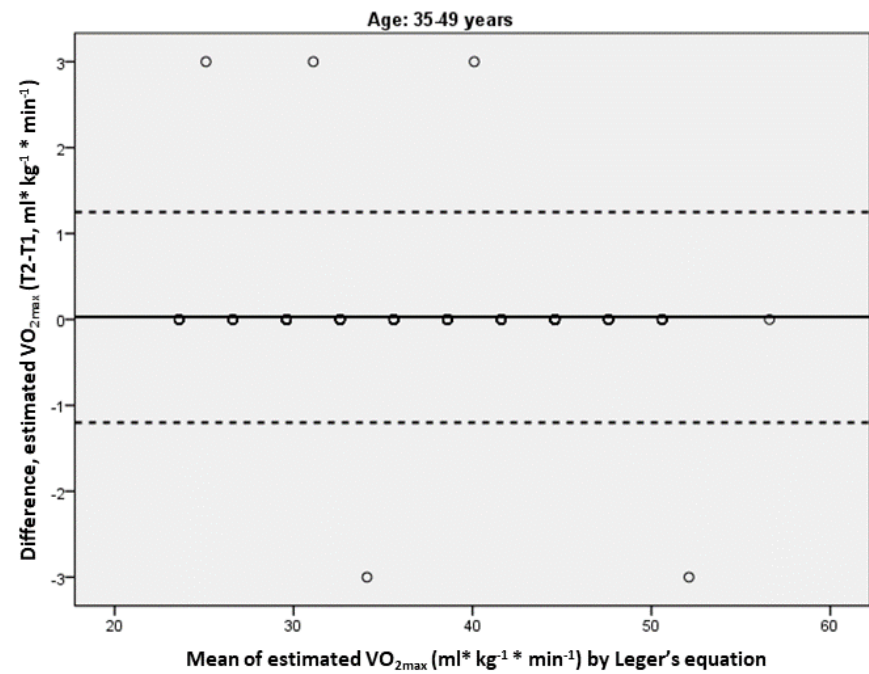
(C.2)



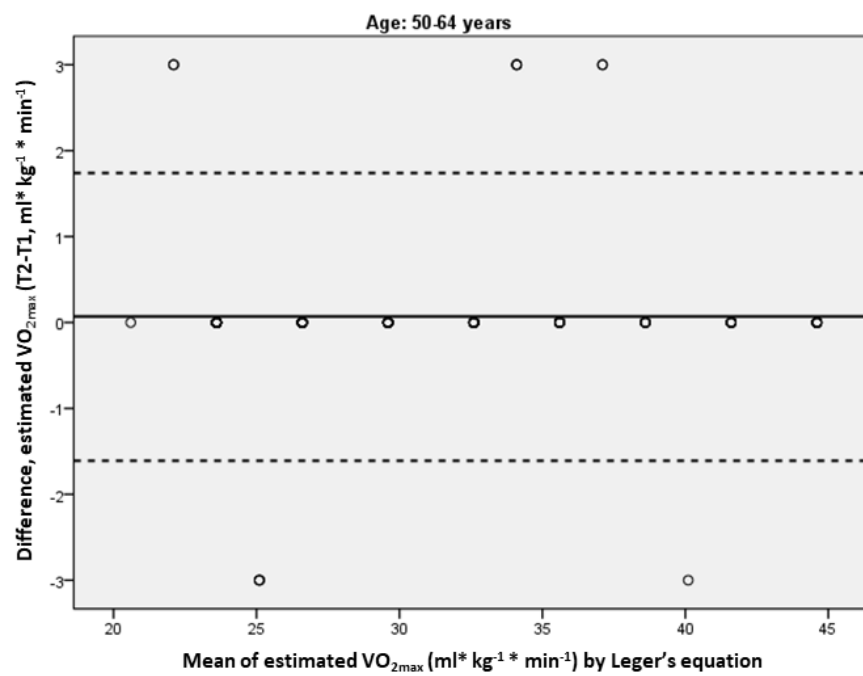
(C.3)

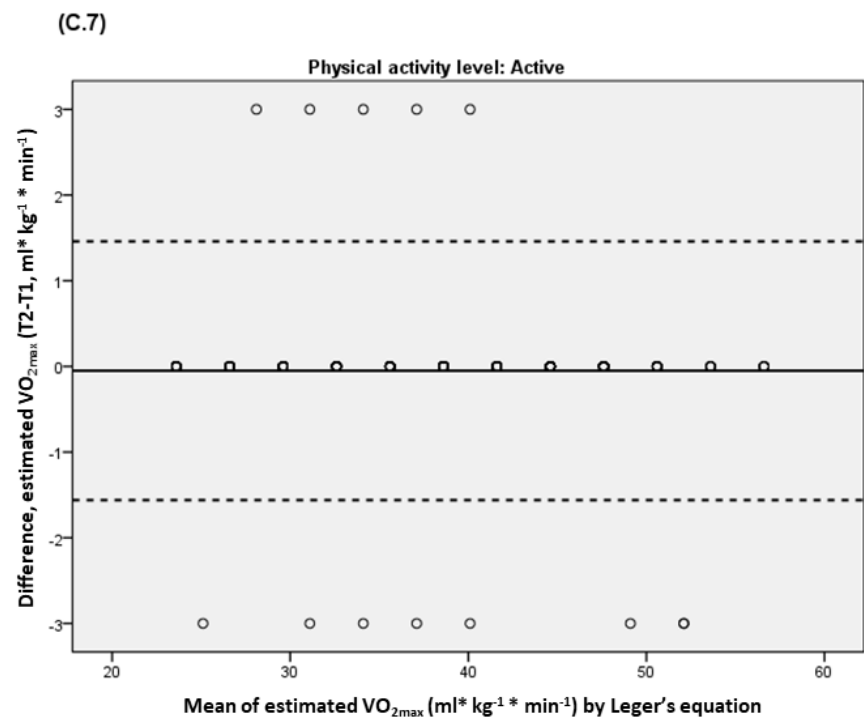
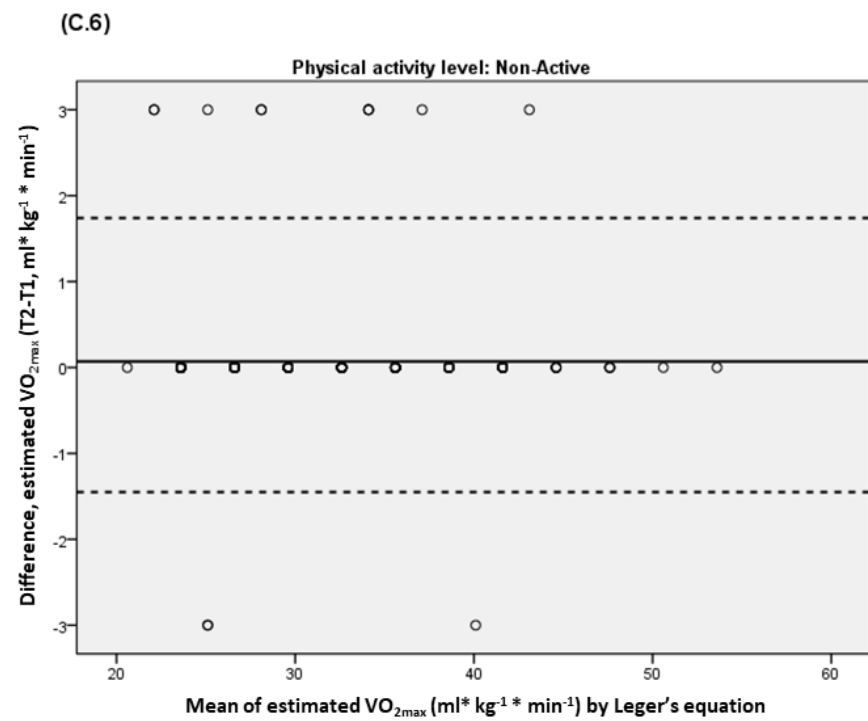


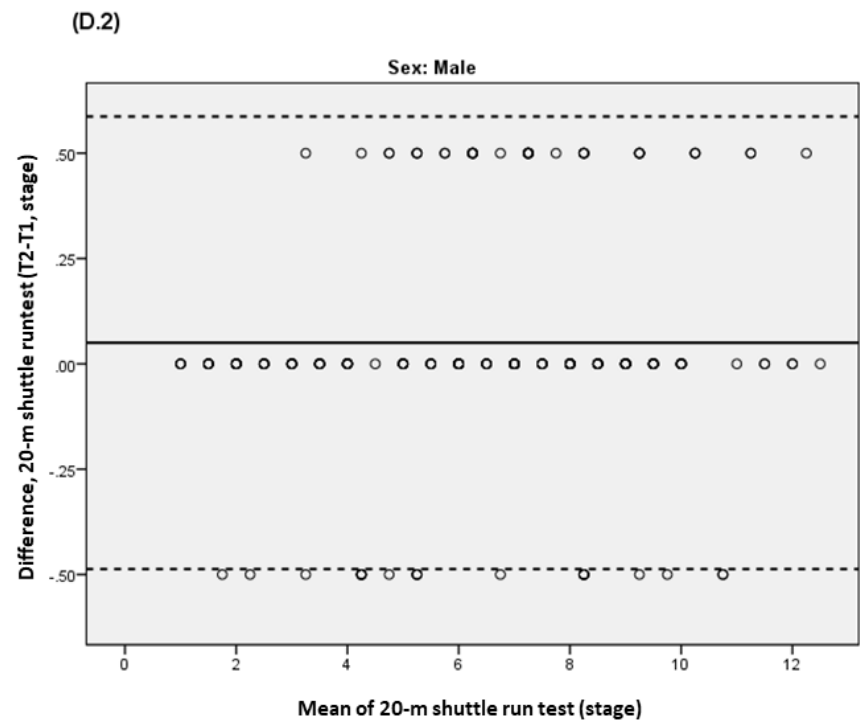
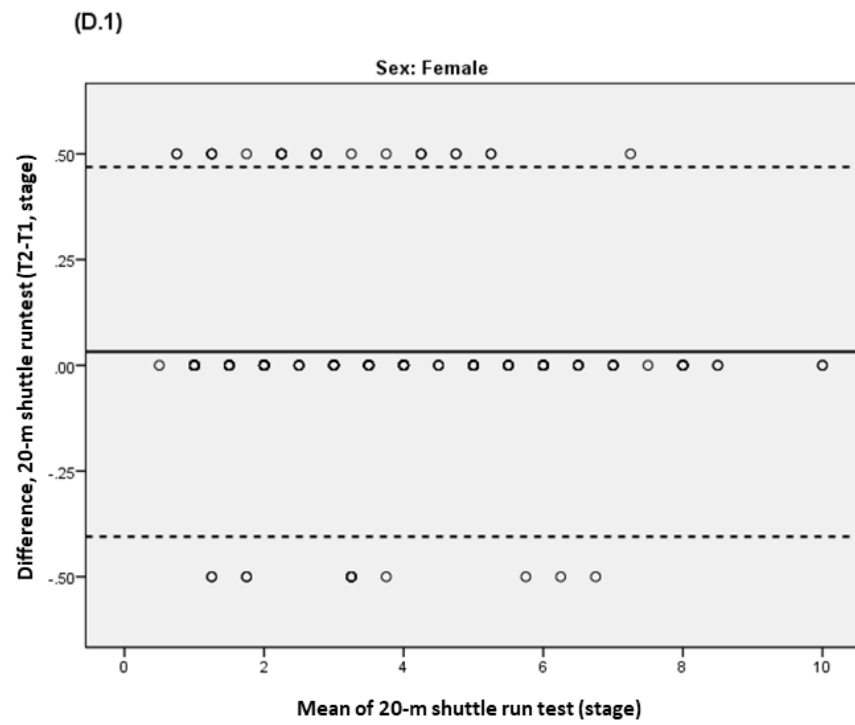
(C.4)



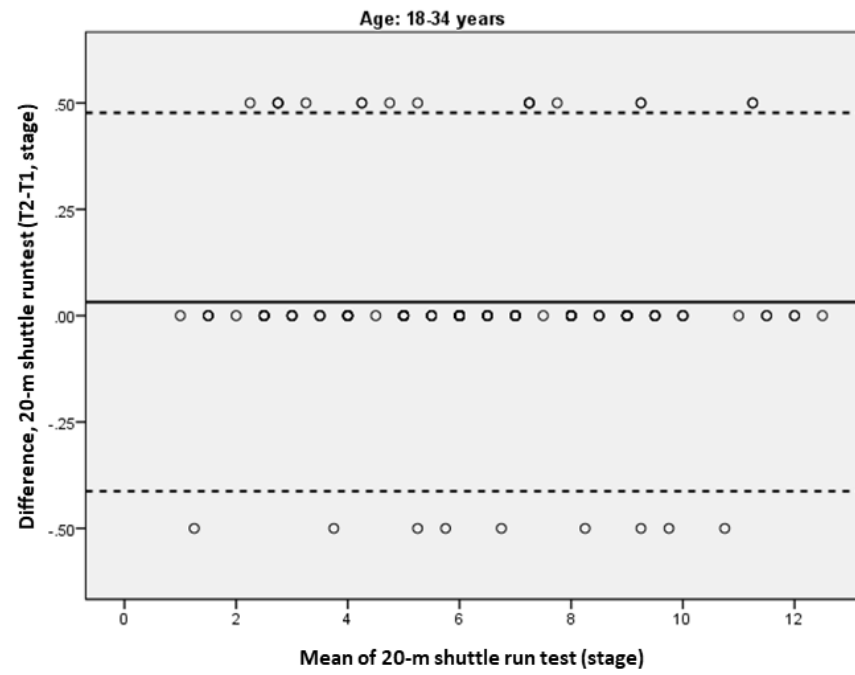
(C.5)



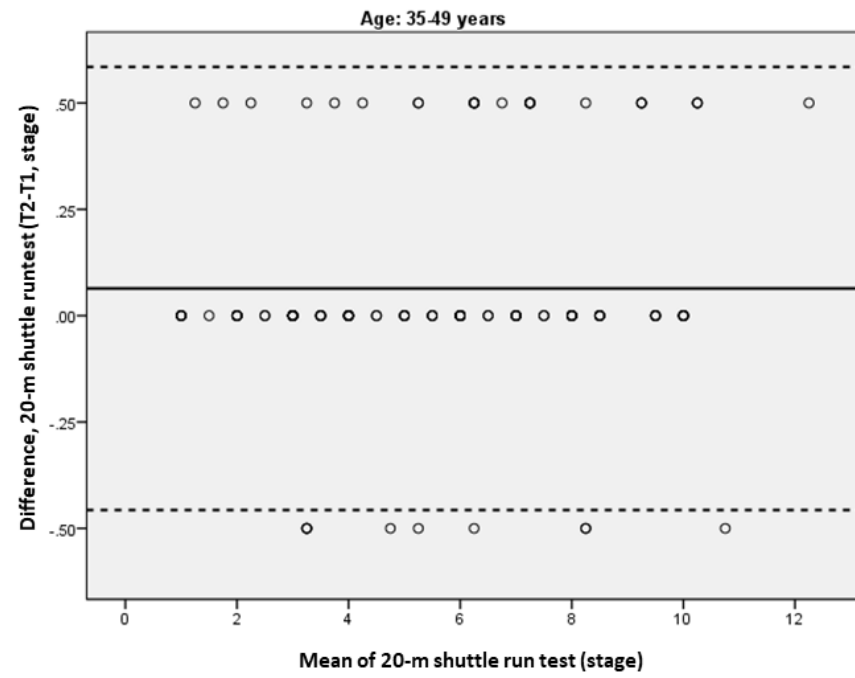




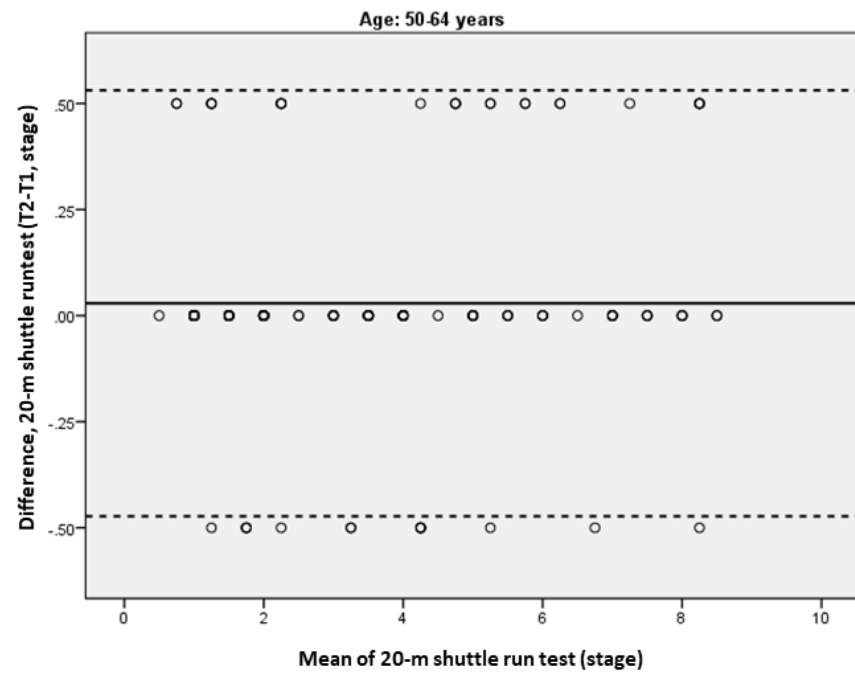
(D.3)



(D.4)



(D.5)



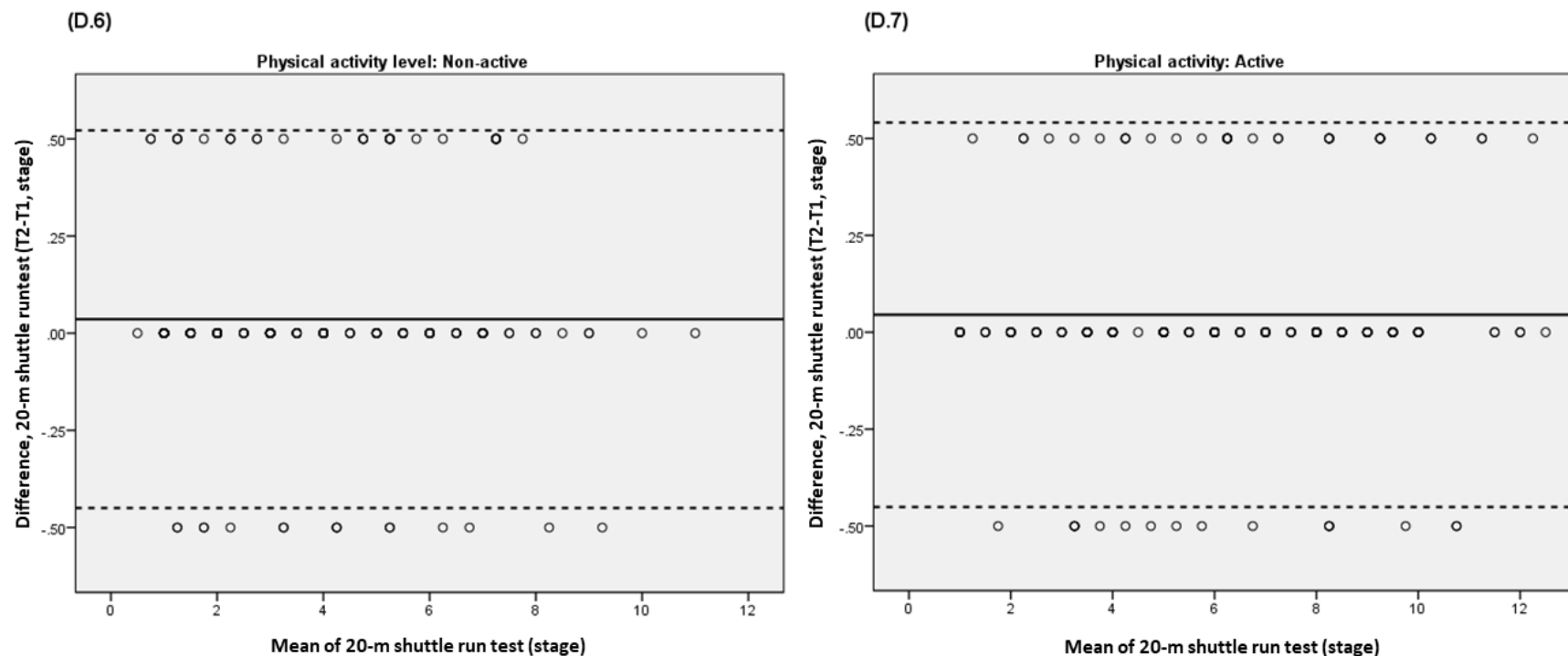


Figure 18. Bland-Altman plot of the estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Oja's equation, and Leger's equation, the 2-km walk test (seconds), and the 20-m shuttle run test (final stage), distributed by sex, age, and physical activity levels.

(A), agreement between mean and difference in estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Oja's equation; (B), agreement between mean and difference in total time (sec) in the 2-km walk test; (C), agreement between mean and difference in estimated VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) by Leger's equation; (D), agreement between mean and difference in final stage in the 20-m shuttle run test.

(A.1-A.2), distributed by sex; (A.3-A.5), distributed by age; (A.6-A.7), distributed by physical activity level.

(B.1-B.2), distributed by sex; (B.3-B.5), distributed by age; (B.6-B.7), distributed by physical activity level.

(C.1-C.2), distributed by sex; (C.3-C.5), distributed by age; (C.6-C.7), distributed by physical activity level.

(D.1-D.2), distributed by sex; (D.3-D.5), distributed by age; (D.6-D.7), distributed by physical activity level.

The central line represents the mean differences between retest (T2) and test (T1). The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

Study VI: The standing long jump test.

The final sample size was composed of 410 healthy adults aged 18-64 years (49.5% females). The descriptive characteristics of the participants, distributed by sex, age groups and physical activity levels are shown in **Table 40**. The mean age of the sample was 42 (± 13.06) years old. Overall, significant differences were found according to sex, age groups, and physical activity levels. Regarding sex, males presented higher anthropometric values than females (all, $p < 0.001$), except for %BF and tricipital skinfold, being higher in females (both, $p < 0.001$). Male sample performed greater than female sample in all tests (all, $p < 0.005$). The 18-35 years old group performed greater than the older groups in all tests (all, $p < 0.005$), as well that the active group, compared with the non-active group (all, $p < 0.001$).

Criterion-related validity

Bivariate correlation analysis between SLJ field-based test and the laboratory lower-body muscular tests and anthropometrics variables are displayed in **Table 41**. The SLJ distance was associated with all studied laboratory variables (all, $p < 0.001$). Maximal horizontal power ($r = 0.826$), maximal horizontal force ($r = 0.815$), jump take off velocity ($r = 0.768$) and horizontal jump take off velocity ($r = 0.766$) presented the higher associations. All the isokinetic variables showed moderate associations ($r = 0.658$ to 0.672 ; all, $p < 0.001$). The SLJ distance was associated with all anthropometric variables (all, $p < 0.001$). %BF, sex, height, tricipital skinfold and sum triceps+subscapular skinfold had moderate associations with SLJ distance ($r = -0.495$ to -0.665 ; all, $p < 0.001$). Similar results were found when the sample was distributed by sex and physical activity levels. Regarding age groups, in the 50-64 years group, jump take off velocity and horizontal jump take off velocity presented moderate associations ($r = 0.620$ and $r = 0.608$, respectively; both $p < 0.001$) (**Table 42**).

Table 43 shows the results from the stepwise lineal regression analysis of the SLJ test. Maximal horizontal power accounted for 68% of the explained variance. When sex, %BF, maximal horizontal force and lean mass were added to the model, the explained variance went up to 78%. Age, physical activity levels and horizontal jump take off velocity were not included in the model as no significant associations were found.

Reliability

The test-retest reliability of the SLJ test is shown in **Table 44**. The mean difference between the SLJ test-retest presented an overestimation of T2 compared to T1 (12.14 ± 14.46 cm, $p < 0.001$). The effect size (Cohen d) of the mean differences was 0.309. The ICC showed a high reproducibility (ICC= 0.94, $p < 0.001$), and the analyzed error measurements were: RMSE= 18.94 cm; %CV= 7.06; SEE= 14.48 cm. **Table 45** shows test-retest reliability of the SLJ test, distributed by sex, age groups and physical activity levels. Overall, similar results to those of the whole sample were found.

Figure 19 shows the Bland-Altman plots of the SLJ test. It can be observed that the systematic error was not close to 0, nor narrow LoA [12.14 cm (95% LoA= -16.20 to 40.48)]. No heteroscedasticity was observed ($p = 0.087$). When the sample was distributed by sex, age groups, and physical activity levels, similar results to those of the whole sample were found (see **Figure 20**).

Table 44 and **Figure 21** show the test-retest reliability of the SLJ test after a learning period. Significant mean differences between T1 and T2 were found in the SLJ distance (0.45 ± 1.04 cm, $p = 0.001$, $d = 0.013$). The ICC reported a high reproducibility (ICC= 1.00, $p < 0.001$), and the analyzed error measurements were: RMSE= 1.13; %CV= 0.53; SEE= 1.03. The systematic error was nearly 0 [0.45 cm (95% LoA= -1.58 to 2.49, $p = 0.361$)]. No heteroscedasticity was observed.

Table 40. Descriptive characteristics of the participants, stratified by sex, age and physical activity levels.

	All (n= 410)	Sex		Age groups			Physical activity levels	
		Females (n= 203)	Males (n= 207)	18-34 yr (n= 136)	35-49 yr (n= 131)	50-64 yr (n= 143)	Non-Active (n= 195)	Active (n= 215)
Age (years)	41.86 (13.06)	41.48 (12.84)	42.24 (13.28)	26.22 (0.36)	42.49 (0.36)	56.17 (0.35)	43.18 (13.10)*	40.62 (12.95)
Weight (kg)	71.16 (15.49)	62.06 (9.36)***	80.13 (15.12)	68.80 (16.24)	72.19 (16.29)	72.48 (13.77)	71.99 (16.82)	70.32 (14.14)
Height (cm)	168.25 (9.19)	161.38 (5.98)***	175.00 (6.36)	168.46 (0.79)	168.42 (0.80)	167.89 (0.77)	167.33 (9.17)	169.02 (9.12)
Body Mass Index	24.99 (4.10)	23.85 (3.45)***	26.12 (4.38)	24.07 (4.27) ^{a^}	25.32 (4.48)	25.56 (3.38) ^{c^^}	25.55 (4.62)**	24.47 (3.50)
Waist circumference (cm)	83.14 (12.66)	76.90 (9.25)***	89.26 (12.58)	78.11 (11.68) ^{a^^}	83.28 (13.11) ^{b^^}	87.80 (11.34) ^{c^^^}	85.54 (14.21)***	80.91 (10.64)
Hip circumference (cm)	99.38 (8.69)	98.80 (7.27)	99.95 (9.87)	98.69 (9.75)	100.30 (9.19)	99.19 (6.96)	101.13 (9.10)***	97.77 (8.00)
Tricipital skinfold (mm)	16.92 (7.48)	19.69 (6.42)***	14.23 (7.46)	16.13 (0.64)	17.03 (0.65)	17.58 (0.63)	19.39 (7.67)***	14.72 (6.56)
Subscapular skinfold (mm)	19.50 (10.27)	20.33 (9.19)	18.68 (11.19)	17.37 (11.72)	19.46 (9.83)	21.57 (8.72) ^{c^^}	22.74 (10.99)***	16.57 (8.62)
Body fat (%)	24.85 (7.47)	28.64 (6.30)***	21.11 (6.62)	22.77 (7.69) ^{a^}	24.99 (7.08)	26.70 (7.15) ^{c^^^}	27.38 (7.31)***	22.57 (6.89)
Lean mass (kg)	50.37 (10.58)	41.56 (3.92)***	59.06 (7.41)	49.85 (10.13)	51.09 (10.81)	50.22 (10.83)	49.10 (10.72)*	51.45 (10.31)
Standing long jump test, field-based measurement								
Maximal long jump test (cm)	143.03 (38.53)	121.05 (28.86)***	164.37 (34.57)	159.64 (35.01) ^{a^}	146.26 (36.66) ^{b^^}	121.41 (33.31) ^{c^^^}	131.48 (36.18)***	153.88 (37.52)
Maximal long jump retest (cm)	155.68 (39.95)	133.00 (29.56)***	177.67 (36.30)	172.20 (37.30)	163.10 (37.75) ^{b^^}	133.55 (34.10) ^{c^^^}	143.86 (37.10)***	166.73 (38.70)
Standing long jump test, force platform measurement								
Maximal force (%BF)	194.97 (24.47)	192.39 (23.07)*	197.48 (25.58)	202.94 (25.49) ^{a^^}	194.12 (21.89) ^{b^^}	188.16 (23.67)	189.62 (23.61)***	200.01 (24.26)
Maximal horizontal force (%BF)	63.96 (16.90)	57.84 (14.60)***	69.92 (16.89)	73.79 (15.96) ^{a^^^}	65.60 (14.32) ^{b^^}	53.10 (13.37) ^{c^^^}	59.37 (15.69)***	68.34 (16.77)
Maximal vertical force (%BF)	186.49 (22.68)	185.96 (21.92)	187.02 (23.45)	192.16 (23.23) ^{a^}	185.01 (20.94)	182.46 (22.78) ^{c^^^}	182.70 (22.50)***	190.07 (22.36)
Maximal power (W/kg)	21.23 (10.70)	19.12 (9.28)***	23.30 (11.57)	25.47 (12.24) ^{a^^}	21.57 (9.37) ^{b^^}	16.90 (8.39) ^{c^^^}	18.31 (9.60)***	24.00 (10.94)
Maximal horizontal power (W/kg)	12.36 (5.96)	10.02 (4.52)***	14.63 (6.31)	15.84 (6.09) ^{a^^^}	12.82 (5.32) ^{b^^}	8.62 (3.89) ^{c^^^}	10.60 (5.05)***	14.02 (6.26)
Maximal vertical power (W/kg)	16.73 (8.68)	15.61 (7.88)**	17.82 (9.28)	19.46 (10.01) ^{a^}	16.81 (7.65) ^{b^}	14.06 (7.33) ^{c^^^}	14.64 (7.87)***	18.70 (8.95)
Jump take off velocity (m/s)	2.38 (0.61)	2.15 (0.50)***	2.61 (0.62)	2.68 (0.56) ^{a^^}	2.45 (0.55) ^{b^^}	2.04 (0.53) ^{c^^^}	2.18 (0.54)***	2.57 (0.61)
Horizontal jump take off velocity (m/s)	2.19 (0.57)	1.98 (0.48)***	2.40 (0.57)	2.47 (0.50) ^{a^^}	2.26 (0.52) ^{b^^}	1.86 (0.51) ^{c^^^}	2.03 (0.53)***	2.35 (0.56)
Vertical jump take off velocity (m/s)	0.73 (0.61)	0.62 (0.60)***	0.84 (0.61)	0.89 (0.59)	0.71 (0.63)	0.60 (0.59) ^{c^^^}	0.59 (0.55)***	0.86 (0.64)
Push off force impulse (Ns)	79.18 (46.47)	61.51 (36.24)***	96.42 (48.91)	93.24 (48.95)	82.86 (43.16) ^{b^^}	62.45 (41.88) ^{c^^^}	67.85 (44.18)***	89.92 (46.06)
Horizontal push off force impulse (Ns)	155.43 (58.77)	122.64 (33.07)***	187.42 (60.76)	171.14 (55.25)	159.46 (66.60) ^{b^^}	136.80 (48.84) ^{c^^^}	144.97 (59.08)***	165.10 (57.08)
Vertical push off force impulse (Ns)	51.90 (43.24)	37.31 (35.78)***	66.14 (45.18)	61.34 (41.78)	52.49 (44.13)	42.39 (12.03) ^{c^^^}	42.92 (39.76)	60.36 (44.74)
Isokinetic								
Maximal isometric peak torque (Nm)	146.83 (51.82)	113.14 (31.16)***	179.86 (46.53)	162.89 (52.74) ^{a^}	147.64 (50.62) ^{b^}	130.80 (47.31) ^{c^^^}	136.05 (47.97)***	156.08 (52.91)
Maximal 60/60 flexion peak torque (Nm)	73.73 (29.73)	60.07 (14.44)***	92.48 (26.10)	84.72 (28.36)	78.59 (25.65) ^{b^^}	66.57 (22.53) ^{c^^^}	70.74 (23.41)***	81.34 (28.15)

Maximal 60/60 extension peak torque (Nm)	103.57 (44.28)	82.36 (23.57)***	131.88 (38.23)	121.39 (40.24) ^{a^}	109.23 (39.83) ^{b^^}	92.31 (40.31) ^{c^^}	98.07 (35.43)***	115.79 (42.50)
Maximal 60/60 flexion peak torque, per repetition (Nm)	84.68 (34.47)	69.40 (16.84)***	105.81 (31.09)	97.25 (32.86)	90.45 (29.53) ^{b^^}	76.32 (26.71) ^{c^^}	81.45 (27.05)***	93.21 (32.95)
Maximal 60/60 extension peak torque, per repetition (Nm)	105.94 (46.32)	83.98 (24.89)***	135.16 (40.72)	123.88 (42.81)	112.03 (42.40) ^{b^^}	94.42 (36.83) ^{c^^}	100.95 (37.09)***	117.51 (45.12)

Differences between sex, and between physical activity levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; differences between age groups: ^a = 18-34 years and 35-49 years, ^b = 35-49 years and 50-64 years, ^c = 18-35 years and 50-64 years ([^] $p < 0.05$; ^{^^} $p < 0.01$, ^{^^^} $p < 0.001$).

Results are expressed as mean (standard deviation).

Difference between sex and physical activity levels measured with an independent t-test, difference between age groups measured with one-way repeated measures analysis of variance (ANOVA).

Table 41. Bivariate correlation analysis between standing long jump field-based test, force platform and isokinetic measurements, and anthropometrics variables.

	SLJ	Sex	Age	PA levels	Weight	Height	BMI	Waist circ.	Hip circ.	Tricipital SKF	Subscapular SKF	SUM tricipital + subscapular SKF	Body fat (%)	Lean mass
SLJ	1	0.560**	-0.421**	0.286**	0.176**	0.531**	-0.137**	-0.079	-0.208**	-0.518**	-0.427**	-0.495**	-0.665**	0.481**
Max. force	0.396**	0.104*	-0.288**	0.212**	-0.090	0.112*	-0.197**	-0.198**	-0.205**	-0.302**	-0.304**	-0.322**	-0.344**	0.069
Max. H force	0.815**	0.358**	-0.535**	-0.266**	0.018	0.317**	-0.182**	-0.191**	-0.215**	-0.445**	-0.386**	-0.437**	-0.582**	0.280**
Max.V force	0.259**	0.023	-0.212**	0.163**	-0.119*	0.044	-0.191**	-0.193**	-0.191**	-0.235**	-0.255**	-0.262**	-0.258**	0.002
Max. power	0.460**	0.196**	-0.355**	0.266**	-0.072	0.196**	-0.221**	-0.228**	-0.221**	-0.356**	-0.349**	-0.374**	-0.400**	0.117*
Maximal H power	0.826**	0.387**	-0.531**	0.287**	0.069	0.376**	-0.151**	-0.144**	-0.171**	-0.435**	-0.376**	-0.426**	-0.567**	0.332**
Maximal Vpower	0.317**	0.127*	-0.279**	0.234**	-0.095	0.131**	-0.211**	-0.216**	0.206**	-0.300**	-0.306**	-0.323**	-0.312**	0.053
Jump take off velocity	0.768**	0.373**	-0.458**	0.324**	0.052	0.389**	-0.184**	-0.162**	-0.202**	-0.446**	-0.398**	-0.444**	-0.562**	0.321**
H jump take off velocity	0.766**	0.37**	-0.458**	0.287**	0.102*	0.402**	-0.127*	-0.106*	-0.133**	-0.392**	-0.337**	-0.382**	-0.518**	0.349**
V jump take off velocity	0.294**	0.183**	-0.218**	0.214**	-0.044	0.178**	-0.171**	-0.147**	-0.172**	-0.267**	-0.254**	-0.275**	-0.287**	0.096
Push off force impulse	0.528**	0.376**	-0.293**	0.238**	0.212**	0.387**	0.021	0.029	-0.010	-0.253**	-0.168**	-0.216**	-0.330**	0.369**
H push off force impulse	0.659**	0.552**	-0.260**	0.171**	0.485**	0.614**	0.234**	0.339**	0.198**	-0.142**	0.045	-0.036	-0.279**	0.644**
V push off force impulse	0.356**	0.334**	-0.202**	0.202**	0.237**	0.335**	0.087	0.089	0.050	-0.188**	-0.115*	-0.155**	-0.220**	0.323**
Maxi. isometric peak torque	0.658**	0.645**	0.293**	0.194**	0.466**	0.669**	0.169**	0.258**	0.095	-0.316**	-0.156**	-0.237**	-0.434**	0.679**
Max. 60/60 flex. peak torque	0.671**	0.609**	-0.318**	0.200**	0.446**	0.648**	0.159**	0.233**	0.092	-0.332**	-0.169**	-0.253**	-0.434**	0.667**
Max. 60/60 ext. peak torque	0.672**	0.615**	-0.328**	0.217**	0.488**	0.661**	0.199**	0.270**	0.141**	-0.279**	-0.122*	-0.200**	-0.396**	0.678**

Max. 60/60 flex. peak torque, P/R	0.668**	0.588**	-0.306**	0.191**	0.427**	0.627**	0.147**	0.212**	0.078	-0.329**	-0.174**	-0.254**	-0.433**	0.647**
Max. 60/60 ext. peak torque, P/R	0.661**	0.604**	-0.310**	0.196**	0.490**	0.653**	0.208**	0.276**	0.153**	-0.271**	-0.112*	-0.191**	-0.379**	0.673**

SLJ, Standing long jump; Max., Maximal; H, Horizontal, V, Vertical, Flex., Flexion; Ext., Extension; P/R, Per Repetition; SKF, Skinfolts. * p<0.05, ** p<0.01.

Table 42. Bivariate correlation analysis between standing long jump field-based test, force platform and isokinetic measurements, and anthropometrics variables, distributed by sex, age groups, and physical activity levels.

	SLJ	Age	PA levels	Weight	Height	BMI	Waist circ.	Hip circ.	Tricipital SKF	Subscapular SKF	SUM tricipital + subscapular SKF	Body fat (%)	Lean mass
Female													
SLJ	1	-0.543**	0.254**	0.206**	0.260**	-0.352**	-0.479**	-0.262**	-0.457**	-0.422**	-0.462**	-0.562**	0.180*
Max. force	0.378**	-0.216**	0.192**	-0.233**	0.017	-0.253**	-0.288**	-0.242**	-0.302**	-0.309**	-0.324**	-0.354**	-0.035
Max. H force	0.778**	-0.544**	-0.205**	-0.211**	0.134	-0.290**	-0.426**	-0.239**	-0.360**	-0.354**	-0.377**	-0.533**	0.137
Max.V force	0.288**	-0.147*	0.153*	-0.218**	0.004	-0.230**	-0.242**	-0.224**	-0.265**	-0.271**	-0.284**	-0.292**	-0.066
Max. power	0.358**	-0.267**	0.239**	-0.211**	0.063	-0.255**	-0.313**	-0.211**	-0.298**	-0.327**	-0.334**	-0.338**	-0.039
Maximal H power	0.803**	-0.543**	0.215**	-0.145*	0.192**	-0.252**	-0.381**	-0.184**	-0.316**	-0.337**	-0.348**	-0.486**	0.207**
Maximal V power	0.227**	-0.191**	0.208**	-0.199**	0.038	-0.230**	-0.264**	-0.194**	-0.256**	-0.288**	-0.291**	-0.266**	-0.084
Jump take off velocity	0.723**	-0.468**	0.228**	-0.174*	0.185**	-0.282**	-0.387**	-0.213**	-0.302**	-0.355**	-0.352**	-0.456**	0.133
H jump take off velocity	0.733**	-0.488**	0.190**	-0.102	0.221**	-0.224**	-0.329**	-0.141*	-0.251**	-0.294**	-0.293**	-0.412**	0.208**
V jump take off velocity	0.169*	-0.139*	0.134	-0.099	0.134	-0.170*	-0.180*	-0.115	-0.144*	-0.191**	-0.182*	-0.155*	-0.004
Push off force impulse	0.359**	-0.256**	0.201**	-0.020	0.135	-0.095	-0.172*	-0.039	-0.169*	-0.193**	-0.194**	-0.199**	-0.140*

H push off force impulse	0.540**	-0.374**	0.111	0.431**	0.384**	0.244**	0.139*	0.356**	0.160*	0.146*	-0.161*	0.045	0.630**
V push off force impulse	0.169*	-0.138	0.129	0.031	0.191**	-0.066	-0.077	-0.003	-0.065	-0.103	-0.093	-0.069	0.124
Maxi. isometric peak torque	0.514**	-0.419**	0.178*	0.169*	0.344**	-0.006	-0.115	0.070	-0.114	-0.114	-0.121	-0.192**	0.412**
Max. 60/60 flex. peak torque	0.515**	-0.410**	0.136	0.222**	0.362**	0.036	-0.095	0.142*	-0.061	-0.048	-0.057	-0.168*	0.481**
Max. 60/60 ext. peak torque	0.557**	-0.470**	0.173*	0.222**	0.338**	0.051	-0.088	0.139*	-0.044	-0.031	-0.039	-0.166*	0.476**
Max. 60/60 flex. peak torque, P/R	0.515**	-0.403**	0.138	0.189**	0.351**	0.006	-0.125	0.099	-0.095	-0.083	-0.093	-0.213**	0.469**
Max. 60/60 ext. peak torque, P/R	0.514**	-0.435**	0.142*	0.250**	0.345**	0.073	-0.044	0.157*	-0.020	-0.003	-0.011	-0.134	0.488**
Male													
SLJ	1	-0.544**	0.254**	-0.253**	0.179*	-0.376**	-0.488**	-0.289**	-0.363**	-0.472**	-0.450**	-0.510**	-0.025
Max. force	0.429**	-0.359**	0.209**	-0.173*	0.080	-0.229**	-0.288**	-0.197**	-0.273**	-0.291**	-0.298**	-0.325**	-0.033
Max. H force	0.804**	-0.623**	-0.257**	-0.281**	0.042	-0.331**	-0.464**	-0.269**	-0.365**	-0.404**	-0.408**	-0.472**	-0.108
Max.V force	0.300**	-0.271**	0.168*	-0.142*	0.071	-0.191**	-0.232**	-0.175**	-0.227**	-0.242**	-0.248**	-0.276**	-0.019
Max. power	0.471**	-0.450**	0.253**	-0.251**	0.089	-0.318**	-0.415**	-0.256**	-0.323**	-0.350**	-0.356**	-0.368**	-0.109
Maximal H power	0.798**	-0.635**	0.288**	-0.241**	0.113	-0.316**	-0.430**	-0.231**	-0.361**	-0.398**	-0.402**	-0.462**	-0.046
Maximal V power	0.339**	-0.365**	0.231**	-0.222**	0.070	-0.278**	-0.357**	-0.231**	-0.290**	-0.308**	-0.315**	-0.305**	-0.108
Jump take off velocity	0.728**	-0.543**	0.351**	-0.245**	0.180**	-0.349**	-0.446**	-0.264**	-0.399**	-0.425**	-0.435**	-0.479**	-0.022
H jump take off velocity	0.723**	-0.527**	0.310**	-0.180**	0.191**	-0.280**	-0.371**	-0.191**	-0.330**	-0.356**	-0.363**	-0.416**	0.027
V jump take off velocity	0.284**	-0.309**	0.251**	-0.254**	0.002	-0.287**	-0.347**	-0.243**	-0.283**	-0.287**	-0.300**	-0.295**	-0.162*
Push off force impulse	0.442**	-0.389**	0.204**	-0.007	0.202**	-0.094	-0.201**	-0.038	-0.115	-0.124	-0.126	-0.161*	0.096

H push off force impulse	0.471**	-0.336**	0.120	0.180**	0.375**	0.043	0.080	0.130	0.039	0.094	0.075	-0.029	0.338**
V push off force impulse	0.256**	-0.297**	0.196**	0.065	0.101	0.029	-0.097	0.047	-0.083	-0.088	-0.090	-0.060	0.072
Maxi. isometric peak torque	0.457**	-0.412**	0.116	0.137	0.398**	-0.016	-0.071	0.069	-0.115	-0.149*	-0.142*	-0.152*	0.318**
Max. 60/60 flex. peak torque	0.510**	-0.453**	0.162*	0.115	0.389**	-0.033	-0.094	0.036	-0.194**	-0.201**	-0.208**	-0.204**	0.334**
Max. 60/60 ext. peak torque	0.483**	-0.438**	0.173*	0.194**	0.428**	0.032	-0.026	0.123	-0.093	-0.123	-0.116	-0.109	0.355**
Max. 60/60 flex. peak torque, P/R	0.515**	-0.421**	0.145*	0.109	0.370**	-0.032	-0.101	0.029	-0.182**	-0.193**	-0.198**	-0.196**	0.320**
Max. 60/60 ext. peak torque, P/R	0.486**	-0.413**	0.149*	0.200**	0.417**	0.045	-0.021	0.136*	-0.095	-0.119	-0.115	-0.097	0.356**

	SLJ	Sex	PA levels	Weight	Height	BMI	Waist circ.	Hip circ.	Tricipital SKF	Subscapular SKF	SUM tricipital + subscapular SKF	Body fat (%)	Lean mass
18-34 years													
SLJ	1	-0.654**	0.266**	0.177*	0.610**	-0.103	0.097	-0.259**	-0.541**	-0.389**	-0.70**	-0.603**	0.522**
Max. force	0.507**	0.298**	0.175*	-0.045	0.246**	-0.191*	-0.060	-0.282**	-0.418**	-0.315**	-0.372**	-0.450**	0.181*
Max. H force	0.781**	0.490**	0.205*	0.015	0.369**	-0.167	-0.015	-0.293**	-0.501**	-0.350**	-0.428**	-0.585**	0.316**
Max.V force	0.395**	0.215*	0.128	-0.069	0.188*	-0.193*	-0.077	-0.275**	-0.359**	-0.279**	-0.324**	-0.384**	0.121
Max. power	0.405**	0.327**	0.224**	0.031	0.363**	-0.150	-0.064	-0.218*	-0.402**	-0.299**	-0.355**	-0.434**	0.254**
Maximal H power	0.823**	0.524**	0.252**	0.101	0.445**	-0.095	0.067	-0.217*	-0.465**	-0.326**	-0.399**	-0.541**	0.399**
Maximal V power	0.275**	-0.251**	0.198*	0.019	0.304**	-0.136	-0.070	-0.187*	-0.340**	-0.256**	-0.302**	-0.353**	0.198*
Jump take off velocity	0.780**	0.512**	0.270**	0.129	0.512**	-0.097	0.062	-0.205*	-0.453**	-0.330**	-0.396**	-0.531**	0.430**
H jump take off velocity	0.784**	0.492**	0.232**	0.160	0.486**	-0.046	0.119	-0.147	-0.402**	-0.279**	-0.343**	-0.473**	0.435**

V jump take off velocity	0.335**	0.284**	0.175*	0.046	0.347**	-0.120	-0.051	-0.194*	-0.317**	-0.212*	-0.265**	-0.332**	0.220*
Push off force impulse	0.436**	0.475**	0.203*	0.424**	0.579**	0.220*	0.292**	0.106	-0.200*	-0.012	-0.091	-0.242**	0.574**
H push off force impulse	0.623**	0.648**	0.152	0.793**	0.745**	0.599**	0.733**	0.421**	0.020	0.276**	0.180*	-0.026	0.879**
V push off force impulse	0.411**	0.440**	0.195*	0.359**	0.548**	0.160	0.245**	0.038	-0.185*	-0.009	-0.084	-0.219*	0.500**
Maxi. isometric peak torque	0.664**	0.679**	0.201*	0.475**	0.693**	0.239**	0.412**	0.057	-0.313**	-0.098	-0.193*	-0.319**	0.710**
Max. 60/60 flex. peak torque	0.652**	0.692**	0.248**	0.532**	0.734**	0.289**	0.474**	0.086	-0.307**	-0.048	-0.159	-0.305**	0.765**
Max. 60/60 ext. peak torque	0.659**	0.668**	0.258**	0.565**	0.723**	0.332**	0.496**	0.149	-0.266**	-0.003	-0.114	-0.255**	0.765**
Max. 60/60 flex. peak torque, P/R	0.649**	0.671**	0.243**	0.526**	0.710**	0.292**	0.463**	0.085	-0.304**	-0.043	-0.155	-0.301**	0.756**
Max. 60/60 ext. peak torque, P/R	0.644**	0.652**	0.234**	0.580**	0.718**	0.354**	0.509**	0.172*	-0.249**	0.019	-0.093	-0.229**	0.768**
35-49 years													
SLJ	1	0.652**	0.219*	0.243**	0.600**	-0.135	0.038	-0.242**	-0.612**	-0.438**	-0.537**	-0.685**	0.554**
Max. force	0.241**	-0.071	0.292**	-0.216*	-0.038	-0.271**	-0.291**	-0.321**	-0.318**	-0.342**	-0.348**	-0.321**	-0.092
Max. H force	0.801**	0.398**	0.277**	0.026	0.360**	-0.202**	-0.125	-0.291**	-0.550**	-0.430**	-0.504**	-0.632**	0.315**
Max.V force	0.092	-0.173*	0.237**	-0.262**	-0.133	-0.268**	-0.309**	-0.291**	-0.233**	-0.285**	-0.275**	-0.230**	-0.184*
Max. power	0.399**	0.195*	0.316**	-0.105	0.214*	-0.275**	-0.230**	-0.268**	-0.406**	-0.423**	-0.436**	-0.398**	0.102
Maximal H power	0.823**	0.470**	0.306**	0.121	0.470**	-0.146	-0.049	-0.197**	-0.521**	-0.404**	-0.476**	-0.607**	0.416**
Maximal V power	0.232*	0.112	0.280**	-0.133	0.127	-0.257**	-0.226**	0.250**	-0.336**	-0.362**	-0.368**	-0.299**	0.027
Jump take off velocity	0.750**	-0.462**	0.358**	0.064	0.466**	-0.217*	-0.094	-0.258**	-0.545**	-0.446**	-0.511**	-0.610**	0.374**
H jump take off velocity	0.752**	-0.474**	0.314**	0.140	0.501**	-0.137	-0.017	-0.165	-0.462**	-0.360**	-0.423**	-0.551**	0.424**

V jump take off velocity	0.197*	0.232**	0.305**	-0.020	0.239**	-0.166	-0.095	-0.180*	-0.302**	-0.252**	-0.287**	-0.279**	0.141
Push off force impulse	0.550**	0.429**	0.270**	0.094	0.427**	-0.141	-0.034	-0.162	-0.366**	-0.326**	-0.360**	-0.416**	0.346**
H push off force impulse	0.623**	0.521**	0.150	0.222*	0.552*	-0.035	0.228**	-0.008	-0.205*	-0.044	-0.118	-0.341**	0.504**
V push off force impulse	0.289**	0.407**	0.260**	0.313**	0.380**	0.159	0.171	0.101	-0.216**	-0.138	-0.179*	-0.182*	0.397**
Maxi. isometric peak torque	0.645**	0.693**	0.100	0.479**	0.704**	0.163	0.341**	0.161	-0.299**	-0.129	-0.211*	-0.390**	0.703**
Max. 60/60 flex. peak torque	0.665**	0.621**	0.136	0.370**	0.644**	0.070	0.220*	0.081	-0.385**	-0.233**	-0.313**	-0.448**	0.639**
Max. 60/60 ext. peak torque	0.651**	0.371**	0.148	0.439**	0.694**	0.120	0.302**	0.136	-0.318**	-0.183*	-0.252**	-0.420**	0.690**
Max. 60/60 flex. peak torque, P/R	0.671**	0.601**	0.129	0.332**	0.618**	0.038	0.174*	0.042	-0.401**	-0.258**	-0.334**	-0.463**	0.603**
Max. 60/60 ext. peak torque, P/R	0.641**	0.667**	0.135**	0.423**	0.670**	0.117	0.295**	0.126	-0.325**	-0.188*	-0.259**	-0.415**	0.673**
50-64 years													
SLJ	1	0.644**	0.313**	0.334**	0.543**	0.003	0.076	-0.151	-0.457**	-0.373**	-0.455**	-0.681**	0.550**
Max. force	0.228**	0.107	0.145	0.052	0.111	-0.024	-0.046	-0.018	-0.127	-0.160	-0.160	-0.139	0.122
Max. H force	0.739**	0.461**	0.273**	0.220**	0.349**	0.017	0.035	-0.105	-0.361**	-0.281**	-0.351**	-0.479**	0.375**
Max.V force	0.124	0.032	0.110	0.017	0.059	-0.030	-0.060	0.037	-0.075	-0.128	-0.115	-0.067	0.066
Max. power	0.313**	0.134	0.227**	-0.089	0.008	-0.138	-0.140	-0.200*	-0.221**	-0.225**	-0.247**	-0.212*	0.003
Maximal H power	0.723**	0.464**	0.283**	0.223**	0.379**	-0.004	0.042	-0.126	-0.390**	-0.292**	-0.373**	-0.505**	0.388**
Maximal Vpower	0.194*	0.049	0.192*	-0.146	-0.054	-0.167*	-0.170*	-0.198*	-0.179*	-0.221**	-0.224**	-0.139	-0.076
Jump take off velocity	0.620**	0.361**	0.306**	0.119	0.313**	-0.098	-0.047	-0.203*	-0.391**	-0.341**	-0.403**	-0.477**	0.294**
H jump take off velocity	0.608**	0.375**	0.260**	0.185*	0.349**	-0.034	0.024	-0.125	-0.349**	-0.266**	-0.336**	-0.446**	0.339**

V jump take off velocity	0.178*	0.080	0.129	-0.110	-0.022	-0.148	-0.127	-0.128	-0.149	-0.228**	-0.213*	-0.153	-0.042
Push off force impulse	0.427**	0.319**	0.186*	0.191*	0.201*	0.091	0.115	0.007	-0.177*	-0.096	-0.146	-0.233**	0.237**
H push off force impulse	0.636**	0.612**	0.151	0.637**	0.631**	0.356**	0.465**	0.243**	-0.212*	0.010	-0.099	-0.375**	0.675**
V push off force impulse	0.245**	0.211*	0.114	0.092	0.125	0.014	0.057	0.020	-0.132	-0.135	-0.148	-0.166*	0.122
Maxi. isometric peak torque	0.590**	0.673**	0.226**	0.594**	0.671**	0.260**	0.358**	0.104	-0.308**	-0.137	-0.237**	-0.505**	0.713**
Max. 60/60 flex. peak torque	0.580**	0.646**	0.134	0.598**	0.645**	0.287**	0.373**	0.147	-0.287**	-0.134	-0.225**	-0.475**	0.702**
Max. 60/60 ext. peak torque	0.589**	0.647**	0.177*	0.643**	0.651**	0.343**	0.417**	0.186*	-0.221**	-0.060	-0.147	-0.410**	0.695**
Max. 60/60 flex. peak torque, P/R	0.568**	0.619**	0.116	0.575**	0.628**	0.269**	0.357**	0.139	-0.262**	-0.137	-0.214*	-0.459**	0.679**
Max. 60/60 ext. peak torque, P/R	0.581**	0.627**	0.149	0.645**	0.650**	0.647**	0.417**	0.209*	-0.206*	-0.054	-0.136	-0.394**	0.687**

	SLJ	Sex	Age	Weight	Height	BMI	Waist circ.	Hip circ.	Tricipital SKF	Subscapular SKF	SUM tricipital + subscapular SKF	Body fat (%)	Lean mass
Non-Active													
SLJ	1	0.531**	-0.428**	0.215**	0.500**	-0.058	-0.014	-0.099	-0.333**	-0.255**	-0.309**	-0.588**	0.456**
Max. force	0.320**	0.059	-0.283**	-0.113	0.090	-0.214**	-0.228**	-0.213**	-0.250**	-0.264**	-0.277**	-0.331**	0.020
Max. H force	0.800**	0.311**	-0.561**	-0.005	0.278**	-0.177*	-0.198**	-0.193**	-0.310**	-0.255**	-0.298**	-0.551**	0.216**
Max.V force	0.192**	-0.012	0.216**	-0.141	0.037	-0.217**	-0.233**	-0.205**	-0.214**	-0.245**	-0.249**	-0.262**	-0.037
Max. power	0.411**	0.142	-0.344**	-0.116	0.159*	-0.249**	-0.250**	-0.222**	-0.321**	-0.322**	-0.345**	-0.412**	0.052
Maximal H power	0.799**	0.331**	-0.552**	0.037	0.323**	-0.147*	-0.154*	-0.139	-0.270**	-0.234**	-0.267**	-0.522**	0.260**
Maximal V power	0.277**	0.073	-0.270**	-0.146*	0.103	-0.254**	-0.250**	0.215**	-0.300**	-0.316**	-0.332**	-0.341**	-0.011

Jump take off velocity	0.728**	0.276**	-0.510**	-0.005	0.317**	-0.200**	-0.196**	-0.157*	-0.267**	-0.267**	-0.286**	-0.509**	0.217**
H jump take off velocity	0.701**	0.275**	-0.485**	0.050	0.311**	-0.125	-0.126	-0.085	-0.185**	-0.187**	-0.200**	-0.442**	0.247**
V jump take off velocity	0.251**	0.105	-0.210**	-0.093	0.184*	-0.222**	-0.188**	-0.156*	-0.226**	-0.227**	-0.243**	-0.290**	0.025
Push off force impulse	0.500**	0.236**	-0.277**	0.137	0.346**	-0.035	-0.001	-0.032	-0.169*	-0.110	-0.144*	-0.337**	0.290**
H push off force impulse	0.623**	0.500**	-0.227**	0.350**	0.539**	0.137	0.298**	0.165*	0.061	0.184*	0.143*	-0.224**	0.521**
V push off force impulse	0.306**	0.282**	-0.181*	0.250**	0.334**	0.111	0.099	0.149*	-0.087	-0.033	-0.059	-0.166*	0.285**
Maxi. isometric peak torque	0.662**	0.671**	-0.295**	0.512**	0.651**	0.264**	0.350**	0.237**	-0.068	0.038	-0.006	-0.293**	0.672**
Max. 60/60 flex. peak torque	0.657**	0.613**	-0.254**	0.475**	0.621**	0.237**	0.312**	0.233**	-0.065	0.028	-0.011	-0.288**	0.648**
Max. 60/60 ext. peak torque	0.671**	0.631**	-0.300**	0.540**	0.646**	0.298**	0.375**	0.293**	0.009	0.099	0.066	-0.235**	0.669**
Max. 60/60 flex. peak torque, P/R	0.674**	0.603**	-0.248**	0.444**	0.610**	0.204**	0.275**	0.191**	-0.084	0.002	-0.036	-0.326**	0.631**
Max. 60/60 ext. peak torque, P/R	0.666**	0.627**	-0.282**	0.540**	0.645**	0.301**	0.370**	0.301**	0.000	0.091	0.058	-0.227**	0.667**
Active													
SLJ	1	0.567**	-0.398**	0.189**	0.566**	-0.155*	-0.041	-0.219**	-0.625**	-0.518**	-0.600**	-0.675**	0.494**
Max. force	0.390**	0.097	-0.264**	-0.043	0.107	-0.133	-0.093	-0.128	-0.266**	-0.259**	-0.278**	-0.270**	0.079
Max. H force	0.804**	0.370**	-0.503**	0.087	0.350**	-0.130	-0.095	-0.153*	-0.503**	-0.435**	-0.494**	-0.548**	0.320**
Max.V force	0.251**	0.015	-0.182**	-0.077	0.029	-0.124	-0.090	-0.123	-0.183**	-0.194**	-0.201**	-0.181**	0.009
Max. power	0.420**	0.191**	-0.341**	-0.001	0.208**	-0.145*	-0.130	-0.141*	-0.290**	-0.277**	-0.300**	-0.291**	0.133
Maximal H power	0.829**	0.403**	-0.513**	0.144*	0.417**	-0.098	-0.043	-0.112	-0.496**	-0.417**	-0.479**	-0.537**	0.374**
Maximal V power	0.260**	0.123	-0.259**	-0.023	0.132	-0.122	-0.113	-0.127	-0.207**	-0.205**	-0.219**	-0.189**	0.067

Jump take off velocity	0.762**	0.417**	-0.405**	0.158*	0.450**	-0.104	-0.020	-0.144*	-0.510**	-0.419**	-0.487**	-0.521**	0.388**
H jump take off velocity	0.787**	0.420**	-0.419**	0.208**	0.480**	-0.060	0.024	-0.080	-0.494**	-0.382**	-0.457**	-0.499**	0.422**
V jump take off velocity	0.248**	0.205**	-0.194**	0.025	0.157*	-0.076	-0.036	-0.120	-0.218**	-0.190**	-0.215**	-0.195**	0.121
Push off force impulse	0.492**	0.381**	-0.278**	0.344**	0.420**	0.161*	0.173*	0.111	-0.230**	-0.105	-0.169*	-0.222**	0.426**
H push off force impulse	0.676**	0.582**	-0.269**	0.679**	0.683**	0.424**	0.495**	0.319**	-0.269**	0.003	-0.121	-0.252**	0.757**
V push off force impulse	0.324**	0.347**	-0.191**	0.267**	0.327**	0.128	0.174*	0.044	-0.186**	-0.090	-0.139*	-0.171*	0.338**
Maxi. isometric peak torque	0.643**	0.606**	-0.281**	0.465**	0.675**	0.137*	0.261**	0.044	-0.469**	-0.256**	-0.371**	-0.494**	0.674**
Max. 60/60 flex. peak torque	0.672**	0.591**	-0.359**	0.468**	0.662**	0.154*	0.257**	0.053	-0.490**	-0.260**	-0.383**	-0.493**	0.676**
Max. 60/60 ext. peak torque	0.654**	0.587**	-0.340**	0.499**	0.669**	0.185**	0.283**	0.105	-0.438**	-0.217**	-0.333**	-0.450**	0.681**
Max. 60/60 flex. peak torque, P/R	0.657**	0.562**	-0.343**	0.456**	0.634**	0.157*	0.246**	0.058	-0.469**	-0.250**	-0.367**	-0.465**	0.651**
Max. 60/60 ext. peak torque, P/R	0.645**	0.572**	-0.321**	0.498**	0.654**	0.195**	0.291**	0.114	-0.423**	-0.201**	-0.316**	-0.434**	0.672**

SLJ, Standing long jump; Max., Maximal; H, Horizontal, V, Vertical, Flex., Flexion; Ext., Extension; P/R, Per Repetition; SKF, Skinfolde. * p<0.05, ** p<0.01

Table 43. Stepwise regression models assessing the association of the force platform variables with the standing long jump distance.

	Step	β	r	R²	R² change	p value
1	Maximal horizontal power	0.827	0.827	0.682	0.683	<0.001
2	Maximal horizontal power	0.716	0.866	0.751	0.067	<0.001
	Sex	0.282				<0.001
3	Maximal horizontal power	0.632	0.878	0.771	0.021	<0.001
	Sex	0.216				<0.001
	% Body fat	-0.191				<0.001
4	Maximal horizontal power	0.401	0.882	0.778	0.006	<0.001
	Sex	0.225				<0.001
	% Body fat	-0.171				<0.001
	Maximal horizontal force	0.252				0.001
5	Maximal horizontal power	0.368	0.885	0.782	0.005	<0.001
	Sex	0.107				0.032
	% Body fat	-0.203				<0.001
	Maximal horizontal force	0.269				0.001
	Lean mass	0.131				0.004

β , Standardized regression coefficient; r, correlation coefficients; R², adjusted coefficients of determination.

Table 44. Test-retest reliability of standing long jump in the whole sample, and in a subsample after a learning period.

	T1*	T2*	Intertrial difference (T2-T1)*	p value	Cohen's <i>d</i>	ICC (95% CI)**	SSE	MSE	RMSE	% Error	% SEM	% CV	SEE
Whole sample (n= 410)													
Standing long jump (cm)	143.03 ± 38.53	155.68 ± 39.95	12.14 ± 14.46	<0.001	0.309	0.94 (0.75-0.97)	139972.00	358.90	18.94	9.47	8.46	7.06	14.48
Subsample after a learning period (n= 66)													
Standing long jump (cm)	142.41 ± 35.12	142.86 ± 35.26	0.45 ± 1.04	0.001	0.013	1.00 (0.99-1.00)	0.00	1.27	1.13	0.80	1.04	0.53	1.03

ICC, intraclass correlation coefficients; CI, confident interval; SSE, sum of squared errors; MSE, mean sum of squared errors; RMSE, root mean sum of squared errors; %Error, percentage error; %SEM, standard error of measurement; %CV, percentage coefficient of variation; SEE, standard error of estimate.

*T1 refers to test (trial 1) and T2 to retest (trial 2). T2-T1 refers to retest (trial 2) minus test (trial 1). Values are displayed as mean ± SD.

**ICC was significant at p<0.001

Table 45. Test-retest reliability of standing long jump, distributed by sex, age groups, and physical activity levels.

	T1*	T2*	Intertrial difference (T2-T1)*	p value	Cohen’s <i>d</i>	ICC (95% CI)**	SSE	MSE	RMSE	% Error	% SEM	% CV	SEE
Female													
Standing long jump (cm)	121.05 ± 28.86	133.00 ± 29.56	11.56 ± 13.38	<0.001	0.395	0.91 (0.61-0.96)	59910.00	312.03	17.66	10.39	9.64	7.63	13.14
Male													
Standing long jump (cm)	164.37 ± 34.57	177.67 ± 36.30	12.70 ± 15.63	<0.001	0.358	0.92 (0.69-0.96)	80062.00	404.35	20.11	10.05	7.69	6.68	15.51
18-34 years													
Standing long jump (cm)	159.64 ± 35.01	172.20 ± 37.30	12.03 ± 13.34	<0.001	0.333	0.94 (0.69-0.98)	41783.00	321.41	17.93	10.42	7.52	5.91	13.33
35-49 years													
Standing long jump (cm)	149.26 ± 36.66	163.10 ± 37.75	12.56 ± 15.00	<0.001	0.338	0.93 (0.70-0.97)	46499.00	381.14	19.52	11.35	8.36	6.35	14.87
50-64 years													
Standing long jump (cm)	121.42 ± 33.31	133.55 ± 34.10	11.87 ± 15.34	<0.001	0.351	0.92 (0.70-0.96)	67819.00	487.91	22.09	15.13	8.94	8.85	15.05
Non-Active													
Standing long jump (cm)	131.48 ± 36.18	143.86 ± 38.00	11.30 ± 15.02	<0.001	0.305	0.93 (0.78-0.97)	65864.00	352.21	18.77	10.48	8.35	8.01	14.97
Active													
Standing long jump (cm)	153.88 ± 37.52	166.73 ± 38.70	12.80 ± 14.04	<0.001	0.333	0.94 (0.68-0.98)	72739.00	360.09	18.98	10.48	8.21	6.46	13.89

ICC, intraclass correlation coefficients; CI, confident interval; SSE, sum of squared errors; MSE, mean sum of squared errors; RMSE, root mean sum of squared errors; %Error, percentage error; %SEM, standard error of measurement; %CV, percentage coefficient of variation; SEE, standard error of estimate.
*T1 refers to test (trial 1) and T2 to retest (trial 2). T2-T1 refers to retest (trial 2) minus test (trial 1). Values are displayed as mean ± SD.
**ICC was significant at p<0.001

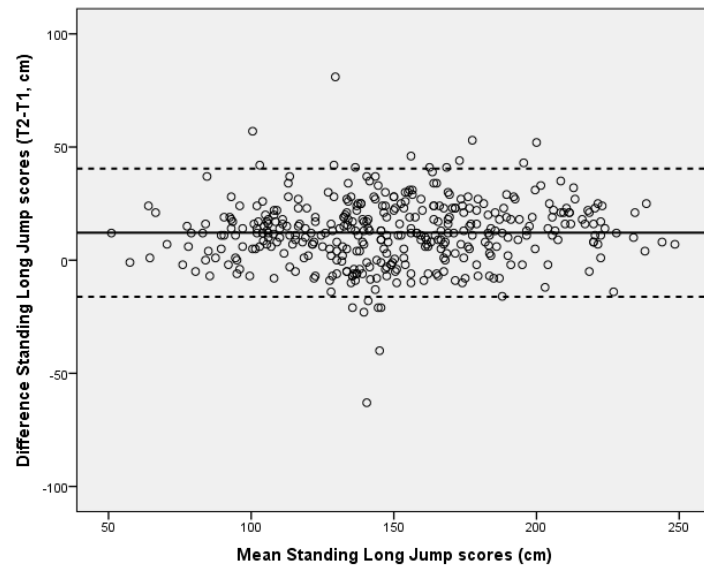
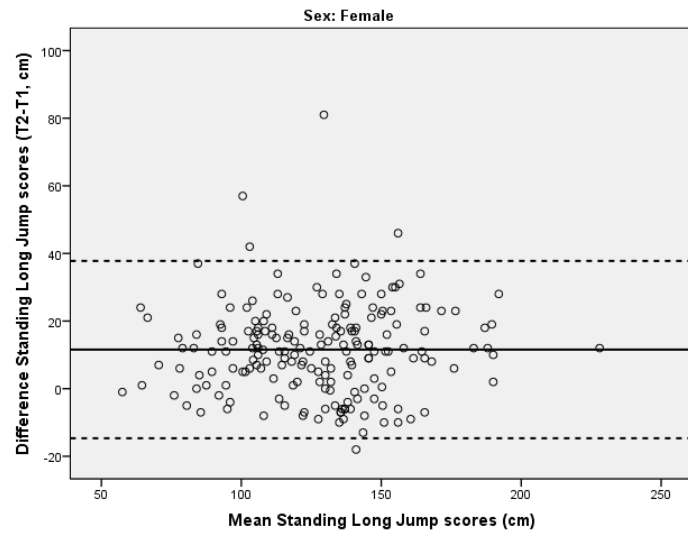


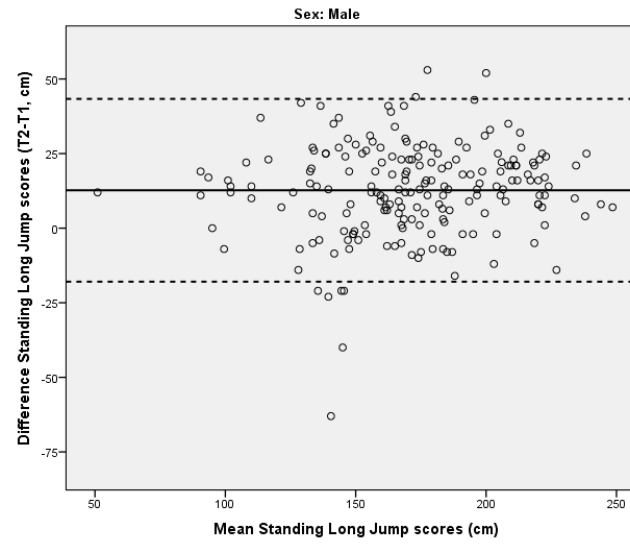
Figure 19. Bland-Altman plot of standing long jump test. The central line represents the mean differences between the second trial (T2) and the first trial (T1).

The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

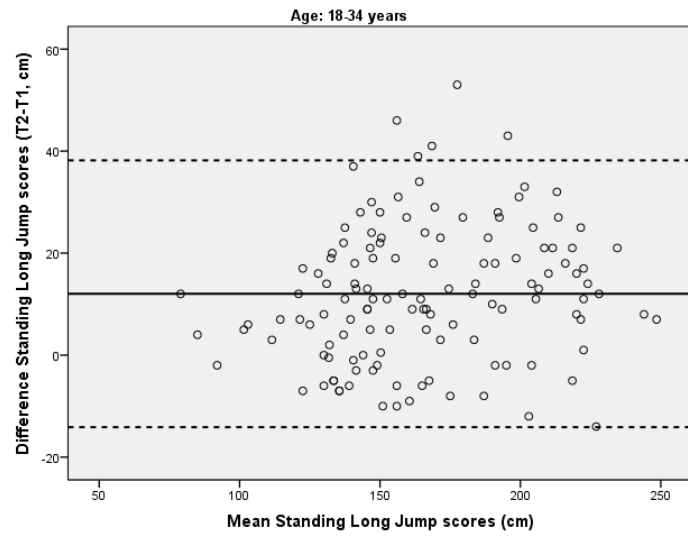
(A.1)



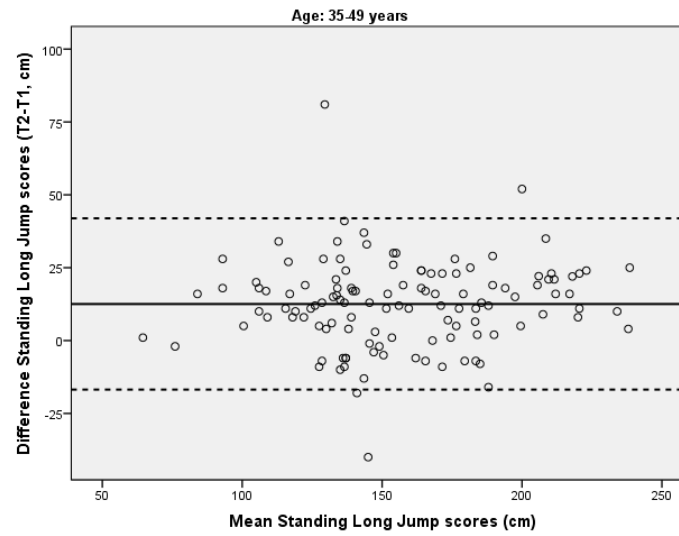
(A.2)



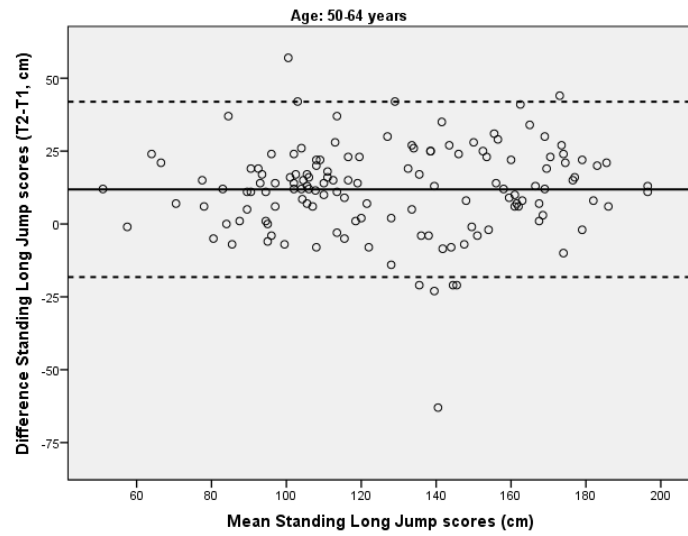
(A.3)



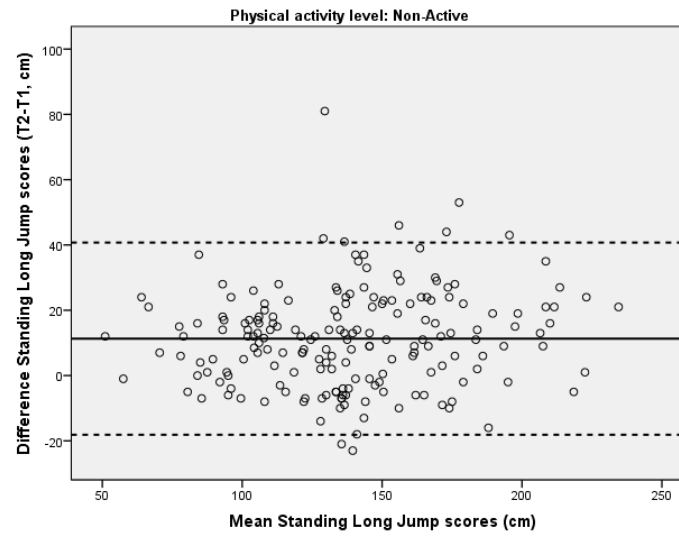
(A.4)



(A.5)



(A.6)



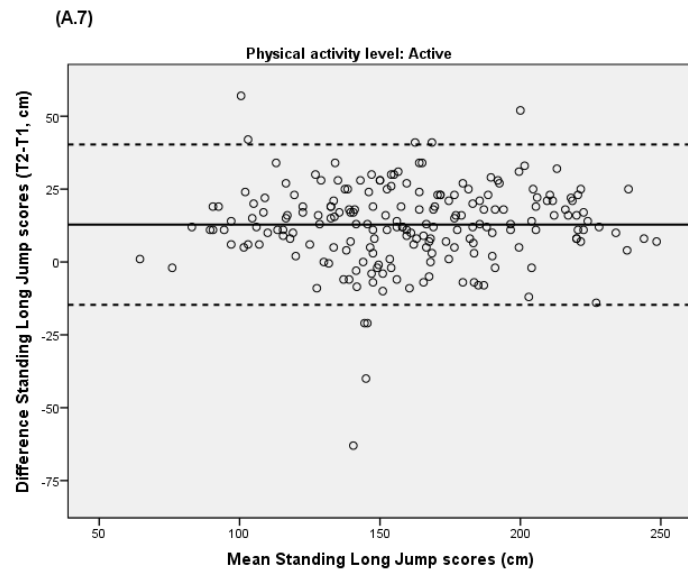


Figure 20. Bland-Altman plot of standing long jump* distributed by sex, age groups, and physical activity levels.

The central line represents the mean differences between the second trial (T2) and the first trial (T1). The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

*(A.1-A.2), distributed by sex; (A.3-A.5), distributed by age groups; (A.6-A.7), distributed by physical activity levels.

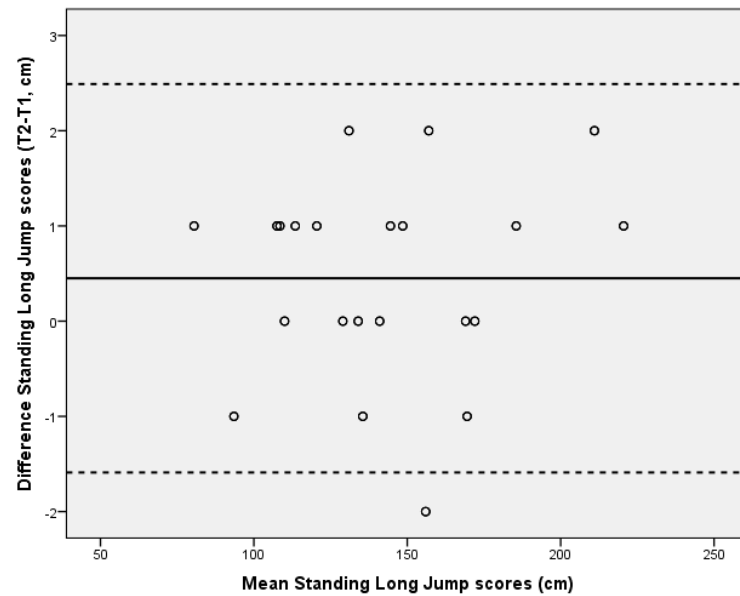


Figure 21. Bland-Altman plot of standing long jump test after a learning period.

The central line represents the mean differences between the second trial (T2) and the first trial (T1). The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

REFERENCES

1. Peel NM, Alapatt LJ, Jones LV, Hubbard RE. The Association Between Gait Speed and Cognitive Status in Community-Dwelling Older People: A Systematic Review and Meta-analysis. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2019;74(6):943-8.
2. Cavanaugh EJ, Richardson J, McCallum CA, Wilhelm M. The Predictive Validity of Physical Performance Measures in Determining Markers of Preclinical Disability in Community-Dwelling Middle-Aged and Older Adults: A Systematic Review. *Physical Therapy*. 2018;98(12):1010-21.
3. Grande G, Triolo F, Nuara A, Welmer AK, Fratiglioni L, Vetrano DL. Measuring gait speed to better identify prodromal dementia. *Experimental Gerontology*. 2019;124.
4. Quan MH, Xun PC, Chen C, Wen J, Wang YY, Wang R, et al. Walking Pace and the Risk of Cognitive Decline and Dementia in Elderly Populations: A Meta-analysis of Prospective Cohort Studies. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2017;72(2):266-70.
5. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *Journal of Cachexia Sarcopenia and Muscle*. 2020;11(1):3-25.
6. Van Kan GA, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. *Journal of Nutrition Health & Aging*. 2009;13(10):881-9.
7. Abe T, Kitamura A, Taniguchi Y, Amano H, Seino S, Yokoyama Y, et al. Pathway from gait speed to incidence of disability and mortality in older adults: A mediating role of physical activity. *Maturitas*. 2019;123:32-6.
8. Abu Samah Z, Singh DKA, Murukesu RR, Shahar S, Nordin NAM, Omar MA, et al. Discriminative and Predictive Ability of Physical Performance Measures in Identifying Fall Risk among Older Adults. *Sains Malaysiana*. 2018;47:2769-76.
9. Adachi T, Kamiya K, Kono Y, Iwatsu K, Shimizu Y, Honda I, et al. Estimation of reduced walking speed using simple measurements of physical and psychophysiological function in community-dwelling elderly people: a cross-sectional and longitudinal study. *Aging Clinical and Experimental Research*. 2019;31:59-66.
10. Andrasfay T. Changes in Physical Functioning as Short-Term Predictors of Mortality. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*. 2020;75(3):630-9.
11. Asai T, Oshima K, Fukumoto Y, Yonezawa Y, Matsuo A, Misu S. Does dual-tasking provide additional value in timed "up and go" test for predicting the occurrence of falls? A longitudinal observation study by age group (young-older or old-older adults). *Aging Clinical and Experimental Research*. 2020.
12. Austin N, Devine A, Dick I, Prince R, Bruce D. Fear of falling in older women: a longitudinal study of incidence, persistence, and predictors. *J Am Geriatr Soc*. 2007;55(10):1598-603.
13. Blain H, Carriere I, Sourial N, Berard C, Favier F, Colvez A, et al. Balance and walking speed predict subsequent 8-year mortality independently of current and intermediate events in well-functioning women aged 75 years and older. *Journal of Nutrition Health & Aging*. 2010;14:595-600.
14. Brach JS, Wert D, VanSwearingen JM, Newman AB, Studenski SA. Use of Stance Time Variability for Predicting Mobility Disability in Community-Dwelling Older Persons: A Prospective Study. *Journal of Geriatric Physical Therapy*. 2012;35:112-7.

15. Bravell ME, Finkel D, Aslan AD, Reynolds CA, Hallgren J, Pedersen NL. Motor functioning differentially predicts mortality in men and women. *Archives of Gerontology and Geriatrics*. 2017;72:6-11.
16. Breton E, Beloin F, Fortin C, Martin A, Ouellet ME, Payette H, et al. Gender-specific associations between functional autonomy and physical capacities in independent older adults: Results from the NuAge study. *Archives of Gerontology and Geriatrics*. 2014;58:56-62.
17. Briggs R, Carey D, Claffey P, McNicholas T, Donoghue O, Kennelly SP, et al. Do Differences in Spatiotemporal Gait Parameters Predict the Risk of Developing Depression in Later Life? *Journal of the American Geriatrics Society*. 2019;67(5):1050-6.
18. Bullain SS, Corrada MM, Perry SM, Kawas CH. Sound Body Sound Mind? Physical Performance and the Risk of Dementia in the Oldest-Old: The 90+Study. *Journal of the American Geriatrics Society*. 2016;64:1408-15.
19. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The Trajectory of Gait Speed Preceding Mild Cognitive Impairment. *Archives of Neurology*. 2010;67(8):980-6.
20. Callisaya ML, Ayers E, Barzilai N, Ferrucci L, Guralnik JM, Lipton RB, et al. Motoric Cognitive Risk Syndrome and Falls Risk: A Multi-Center Study. *Journal of Alzheimers Disease*. 2016;53:1043-52.
21. Camargo EC, Weinstein G, Beiser AS, Tan ZS, DeCarli C, Kelly-Hayes M, et al. Association of Physical Function with Clinical and Subclinical Brain Disease: The Framingham Offspring Study. *Journal of Alzheimer's Disease*. 2016;53:1597-608.
22. Clemson L, Kendig H, Mackenzie L, Browning C. Predictors of Injurious Falls and Fear of Falling Differ: An 11-Year Longitudinal Study of Incident Events in Older People. *Journal of Aging and Health*. 2015;27:239-56.
23. Cooper R, Strand BH, Hardy R, Patel KV, Kuh D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. *BMJ (Clinical research ed)*. 2014;348:2219.
24. Dargent-Molina P, Schott AM, Hans D, Favier F, Grandjean H, Baudoin C, et al. Separate and combined value of bone mass and gait speed measurements in screening for hip fracture risk: Results from the EPIDOS study. *Osteoporosis International*. 1999;9(2):188-92.
25. Davis JC, Best JR, Bryan S, Li LC, Hsu CL, Gomez C, et al. Mobility Is a Key Predictor of Change in Well-Being Among Older Adults Who Experience Falls: Evidence From the Vancouver Falls Prevention Clinic Cohort. *Arch Phys Med Rehabil*. 2015;96(9):1634-40.
26. Deshpande N, Metter EJ, Guralnik J, Bandinelli S, Ferrucci L. Predicting 3-Year Incident Mobility Disability in Middle-Aged and Older Adults Using Physical Performance Tests. *Archives of Physical Medicine and Rehabilitation*. 2013;94:994-7.
27. Doi T, Hirata S, Ono R, Tsutsumimoto K, Misu S, Ando H. The harmonic ratio of trunk acceleration predicts falling among older people: results of a 1-year prospective study. *Journal of Neuroengineering and Rehabilitation*. 2013;10.
28. Doi T, Nakakubo S, Tsutsumimoto K, Kim MJ, Kurita S, Ishii H, et al. Spatio-temporal gait variables predicted incident disability. *Journal of Neuroengineering and Rehabilitation*. 2020;17(1).
29. Doi T, Tsutsumimoto K, Nakakubo S, Kim MJ, Kurita S, Hotta R, et al. Physical Performance Predictors for Incident Dementia Among Japanese Community-Dwelling Older Adults. *Physical Therapy*. 2019;99(9):1132-40.

30. Donoghue O, Feeney J, O'Leary N, Kenny RA. Baseline Mobility is Not Associated with Decline in Cognitive Function in Healthy Community-Dwelling Older Adults: Findings From The Irish Longitudinal Study on Ageing (TILDA). *American Journal of Geriatric Psychiatry*. 2017;26:438-48.
31. Elbaz A, Sabia S, Brunner E, Shipley M, Marmot M, Kivimaki M, et al. Association of walking speed in late midlife with mortality: results from the Whitehall II cohort study. *Age*. 2013;35:943-52.
32. Ersoy Y, MacWalter RS, Durmus B, Altay ZE, Baysal O. Predictive Effects of Different Clinical Balance Measures and the Fear of Falling on Falls in Postmenopausal Women Aged 50 Years and Over. *Gerontology*. 2009;55:660-5.
33. Frames CW, Soangra R, Lockhart TE, Lach J, Ha DS, Roberto KA, et al. Dynamical Properties of Postural Control in Obese Community-Dwelling Older Adults. *Sensors*. 2018;18.
34. Georgiopoulou VV, Kalogeropoulos AP, Chowdhury R, Binongo JNG, Bibbins-Domingo K, Rodondi N, et al. Exercise Capacity, Heart Failure Risk, and Mortality in Older Adults: The Health ABC Study. *American Journal of Preventive Medicine*. 2016;52:144-53.
35. Heiland EG, Welmer AK, Wang R, Santoni G, Fratiglioni L, Qiu CX. Cardiovascular Risk Factors and the Risk of Disability in Older Adults: Variation by Age and Functional Status. *Journal of the American Medical Directors Association*. 2018;20:208.
36. Hoogendijk EO, Rijnhart JJM, Skoog J, Robitaille A, van den Hout A, Ferrucci L, et al. Gait speed as predictor of transition into cognitive impairment: Findings from three longitudinal studies on aging. *Experimental Gerontology*. 2020;129.
37. Idland G, Pettersen R, Avlund K, Bergland A. Physical performance as long-term predictor of onset of activities of daily living (ADL) disability: A 9-year longitudinal study among community-dwelling older women. *Archives of Gerontology and Geriatrics*. 2013;56:501-6.
38. Jung HW, Jang IY, Lee CK, Yu SS, Hwang JK, Jeon C, et al. Usual gait speed is associated with frailty status, institutionalization, and mortality in community-dwelling rural older adults: a longitudinal analysis of the Aging Study of Pyeongchang Rural Area. *Clinical Interventions in Aging*. 2018;13:1079-89.
39. Kang L, Han PP, Wang JZ, Ma YX, Jia LY, Fu LY, et al. Timed Up and Go Test can predict recurrent falls: a longitudinal study of the community-dwelling elderly in China. *Clinical Interventions in Aging*. 2017;12:2009-16.
40. Kauppi M, Stenholm S, Impivaara O, Maki J, Heliovaara M, Jula A. Fall-related risk factors and heel quantitative ultrasound in the assessment of hip fracture risk: a 10-year follow-up of a nationally representative adult population sample. *Osteoporosis International*. 2014;25:1685-95.
41. Kwan MM, Lin SI, Close JC, Lord SR. Depressive symptoms in addition to visual impairment, reduced strength and poor balance predict falls in older Taiwanese people. *Age Ageing*. 2012;41(5):606-12.
42. Laukkanen P, Leskinen E, Kauppinen M, Sakari-Rantala R, Heikkinen E. Health and functional capacity as predictors of community dwelling among elderly people. *Journal of Clinical Epidemiology*. 2000;53:257-65.
43. Lee WJ, Peng LN, Chiou ST, Chen LK. Physical Health Indicators Improve Prediction of Cardiovascular and All-cause Mortality among Middle-Aged and Older People: a National Population-based Study. *Scientific Reports*. 2017;7.
44. Looker AC. Dysmobility syndrome and mortality risk in US men and women age 50 years and older. *Osteoporosis International*. 2015;26:93-102.

45. Luukinen H, Koski K, Laippala P, Kivela SL. Predictors for recurrent falls among the home-dwelling elderly. *Scandinavian Journal of Primary Health Care*. 1995;13(4):294-9.
46. Maki BE, Holliday PJ, Topper AK. A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population. *Journals of Gerontology*. 1994;49(2):M72-M84.
47. Makizako H, Furuta T, Yoshida H, Shimada H, Satoh K, Akanuma T, et al. Usual Walking Speed Predicts Decline of Functional Capacity among Community-Dwelling Older Japanese Women: a 4-year Longitudinal Study. *Journal of Physical Therapy Science*. 2010;22(4):405-12.
48. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Lee S, Hotta R, et al. Cognitive Functioning and Walking Speed in Older Adults as Predictors of Limitations in Self-Reported Instrumental Activity of Daily Living: Prospective Findings from the Obu Study of Health Promotion for the Elderly. *International Journal of Environmental Research and Public Health*. 2015;12:3002-13.
49. Muir SW, Berg K, Chesworth BM, Klar N, Speechley M. Modifiable Risk Factors Identify People Who Transition from Non-fallers to Fallers in Community-Dwelling Older Adults: A Prospective Study. *Physiotherapy Canada*. 2010;62:358-67.
50. Mulasso A, Roppolo M, Gobbens RJ, Rabaglietti E. Mobility, balance and frailty in community-dwelling older adults: at is the best 1-year predictor of falls? *Geriatrics & Gerontology International*. 2017;17:1463-9.
51. Muraki S, Akune T, Oka H, Ishimoto Y, Nagata K, Yoshida M, et al. Physical performance, bone and joint diseases, and incidence of falls in Japanese men and women: a longitudinal cohort study. *Osteoporosis International*. 2013;24:459-66.
52. Nakamoto M, Otsuka R, Yuki A, Nishita Y, Tange C, Tomida M, et al. Higher gait speed and smaller sway area decrease the risk for decline in higher-level functional capacity among middle-aged and elderly women. *Archives of Gerontology and Geriatrics*. 2015;61:429-36.
53. Niiranen TJ, Enserro DM, Larson MG, Vasan RS. Multisystem Trajectories Over the Adult Life Course and Relations to Cardiovascular Disease and Death. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2019;74(11):1778-85.
54. Nitz JC, Stock L, Khan A. Health-related predictors of falls and fractures in women over 40. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2013;24:613-21.
55. Nofuji Y, Shinkai S, Taniguchi Y, Amano H, Nishi M, Murayama H, et al. Associations of Walking Speed, Grip Strength, and Standing Balance With Total and Cause-Specific Mortality in a General Population of Japanese Elders. *Journal of the American Medical Directors Association*. 2016;17.
56. Ojagbemi A, D'Este C, Verdes E, Chatterji S, Gureje O. Gait speed and cognitive decline over 2 years in the Ibadan study of aging. *Gait & Posture*. 2015;41(2):736-40.
57. Pajala S, Era P, Koskenvuo M, Kaprio J, Tormakangas T, Rantanen T. Force platform balance measures as predictors of indoor and outdoor falls in community-dwelling women aged 63-76 years. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2008;63(2):171-8.
58. Quach L, Galica AM, Jones RN, Procter-Gray E, Manor B, Hannan MT, et al. The Nonlinear Relationship Between Gait Speed and Falls: The Maintenance of Balance, Independent Living, Intellect, and Zest in the Elderly of Boston Study. *Journal of the American Geriatrics Society*. 2011;59:1069-73.

59. Rosso AL, Metti AL, Faulkner K, Brach JS, Studenski SA, Redfern M, et al. Associations of Usual Pace and Complex Task Gait Speeds With Incident Mobility Disability. *Journal of the American Geriatrics Society*. 2019;67(10):2072-6.
60. Sabia S, Dumurgier J, Tavernier B, Head J, Tzourio C, Elbaz A. Change in Fast Walking Speed Preceding Death: Results From a Prospective Longitudinal Cohort Study. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2014;69:354-62.
61. Sakurai R, Suzuki H, Ogawa S, Kawai H, Yoshida H, Hirano H, et al. Fear of falling, but not gait impairment, predicts subjective memory complaints in cognitively intact older adults. *Geriatrics & Gerontology International*. 2017;17:1125-31.
62. Sanders JB, Bremmer MA, Comijs HC, van de Ven PM, Deeg DJH, Beekman ATF. Gait Speed and Processing Speed as Clinical Markers for Geriatric Health Outcomes. *American Journal of Geriatric Psychiatry*. 2016;25:374-85.
63. Sanders JB, Bremmer MA, Deeg DJH, Beekman ATF. Do Depressive Symptoms and Gait Speed Impairment Predict Each Other's Incidence? A 16-Year Prospective Study in the Community. *Journal of the American Geriatrics Society*. 2012;60:1673-80.
64. Savva GM, Donoghue OA, Horgan F, O'Regan C, Cronin H, Kenny RA. Using Timed Up-and-Go to Identify Frail Members of the Older Population. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2013;68:441-6.
65. Schroll M, Avlund K, Davidsen M. Predictors of five-year functional ability in a longitudinal survey of men and women aged 75 to 80. The 1914-population in Glostrup, Denmark. *Aging-Clinical and Experimental Research*. 1997;9(1-2):143-52.
66. Stenhagen M, Ekstrom H, Nordell E, Elmstahl S. Falls in the general elderly population: a 3- and 6-year prospective study of risk factors using data from the longitudinal population study 'Good ageing in Skane'. *Bmc Geriatrics*. 2013;13.
67. Stijntjes M, Aartsen MJ, Taekema DG, Gussekloo J, Huisman M, Meskers CGM, et al. Temporal Relationship Between Cognitive and Physical Performance in Middle-Aged to Oldest Old People. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2017;72:662-8.
68. Swanenburg J, de Bruin ED, Uebelhart D, Mulder T. Falls prediction in elderly people: A 1-year prospective study. *Gait & Posture*. 2010;31(3):317-21.
69. Tian Q, Resnick SM, Studenski SA, Ferrucci L. Lap Time Variability From a 400-m Walk Is Associated With Future Mild Cognitive Impairment and Alzheimer's Disease. *Journal of the American Medical Directors Association*. 2019;20(12):1535-+.
70. Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. *Journal of the American Geriatrics Society*. 1997;45:735-8.
71. Veronese N, Solmi M, Maggi S, Noale M, Sergi G, Manzato E, et al. Frailty and incident depression in community-dwelling older people: results from the ELSA study. *International Journal of Geriatric Psychiatry*. 2017;32:E141-E9.
72. Veronese N, Stubbs B, Trevisan C, Bolzetta F, De Rui M, Solmi M, et al. Poor Physical Performance Predicts Future Onset of Depression in Elderly People: Progetto Veneto Anziani Longitudinal Study. *Physical Therapy*. 2017;97:659-68.
73. Ward RE, Beauchamp MK, Latham NK, Leveille SG, Percac-Lima S, Kurlinski L, et al. Neuromuscular Impairments Contributing to Persistently Poor and Declining Lower-Extremity Mobility Among Older Adults: New Findings Informing Geriatric Rehabilitation. *Archives of Physical Medicine and Rehabilitation*. 2016;97:1316-22.
74. Wihlborg A, Englund M, Akesson K, Gerdhem P. Fracture predictive ability of physical performance tests and history of falls in elderly women: a 10-year prospective study. *Osteoporosis International*. 2015;26(8):2101-9.

75. Zucchelli A, Vetrano DL, Grande G, Calderon-Larranaga A, Fratiglioni L, Marengoni A, et al. Comparing the prognostic value of geriatric health indicators: a population-based study. *Bmc Medicine*. 2019;17(1).
76. Osuka Y, Kim H, Watanabe Y, Taniguchi Y, Kojima N, Seino S, et al. A combined stepping and visual tracking task predicts cognitive decline in older adults better than gait or visual tracking tasks alone: a prospective study. *Aging Clinical and Experimental Research*. 2020;1-9.
77. . !!! INVALID CITATION !!! 8, 11, 12, 20, 22, 24, 27, 32, 33, 39-41, 45, 46, 49-51, 54, 57, 58, 62, 66, 68, 70, 74.
78. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society*. 1991;39(2):142-8.
79. Mayorga-Vega D, Aguilar-Soto P, Viciano J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med*. 2015;14(3):536-47.
80. Mayorga-Vega D, Bocanegra-Parrilla R, Ornelas M, Viciano J. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: a systematic review and Meta-analysis. *PLoS One*. 2016;11(3):e0151671-e.
81. Bennett H, Parfitt G, Davison K, Eston R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. *Sports Med*. 2016;46(5):737-50.
82. Mayorga-Vega D, Merino-Marban R, Viciano J. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: a meta-analysis. *J Sports Sci Med*. 2014;13(1):1-14.
83. Mayorga-Vega D, Viciano J, Cocca A, Merino-Marban R. Criterion-related validity of toe-touch test for estimating hamstring extensibility: A metaanalysis. *J Hum Sport Exerc*. 2014;9(1):188-200.
84. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *Int J Sports Med*. 1991;12(4):356-62.
85. Laukkanen R, Oja P, Pasanen M, Vuori I. Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults. *Int J Obes Relat Metab Disord*. 1992;16(4):263-8.
86. Laukkanen RMT, Oja P, Pasanen ME, Vuori IM. Criterion validity of a two-kilometer walking test for predicting the maximal oxygen uptake of moderately to highly active middle-aged adults. *Scand J Med Sci Sports*. 1993(3): 267-72.
87. Laukkanen RMT, Kukkonen-Harjula TK, Oja P, Pasanen ME, Vuori IM. Prediction of change in maximal aerobic power by the 2-km walk test after walking training in middle-aged adults. *Int J Sports Med*. 2000;21(2):113-6.
88. Larsen GE, George JD, Alexander JL, Fellingham GW, Aldana SG, Parcell AC. Prediction of maximum oxygen consumption from walking, jogging, or running. *Res Q Exerc Sport*. 2002;73(1):66-72.
89. McNaughton L, Hall P, Cooley D. Validation of several methods of estimating maximal oxygen uptake in young men. *Percept Mot Skills*. 1998;87(2):575-84.
90. Kline C, Porcari JP, Hintermeister R, Freedson PS, Ward A, McCarron RF, et al. Estimation of from a one-mile track walk, gender, age and body weight. *Med Sports Exerc*. 1987;19:253-9.
91. George JD, Fellingham GW, Fisher AG. A modified version of the Rockport Fitness Walking Test for college men and women. *Res Q Exerc Sport*. 1998;69(2):205-9.

92. Lunt H, Roiz De Sa D, Roiz De Sa J, Allsopp A. Validation of one-mile walk equations for the estimation of aerobic fitness in British military personnel under the age of 40 years. *Mil Med.* 2013;178(7):753-9.
93. Greenhalgh HA, George JD, Hager RL. Cross-validation of a quarter-mile walk test using two VO₂ max regression models. *Meas Phys Educ Exerc Sci.* 2001;5(3):139-51.
94. Dolgener FA, Hensley LD, Marsh JJ, Fjelstul JK. Validation of the Rockport Fitness Walking Test in college males and females. *Res Q Exerc Sport.* 1994;65(2):152-8.
95. Seneli RM, Ebersole KT, O'Connor KM, Snyder AC. Estimated VO₂max from the Rockport Walk Test on a Nonmotorized Curved Treadmill. *J Strength Cond Res.* 2013;27(12):3495-505.
96. George JD, Vehrs PR, Allsen PE, Fellingham GW, Fisher AG. Development of a submaximal treadmill jogging test for fit college-aged individuals. *Med Sci Sports Exerc.* 1993;25(5):643-7.
97. Cao Z-B, Miyatake N, Aoyama T, Higuchi M, Tabata I. Prediction of maximal oxygen uptake from a 3-minute walk based on gender, age, and body composition. *J Phys Act Health.* 2013;10(2):280-7.
98. Di Thommazo-Luporini L, Pinheiro Carvalho L, Luporini R, Trimer R, Falasco Pantoni CB, Catai AM, et al. The 6-minute step test as a predictor of cardiorespiratory fitness in obese women. *Eur J Phys Rehabil Med.* 2015;51(6):793-802.
99. Di Thommazo-Luporini L, Carvalho LP, Luporini RL, Trimer R, Falasco Pantoni CB, Martinez AF, et al. Are cardiovascular and metabolic responses to field walking tests interchangeable and obesity-dependent? *Disabil Rehabil.* 2016;38(18):1820-9.
100. Manttari A, Suni J, Sievanen H, Husu P, Vaha-Ypya H, Valkeinen H, et al. 6-minute walk test: a tool for predicting maximal aerobic power (VO₂ max) in healthy adults. *Clin Physiol Funct Imaging.* 2018.
101. Bonet JB, Magalhaes J, Viscor G, Pages T, Javierre CF, Torrella JR. A field tool for the aerobic power evaluation of middle-aged female recreational runners. *Women Health.* 2020.
102. Leger L, Boucher R. An indirect continuous running multistage field test: the Universite de Montreal track test. *Can J Appl Sport Sci.* 1980;5:77-84.
103. Gabriel KKP, Rankin RL, Lee C, Charlton ME, Swan PD, Ainsworth BE, et al. Test-retest reliability and validity of the 400-meter walk test in healthy, middle-aged women. *J Phys Act Health.* 2010;7(5):649-57.
104. Burr JF, Bredin SS, Faktor MD, Warburton DE. The 6-minute walk test as a predictor of objectively measured aerobic fitness in healthy working-aged adults. *Phys Sportsmed.* 2011;39(2):133-9.
105. Mikawa K, Senjyu H. Development of a field test for evaluating aerobic fitness in middle-aged adults: Validity of a 15-m Incremental Shuttle Walk and Run Test. *J Sports Sci Med.* 2011;10(4):712-7.
106. Azmi SH, Sulaiman N. Validity of selected cardiovascular field-based test among Malaysian healthy female adult. *J Fundam Appl Sci.* 2017;9:1227-35.
107. McGawley K. The Reliability and Validity of a Four-Minute Running Time-Trial in Assessing VO₂max and Performance. *Front Physiol.* 2017;8:270.
108. Cooper SM, Baker JS, Tong RJ, Roberts E, Hanford M. The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *Br J Sports Med.* 2005;39(4):e19.

109. Flouris AD, Koutedakis Y, Nevill A, Metsios GS, Tsiotra G, Parasiris Y. Enhancing specificity in proxy-design for the assessment of bioenergetics. *J Sci Med Sport*. 2004;7(2):197-204.
110. Flouris AD, Metsios GS, Koutedakis Y. Enhancing the efficacy of the 20 m multistage shuttle run test. *Br J Sports Med*. 2005;39(3):166-70.
111. Metsios GS, Flouris AD, Koutedakis Y, Nevill A. Criterion-related validity and test-retest reliability of the 20m square shuttle test. *J Sci Med Sport*. 2008;11(2):214-7.
112. Kim J, Jung SH, Cho HC. Validity and Reliability of Shuttle-Run Test in Korean Adults. *Int J Sports Med*. 2011;32(8):580-5.
113. Jurio-Iriarte B, Gorostegi-Anduaga I, Rodrigo Aispuru G, Perez-Asenjo J, Brubaker PH, Maldonado-Martin S. Association between Modified Shuttle Walk Test and cardiorespiratory fitness in overweight/obese adults with primary hypertension: EXERDIET-HTA study. *J Am Soc Hypertens*. 2017;11(4):186-95.
114. Jurio-Iriarte B, Brubaker PH, Gorostegi-Anduaga I, Corres P, Martinez Aguirre-Betolaza A, Maldonado-Martin S. Validity of the modified shuttle walk test to assess cardiorespiratory fitness after exercise intervention in overweight/obese adults with primary hypertension. *Clin Exp Hypertens*. 2018:1-6.
115. Lima LP, Leite HR, Matos MA, Neves CDC, Lage V, Silva GPD, et al. Cardiorespiratory fitness assessment and prediction of peak oxygen consumption by Incremental Shuttle Walking Test in healthy women. *PLoS One*. 2019;14(2):e0211327.
116. Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *Journal of sports sciences*. 1988;6(2):93-101.
117. Aadahl M, Zacho M, Linneberg A, Thuesen BH, Jorgensen T. Comparison of the Danish step test and the watt-max test for estimation of maximal oxygen uptake: the Health2008 study. *Eur J Prev Cardiol*. 2013;20(6):1088-94.
118. Kumar SK, Khare P, Jaryal AK, Talwar A. Validity of heart rate based nomogram for estimation of maximum oxygen uptake in Indian population. *Indian J Physiol Pharmacol*. 2012;56(3):279-83.
119. Ricci PA, Cabiddu R, Jørgensen SP, André LD, Oliveira CR, Di Thommazo-Luporini L, et al. Validation of the two-minute step test in obese with comorbidities and morbidly obese patients. *Braz J Med Biol Res*. 2019;52(9):e8402.
120. Weller IM, Thomas SG, Cox MH, Corey PN. A study to validate the Canadian Aerobic Fitness Test. *Can J Public Health*. 1992;83(2):120-4.
121. Carvalho LP, Di Thommazo-Luporini L, Aubertin-Leheudre M, Bonjorno Junior JC, de Oliveira CR, Luporini RL, et al. Prediction of cardiorespiratory fitness by the 6-minute step test and its association with muscle strength and power in sedentary obese and lean young women: a cross-sectional study. *PLoS One*. 2015;10(12):e0145960.
122. Teren A, Zachariae S, Beutner F, Ubrich R, Sandri M, Engel C, et al. Incremental value of veterans specific activity questionnaire and the ymca-step test for the assessment of cardiorespiratory fitness in population-based studies. *Eur J Prev Cardiol*. 2016;23(11):1221-7.
123. Beutner F, Ubrich R, Zachariae S, Engel C, Sandri M, Teren A, et al. Validation of a brief step-test protocol for estimation of peak oxygen uptake. *Eur J Prev Cardiol*. 2015;22(4):503-12.
124. Lee O, Lee S, Kang M, Mun J, Chung J. Prediction of maximal oxygen consumption using the Young Men's Christian Association-step test in Korean adults. *Eur J Appl Physiol*. 2019;119(5):1245-52.
125. Hong SH, Yang HI, Kim DI, Gonzales TI, Brage S, Jeon JY. Validation of submaximal step tests and the 6-min walk test for predicting maximal oxygen

consumption in young and healthy participants. *Int J Environ Res Public Health*. 2019;16(23).

126. Kieu NTV, Jung SJ, Shin SW, Jung HW, Jung ES, Won YH, et al. The validity of the YMCA 3-minute step test for estimating maximal oxygen uptake in healthy Korean and Vietnamese adults. *J Lifestyle Med*. 2020;10(1):21-9.

127. Hansen D, Jacobs N, Thijs H, Dendale P, Claes N. Validation of a single-stage fixed-rate step test for the prediction of maximal oxygen uptake in healthy adults. *Clin Physiol Funct Imaging*. 2016;36(5):401-6.

128. Giacomantonio N, Morrison P, Rasmussen R, MacKay-Lyons MJ. Reliability and validity of the 6-minutestep test for clinical assessment of cardiorespiratory fitness in people at risk of cardiovascular disease. *J Strength Cond Res*. 2020;34(5):1376-82.

129. Arcuri JF, Borghi-Silva A, Labadessa IG, Sentanin AC, Candolo C, Pires Di Lorenzo VA. Validity and reliability of the 6-minute step test in healthy individuals: a cross-sectional study. *Clin J Sport Med*. 2016;26(1):69-75.

130. Siconolfi SF, Garber CE, Lasater TM, Carleton RA. A simple, valid step test for estimating maximal oxygen uptake in epidemiologic studies. *Am J Epidemiol*. 1985;121(3):382-90.

131. Ritchie C, Trost SG, Brown W, Armit C. Reliability and validity of physical fitness field tests for adults aged 55 to 70 years. *J Sci Med Sport*. 2005;8(1):61-70.

132. Santa Maria D, Kinnear GR, Kearney JT, Martin TP. The objectivity, reliability, and validity of the OSU step test for college males. *Res Q*. 1976;47(3):445-52.

133. Hansen D, Jacobs N, Bex S, D'Haene G, Dendale P, Claes N. Are fixed-rate step tests medically safe for assessing physical fitness? *European Journal of Applied Physiology*. 2011;111(10):2593-9.

134. Lerche L, Olsen A, Petersen KEN, Rostgaard-Hansen AL, Dragsted LO, Nordsborg NB, et al. Validity of physical activity and cardiorespiratory fitness in the Danish cohort "Diet, Cancer and Health-Next Generations". *Scandinavian Journal of Medicine & Science in Sports*. 2017;27(12):1864-72.

135. Espana-Romero V, Artero EG, Santaliestra-Pasias AM, Gutierrez A, Castillo MJ, Ruiz JR. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *J Hand Surg [Am]*. 2008;33(3):378-84.

136. Cadenas-Sanchez C, Sanchez-Delgado G, Martinez-Tellez B, Mora-Gonzalez J, Löf M, España-Romero V, et al. Reliability and validity of different models of TKK hand dynamometers. *Am J Occup Ther*. 2016;70(4):7004300010.

137. Kolimechkov S, Castro-Piñero J, Petrov A, Alexandrova A. The effect of elbow position on the handgrip strength test in children: validity and reliability of TKK 5101 and DynX dynamometers. *Pedagogy Phys Cult Sports*. 2020;24(5):240-7.

138. Shechtman O, Gestewitz L, Kimble C. Reliability and validity of the DynEx dynamometer. *J Hand Ther*. 2005;18(3):339-47.

139. Mathiowetz V. Comparison of Rolyan and Jamar dynamometers for measuring grip strength. *Occup Ther Int*. 2002;9(3):201-9.

140. Härkönen R, Harju R, Alaranta H. Accuracy of the Jamar dynamometer. *J Hand Ther*. 1993;6(4):259-62.

141. Clemons JM. Construct validity of a modification of the flexed arm hang test. *Journal of strength and conditioning research*. 2014;28(12):3523-30.

142. Mannion AF, Dolan P. Electromyographic median frequency changes during isometric contraction of the back extensors to fatigue. *Spine (Phila Pa 1976)*. 1994;19(11):1223-9.

143. Coorevits P, Danneels L, Cambier D, Ramon H, Vanderstraeten G. Assessment of the validity of the Biering-Sørensen test for measuring back muscle fatigue based on

EMG median frequency characteristics of back and hip muscles. *J Electromyogr Kinesiol.* 2008;18(6):997-1005.

144. Kankaanpää M, Laaksonen D, Taimela S, Kokko SM, Airaksinen O, Hänninen O. Age, sex, and body mass index as determinants of back and hip extensor fatigue in the isometric Sørensen back endurance test. *Arch Phys Med Rehabil.* 1998;79(9):1069-75.

145. De Blaiser C, De Ridder R, Willems T, Danneels L, Roosen P. Reliability and validity of trunk flexor and trunk extensor strength measurements using handheld dynamometry in a healthy athletic population. *Phys Ther Sport.* 2018;34:180-6.

146. Baumgartner TA, Gaunt SJ. Construct related validity for the Baumgartner Modified pull-up test. *Meas Phys Educ Exerc Sci.* 2005;9(1):51-60.

147. Ten Hoor GA, Musch K, Meijer K, Plasqui G. Test-retest reproducibility and validity of the back-leg-chest strength measurements. *Isokinet Exerc Sci.* 2016;24(3):209-16.

148. Clemons J. Construct Validity of Two Different Methods of Scoring and Performing Push-ups. *J Strength Cond Res.* 2019;33(11):2971-80.

149. Diener MH, Golding LA, Diener D. Validity and reliability of a one-minute half sit-up test of abdominal strength and endurance. *Res Sports Med.* 1995;6(2):105-19.

150. Knudson D. The validity of recent curl-up tests in young adults. *J Strength Cond Res.* 2001;15(1):81-5.

151. Hall GL, Hetzler RK, Perrin D, Weltman A. Relationship of timed sit-up tests to isokinetic abdominal strength. *Res Q Exerc Sport.* 1992;63(1):80-4.

152. Mannion AF, Connolly B, Wood K, Dolan P. The use of surface EMG power spectral analysis in the evaluation of back muscle function. *J Rehabil Res Dev.* 1997;34(4):427-39.

153. Sparto PJ, Parnianpour M, Reinsel TE, Simon S. Spectral and temporal responses of trunk extensor electromyography to an isometric endurance test. *Spine (Phila Pa 1976).* 1997;22(4):418-25; discussion 25-26.

154. Applegate ME, France CR, Russ DW, Leitkam ST, Thomas JS. Sorensen test performance is driven by different physiological and psychological variables in participants with and without recurrent low back pain. *J Electromyogr Kinesiol.* 2019;44:1-7.

155. Bohannon RW, Bubela DJ, Magasi SR, Wang Y-C, Gershon RC. Sit-to-stand test: Performance and determinants across the age-span. *Isokinet Exerc Sci.* 2010;18(4):235-40.

156. Silva P, Franco J, Gusmao A, Moura J, Teixeira-Salmela L, Faria C. Trunk strength is associated with sit-to-stand performance in both stroke and healthy subjects. *Eur J Phys Rehabil Med.* 2015;51(6):717-24.

157. Taylor JD, Fletcher JP. Correlation between the 8-repetition maximum test and isokinetic dynamometry in the measurement of muscle strength of the knee extensors: A concurrent validity study. *Physiother Theory Pract.* 2013;29(4):335-41.

158. Clemons JM, Duncan CA, Blanchard OE, Gatch WH, Hollander DB, Doucet JL. Relationships between the flexed-arm hang and select measures of muscular fitness. *Journal of strength and conditioning research.* 2004;18(3):630-6.

159. Wood HM, Baumgartner TA. Objectivity, reliability, and validity of the bent-knee push-up for college-age women. *Meas Phys Educ Exerc Sci.* 2004;8(4):203-12.

160. Harvey VP, Scott GD. An investigation of the curl-down test as a measure of abdominal strength. *Res Q.* 1967;38(1):22-7.

161. Berger R. Evaluation of the 2-minute sit-up test as a measure of muscular endurance and strength. *Journal of the Association for Physical and Mental Rehabilitation.* 1966;20(4):140.

162. Knudson D, Johnston D. Validity and reliability of a bench trunk-curl test of abdominal endurance. *J Strength Cond Res*. 1995;9(3):165-9.
163. DeWitt R. A study of the sit-up Type of test as a means of measuring strength and endurance of the abdominal muscles. *Res Q*. 1944;15(1):60-3.
164. Bui HT, Farinas M-I, Fortin A-M, Comtois A-S, Leone M. Comparison and analysis of three different methods to evaluate vertical jump height. *Clin Physiol Funct Imaging*. 2015;35(3):203-9.
165. Clemons JM, Campbell B, Jeansonne C. Validity and reliability of a new test of upper body power. *J Strength Cond Res*. 2010;24(6):1559-65.
166. Kawano MM, Ambar G, Oliveira BIR, Boer MC, Cardoso APRG, Cardoso JR. Influence of the gastrocnemius muscle on the sit-and-reach test assessed by angular kinematic analysis. *Braz J Phys Ther*. 2010;14(1):10-5.
167. Jackson A, Langford NJ. The criterion-related validity of the sit and reach test: replication and extension of previous findings. *Res Q Exerc Sport*. 1989;60(4):384-7.
168. Youdas JW, Krause DA, Hollman JH. Validity of hamstring muscle length assessment during the sit-and-reach test using an inclinometer to measure hip joint angle. *J Strength Cond Res*. 2008;22(1):303-9.
169. Mier CM, Shapiro BS. Sex differences in pelvic and hip flexibility in men and women matched for sit-and-reach score. *J Strength Cond Res*. 2013;27(4):1031-5.
170. López-Miñarro P, Vaquero-Cristóbal R, Muyor JM, Espejo-Antúnez L. Criterion-related validity of sit-and-reach test as measure of hamstring extensibility in older women. *Nutr Hosp*. 2015;32(1):312-7.
171. Liemohn W, Sharpe GL, Wasserman JF. Criterion related validity of the sit-and-reach test. *J Strength Cond Res*. 1994;8(2):91-4.
172. Cotten DJ. A comparison of selected trunk flexibility tests. *Am Correct Ther J*. 1972;26(1):24-6.
173. Broer MR, Galles NRG. Importance of relationship between various body measurements in performance of the toe-touch test. *Res Q*. 1958;29(3):253-63.
174. Kippers V, Parker AW. Toe-touch test: A measure of its validity. *Physical Therapy*. 1987;67(11):1680-4.
175. Ayala F, Sainz de Baranda P, De Ste Croix M, Santonja F. Reproducibility and concurrent validity of hip joint angle test for estimating hamstring flexibility in recreationally active young men. *J Strength Cond Res*. 2012;26(9):2372-82.
176. Wells KF, Dillon EK. The sit and reach—a test of back and leg flexibility. *Res Q Exerc Sport*. 1952;23(1):115-8.
177. Mathews DK, Shaw V, Bohnen M. Hip flexibility of college women as related to length of body segments. *Res Q*. 1957;28(4):352-6.
178. Miyamoto K, Takebayashi H, Takimoto K, Miyamoto S, Inoue Y, Takuma Y, et al. The criterion-related validity of the ten step test compared with motor reaction time. *J Phys Ther Sci*. 2008;20(4):261-5.
179. Miyamoto K, Takebayashi H, Takimoto K, Miyamoto S, Morioka S, Yagi F. A new simple performance test focused on agility in elderly people: The Ten Step Test. *Gerontology*. 2008;54(6):365-72.
180. Sun G, Cahill F, Gulliver W, Yi Y, Xie Y, Bridger T, et al. Concordance of BAI and BMI with DXA in the Newfoundland population. *Obesity (Silver Spring)*. 2013;21(3):499-503.
181. Hodgdon JA, Beckett MB. Prediction of percent body fat for US Navy women from body circumferences and height. NAVAL HEALTH RESEARCH CENTER SAN DIEGO CA, 1984.

182. Sloan A, Burt J, Blyth C. Estimation of body fat in young women. *Journal of applied physiology*. 1962;17(6):967-70.
183. Durnin J, Rahaman MM. The assessment of the amount of fat in the human body from measurements of skinfold thickness. *British journal of Nutrition*. 1967;21(3):681-9.
184. Erselcan T, Candan F, Saruhan S, Ayca T. Comparison of body composition analysis methods in clinical routine. *Annals of Nutrition and Metabolism*. 2000;44(5-6):243-8.
185. Eston RG, Rowlands AV, Charlesworth S, Davies A, Hoppitt T. Prediction of DXA-determined whole body fat from skinfolds: importance of including skinfolds from the thigh and calf in young, healthy men and women. *European Journal of Clinical Nutrition*. 2005;59(5):695-702.
186. Friedl KE, Vogel JA. Validity of percent body fat predicted from circumferences: classification of men for weight control regulations. *Mil Med*. 1997;162(3):194-200.
187. Haisman M. The assessment of body fat content in young men from measurements of body density and skinfold thickness. *Human biology*. 1970;679-88.
188. Katch FI, McArdle WD. Prediction of body density from simple anthropometric measurements in college-age men and women. *Human biology*. 1973;445-55.
189. Katch FI, McArdle WD. VALIDITY OF BODY COMPOSITION PREDICTION EQUATIONS FOR COLLEGE MEN AND WOMEN. *American Journal of Clinical Nutrition*. 1975;28(2):105-9.
190. Temple D, Denis R, Walsh MC, Dicker P, Byrne AT. Comparison of anthropometric-based equations for estimation of body fat percentage in a normal-weight and overweight female cohort: validation via air-displacement plethysmography. *Public Health Nutr*. 2015;18(3):446-52.
191. Wilmore JH, Behnke AR. An anthropometric estimation of body density and lean body weight in young men. *Journal of Applied Physiology*. 1969;27(1):25-31.
192. Wilmore JH, Behnke AR. An anthropometric estimation of body density and lean body weight in young women. *The American journal of clinical nutrition*. 1970;23(3):267-74.
193. Morabia A, Ross A, Curtin F, Pichard C, Slosman DO. Relation of BMI to a dual-energy X-ray absorptiometry measure of fatness. *Br J Nutr*. 1999;82(1):49-55.
194. Sun Q, van Dam RM, Spiegelman D, Heymsfield SB, Willett WC, Hu FB. Comparison of dual-energy x-ray absorptiometric and anthropometric measures of adiposity in relation to adiposity-related biologic factors. *Am J Epidemiol*. 2010;172(12):1442-54.
195. Chang H, Simonsick EM, Ferrucci L, Cooper JA. Validation study of the body adiposity index as a predictor of percent body fat in older individuals: findings from the BLSA. *J Gerontol A Biol Sci Med Sci*. 2014;69(9):1069-75.
196. McLean KP, Skinner JS. Validity of Futrex-5000 for body composition determination. *Med Sci Sports Exerc*. 1992;24(2):253-8.
197. Aandstad A, Holtberget K, Hageberg R, Holme I, Anderssen SA. Validity and reliability of bioelectrical impedance analysis and skinfold thickness in predicting body fat in military personnel. *Mil Med*. 2014;179(2):208-17.
198. Ball S, Swan PD, DeSimone R. Comparison of anthropometry to dual energy X-ray absorptiometry: a new prediction equation for women. *Res Q Exerc Sport*. 2004;75(3):248-58.
199. Cui Z, Truesdale KP, Cai J, Stevens J. Evaluation of anthropometric equations to assess body fat in adults: NHANES 1999-2004. *Med Sci Sports Exerc*. 2014;46(6):1147-58.

200. Davidson LE, Wang J, Thornton JC, Kaleem Z, Silva-Palacios F, Pierson RN, et al. Predicting Fat Percent by Skinfolds in Racial Groups: Durnin and Womersley Revisited. *Medicine and Science in Sports and Exercise*. 2011;43(3):542-9.
201. Friedl KE, Westphal KA, Marchitelli LJ, Patton JF, Chumlea WC, Guo SS. Evaluation of anthropometric equations to assess body-composition changes in young women. *American Journal of Clinical Nutrition*. 2001;73(2):268-75.
202. Lee DH, Keum N, Hu FB, Orav EJ, Rimm EB, Sun Q, et al. Development and validation of anthropometric prediction equations for lean body mass, fat mass and percent fat in adults using the National Health and Nutrition Examination Survey (NHANES) 1999-2006. *Br J Nutr*. 2017;118(10):858-66.
203. Tucker LA, Greenwell SD. Using simple measures to estimate body fat percentage in college men. *American Journal of Health Behavior*. 2001;25(5):460-7.
204. Al-Gindan YY, Hankey CR, Govan L, Gallagher D, Heymsfield SB, Lean MEJ. Derivation and validation of simple anthropometric equations to predict adipose tissue mass and total fat mass with MRI as the reference method. *British Journal of Nutrition*. 2015;114(11):1852-67.
205. Fedewa MV, Nickerson BS, Esco MR. Associations of body adiposity index, waist circumference, and body mass index in young adults. *Clinical Nutrition*. 2019;38(2):715-20.
206. Garcia AL, Wagner K, Hothorn T, Koebnick C, Zunft HJ, Trippo U. Improved prediction of body fat by measuring skinfold thickness, circumferences, and bone breadths. *Obes Res*. 2005;13(3):626-34.
207. Jackson AS, Ellis KJ, McFarlin BK, Sailors MH, Bray MS. Cross-validation of generalised body composition equations with diverse young men and women: the Training Intervention and Genetics of Exercise Response (TIGER) Study. *Br J Nutr*. 2009;101(6):871-8.
208. Kanellakis S, Kourlaba G, Moschonis G, Vondrou A, Manios Y. Development and validation of two equations estimating body composition for overweight and obese postmenopausal women. *Maturitas*. 2010;65(1):64-8.
209. Kanellakis S, Manios Y. Validation of five simple models estimating body fat in white postmenopausal women: use in clinical practice and research. *Obesity (Silver Spring)*. 2012;20(6):1329-32.
210. Kanellakis S, Skoufas E, Khudokonenko V, Apostolidou E, Gerakiti L, Andrioti M-C, et al. Development and Validation of Two Equations Based on Anthropometry, Estimating Body Fat for the Greek Adult Population. *Obesity*. 2017;25(2):408-16.
211. Leahy S, O'Neill C, Sohun R, Toomey C, Jakeman P. Generalised equations for the prediction of percentage body fat by anthropometry in adult men and women aged 18-81 years. *Br J Nutr*. 2013;109(4):678-85.
212. Manios Y, Kanellakis S, Androutsos O, Maragkopoulou K, Giannopoulou A, Argyri E, et al. Development and validation of a simple model based on anthropometry: estimating fat mass for white postmenopausal women. *Menopause*. 2012;19(4):467-70.
213. Nickerson BS, Esco MR, Bishop PA, Fedewa MV, Snarr RL, Kliszczewicz BM, et al. Validity of BMI-Based Body Fat Equations in Men and Women: A 4-Compartment Model Comparison. *J Strength Cond Res*. 2016;32(1):121-9.
214. O'Connor DP, Bray MS, McFarlin BK, Sailors MH, Ellis KJ, Jackson AS. Generalized Equations for Estimating DXA Percent Fat of Diverse Young Women and Men: The TIGER Study. *Medicine and Science in Sports and Exercise*. 2010;42(10):1959-65.

215. Skoufas E, Kanellakis S, Apostolidou E, Makridi T, Pigiou E, Papassotiriou I, et al. Development and validation of two anthropometric models estimating abdominal fat percentage in Greek adult women and men. *Clin Nutr ESPEN*. 2018;28:239-42.
216. Stout JR, Eckerson JM, Housh TJ, Johnson GO, Betts NM. Validity of percent body fat estimations in males. *Med Sci Sports Exerc*. 1994;26(5):632-6.
217. Ketel IJ, Volman MN, Seidell JC, Stehouwer CD, Twisk JW, Lambalk CB. Superiority of skinfold measurements and waist over waist-to-hip ratio for determination of body fat distribution in a population-based cohort of Caucasian Dutch adults. *Eur J Endocrinol*. 2007;156(6):655-61.
218. Gallagher D, Visser M, Sepulveda D, Pierson RN, Harris T, Heymsfield SB. How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups? *American journal of epidemiology*. 1996;143(3):228-39.
219. Gallagher D, Heymsfield SB, Heo M, Jebb SA, Murgatroyd PR, Sakamoto Y. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *The American journal of clinical nutrition*. 2000;72(3):694-701.
220. Gómez-Ambrosi J, Silva C, Catalán V, Rodríguez A, Galofré JC, Escalada J, et al. Clinical usefulness of a new equation for estimating body fat. *Diabetes Care*. 2012;35(2):383-8.
221. Steinkamp R, Cohen N, Gaffey W, McKey T, Bron G, Siri W, et al. Measures of body fat and related factors in normal adults—II: A simple clinical method to estimate body fat and lean body mass. *Journal of chronic diseases*. 1965;18(12):1291-307.
222. Pasco JA, Nicholson GC, Brennan SL, Kotowicz MA. Prevalence of obesity and the relationship between the body mass index and body fat: cross-sectional, population-based data. *PloS one*. 2012;7(1):e29580.
223. Vogel JA, Kirkpatrick J, Fitzgerald P, Hodgdon JA, Harman E. Derivation of anthropometry based body fat equations for the Army's weight control program. ARMY RESEARCH INST OF ENVIRONMENTAL MEDICINE NATICK MA, 1988.
224. Wang J, Thornton JC, Russell M, Burastero S, Heymsfield S, Pierson Jr RN. Asians have lower body mass index (BMI) but higher percent body fat than do whites: comparisons of anthropometric measurements. *The American journal of clinical nutrition*. 1994;60(1):23-8.
225. Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao D, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *International journal of obesity*. 2002;26(6):789-96.
226. Lean M, Han TS, Deurenberg P. Predicting body composition by densitometry from simple anthropometric measurements. *The American journal of clinical nutrition*. 1996;63(1):4-14.
227. Pascale LR, Grossman MI, Sloane HS, Frankel T. Correlations between thickness of skinfolds and body density in 88 soldiers. *Human Biology*. 1956;28(2):165.
228. Rush EC, Plank LD, Lualaba MS, Robinson SM. Prediction of percentage body fat from anthropometric measurements: comparison of New Zealand European and Polynesian young women. *The American journal of clinical nutrition*. 1997;66(1):2-7.
229. Slaughter MH, Lohman TG, Boileau R, Horswill C, Stillman R, Van Loan M, et al. Skinfold equations for estimation of body fatness in children and youth. *Human biology*. 1988:709-23.
230. Zanovec M, Johnson L, Marx B, Keenan M, Tuuri G. Self-reported physical activity improves prediction of body fatness in young adults. *Medicine+ Science in Sports+ Exercise*. 2009;41(2):328.

231. Dioum A, Gartner A, Maire B, Delpuech F, Wade S. Body composition predicted from skinfolds in African women: a cross-validation study using air-displacement plethysmography and a black-specific equation. *Br J Nutr*. 2005;93(6):973-9.
232. Bergman RN, Stefanovski D, Buchanan TA, Sumner AE, Reynolds JC, Sebring NG, et al. A better index of body adiposity. *Obesity (Silver Spring)*. 2011;19(5):1083-9.
233. Wang J, Deurenberg P. The validity of predicted body composition in Chinese adults from anthropometry and bioelectrical impedance in comparison with densitometry. *Br J Nutr*. 1996;76(2):175-82.
234. Zhang Z-Q, Liu Y-H, Xu Y, Dai X-W, Ling W-h, Su Y-X, et al. The validity of the body adiposity index in predicting percentage body fat and cardiovascular risk factors among Chinese. *Clinical Endocrinology*. 2014;81(3):356-62.
235. Demura S, Yamaji S, Goshi F, Kobayashi H, Sato S, Nagasawa Y. The validity and reliability of relative body fat estimates and the construction of new prediction equations for young Japanese adult males. *J Sports Sci*. 2002;20(2):153-64.
236. Pongchaiyakul C, Kosulwat V, Rojroongwasinkul N, Charoenkiatkul S, Thepsuthammarat K, Laopaiboon M, et al. Prediction of percentage body fat in rural Thai population using simple anthropometric measurements. *Obes Res*. 2005;13(4):729-38.
237. Eston RG, Fu F, Fung L. VALIDITY OF CONVENTIONAL ANTHROPOMETRIC TECHNIQUES FOR PREDICTING BODY-COMPOSITION IN HEALTHY CHINESE ADULTS. *British Journal of Sports Medicine*. 1995;29(1):52-6.
238. Kohli S, Gao M, Lear SA. Using simple anthropometric measures to predict body fat in South Asians. *Appl Physiol Nutr Metab*. 2009;34(1):40-8.
239. Lam BC, Lim SC, Wong MT, Shum E, Ho CY, Bosco JJ, et al. A method comparison study to validate a novel parameter of obesity, the body adiposity index, in Chinese subjects. *Obesity (Silver Spring)*. 2013;21(12):E634-9.
240. Lee G, Chang J, Hwang SS, Son JS, Park SM. Development and validation of prediction equations for the assessment of muscle or fat mass using anthropometric measurements, serum creatinine level, and lifestyle factors among Korean adults. *Nutrition Research and Practice*. 2021;15(1):95-105.
241. Yao M, Roberts SB, Ma G, Pan H, McCrory MA. Field methods for body composition assessment are valid in healthy Chinese adults. *J Nutr*. 2002;132(2):310-7.
242. Balas-Nakash M, Legorreta-Legorreta J, Rodriguez-Cano A, Aguilera-Perez R, Perichart-Perera O. Validation of body composition estimation equations by bioelectric impedance in postmenopausal women with metabolic syndrome. *Rev Invest Clin*. 2010;62(6):538-45.
243. Macias N, Alemán-Mateo H, Esparza-Romero J, Valencia ME. Body fat measurement by bioelectrical impedance and air displacement plethysmography: a cross-validation study to design bioelectrical impedance equations in Mexican adults. *Nutrition Journal*. 2007;6(1):1-7.
244. Bhat DS, Yajnik CS, Sayyad MG, Raut KN, Lubree HG, Rege SS, et al. Body fat measurement in Indian men: comparison of three methods based on a two-compartment model. *Int J Obes (Lond)*. 2005;29(7):842-8.
245. Goel K, Gupta N, Misra A, Poddar P, Pandey RM, Vikram NK, et al. Predictive equations for body fat and abdominal fat with DXA and MRI as reference in Asian Indians. *Obesity (Silver Spring)*. 2008;16(2):451-6.
246. Hicks VL, Stolarczyk LM, Heyward VH, Baumgartner RN. Validation of near-infrared interactance and skinfold methods for estimating body composition of American Indian women. *Med Sci Sports Exerc*. 2000;32(2):531-9.

247. Cerqueira M, Amorim P, Magalhaes F, Castro E, Franco F, Franceschini S, et al. Validity of body adiposity index in predicting body fat in a sample of brazilian women. *Obesity*. 2013;21(12):E696-E9.
248. Segheto W, Coelho FA, Guimaraes da Silva DC, Hallal PC, Bouzas Marins JC, Ribeiro AQ, et al. Validity of body adiposity index in predicting body fat in Brazilians adults. *American Journal of Human Biology*. 2017;29(1).
249. Aristizabal JC, Restrepo MT, Amalia L. Validation by hydrodensitometry of skinfold thickness equations used for female body composition assessment. *Biomedica*. 2008;28(3):404-13.
250. Aristizabal JC, Estrada-Restrepo A, García AG. Desarrollo y validación de ecuaciones antropométricas para estimar la composición corporal en mujeres adultas. *Revista Colombia Médica*. 2018;49(2):154-9.
251. Wattanapenpaiboon N, Lukito W, Strauss BJG, Hsu-Hage BH, Wahlqvist ML, Stroud DB. Agreement of skinfold measurement and bioelectrical impedance analysis (BIA) methods with dual energy X-ray absorptiometry (DEXA) in estimating total body fat in Anglo-Celtic Australians. *International Journal of Obesity*. 1998;22(9):854-60.
252. Lanham DA, Stead MA, Tsang K, Davies PSW. The prediction of body composition in Chinese Australian females. *International Journal of Obesity*. 2001;25(2):286-91.
253. Kagawa M, Byrne NM, Hills AP. Comparison of body fat estimation using waist: height ratio using different 'waist' measurements in Australian adults. *British Journal of Nutrition*. 2008;100(5):1135-41.
254. Johnson W, Chumlea WC, Czerwinski SA, Demerath EW. Concordance of the recently published body adiposity index with measured body fat percent in European-American adults. *Obesity (Silver Spring)*. 2012;20(4):900-3.
255. Al-Bachir M, Ahmad H. VALIDITY OF USING WAIST AND HIP CIRCUMFERENCE MEASUREMENTS TO DETERMINE BODY COMPOSITION OF YOUNG SYRIAN MEN. *J Biosoc Sci*. 2016;48(5):647-57.
256. Ramirez-Zea M, Torun B, Martorell R, Stein AD. Anthropometric predictors of body fat as measured by hydrostatic weighing in Guatemalan adults. *The American journal of clinical nutrition*. 2006;83(4):795-802.
257. Gomez Campos R, Pacheco Carrillo J, Almonacid Fierro A, Urra Albornoz C, Cossio-Bolanos M. Validation of equations and proposed reference values to estimate fat mass in Chilean university students. *Endocrinol Diabetes Nutr*. 2018;65(3):156-63.
258. Sun G, French CR, Martin GR, Younghusband B, Green RC, Xie Y-g, et al. Comparison of multifrequency bioelectrical impedance analysis with dual-energy X-ray absorptiometry for assessment of percentage body fat in a large, healthy population. *The American journal of clinical nutrition*. 2005;81(1):74-8.
259. Simoes M, Severo M, Oliveira A, Ferreira I, Lopes C. Predictive equations for estimating regional body composition: a validation study using DXA as criterion and associations with cardiometabolic risk factors. *Ann Hum Biol*. 2016;43(3):219-28.
260. Shaw KA, Srikanth VK, Fryer JL, Blizzard L, Dwyer T, Venn AJ. Dual energy X-ray absorptiometry body composition and aging in a population-based older cohort. *Int J Obes (Lond)*. 2007;31(2):279-84.
261. Noppa H, Andersson M, Bengtsson C, Bruce Å, Isaksson B. Body composition in middle-aged women with special reference to the correlation between body fat mass and anthropometric data. *The American journal of clinical nutrition*. 1979;32(7):1388-95.
262. Flanagin A, Frey T, Christiansen SL, Bauchner H. The reporting of race and ethnicity in medical and science journals: comments invited. *JAMA*. 2021;325(11):1049-52.

263. Lahav Y, Epstein Y, Kedem R, Schermann H. A novel body circumferences-based estimation of percentage body fat. *Br J Nutr.* 2018;119(6):720-5.
264. Thomas EL, Saeed N, Hajnal JV, Brynes A, Goldstone AP, Frost G, et al. Magnetic resonance imaging of total body fat. *Journal of Applied Physiology.* 1998;85(5):1778-85.
265. Benito PJ, Gomez-Candela C, Dolores Cabanas M, Szendrei B, Aparecida Castro E, Grp PS. COMPARISON BETWEEN DIFFERENT METHODS FOR MEASURING BODY FAT AFTER A WEIGHT LOSS PROGRAM. *Revista Brasileira De Medicina Do Esporte.* 2019;25(6):474-9.
266. Shafer KJ, Siders WA, Johnson LK, Lukaski HC. Body density estimates from upper-body skinfold thicknesses compared to air-displacement plethysmography. *Clin Nutr.* 2010;29(2):249-54.
267. Lukaski HC, Bolonchuk WW, Hall CB, Siders WA. Validation of tetrapolar bioelectrical impedance method to assess human body composition. *Journal of applied physiology.* 1986;60(4):1327-32.
268. Brožek J, Keys A. The evaluation of leanness-fatness in man: norms and interrelationships. *British Journal of Nutrition.* 1951;5(2):194-206.
269. Durnin JV, Womersley J. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *British journal of nutrition.* 1974;32(1):77-97.
270. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. *British journal of nutrition.* 1978;40(3):497-504.
271. Jackson AS, Pollock ML, Ward A. Generalized equations for predicting body density of women. *Medicine and science in sports and exercise.* 1980;12(3):175-81.
272. Smith DP, Boyce RW. Prediction of body density and lean body weight in females 25 to 37 years old. *The American Journal of Clinical Nutrition.* 1977;30(4):560-4.
273. Pollock ML, Hickman T, Kendrick Z, Jackson A, Linnerud A, Dawson G. Prediction of body density in young and middle-aged men. *Journal of Applied Physiology.* 1976;40(3):300-4.
274. Pollock ML, Laughridge EE, Coleman B, Linnerud A, Jackson A. Prediction of body density in young and middle-aged women. *Journal of applied physiology.* 1975;38(4):745-9.
275. Womersley J, Durnin J. A comparison of the skinfold method with extent of 'overweight' and various weight-height relationships in the assessment of obesity. *British journal of nutrition.* 1977;38(2):271-84.
276. Deurenberg P, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. *British journal of nutrition.* 1991;65(2):105-14.
277. Hassager C, Gotfredsen A, Jensen J, Christiansen C. Prediction of body composition by age, height, weight, and skinfold thickness in normal adults. *Metabolism.* 1986;35(12):1081-4.
278. Day K, Kwok A, Evans A, Mata F, Verdejo-Garcia A, Hart K, et al. Comparison of a Bioelectrical Impedance Device against the Reference Method Dual Energy X-Ray Absorptiometry and Anthropometry for the Evaluation of Body Composition in Adults. *Nutrients.* 2018;10(10).
279. Ramos-Jimenez A, Hernandez-Torres RP, Murguía-Romero M. Anthropometric equations for calculating body fat in young adults. *Archivos Latinoamericanos De Nutricion.* 2018;68(2):111-21.
280. Kim SG, Ko K, Hwang IC, Suh HS, Kay S, Caterson I, et al. Relationship between indices of obesity obtained by anthropometry and dual-energy X-ray absorptiometry: The

- Fourth and Fifth Korea National Health and Nutrition Examination Survey (KNHANES IV and V, 2008-2011). *Obes Res Clin Pract*. 2015;9(5):487-98.
281. Weir CB, Jan A. BMI classification percentile and cut off points. 2019.
282. Ceniccola GD, Castro MG, Piovacari SMF, Horie LM, Corrêa FG, Barrere APN, et al. Current technologies in body composition assessment: advantages and disadvantages. *Nutrition*. 2019;62:25-31.
283. Fosbøl MØ, Zerah B. Contemporary methods of body composition measurement. *Clinical physiology and functional imaging*. 2015;35(2):81-97.
284. Gatterer H, Schenk K, Burtscher M. Assessment of Human Body Composition Methods and Limitations. *Body Composition*: CRC Press; 2017. p. 13-26.
285. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nature Reviews Endocrinology*. 2020;16(3):177-89.

DISCUSSION



DISCUSSION

The specific discussions for each study are shown below.

Chapter I: Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study I*).

The purpose of this study was to systematically examine the predictive validity of existing motor fitness and flexibility tests on health outcomes in adults and older adults.

Results from this systematic review revealed that there exists **strong evidence** indicating that: (i) slower **gait speed** predicts falls in adults over 60 years old, cognitive decline and impairment (including development of dementia and Alzheimer's disease) in adults over 55 years old, mobility disability in adults over 50 years old, Disability in Instrumental Activities of Daily Living (IADL) in adults over 54 years old, CVD risk in adults over 45 years old, all-cause mortality in adults over 35 years old, and risk of institutionalization or hospitalization in adults over 60 years old; (ii) impaired **balance** predicts falls in adults over 40 years old, disability in IADL or mobility disability in adults over 40 years old, and all-cause mortality in adults over 53 years old; (iii) worse **TUG** performance predicts falls and fear of falling in adults over 40 years old.

We have also found that there exists **moderate evidence** indicating that: (i) slower **gait speed** predicts hip fracture in older females over 75 years, depressive symptoms in adults over 50 and older males over 65 years, and other-cause mortality in adults over 50 years old; (ii) worse **TUG** performance predicts falls in females over 40 years old.

Due to a limited number of studies, the results also suggested **limited evidence** showing that: (i) slower **gait speed** predicts hip fracture in adults over 55 years old, frailty status in older adults over 65 years, decrease in functional autonomy in older adults over 68 years, and that does not predict stroke risk in adults over 35 years old; (ii) impaired **balance** predicts hip fracture in females over 75 years old, incident of dementia in adults over 90 years old, functional decline in older adults over 68 years, CVD risk in older adults over 65 years, other-cause mortality in older adults over 65 years, and risk of hospitalization in adults over 60 years old; (iii) worse **TUG** performance predicts cognitive decline and dementia in older adults over 65 years, disability in IADL in older adults over 65 years, decline in well-being in older adults over 70 years, and frailty status and loss of functional autonomy in older adults over 65 years; (iv) poor **flexibility** predicts falls in older adults over 70 years, mobility disability in older adults 65 years, and does not predict all-cause mortality in adults over 60 years old; (v) a shorter **step length** predicts risk of falling in older adults over 70 years, slower speed of movement predicts mobility disability in older adults over 65 years, shorter maximum step length predicts

disability in IADL in older adults over 76 years, and worse performance in the stair mounting test predicts disability in IADL in older males over 75 years.

At the age of 30, the biological system functioning reflects a critical transition point into decline and ageing¹. Adults over 60 years old suffer from more diseases related to aging² and they are users of more healthcare resources³. Ageing is characterized by gait changes, which may be a manifestation of compromised motor executive function⁴, then these results imply the neurological ability to coordinate motor tasks, reflecting motor fitness tests, such as gait speed, balance or the TUG tests.

The results derived from the studies included in the current systematic review suggest a predictive value of motor fitness tests related to diverse adverse health outcomes mainly for adults aged over 60, which is in agreement with this fact. Indeed, previous reviews are in accordance with our findings in this population⁵⁻¹⁰.

Moreover, predictive value has also been found for motor fitness tests related to health outcomes in younger adults. Thus, an early screening of physical fitness performance could help prevent adults from more harmful aging deterioration. For instance, evidence extracted from the present review seems to indicate adverse health outcomes from 35 years old related to slower gait speed (mobility disability, disability in IADL, CVD risk, all-cause mortality), or related to impaired balance or worse TUG performance (i.e., risk of falling, disability). Therefore, knowing the predictive ability of each test to detect health outcomes can facilitate to incorporate an early screening into clinical and practical settings.

Predictive validity of gait speed tests

Walking ability is a global measure of mobility that reflects a basic aspect of daily activity¹¹. Gait speed seems to begin to decrease in old age⁶ and is related to difficulty in walking, which may indicate more healthcare needs, higher incident disability, and shorter life expectancy¹¹. Our review supports the idea that gait speed assessment might be used for risk stratification and to guide health practitioners and clinicians in the management of altered health outcomes.

Although physical assessment is proposed mainly for older adults¹², and all the selected reviews corroborated it, as they included only older adults⁵⁻¹⁰, based on the present review we propose the inclusion in younger populations. Specifically, gait speed assessment is a good tool to identify adults at risk of cognitive decline and impairment, mobility disability/IADL, CVD and all-cause mortality in adults from 35 years old.

Moreover, gait speed may also identify depressive symptoms in adults over 50 years old. However, further research is needed because only 2 studies have explored this relationship^{13, 14}.

The elderly population shows a wide range of adverse health outcomes, including those mentioned above. In this sense, assessment of gait speed is currently recommended for older adults in the health practice environment^{15, 16}. Gait speed is a good measure of overall ability to compensate for decline in multiple body systems, including sensorimotor and cognitive function, which are common risk factors for falls¹⁵. The reduction in gait speed may be due to loss of physical functioning or the deterioration of brain motor control centers^{6, 17}. Moreover, evidence suggests that inflammatory markers^{6, 18}, play a role in the association between gait speed and mortality events are since they are associated with disability, worse cognitive performance and motor functioning, frailty and death^{6, 18}.

Different gait speed cut points have been reported for risk stratification of adverse health outcomes, but a single threshold was not yet evident. Nevertheless, and independent of the distance of the chosen test, cut points of <0.8m/s and <1.0m/s in gait speed appear to be sensitive to predict the risk of most of these adverse health outcomes in adults and older adults (see page 138, **Figure 4**). In fact, these cut points have been previously proposed^{6, 9}.

Predictive validity of balance tests

Maintaining balance is a complex task that demands good functioning of multiple organ systems with accurate coordination between them¹⁹. The impaired balance may be due to a lack of strength or age-related deterioration of sensory and neuromuscular control mechanisms^{16, 20}.

Although gait speed assessment is well-established, strong evidence indicates that impaired balance is a good tool to identify adults over 40 years old at risk of falling, disability in IADL or mobility disability, and all-cause mortality in adults over 53 years old.

Limited evidence was found for balance and other health outcomes in older adults, such as hip fracture, incident of dementia, functional decline, CVD risk, and other-cause mortality risk. This limited evidence may be primarily due to a lack of consensus on which balance test is most accurate in measuring different health outcomes or populations. A recent meta-analysis has suggested that the one-leg balance test is a good

predictor of disability in IADL in older adults¹⁰, although only three studies pooled the analysis. In fact, establishing cut points for balance measurement is also challenging, due to the wide range of existing protocols, although some studies have identified that inability to stand on one leg for less than 5 seconds may be sensitive to changes in risk of falls²¹⁻²³. Therefore, more studies are still needed to clarify which protocol is the most accurate in order to identify the most sensitive cut-off in balance performance related to different health outcomes.

Predictive validity of multidimensional measurement tests

The TUG test was a reliable and valid test developed by Podsiadlo and Richardson²⁴ in 1991 to assess “basic mobility skills” in older adults (70–84 years). This test was intended to be a simple and useful compendium between the measures of gait speed, balance, and functional capacity²⁴. Although various factors are associated with falls, mobility problems and impaired balance have been consistently identified as the main risk factors^{23, 25-27}.

The TUG test has been identified as a good tool to identify adults over 40 years old at risk of falling and fear of falling^{21, 28-32}. Moreover, worse TUG performance may also predict falls in females over 40 years old^{23, 25, 26}.

Therefore, the TUG test could be proposed as a predictive test for risk of falling in community dwelling adults from 40 years old^{21, 23, 25, 26, 28-32}. Since the risk of falling is associated with loss of independence, injuries, disability, long-term health care and premature mortality^{23, 25, 26, 31}, an available predictive tool for risk of falling is advisable. The adverse health outcomes related to health have been especially identified for females^{25, 26}, due to the loss of estrogens derived from perimenopause (around the age of 40) that can affect bone quality, being more prone to bone deterioration and fractures (such as hip fracture)³³.

Hence, falls and their consequences could be prevented with early screening by evaluating the TUG test among the aging adult population.

Nevertheless, this motor fitness test presents a series of drawbacks, resulting in limited and inconclusive predictive value in relation to other health outcomes.

Since the TUG test combines gait speed plus balance skills, it is not surprising that in those studies where some of these skills have been evaluated simultaneously (TUG performance plus gait speed or balance performance), all of them proved predictive capacity. In fact, the TUG test, gait speed and balance test were related to falls³⁴ and to

decline in functional autonomy³⁵ in older adults; the TUG and a gait speed tests were related to falls in older adults²⁹; or the TUG and balance tests were related to falls in adult females over 40 years old^{23, 26}.

Conversely, Mulasso et al.²⁷ found that neither the TUG nor a one-leg standing balance tests were predictive of falls. Same results derived from the systematic review conducted by Cavanaugh et al.⁵ between the TUG and the FRT (a balance test) related to falls and mobility disability. Nevertheless, only a few participants reported experiencing at least one fall in the previous 12 months, 30²⁷ and 42⁵ participants respectively, which could imply such a lack of significance.

Donoghue et al. also failed to find any association between the TUG and gait speed tests related to cognitive decline³⁶. However, trends of significance were found between slower TUG performance and increased Mini-Mental State Examination errors. They also reported that their participants had a mean of 10.5 seconds in TUG performance and the cognitive performance did not change substantially over the follow-up. Thus, that could partially explain its lack of predictive validity.

On the other hand, these simultaneous results have not always been given among tests. Thus, some studies found predictive value of the TUG test and failed to find association regarding falls risk for gait speed tests^{21, 28, 31}. One possible explanation for that fact could be the minimal difference in gait speed tests performance between fallers and non-fallers, since its difference was only of 0.05 m/s³¹, 0.12 m/s²⁸ and 0.13 m/s²¹ respectively between groups, being its difference higher in TUG performance (for fallers vs non-fallers).

The fact that the health outcome “falls” had the highest number of studies measured with the TUG test, gait speed, and balance tests measured simultaneously may be partly due to that, precisely, avoiding falls require of good mobility functions, such as walking ability, maintaining balance, strength, and agility³¹.

Inconclusive evidence was found for the predictive validity of the TUG test related to IADL. In the meta-analysis conducted by Wang et al.¹⁰, they suggested that the TUG test is a good tool to identify disability in IADL. However, Cavanaugh et al.⁵ found insignificant results. This fact could be partially explained by the number of studies explored in each review. While Wang et al.¹⁰ included eight studies analyzing TUG, Cavanaugh et al.⁵ only included 2, which may limit the predictive evidence found for such test.

Cut-off points in TUG performance are still unclear. Therefore, different values have been previously reported. For example, cut-off points of 8.1, 9.2, and 11.3 s for adults aged 60–69, 70–79, and 80–99 years old, respectively, has been described by Bohannon³⁷ in 2006 to identify elders with mobility deficits. We have summarized the cut-off points found in TUG to identify diverse health outcomes, such as falls, functional autonomy decline, disability in IADL, frailty, or well-being decline, ranging from faster performance (7s) to slower (21s) (see page 139, **Figure 5**).

Predictive validity of flexibility tests

Although several cross-sectional studies have found associations between flexibility and diverse health outcomes, such as cardiometabolic risk, well-being, etc., its predictive value (through a longitudinal design) is still scarce^{34, 38, 39}. These three studies were carried out in older adults with different flexibility protocols. Furthermore, only Ward et al.³⁹ found that flexibility was predictive of mobility disability³⁹, while the other two studies did not confirm its predictive capacity for falls³⁴ or all-cause mortality³⁸.

Chapter II: Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

The present systematic review comprehensively studied the criterion-related validity of the existing field-based physical fitness tests used in adults. The findings of this review provide an evidence-based proposal for most valid field-based physical fitness tests for adult population.

Cardiorespiratory fitness

The gold standard to assess VO_{2max} is the Douglas bag method, although there is agreement on that respiratory gas analyser is a valid method to assess oxygen uptake⁴⁰. All high quality studies measured VO_{2max} or peak oxygen consumption when performing a submaximal/maximal treadmill or cycle test, except Manttari et al.⁴¹, that measured VO_{2max} directly when performing the 6-minute walk test.

Distance and time-based run/walk tests

The run/walk field tests are probably the most widely used^{42, 43}, however, until recently, there was no consensus regarding the most appropriate distance or time to use on these tests⁴⁴. Mayorga et al.⁴⁵ performed a meta-analysis where examined the criterion-related validity of the 5000-m, 3-miles, 2-miles, 3.000-m, 1.5-mile, 1-mile, 1000-m, ½-mile, 600-m, 600-yd, ¼-mile, 15-minute, 12-minute, 9-minute, 6-minute run/walk tests. They found that the criterion-related validity of the run/walk tests, considering only the performance score, ranged from low to high, being the 1.5-mile and the 12-minute run/walk tests the most appropriate tests for estimating cardiorespiratory fitness in adults aged 18-64 years. Sex, age or VO_{2max} level did not affect the criterion-related validity, whereas when multiple predictors (i.e. performance score, sex, age or body mass) were considered, the criterion-related validity values were higher. In this sense, two high quality original studies reinforce these results, and showed that the 12-minute⁴⁶ and the 1.5-mile^{46, 47} run/walk tests were fairly accurate for estimating cardiorespiratory fitness in adults from 18 to 26 years ($r = 0.87-0.93$, $p < 0.05$).

Overall, the run/walk tests are not user-friendly tests, due to the difficulty to develop an appropriate pace, which may affect test outcome (some participants start too fast, so that they are unable to maintain their speed throughout the test; others, start too

slow so that when they wish to increase their speed the test has already finished). These problems are more likely to occur in longer distance tests. Other factors affecting test outcome include the individual's willingness to endure the discomfort of strenuous exercise, a short attention span, poor motivation, and limited interest in a monotonous task⁴⁸⁻⁵⁰.

The 2-km and 6-minute walk tests are probably the most widely used walk tests in adults^{51, 52}. Both tests require a submaximal effort, thus avoiding the problem of enduring the discomfort of strenuous exercise. In addition, it allows evaluating those people with a low level of physical fitness or unable to run. Three high-quality studies^{51, 53, 54} observed that the Oja's equation derived from the 2-km walk test has high validity ($r = 0.80-0.87$, all $p < 0.05$) in untrained and/or overweight/obese adults aged 20-64 years. One high quality study reported that the 2-km walk test⁵⁵ is a reasonably valid field test for estimating cardiorespiratory fitness of moderately active adults aged 35-45 years, but not in adults with very high maximal aerobic power.

Many studies have developed prediction equations for the 6-minute test based on spirometry⁵⁶. However, only three high-quality studies^{41, 52, 57} analysed the criterion-related validity of the 6-minute test based on VO_{2max} in adults. They showed a moderate to high validity ($r = 0.70-0.93$, all $p < 0.001$) in obese and healthy adults aged 18-64 years. Burr et al.⁵⁸ suggested that, on its own, the 6-minute walk test can be useful to discriminate between broad categories of high, moderate and low fitness, but that this approach may be associated with a degree error, especially in the high fitness group. According to these findings, the 2-km and 6-minute walk tests are valid to use in adults aged 18-64 years with low or moderate fitness levels, but not in adults with a high fitness level.

Regarding the 1-mile walk test, conflicting results were found, especially when examining the accuracy of the Kline's⁵⁹ and Dolgener's⁶⁰ equations in adults aged 18-64 years.

20-m shuttle run test

The 20-m SRT was developed by Leger et al.⁶¹, to solve the pace issue of the run/walk tests. The test consists of 1-minute stages of continuous, incremental speed running. Lately, a meta-analysis⁶² showed that the performance score of the 20-m SRT had a moderate to high criterion-related validity for estimating VO_{2max} ($r_p = 0.66-0.84$) in youth and adults aged 18-64 years, being higher when other variables (i.e. sex, age or body mass) were accounted ($r_p = 0.78-0.95$). This study also reported that Leger's protocol

had greater average criterion-related validity coefficient ($r_p = 0.84$; 95%CI: 0.80-0.89) than Eurofit, QUB and Dong-HO protocols; and Leger's protocol was statistically higher for adults ($r_p = 0.94$, 0.87-1.00) than for children ($r_p = 0.78$; 95% CI: 0.72-0.85). These values are higher than reported the 1500-m and 12-minute run/walk tests⁴⁵. Moreover, the meta-analysis showed that sex did not seem to affect the criterion-related validity values. On the other hand, Cooper et al.⁶³ showed that the Brewer's protocol and equation was not valid to assess active young people aged 18-26 years (mean difference= 1.8 ± 6.3 ml/kg/min; $p = 0.004$). In line with these findings, Kim et al.⁶⁴ observed that Leger's protocol and equation was more accurate than Brewer's protocol and equation (mean difference -0.54 ml/kg/min; %CV: 1.39 vs mean difference -2.944 ml/kg/min; %CV: 8.87) in Korean adults, especially in females. Nonetheless the authors suggested the need to develop new equations for Korean adults.

It is important to note that the 20-m square SRT^{65, 66} has been proposed as an alternative to the 20-m SRT to reduce the test's turning angle from 180 to 90. This test was best predictor of VO_{2max} than the 20-m SRT in young male adults aged 18-25 years.

Step tests

The step tests are safe, simple, inexpensive and practical method of assessing cardiorespiratory fitness under submaximal conditions, which require a minimum space⁶⁷; and they are a great alternative to laboratory tests in clinical setting. There are a wide variety of step-test protocols which differ in stepping frequency, test duration and number of test stages. Bennett et al.⁶⁷ analysed the criterion-related validity of different step tests (the Chester step test, a personalised step test, the STEP tool step test, the Queen's College step test, the Skubic and Hodgkins step test, a height-adjusted, rate-specific, single-state step test, the Astrand-Ryhming step test, and a modified YMCA 3-minute step test) in adults aged 18-64 years. The validity of these tests ranged from moderate to high, and they suggested that the Chester step test was the most valid step test to evaluate cardiorespiratory fitness in adults. However, this systematic review only included two studies with contradictory results, similarly to the Queen's College step test. Analysing the 12 high-quality studies that examined the criterion-related validity of the step tests in adults aged 18-64 years, we can conclude that the YMCA step test⁶⁸⁻⁷² seemed to be the most appropriate step test to estimate VO_{2max} in adults aged 18-64 years. However, it is important to note that there is no single equation, since the result of the equation depends on the sample used. Even, Santos and Golding⁷³ altered the protocol

adjusting the step height to the individual participant's height, in order to increase the accuracy of this test.

Levels of evidence

There was **strong evidence** indicating that: (i) the 20-m SRT using the Leger's equation, the 2-km walk using the Oja's equation, the 6-minute and the YMCA step tests are valid for estimating cardiorespiratory fitness; (ii) the criterion-related validity of the distance and time-based run/walk tests range from low to high, being the best predictors the 1.5-mile and 12-minute run/walk tests. There was **moderate evidence** indicating that: (i) the 20-m square SRT is valid for estimating cardiorespiratory fitness. Due to the inconsistent results found in high quality studies, there was **limited evidence** about the validity of the 1-mile walk, treadmill jogging, incremental shuttle walking, Chester, and Queen's College step tests. Due to the low number of high quality studies, we found **limited evidence** that: (i) the 3-minute walk, the ¼-mile walk, Mankato submaximal, Modified Astrand-Ryhming, University Montreal, Modified Canadian aerobic fitness step, 6-minute single 15-cm step, Tecumseh step, modified Harvard step and Astrand–Ryhming Step tests are valid for estimating cardiorespiratory fitness; (ii) the YMCA cycle, Ruffier, Danish step, and 2-minute step tests are not valid for estimating cardiorespiratory fitness. Due to the consistent results found in multiple low quality studies, we found **limited evidence** that the 6-minute step test is valid for estimating cardiorespiratory fitness.

Muscular fitness

The specificity of the type of muscular work performed and the use of different energy systems are both major challenges for establishing a gold standard method for maximal, endurance and explosive muscular fitness tests⁷⁴. One repetition maximum (1RM) and repetitions to a certain percentage of 1RM (i.e. 50% of 1RM or 70% of 1RM),⁴² isokinetic dynamometer strength⁷⁵⁻⁷⁷, and electromyography^{78, 79} have been used as gold standards.

Maximal isometric muscular fitness

The TKK dynamometer⁸⁰⁻⁸² seemed the most appropriate test to assess maximal isometric muscular fitness in adults. All the studies used the “known weights” as criterion reference.

Several studies have examined whether the elbow position (extended or flexed at 90 degrees) affects the handgrip test score in children⁸², adolescents⁸³, and young adults⁸⁴. They observed that performing the handgrip test with the elbow extended seems the most appropriate protocol to evaluate maximal handgrip muscular fitness in these populations; what is in accordance with the protocol recommended by the American Center for Disease Control and Prevention⁸⁵.

Ruiz et al.⁸⁶, also investigated whether the position (grip span) on the standard grip dynamometer determined the maximum handgrip muscular fitness in adults. They found that when measuring handgrip muscular fitness in females, hand size must be taken into consideration, providing mathematical equation ($y = x/5 + 1.5$ cm) to adapt optimal grip span (y) to hand size (x). In adult males, optimal grip span could be set at a fixed value (5.5 cm) and is not influenced by hand size.

Importantly, just like the step test, the handgrip test can be very useful in clinical setting, because it requires minimal equipment and space, is time-efficient and easy to administer.

Endurance muscular fitness

The Biering-Sørensen test, a trunk holding test in an antigravity prone position, is commonly used to measure the back and hip muscle endurance muscular fitness, which is associated with low back pain⁸⁷. Mannion et al.⁸⁸ and Coorevits et al.⁷⁸, showed that the test endurance time was highly associated with isometric/endurance hip and back muscular fitness ($r = 0.84-0.98$, $p < 0.01$). On the other hand, Kankaanpää et al.⁸⁹, found that this association was moderate ($r = 0.60-0.71$, $p < 0.05$). However, when BMI ($r = -0.49-0.51$, $p < 0.001$) in females and age ($r = 0.25-0.29$, $p < 0.05$) in males were accounted in the prediction model, the variance explained increased considerably. Thus, the Biering-Sørensen test might be considered as valid for measuring back muscle endurance muscular fitness.

Assessing abdominal muscle functionality is clinically relevant since it is considered to be related to low back pain^{90, 91}. The curl-up test, or its different versions, is originally the field test used to assess this capacity. In the present review, no original studies evaluating the criterion-related validity of this test were classified as high quality. An alternative of the curl-up test, could be the prone bridging test, an isometric holding test in prone position, which is currently being used to supposedly measure the abdominal endurance muscular fitness. The prone bridging test time is inversely associated with low back pain^{92, 93}. In relation to the validity of this test, De Blaiser et al.⁷⁹ found higher

activation of the abdominal core musculature during the test than the back and hip musculature, showing a high association between test time and abdominal endurance muscular fitness. Future high quality studies are necessary to clarify the validity of this test.

It should be noted that no study that analysed the criterion-related validity of lower and upper-body endurance muscular fitness test were classified as high quality study.

Explosive muscular fitness

The SLJ has been proposed in health-related fitness test batteries in preschools⁹⁴ and children and adolescents⁹⁵ to assess body lower explosive muscular fitness, given its criterion-related and predictive validity. However, to our knowledge, the criterion-related validity of this test has not been studied. Bui et al.⁹⁶ observed that the Sargent jump test is not appropriate to evaluate lower-body explosive muscular fitness, because overestimated the height of a vertical jump and its accuracy is reduced as jump height increase (mean difference: 4.4 ± 5.1 , $p < 0.001$). Due to the close relationship that lower-body maximal/explosive muscular fitness has on adult health^{97, 98}, more high-quality studies are required to analyse the criterion-related validity of these tests in future research.

Levels of evidence

There was **strong evidence** indicating that: (i) the handgrip test with the elbow extended and with the grip span adapted to the hand size and sex (using the TKK dynamometer) is a valid test to assess maximal isometric muscular fitness; (ii) the Biering-Sørensen test is valid for estimating endurance muscular fitness of hip/back muscles. There was **moderate evidence** indicating that: (i) handgrip test (Jamar) has acceptable validity to assess maximal isometric muscular fitness. Due to: (i) the low number of high quality studies, we found **limited evidence** about the validity of the prone bridging and Sargent jump tests to assess abdominal endurance muscular fitness and lower-body explosive muscular fitness, respectively; (ii) the inconsistent results found in multiple high quality studies, there was **limited evidence** about the validity of the handgrip test (DynEx) to assess maximal isometric muscular fitness; (iii) the consistent results found in multiple low or very low quality studies, we found **limited evidence** that the curl-up test or its different versions are not valid to assess abdominal endurance muscular fitness.

Flexibility

Radiography seems to be the best criterion measurement of flexibility, but goniometry has also been used as a criterion measure^{99, 100}. Goniometers are relatively easy to obtain, nevertheless, their use requires a certain technical qualification since it is a sensitive method, and thus it is not feasible for use in all settings¹⁰¹. Traditionally, the sit-and-reach test, originally designed by Wells and Dillon¹⁰² and its different versions, have been included in the fitness test batteries to measure hamstring and low back flexibility; and probably they are most widely used measures of flexibility⁴².

Mayorga et al.¹⁰³ performed a meta-analysis to analyse the criterion-related validity of the sit-and-reach and its different versions (modified sit-and-reach, back-saver sit-and-reach, modified back-saver sit-and-reach, V sit-and-reach, modification V sit-and-reach, unilateral sit-and-reach and chair sit-and-reach). These tests showed moderate validity for estimating hamstring extensibility, but a low validity for estimating lumbar extensibility. They also found that the classic sit-and-reach test had the highest criterion-related validity coefficient in both, hamstring and lumbar extensibility, than the other test, which does not seem to justify the use of the classic protocol modifications in order to solve the problems attributed to itself (i.e. the length proportion between the upper and lower limbs or the position of the head and ankles).

The toe-to-touch test is another field-based test to measure hamstring flexibility, in which the individuals were assessed standing instead of sitting on the floor¹⁰⁴. Although this test is easy to administer and can be an alternative to the sit-and-reach test, when the participant has problems being measured sitting, it has not been proposed for any field-based fitness test battery. A meta-analysis¹⁰⁵ analysed the criterion-related validity of the toe-to-touch test to measure hamstring flexibility, reporting similar validity coefficients that the classic sit-and-reach.

It is interesting to highlight that recently, Nuzzo¹⁰⁶ has suggested that flexibility should be retired as a major component of fitness, due to the lack of predictive and concurrent validity with meaningful health and performance outcomes.

Levels of evidence

There was **strong evidence** indicating that: (i) the sit-and-reach test and its modified versions has moderate validity for estimating hamstring extensibility, but a low validity for estimating lumbar extensibility; (ii) the toe-to-touch test has moderate validity for estimating hamstring extensibility.

Motor fitness

The validity of motor fitness tests has been the least studied in adults. None of the three studies that analysed the criterion-related validity in motor fitness tests was classified as high quality. Given that the motor fitness tests (i.e. gait/walking speed, balance, TUG) are associated with all-cause mortality¹⁰⁷⁻¹⁰⁹, falls and fractures¹¹⁰, disability in activities of daily living¹¹¹, and depression¹¹², it would be appropriate to know their criterion-related validity.

Levels of evidence

Due to the consistent results found in multiple low quality studies, we found **limited evidence** that the ten step test has moderate validity to assess agility.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

As obesity data is rising, practitioners and researchers are more frequently interested in evaluating body composition to better manage it, improve treatment decisions, and optimize health outcomes. Therefore, we have systematically examined the validity of different field-based methods and equations for body composition estimation in the adult population that may provide accurate results, minimizing prediction error, compared with “gold standard” methods.

The main findings of this systematic review were: (i) **strong evidence** indicated that WC, BAI, and BMI are valid to estimate body adiposity; (ii) in estimation equations studies, **strong evidence** indicated that Durnin/Womersley equation, Jackson/Pollock equation (males), and Jackson, Pollock and Ward equation (females), BMI equations, BIA equations, equations implying sum of SKF, equations combining circumferences+SKF, equations based on WC, equations combining WC+SKF, BAI equations, equations combining various circumferences, WHtR equations, equations combining abdominal circumference+SKF, equations combining WC+SKF+BMI, equations combining circumferences+diameters+SKF, and equations combining SKF+BMI are valid to estimate total body fat mass or %BF.

The findings of this systematic analysis are largely heterogeneous. The methods to estimate body composition, both field-based methods and criterion methods, and the estimation equations used in the studies are also variable, further contributing to the heterogeneity of the collected data. Thus, it is still challenging to establish valid methods for the field assessment of human body composition due to specific differences in age, sex, race-ethnicity, health status, weight status, etc. Despite the difficulty in drawing meaningful conclusions from this data set, the most relevant findings (i.e., those with moderate and strong evidence of validity), including main advantages and limitations, are discussed below (see pages 312-315, **Table 30**).

Criterion-related validity of field-based body composition estimation in adults

The measurement of field-based body composition is very easy and cost-effective, and therefore, frequently used in epidemiologic studies. Numerous obesity indexes have been applied to characterize obesity, being BMI by far the most common approach in clinical practice. The main advantage of using BMI as a overall index of adiposity is its

predictive power for clinical outcomes, such as cardiovascular mortality¹¹³. However, despite its worthiness, BMI can vary between individuals, making it an insufficient method to measure health-related body composition, as increased muscle mass and weight may falsely increase its BMI, as in athlete population. Moreover, it is also known its limitations regarding age, sex, and race-ethnicity¹¹³. Although same BMI classification is used for both female and male adults, at the same BMI, females tend to have more body fat than males and to have abdominal obesity ($WC \geq 88$ cm in females or ≥ 102 cm in men)¹¹⁴, especially during the stages of pregnancy or menopause. Regarding age, at the same BMI, older people tend to have more body fat than younger adults¹¹⁴. Likewise, different cutoffs are established for White, Hispanic, and Black individuals and the Asian and South Asian populations¹¹⁴. For the first group, a BMI > 25.0 kg/m² is defined as overweight, while in the second group, overweight status is pointed out from 23.0 kg/m²¹¹⁴. Thus, at the same BMI, White/Hispanic have more body fat than Blacks, and Asians have more body fat than do Whites¹¹⁴.

Therefore, other indexes have emerged trying to overcome these limitations, such as BAI¹¹⁵, WC, WHR, or WHtR. The main advantage of these indexes is the ease and speed of measurement (just needing a non-elastic tape) and calculation, being suitable for clinical settings and epidemiological studies. Moreover, all of them imply central adiposity measurement, which provides independent and additive information to BMI to predict morbidity and mortality risk¹¹⁶. Likewise, height may influence the fat distribution, being more accurate than BMI (a decrease in height will increase BMI without an increase in fat mass)¹¹⁷, and short stature is associated with an increased risk of cardiometabolic health complications¹¹⁶. Notwithstanding, height is only marginally associated with WC and WHtR seems less useful as, in adults, the height is generally fixed and the value can only be changed by changes in WC¹¹⁶. One disadvantage of BAI is that HC by itself does not seem to be a good estimator of %BF^{118, 119}. Therefore, although strong evidence has been found, it is possible that others studies failed to find validity based on this fact. Thereby, some caution is still warranted when using this novel index as a measure of body composition alone. Similarly, and regarding WHR, the relocation of fat to a more central distribution and loss of muscle mass in the gluteal region will increase WC and WHR without increasing fat mass¹¹⁷. For these reasons, it seems that the simple measurement of WC can provide relevant clinical information in the management of obesity¹²⁰. Indeed, among the numerous existing circumferences measurements, WC is the most used. WC correlates with total body fat and is often used

as a measure of central obesity¹²⁰. In addition, WC can be especially useful for stratifying high-risk people classified as overweight or with obesity class 1, and is also a key part of the definition of metabolic syndrome¹²⁰. The main drawback of WC is that the measure may be influenced by the timing of measurement (e.g., preprandial and postprandial period), and it requires the individual to remove clothing to obtain a more accurate measurement, so some people may be uncomfortable with this measurement¹²¹. Furthermore, population-specific cutoff points have been proposed¹²².

Finally, the BIA device, which can be considered as a field method, is also widely used in clinical, research and field settings, due to its relatively low cost and portable. However, BIA presents some bias, due to under or overestimation in body fat¹²³, especially in measuring individuals with severe obesity¹²⁴.

In summary, the field approach to body composition assessment highlights its applicability in non-laboratory settings, with WC being the most commonly, quickest, simplest and economical method.

Criterion-related validity of estimation equations body composition assessment in adults

In addition to field-based methods, estimation equations help implement information about an individual's body composition. Consequently, the most accurate estimation is the clinical importance. However, due to typical time and resources constraints, a simplified estimation equation is advisable for practical implications.

In early 1951, Brozek and Keys¹²⁵ published the first valid equations for males aged 18-55 years derived from SKF and body density (with UWW as the criteria method), although the SKF chosen were not ideal and their formula has not been widely used. Since body fat needed to be determined from body density, some representative authors, such as Siri¹²⁶ and Brozek¹²⁷ calculated formulas to estimate %BF. They obtained similar results in their conversion formulas (the difference between equations is less than 1% BF error of estimation).

To date, the generalized equations for predicting body fat from the sum of SKF, proposed by Durnin/Womersley¹²⁸ in 1974 for females and males aged 16-72 years; by Jackson/Pollock¹²⁹ in 1978 for males aged 18-61 years and; by Jackson, Pollock and Ward¹³⁰ in 1980 for females aged 18-55 years, are still widely used (see page 229, **Tables 26-27**). Durnin/Womersley¹²⁸ tested different regression equations, based on the sum of one to four SKF. They assumed that there is little error using one, two or three measured

SKF; although the error can be reduced by using the sum of four SKF ($r = 0.84-0.95$). In fact, the Durnin/Womersley sum of four SKF equation, is the most used¹³¹⁻¹⁵⁶. Likewise, Jackson/Pollock¹²⁹, and Jackson, Pollock and Ward¹³⁰ established models with the sum of three, four and seven SKF, concluding that all of them were almost equally valid ($r = 0.91-0.92$ and $r = 0.85-0.87$ for male and female adults, respectively). This validity among different SKF sums was supported by later studies^{134, 157, 158}. Therefore, when time is a handicap (i.e., in epidemiological studies or clinical settings) it seems that the most feasible is to apply one SKF (Durnin/Womersley)¹²⁸ or the sum of three-four SKF (Jackson/Pollock, or Jackson, Pollock and Ward)^{129, 130}.

Particularly, these equations showed high validity in general adult population^{128-130, 134-139, 141, 143-149, 152, 153, 156-162}. Only a few studies found low validity in these equations^{138, 145, 154, 155}. Two of them found low validity in females with overweight¹³⁸ and adults with obesity^{138, 145}, in addition to Wattanapenpaiboon et al.¹⁵⁵, who suggest that Durnin/Womersley equation tends to increasingly underestimate %BF with increasing subcutaneous fat, being aware of some methodological limitations in adults with overweight or obesity, especially in females. Shafer et al.¹⁵⁴ reached the same conclusion regarding overestimation in female adults, although they found low validity even in adults with normal weight. Moreover, these seem to need a reevaluation within different race/ethnicity groups^{136, 137, 139} or body fat status^{134, 138}. Accordingly, although these equations are widely applicable, the SKF method has some drawbacks. Principally, SKF results may be affected by the expertise of the technician, the type of caliper used (e.g., most economic calipers have less scale precision), and some individual factors (e.g., adults with overweight or obesity, sex or race/ethnicity).

Since then, several new equations based on anthropometric measurements or BIA analysis are still emerging and re-evaluated, resulting in an endless demand for comparisons between them, since regression equations seem to be population-specific, especially regarding sex, age, body fat status, or race-ethnicity^{139, 140, 151, 153, 158, 163, 164}. Specifically, BMI alone cannot be either generalized among different race-ethnic groups^{115, 148, 165, 166}, and is age and sex dependent^{163, 166-168}. Therefore, BMI equations used in combination with SKF^{159, 164, 169}, or WC and SKF^{139, 150, 159, 170, 171} have been developed to improve its accuracy to estimate %BF. Regarding BAI¹¹⁵, and despite race-ethnicity or weight status differences, the studies in this systematic review concluded that BAI equations are more accurate than BMI alone to estimate %BF^{115, 118, 119, 162, 172, 173}. Finally, although BIA is a method of easy applicability (as a field method), some studies

have found that it underestimates or overestimates the estimation of %BF, mainly caused by derived equations based on race-ethnicity, or specific calibration machine^{143, 148, 174}.

Finally, it seems that other lifestyle characteristics are rising to create the most specific equation to predict %BF. In 2001, Tucker et al.¹⁷⁵ developed a prediction equation that included diet and fitness and exercise questions, showing high validity ($r=0.89$, $p<0.05$). Similarly, Zanovec et al.¹⁷⁶ examined the role of physical activity in prediction equations based on BMI, suggesting that physical activity relied on the more precise and accurate equation to estimate %BF (in addition to sex and race-ethnicity).

Nowadays, these equations are still being refined. In fact, Lee et al. 2021¹⁷⁷ developed four equations where including from the simplest (i.e., age, height, weight, and WC) to the most complete (i.e., age, height, weight, WC, serum creatinine level, physical activity, smoking habit, and alcohol use), and unsurprisingly, they found the last one the most valid ($r=0.91$, $p<0.001$). Therefore, collecting information on lifestyle behavior should be worthy in designing a more efficient individual prediction equation.

The previous discussions have made clear that valid body composition estimation equations are applicable in clinical and epidemiological research and in health-related environments. Where possible, individualization is desirable. Otherwise, the generalized equation, such as the simplest versions of the Durnin/Wommersley equation (one to three SKF), the Jackson/Pollock equation, and the Jackson, Pollock and Ward equation (three or four SKF) are suitable for body composition estimation in adult population.

Chapter III. Reliability of Field-Based Physical Fitness Test in Adults: A Systematic Review (*study IV*).

The present systematic review summarized the existing reliability for field-based physical fitness tests in adult population. The finding of this review provide an evidence-based proposal for most reliable field-based physical fitness tests for adults.

Cardiorespiratory fitness

Cardiorespiratory fitness reflects the overall capacity of the cardiovascular and respiratory systems to supply oxygen during maintained physical activity¹⁷⁸. It has been traditionally assessed through maximal (e.g., 20-m SRT) and submaximal (e.g., 2-km walk, 6-minute walk, or 6-minute step tests) field-based physical fitness tests. The reliability of the 20-m SRT was investigated in nine high-quality studies^{61, 179-186}. In seven of these studies, the 20-m SRT demonstrated high test–retest reliability based on coefficients of correlation ($ICC = 0.93–0.96^{180, 182, 183}$; $r = 0.85–0.96^{61, 179, 182-184, 186}$ in adults aged 18–45 years. Caution is required when comparing reported coefficients between different samples (mainly of different ages and fitness levels).

In four of these studies, SEM was calculated showing the low dispersion of measurement errors between test–retest (all SEMs < 15%^{180, 182-184}). In one study conducted in active males of mean age 21.8 years, the 20-m SRT returned no significant mean test–retest differences ($p = 0.90$), and a mean difference between test–retest (i.e., bias) and limits of agreement of 0.4 ± 2.7 mlO₂/ kg/min¹⁸¹.

These findings are in agreement with those of other studies performed in young adults^{182, 185}. One of these studies detected a mean difference and limits of agreements between test–retest of 0.6 ± 6.8 mlO₂/ kg/min¹⁸⁵. In the other study, the 20-m SRT was performed three times in active young adults, and the mean difference and limits of agreement were -1.1 ± 4.7 mlO₂/kg/min between sessions 1 and 2, and 0.0 ± 5.5 mlO₂/ kg/min between sessions two and three¹⁸².

These studies concluded that the 20-m SRT offers high reliability with reasonably narrow limits of agreement and consistent results between test–retest in young adults (i.e., no significant mean test–retest differences, $ps > 0.05$). However, another study performed in males of mean age 34.8 years showed (despite an $ICC \geq 0.95$) a lack of significant differences between test–retest ($p \leq 0.05$), and a mean difference between test–retest and limits of agreements of 0.8 ± 3.1 mlO₂/ kg/min¹⁸³. Other studies analyzed the reliability of different versions of the 20-m SRT (i.e., 20-m square SRT¹⁸⁵, 15-m square SRT¹⁸⁷, 15-

m shuttle walk/run test¹⁸⁸ and 10-m shuttle walk test^{189, 190}). The 20-m and 15-m square SRT showed high reliability in adults aged 18–29 years, and could be a good option to reduce the test's turning angle from 180° to 90° ($r \geq 0.78$, $p > 0.05$ between test–retest, $CV \leq 5.2\%$). The 15-m shuttle walk/run and the 10-m shuttle walk tests demonstrated high reliability in adults aged 40–59 years ($ICCs \geq 0.93$, $p > 0.05$ between test–retest).

These tests could be an alternative to the 20-m SRT in adults whose fitness level is low or who find it difficult to run (e.g., obese individuals or those with low back pain).

The reliability of the 6-minute walk test was examined in four high-quality studies^{191–194}. The test–retest reliability of the 6-minute walk test was considered high based on coefficients of correlation ($ICCs = 0.93–0.96$ ^{193, 194}, $r = 0.90$ ¹⁹¹) in three of these studies, and moderate in one ($ICC = 0.82$ ¹⁹²) in adults aged 18–64 years. The 6-minute walk test showed the low dispersion of measurement errors between test–retest (all $SEMs \leq 3.2\%$ ^{193, 194}). In one of these studies performed in obese adults aged 21–62 years, in addition to an ICC of 0.96 and considering mean significant difference between test–retest ($p < 0.001$), the 6-minute walk test showed a mean test–retest difference of 18 m, limits of agreement between –46 m and 80 m, and a CV of 4.7%¹⁹⁴. Thus, the 6-minute walk test demonstrated good reproducibility in obese individuals based on a low CV and high ICC.

The reliability of the 5-minute walk test, a modified version of the 6-minute walk test, was analyzed in 44 participants with low back pain and 48 controls aged 21–63 years. Test retest reliability was moderate in the low back pain group ($ICC = 0.87$) and low in the control group ($ICC = 0.60$). However, $SEMs$ were low in both groups ($\leq 12\%$)¹⁹⁵. Other field-based physical fitness tests for assessing cardiorespiratory fitness based on walking/running in a given time or space used in reliability studies were: 2-km walk¹⁹⁶, 400-m walk¹⁹⁷, 1.5-mile test (walking or running)¹⁹⁸, Université-Montréal¹⁹⁹ and Cooper's 12-minute run¹⁸⁶. The reliability of the 6-minute step test was examined in four high-quality studies^{200–203}. In all of them, the 6-minute step test showed high test–retest reliability (based on $ICCs$ ranging from 0.90 to 0.97^{200–203}) in healthy adults^{200–202} and adults at risk of CVD²⁰³ aged 20–64 years. The 6-minute step test revealed the low dispersion of measurement errors between test–retest (all $SEMs \leq 2.8\%$ ^{200, 202, 203}). The reliability of the Chester step test was investigated in two studies conducted in adults aged 19–52 years^{204, 205}. These studies analyzed reliability through the Bland–Altman method and showed mean differences between test–retest and limits of agreement of 0.8 ± 3.7 mlO₂/kg/min²⁰⁴ and -0.7 ± 4.5 mlO₂/kg/min (mean significant difference between test–

retest $p > 0.05$)²⁰⁵. In both studies, the Chester step test revealed the low dispersion of measurement errors between test–retest (all SEMs < 3%), indicating that this test had a high test–retest reliability. Other field-based tests for assessing cardiorespiratory fitness based on steps used in reliability studies were the 2-level step²⁰⁶ and the OSU step²⁰⁷.

Levels of evidence

There was **strong evidence** indicating that: (i) the 20-m SRT demonstrates high test–retest reliability in young adults; and (ii) the submaximal tests 6-minute walk test and the step tests show high test–retest reliability in adults aged 18–64 years. These submaximal tests might be an alternative in adults with a low level of physical fitness or difficulty in running (e.g., obese individuals). There was **moderate evidence** indicating that the Chester step test provides results with high test–retest reliability in adults aged 19–52 years. Due to the low number of studies (a single study), there was **limited evidence** indicating that: (i) the 400-m walk and 2-km walk tests produce results with high test–retest reliability in adults aged 30–64 years and could be an option to assess cardiorespiratory fitness in adults with a low level of fitness or difficulty in running; (ii) the 1.5-mile test (walking or running) demonstrates high test–retest reliability in young adults aged 19–26 years; (iii) the Cooper’s 12-minute run test produces results with high reliability in young adults, and the Université de Montréal Track test does so in middle-age adults; (iv) the 2-level step test provides highly reliable results in adults aged 55–64 years; and (v) the OSU step test shows low reliability in young adults.

Musculoskeletal fitness

Muscular fitness is defined as the ability of a specific muscle or muscle group to generate force (muscular fitness) to resist repeated contraction over time, to maintain a maximal voluntary contraction for a prolonged period of time (muscular endurance), or to carry out a maximal/dynamic contraction in a short period of time (explosive muscular fitness)¹⁷⁸. Flexibility is defined as the ability of a specific muscle or muscle group to move freely through a full ROM¹⁷⁸.

Maximal isometric muscular fitness

The reliability of the handgrip test was analyzed in 22 high-quality studies in adults aged 18–64 years^{180, 192, 208–227}. Most of them examined handgrip test reliability using the JAMAR dynamometer as the measurement tool (14 studies, aged range 18–64 years). In all of these studies, the handgrip test demonstrated high test–retest reliability

according to coefficients of correlation (ICCs= 0.90–0.99^{208-210, 214, 215, 218, 219, 221-227}, $r=0.80^{214}$). The test–retest reliability of the handgrip test using other models of dynamometers (i.e., analog and digital TKK²¹⁶, DynEx²²⁴, BTE-Primus^{223, 228}, Grippit²²⁶, Rolyan²²⁷, Lode²¹¹, Smedle¹⁹², MicroFET 4²¹⁷, Lafayette²¹², and Grip-ball²¹³) was investigated showing a high reliability based on ICCs > 0.90. While only a few studies have addressed the reliability of the TKK dynamometer, this method has several benefits over JAMAR, for example: (i) the grip span of the TKK dynamometer can be continuously adjusted for differences in hand size using age- and sex-specific equations, whereas JAMAR has five positions²²⁹⁻²³²; and (ii) the TKK dynamometer does not need regular calibration, while JAMAR requires calibration every year.

The reliability of the back-leg test (using as instruments the Takei and the Baseline back-leg-chest dynamometers) was examined in three high-quality studies²³³⁻²³⁵. In all of them, the back-leg test showed high test–retest reliability based on coefficients of correlation (all r s > 0.8²³⁴ and ICCs > 0.9^{233, 235}) and acceptable CV (<10%)²³⁴ in adults aged 19–46 years. In addition, in one of these studies, the back-leg test returned no significant test–retest differences ($p > 0.05$)²³⁴. The SEMs observed in these studies were low (all < 15%), thus reflecting the low dispersion of measurement errors between test–retest.

Levels of evidence

There was **strong evidence** indicating that: (i) irrespective of the type of dynamometer used, the handgrip test shows high test–retest reliability test provides results with high test–retest reliability in adults aged 19–46 years.

Endurance muscular fitness (dynamic or isometric)

Trunk endurance muscular fitness

Numerous tests have been proposed to measure trunk muscle performance. Many of these tests are isometric trunk holding tests (i.e., the Biering-Sørensen, bridge and trunk flexion sustained tests) used to measure the endurance capacity and fatigability of the trunk muscles. Moreover, dynamic trunk tests such as full-range sit-up (mainly foot held), curl-up or partial curl-up (mainly foot free) or flexion-rotation trunk tests have been proposed.

The reliability of the Biering-Sørensen test to assess the isometric endurance trunk extensor muscles was analyzed in 11 high quality studies in adults aged 18–64 years¹⁹⁵,

²³⁶⁻²⁴⁵; four of them examined a modified version of the Biering-Sørensen test with the participant lying on the floor and held by a partner^{238-240, 243}. In healthy adults, the Biering-Sørensen test based on coefficients of correlation demonstrated a high reliability in three studies (ICCs= 0.91–0.93^{195, 240}, $r = 0.97^{238}$); moderate reliability in three (ICCs= 0.80–0.85^{236, 243, 244}) and low reliability in one ($r = 0.74^{242}$). In adults with low back pain, the Biering-Sørensen test provided results with a high reliability based on coefficients of correlation in four studies (ICCs= 0.91–0.96^{195, 239}, $r = 0.97^{238}$), moderate reliability in two (ICCs= 0.88^{241, 244}) and low reliability in one (ICC= 0.59²⁴⁵). In another reliability study utilizing the Bland–Altman method in adults with low back pain aged 30–58 years, the Biering-Sørensen test showed no systematic significant differences between sessions ($p > 0.05$), a mean difference between test–retest of 0.15 s and limits of agreement between – 1.61 and 1.93 s²³⁷. Overall, the SEMs observed in these studies were low (all < 15%^{195, 239, 242, 243, 245}), thus reflecting the low dispersion of measurement errors between test–retest.

The reliability of the prone and bilateral side bridge tests was analyzed in three and four high-quality studies, respectively^{235, 241, 243, 246-249}. The test–retest reliability of the prone bridge test (supported by the elbow and foot) was considered moderate to high, with ICCs ranging from 0.86 to 0.95 in three studies conducted in adults aged 19–45 years^{235, 248, 249}. The bilateral side bridge test demonstrated a high test–retest reliability (ICCs > 0.90) in one study in adults with low back pain²⁴¹ and moderate to high test–retest reliability (ICCs= 0.80–0.91) in three studies in healthy adults aged 18–57 years^{243, 246, 247}. The prone and bilateral side bridge tests showed the low dispersion of measurement errors between test–retest based on low SEMs (all < 15%^{235, 241, 243, 246, 247, 249}). The bridge test requires minimal inexpensive equipment and is simple to implement in different settings and in a wide range of populations according to its easy held isometric position. The reliability of the sit-up test was analyzed in two high-quality studies^{180, 241}, the partial curl-up test in five high-quality studies^{241, 242, 250-252}, the trunk flexion sustained test in 11 high-quality studies^{237-242, 246, 248, 251, 253, 254} and the flexion-rotation trunk test in two high-quality studies^{255, 256}.

The sit-up test showed moderate test–retest reliability based on ICCs of 0.83 and the low dispersion of measurement errors between test–retest (SEMs < 5%) in adults aged 18–55 years^{180, 241}. The test–retest reliability of the partial curl-up test was considered moderate to high based on coefficients of correlation (ICCs= 0.89–0.98^{241, 250, 251}, $r = 0.93^{252}$) in adults aged 18–57 years. The dispersion of measurement errors between test–retest was low (SEMs < 15%^{241, 250, 251}).

Another study performed in adults aged 35–44 years investigated the test–retest and inter-rater reliability of a modified version of the partial curl-up test (i.e., feet without support), finding low levels of agreement (Kappa's ≤ 0.78)²⁴².

The reliability of the trunk flexion sustained test was analyzed in healthy adults^{240, 242, 246, 248, 251, 253, 254} and adults with low back pain^{237-239, 241} aged 18–59 years. In healthy adults aged 18–57 years, the trunk flexion sustained test demonstrated high test–retest reliability in four studies according to coefficients of correlation (ICCs= 0.93–0.95^{240, 246}, r_s = 0.93–0.95^{238, 242}) and low test–retest reliability in three (ICCs= 0.51–0.71^{248, 251}, r = 0.71²⁵⁴). In a further study, the trunk flexion sustained test showed a mean test–retest difference of 7.9 s (95% CI – 5.7 to 21.5) and a low CV (3.7%); but with low inter-rater reliability (ICC= 0.76) and high dispersion of inter-rater measurement errors (SEM= 19%)²⁵³. In adults with low back pain aged 18–58 years, the trunk flexion sustained test returned results with high test–retest reliability in three studies (two of which using a modified version of this test, i.e., lying on the floor with legs elevated and arms folded across the chest) based on ICCs ranging from 0.91 to 0.97 or Pearson's coefficients of 0.91^{238, 239, 241}, and the low dispersion of measurement errors between test–retest based on SEMs< 15%^{239, 241}. In an additional study, the trunk flexion sustained test showed no significant mean test–retest differences ($p > 0.05$), an absolute mean test–retest difference of – 0.1 s, and limits of agreement between – 3.4 and 3.2 s (for inter-rater reliability)²³⁷. Finally, the flexion-rotation trunk test used to assess trunk flexor-rotator dynamic endurance was analyzed in two studies in young adults aged 19–27 years^{255, 256}. The flexion-rotation trunk test provided results with moderate test–retest reliability based on ICCs $\geq$ 0.83 and SEMs $\leq$ 15%^{255, 256}.

Levels of evidence

There was **strong evidence** indicating that: (i) the Biering-Sørensen test and its modified versions and (ii) the trunk flexion sustained test produce results with high test–retest reliability in healthy adults and adults with low back pain aged 18–64 years; and (iii) the prone and bilateral side bridge tests show moderate test–retest reliability in adults aged 18–64 years. There was **moderate evidence** indicating that: (i) the sit-up and partial curl-up tests offer results with moderate to high test–retest reliability in adults aged 18–57 years; and (ii) the flexion-rotation trunk test provides results with moderate test–retest reliability in adults aged 19–37 years. However, a prolonged familiarization period is needed before testing (at least three practice trials).

Lower-body endurance muscular fitness

Lower-body endurance muscular fitness has been traditionally assessed using sit-to-stand and timed stair-ascent tests. The reliability of the sit-to-stand test was determined in seven high-quality studies^{191, 193, 195, 206, 257-259}. Results of different versions of the sit-to-stand test were expressed as maximum repetitions executed in 60 s²⁰⁶ and 30 s²⁵⁹, and time spent on 5-reps^{191, 193, 195, 257, 258} or 10-reps¹⁹¹ in healthy adults and adults with low back pain¹⁹⁵. The 5-reps sit-to-stand test showed moderate to high test–retest reliability based on coefficients of correlations in three studies (ICCs= 0.83–0.94^{193, 258}, $r = 0.80$ ¹⁹¹) in healthy adults aged 18–64 years. Overall, the SEMs observed in these studies were low (all < 15%^{193, 195, 258}). However, one study showed a low test–retest reliability of the 5-reps sit-to-stand test in healthy adults aged 50–64 years (ICC= 0.72²⁵⁷) and another one in adults with low back pain (ICC= 0.45¹⁹⁵) aged 21–63 years.

The reliability of the timed stair ascent (12 steps) test was analyzed in three high-quality studies^{193, 237, 260}. In one study conducted in adults aged 18–43 years, the timed stair ascent test revealed high reliability (ICC= 0.90)¹⁹³. In another study in adults of mean age 38.7 years, the timed stair ascent test provided low and high reliability in unloaded and loaded versions, respectively (ICC= 0.79 and 0.94). Moreover, mean test–retest differences were 0.17 s and limits of agreement were between –0.68 and 1.04 s in the unloaded version [0.21 s (–0.25 and 0.66 s, for the loaded version)], indicating excellent agreement between tests²⁶⁰. A modified version test (18–20 steps) yielded high reliability in adults with back/neck pain aged 30–58 years and healthy adults aged 27–59 years, with no significant differences between test sessions ($p > 0.05$), a mean test–retest difference of 0.30 s, and limits of agreement of – 3.44 to 4.04 s²³⁷. Other lower-body endurance muscular fitness field tests used in reliability studies were: single-leg squat (timed isometric wall squat (hip/knee 90°)²³⁵, dynamic squat knee flex 90°²⁵³ or 60°²⁴¹ and step-up²⁴³.

Levels of evidence

There was **strong evidence** indicating that: (i) the 5-reps sit-to-stand test produces results with moderate to high test–retest reliability in adults aged 19–64 years.

There was **moderate evidence** indicating that: (i) the timed stair ascent test offers high reliability in adults aged 18–58 years. However, comparisons between different testing protocols should be interpreted with caution. Due to the low number of studies, there was **limited evidence** indicating that: (i) the 30-s sit-to-stand test provides results

with high reliability in healthy adults aged 18–55 years and the 60-s sit-to-stand test shows moderate reliability in adults with low back pain aged 55–64 years; (ii) the modified step-up test (20 kg load) produces high test–retest and inter-rater reliability in adults aged 19–46 years; and (iii) the single-leg squat test in isometric (i.e., timed wall squat hip/knee flex 90°) and dynamic positions (squat knee flex 90° or 60°) generates results with moderate reliability in adults aged 19–59 years. However, the timed isometric wall single-leg squat could be the safest test.

Upper-body endurance muscular fitness

Upper-body endurance muscular fitness has been traditionally assessed with push-up, pull-up, and bent-arm hang tests. The reliability of the push-up test was analyzed in four high-quality studies^{235, 253, 261, 262}. The push-up test demonstrated moderate to high test–retest reliability in two studies in adults aged 18–45 years (ICCs ≥ 0.87 ^{235, 262}) and low test–retest reliability in one study in young adults of mean age 21.2 years (ICCs = 0.25–0.52, Kappa = 0.14–0.52²⁶¹). A modified version of the push-up test provided results with moderate inter-rater reliability (ICC = 0.88) and low test–retest reliability (a mean test–retest difference of 3 repetitions with a CI between 2.1 and 3.9 repetitions and SEM of 22%) in adults aged 25–59 years²⁵³.

The reliability of the pull-up test and its modified version (i.e., in horizontal position) was analyzed in two high quality studies showing a high test–retest reliability based on ICCs ranging from 0.94 to 0.98 and the low dispersion of measurement errors between test–retest based on low SEMs (< 15%)^{243, 262} in adults aged 19–45 years. The reliability of the modified version of the bent-arm hang test (i.e., timed until 90° of elbow extension) was investigated in two high quality studies. In adults aged 19–36 years, the bent-arm hang test provided results with moderate to high test–retest reliability (ICCs = 0.89–0.99^{180, 263}; SEM = 8.5%¹⁸⁰).

Finally, the reliability of the lift and lower task test (as a measure of upper-body function, consisting of lifting a weighted basket) was investigated in two high-quality studies^{206, 260}. One of them showed low reliability based on an ICC of 0.66 in adults aged 55–65 years²⁰⁶, and the other generated results with a high reliability based on an ICC of 0.93 in males of an average age of 37.8 years²⁶⁰.

Levels of evidence

There was **moderate evidence** indicating that: (i) the pull-up test and its modified version show high reliability in adults aged 18–45 years; and (ii) the bent arm hang test and its modified version produce results with moderate and high reliability in adults aged 19–45 years. However, we need to be cautious when comparing the reliability of an original test with that of its modified versions.

Finally, there was inconclusive evidence about the reliability of the push-up test and its modified versions and the lift and reach task test. The pull-up test and the bent-arm hang test are highly influenced by body weight. An alternative test to assess upper-body endurance muscular fitness in individuals with overweight or even a low physical fitness level could be the push-up test or the modified pull-up test (i.e., horizontal). In addition, the bent-knee push-up test could be used instead of the full-body push-up test for adults with a low physical fitness level.

Explosive muscular fitness

Lower and upper-body explosive muscular fitness

Lower-body explosive muscular fitness has been traditionally assessed with jump tests. The reliability of the standing long jump test was assessed in two high-quality studies^{180, 235}. The standing long jump test demonstrated moderate to high test–retest reliability, with ICCs of 0.89 to 0.98 and low dispersion of test–retest measurement errors (SEMs < 15%) in adults aged 18–45 years^{180, 235}. The reliability of the hop sequence test (i.e., single leg hop for distance, 6-m timed hop, triple hop for distance, and crossover hop for distance) was analyzed in two high-quality studies in young adults aged 18–24 years^{264, 265}. Based on coefficients of correlation, this test showed moderate to high test–retest reliability (ICCs = 0.80–0.95). SEM values based on test–retest data were reported as 4.5–7.9 cm for the single leg hop, 0.06–0.13 s for the timed hop, 11.2–23.2 cm for the triple hop, and 15.9–21.2 cm for the crossover hop^{264, 265}. Learning effects were observed for all hops (SEMs between trials 1–5 were < 15% vs. < 2.3% between trials 4–5²⁶⁵). Other field-based physical fitness tests to assess low body explosive muscular fitness used in reliability studies were: countermovement vertical jump²⁶⁶, jump and reach²⁵³, figure-eight hop, up-and-down hop, and t side-to-side hop²⁶⁷. Field-based physical fitness tests to assess trunk and upper-body explosive muscular fitness used in reliability studies were: modified versions of the medicine ball put²⁶⁸ and front/side abdominal power²⁶⁹.

Levels of evidence

There was **moderate evidence** indicating that: (i) the standing long jump and the hop sequence tests produce results with moderate to high test–retest reliability in adults aged 18–45 years. Due to the low number of studies, there was **limited evidence** indicating that: (i) the countermovement vertical jump test provides results with moderate to high reliability in young adults; (ii) the jump and reach test offers high reliability in adults aged 25–59 years; (iii) the figure-eight hop and up-and-down hop tests produce results with high reliability and the side-to-side hop test shows moderate reliability in young adults; (iv) the modified version of the medicine ball put test (i.e., utilizing a 45° incline bench to facilitate the optimal trajectory of the shot-put ball) returns results with high test–retest reliability in young adults; and (v) the front/side abdominal power test demonstrates high test–retest reliability in young adults.

Flexibility

Sit-and-reach and toe-to-touch tests are widely used to assess hamstring and lower back flexibility. The reliability of the sit-and-reach test and its modified versions was investigated in 13 high-quality studies^{180, 241, 270-280}. Several protocols of the sit-and-reach test are described in the scientific literature. Ten studies analyzed the reliability of the sit-and-reach test, one the modified sit-and-reach test, four the back-saver sit-and-reach test, three the modified back-saver sit-and-reach test, three the V sit-and-reach test, and one the chair sit-and-reach test. All sit-and-reach tests have the same testing procedure involving maximal trunk flexion, but with differences with respect to position (unilateral or bilateral) and equipment (standard sit-and-reach box, chair, bench or floor). Nine studies in adults aged 18–64 years concluded that the sit-and-reach test provided high test–retest reliability based on coefficients of correlation (ICCs= 0.91–0.99^{180, 241, 270, 271, 274, 276-278, 280}, $r_s = 0.95–0.98$ ^{272, 273}) and one study found moderate reliability (ICC= 0.83²⁷⁵). In four of these studies, the sit-and-reach test showed the low dispersion of measurement errors between test–retest (SEMs< 11%^{180, 241, 270, 276}). Test–retest measurements in the V sit-and-reach test^{274, 277, 278}, the back-saver sit-and-reach test^{274, 277-279} and the modified back-saver sit-and-reach test^{277, 278, 280} demonstrated moderate to high test–retest agreement (ICCs≥ 0.80) in adults aged 18–64 years. Finally, the chair sit-and-reach test showed moderate to high test–retest reliability based on ICCs of 0.89–0.97²⁸⁰ in adults aged 18–48 years. The reliability of the toe-to-touch test was examined in ten studies (eight of high quality^{242, 276, 278, 280-284}) in adults aged 18–64 years. The toe-to-touch

showed high reliability based on coefficients of correlation (ICCs= 0.93–0.99^{278, 283}, r_s = 0.88–0.97^{242, 282}) in five of these studies and moderate reliability in two (ICCs $\geq$ 0.89^{276, 280}). The SEM was calculated in three of these studies, revealing the low dispersion of measurement errors between test–retest in two (SEMs< 15%^{276, 283}).

The reliability of the trunk lift test used to measure trunk flexibility and endurance muscular fitness was analyzed in two high-quality studies in adults aged 18–28 years^{272, 285}. The trunk lift test demonstrated high test–retest reliability (ICC= 0.96, SEM < 15%²⁸⁵, r = 0.96²⁷²) in young adults. Finally, the reliability of the side-bending of the trunk test used to assess lateral trunk flexibility was investigated in two studies in adults aged 25–59 years^{242, 253}. One study found high coefficients of correlation in test–retest and inter-rater reliability (r_s = 0.82–0.88), but significant differences between test–retest (p < 0.001)²⁴². In the other study, the side-bending of the trunk test demonstrated high inter-rater (ICC= 0.92) and test–retest reliability (mean test–retest difference of – 0.5 cm and CI of – 1.3–0.3 cm and a CV of 4.7%)²⁵³. In both studies the side-bending of the trunk test showed the low dispersion of measurement errors between test–retest (SEMs< 5%^{242, 253}). Other musculoskeletal field-based physical fitness tests to assess trunk or upper-body flexibility used in reliability studies were: hand-behind-back²⁸⁶ and shoulder–neck mobility²⁵³.

Levels of evidence

Overall, there was **strong evidence** indicating that: (i) the sit-and-reach test and its modified versions and the toe-to-touch test offer high test–retest reliability in adults aged 18–64 years. There was **moderate evidence** indicating that: (i) the trunk lift test produces results with high reliability in adults aged 18–28 years. Due to the low number of studies, there was **limited evidence** indicating that: (i) the hand behind- back test shows high reliability; and (ii) the shoulder–neck mobility test provides results with low agreement.

Finally, there was **inconclusive evidence** regarding the level of reliability of the side-bending test.

Motor fitness

Motor fitness refers to any component of physical fitness that enables a person to successfully perform a particular motor task, game, or activity²⁸⁷. Specific motor fitness components include agility, balance, coordination, power, reaction time, and speed²⁸⁸.

The reliability of tests assessing motor fitness was investigated in 22 studies (all classified as high quality).

Balance

The reliability of static and dynamic balance tests was analyzed in nine studies^{180, 191, 206, 253, 257, 289-292}. Among the static balance tests, reliability studies analyzed the single-leg stand test using different protocols (i.e., eyes open, closed or with the head turned on the floor or on a bar) and also with the variations feet together, semi-tandem, and tandem stand^{180, 191, 206, 253, 257, 289}. The single leg stand test (eyes open) demonstrated low test–retest reliability in three studies (ICCs= 0.60–0.73; SEMs= 10–18%^{285, 292}, $r = 0.69$ ¹⁹⁸) in adults aged 19–64 years. A fourth study analyzed the single-leg stand test with eyes open, closed, or head turned in adults aged 25–59 years, and obtained results with low inter-rater reliability in all tests (ICCs= 0.18–0.76 and SEMs= 7–38%) and high variability between test–retest for standing on one leg with eyes open (mean differences between testing days of 3.7 s with a confidence interval of – 2.2 to 9.6 s), eyes closed (mean differences of 0.6 s, from – 1.6 to 2.8 s) and head turned (mean differences of 2.6 s, from – 0.7 to 5.9 s), and low CVs ranging from 11 to 5%²⁸⁰. However, the single-leg stand test (on a narrow bar) showed high reliability based on ICCs> 0.90 and SEMs≤ 5.5% in adults aged 36–64 years²⁸⁹. The tests semi-tandem stand with eyes open and tandem stand with eyes open and closed demonstrated low test–retest reliability according to ICCs of 0.27–0.58, but the dispersion of test–retest measurement errors was low based on SEMs≤ 8.5% in adults aged 50–64 years²⁶¹.

The Romberg test (eyes open) showed high reliability based on a percent agreement for the parallel stance of 94.7% and low reliability for the semi-tandem and tandem positions (< 75%) in adults aged 55–70 years²⁶⁶. The reliability of the star excursion balance test was analyzed in three studies in adults aged 18–50 years^{290, 291, 293}. Overall, this test demonstrated a moderate test–retest reliability for the dominant leg (ICCs> 0.80) and low test–retest reliability for the non-dominant leg (ICCs< 0.75). In two of these studies, this test showed the low dispersion of measurement errors between test–retest (SEMs< 5%^{290, 293}). Other dynamic balance field-based physical fitness tests used in reliability studies were: functional balance (i.e., walk forwards and backwards across a wooden plank)²⁹² and tandem walking (forwards and backwards)²⁸⁹.

Levels of evidence

There was **strong evidence** indicating that: (i) the single-leg stand test and its modified versions show low reliability in adults aged 19–64 years. There was **limited evidence**, mainly due to a limited number of studies, indicating that: (i) the tandem walking and functional balance tests provide results showing moderate to high agreement in adults aged 33–64 years. Finally, there was **inconclusive evidence** regarding the level of reliability of the parallel, semi-tandem and tandem stand tests and star excursion balance test.

Gait speed-agility

The TUG is a multidimensional test that measures mobility skills (agility), combining gait speed, balance, and functional capacity²⁹⁴. The reliability of this test was investigated in two high-quality studies. The TUG (3 m) returned high test–retest reliability based on coefficients of correlation (ICCs= 0.90–0.98) and the low dispersion of measurement errors between test–retest (SEMs≤ 6%) in healthy adults and adults with low back pain aged 21–64 years^{195, 295}. The T-test is a measure of four directional agility and body control that assesses the ability to change direction quickly and maintain balance without reducing speed. Its reliability was analyzed in three studies in adults aged 18–39 years. The T-test^{265, 296} and its modified version²⁹⁷ (i.e., reducing the total distance covered) provided results with moderate to high reliability with ICCs ranging between 0.82 and 0.98 and a low dispersion of measurement errors between test–retest (SEMs≤ 3%)^{255, 296, 297}. In addition, the modified T-test demonstrated no systematic differences between test sessions ($p > 0.05$), a mean test–retest difference of 0.03 s, a limit of agreement of ± 0.33 s for females and ± 0.37 s for males, and low CVs ($\leq 2.7\%$)²⁹⁷. The reliability of the hexagon agility test, a measure of agility and foot quickness involving balance and coordination, was analyzed in two studies in adults aged 19–30 years^{267, 298}. The hexagon agility test and its modified version (i.e., hopping single-legged instead of double- legged) showed moderate to high test–retest reliability (ICCs= 0.84–0.94) and the low dispersion of measurement errors between test–retest (SEMs≤ 0.7 s²⁵⁷ ranging from 1 to 9.4% between sessions²⁹⁸). Thus, a practice trial is recommended prior to recording scores to minimize any possible learning effect. The reliability of the gait speed test was investigated in five studies. The gait speed test (4 m/3 m) demonstrated low reliability in two studies in adults aged 35–64 years (ICC= 0.56; SEM= 2%²⁵⁷, $r = 0.57$ ¹⁹¹). However, in one study in healthy adults and adults with low back pain aged 21–63 years,

the gait speed test [15.2 m (50 ft) fast walk or preferred speed walk] showed high reliability (ICCs= 0.91–0.99; SEMs $\leq$ 4.6%¹⁹⁵). In another study also in healthy adults and those with low/high back pain aged 27–59 years, the gait speed test (2 $\times$ 20-m walk, with and without load) showed no significant mean test–retest differences ($p > 0.05$)²³⁷. Finally, in one study in adults of mean age 39 years, the gait speed test (50 m fast walk with and without load) returned results showing a high reliability based on ICCs of 0.98–0.99²⁶⁰. Other gait speed and agility field-based physical fitness tests used in reliability studies were: 4-square step¹⁹³, 10 $\times$ 5 shuttle run¹⁸⁰, pro-agility²³⁵, agility track²⁹⁹, zig-zag run²⁹⁷, Illinois agility, Edgren side step²⁹⁶, and plate tapping²⁹¹.

Levels of evidence

There was **strong evidence** indicating that: (i) the T-test shows moderate to high test–retest and interrater reliability in adults aged 18–39 years. There was **moderate evidence** indicating that: (i) the TUG demonstrates high reproducibility in adults aged 19–64 years with or without low back pain; and (ii) the hexagon agility test and its modified version show moderate to high reliability in adults aged 19–64 years. A large number of other motor fitness tests provided **limited evidence**, mainly due to a low number of studies: (i) the 10 $\times$ 5 shuttle run and zigzag tests provide results with moderate and high reliability, respectively, in young adults; (ii) the agility track and pro agility tests show high test–retest reliability in adults aged 28–55 years; (iii) the 4-square step test returned results with moderate test–retest reliability in adults aged 19–43 years; (iv) the Edgren side step and Illinois agility tests show low test–retest reliability but high interrater reliability in adults aged 19–39 years; and (v) the plate tapping test provides results with low reliable values in young adults of mean age 19.5 years. Finally, there was **inconclusive evidence** regarding the level of reliability of the gait speed tests. The fact that different versions of the gait speed test were used involving sample groups with different characteristics makes it difficult to compare their reliability data.

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

The aim of the present study was to analyze the **criterion-related validity** and the **reliability** of the 2-km walk test and the 20-m SRT for evaluating cardiorespiratory fitness in adult population, according to sex, age, and physical activity level. The results showed that both tests are valid and reliable.

Our study included a homogeneously distributed sample by sex, age, and physical activity level, to analyse whether the criterion-related validity and reliability of the 2-km walk test and the 20-m SRT are dependent of these variables. Overall, we found no significant differences. Hence, the criterion-related validity and reliability of both tests were not determined by sex, age, or physical activity level.

Criterion-related validity

The main results showed a high association between laboratory and field-based cardiorespiratory fitness tests, and between measured and estimated VO_{2max} by Oja's³⁰⁰ ($R^2 = 0.61$, $p < 0.01$) and Leger's⁶¹ ($R^2 = 0.77$, $p < 0.01$) equations. In fact, Bland-Altman analysis also indicated that the 2-km walk and the 20-m SRT test **are valid** to estimate cardiorespiratory fitness by Oja's and Leger's equations, showing a mean of differences close to 0 and narrow LoA, regardless of sex, age, and physical activity level.

The 2-km walk test is considered a user-friendly submaximal cardiorespiratory fitness test, since it allows the assessment of people with low physical fitness level or who are unable to run. In fact, the Oja's equation³⁰⁰ is valid to assess VO_{2max} , in adults aged 20-65 years ($R^2 = 0.73$ to 0.75 , $p < 0.05$)³⁰⁰, with overweight/obesity ($R^2 = 0.56$ to 0.59 , $p < 0.05$)³⁰¹, and with low or moderate fitness levels ($R^2 = 0.30$ to 0.64 , $p < 0.05$), but not in adults with a high fitness level ($R^2 = 0.27$, $p < 0.05$)³⁰².

Oja et al.³⁰⁰ found that the total time performed in the 2-km walk test was highly correlated with measured VO_{2max} in females ($r = 0.74$, $p < 0.001$), but moderately in males ($r = 0.58$, $p < 0.001$). These results are very similar to those found in our sample ($r = 0.67$ and $r = 0.51$ for females and males, respectively; both, $p < 0.001$). Moreover, in our study, that association was also higher in adults aged 50-64 years ($r = 0.71$, $p < 0.001$) and non-active participants ($r = 0.66$, $p < 0.001$).

In the prediction model, Oja et al. found that the total time, heart rate, age and weight predicted 66-76% ($SEE = 6.2$ to 3.0) of the variance in VO_{2max} , and 73-75% ($SEE =$

3.3 to 5.1) when replacing weight by BMI³⁰⁰. In this sense, we tried to develop a more accurate equation to estimate VO_{2max} considering total time, heart rate, sex, age, physical activity level, and anthropometric variables. The total time represented the 42% of explained variance, reaching 47% when heart rate was added. Finally, the explained variance increased until 70% when total time, heart rate, age, physical activity level and %BF were included. We also replaced %BF by BMI, sum tricipital+subscapular SKF, WC, HC and lean mass, obtaining similar results. Based on this, our results did not improve the prediction equation proposed by Oja et al.³⁰⁰. Accordingly, the Oja's equation seems more feasible to estimate VO_{2max} , since it includes a lower number of prediction variables.

Finally, Bland-Altman plots support the criterion-related validity of the 2-km walk test, with differences nearly 0 and narrow LoA, especially for non-active adults [-0.19 ml* kg⁻¹ * min⁻¹ (95% LoA= -13.45 to 13.06, p< 0.001), d= -0.016]. One explanation could be that the cardiovascular system of active people is less able to be stressed enough by regular walking to produce an accurate VO_{2max} prediction (i.e., underprediction)³⁰³. Moreover, the heteroscedasticity analyses indicate that the higher VO_{2max} (i.e., the fitter) the worse the degree of agreement in the VO_{2max} prediction.

On the other hand, the 20-m SRT is an incremental maximal cardiorespiratory test, which has been found **valid** to estimate VO_{2max} by Leger's equation⁶¹ in adults aged 18-64 years ($R^2= 0.81$, p< 0.05). In this sense, we obtained similar results ($R^2= 0.78$, p< 0.01) when testing the Leger's equation in our study sample. These results remained the same after analyzing them by sex, age and physical activity levels. Consequently, these variables seem not to affect the criterion-related validity of the Leger's equation.

There were different attempts to improve the Leger's equation⁶¹. In fact, we also tried to develop a more accurate equation to estimate VO_{2max} considering final stage, sex, age, physical activity, and anthropometric variables. We found that the final stage represented the 78% of explained variance for measured VO_{2max} (SEE= 3.841 ml* kg⁻¹ * min⁻¹). When physical activity, was included, the explained variance only increased 1%. Our final model included final stage, physical activity level and %BF, explaining the 80% of explained variance for measured VO_{2max} , being still lower than those reported by Leger's equation (i.e., 81% of explained variance)⁶¹. Furthermore, replacing %BF by BMI, or sum tricipital+subscapular SKF, WC, HC and lean mass, yielded similar results. Hence, the Leger's equation seemed to be the most precise and feasible, since only final stage is needed.

Finally, Bland-Altman plots nearly 0 and narrow LoA, confirm the criterion-related validity of the 20-m SRT in the whole sample, as well as in the different groups of sex, age and physical activity level.

Mayorga et al.³⁰⁴ reported that the protocol used by Leger et al.⁶¹, only including the final stage reached, was the protocol that presented a greater criterion-related validity with measured $\text{VO}_{2\text{max}}$ ($r = 0.84$; $p < 0.05$), than the EUROFIT protocol ($r = 0.73$, $p < 0.05$), Queen's University Belfast ($r = 0.71$, $p < 0.05$), and Dong-Ho ($r = 0.66$, $p < 0.05$). Moreover, sex and maximum oxygen uptake level did not seem to affect the criterion-related validity which is in line with our results.

A meta-analysis³⁰⁵ highlights that the 20-m SRT has greater criterion-related validity than the distance tests in adults (such as the 2-km walk test) which concurs with our results. Due to the easiness in regulating the pace through an acoustic signal (where participants cannot choose their own pace) and the relatively short duration of this maximal test, the 20-m SRT seems likely to reduce the influence of psychological factors (e.g., self-motivation and monotonous) that may affect the performance, and thus the validity/reliability, when comparing with distance tests. Therefore, scientists and practitioners could use the 20-m SRT over the 2-km walk test, when no known physical impairments are present. Otherwise, the 2-km walk test is also a useful alternative to estimate cardiorespiratory fitness (i.e., participants with low physical fitness level or who is unable to run).

Reliability

The main results showed a **good reproducibility** of the 2-km walk test and 20-m SRT, as well as the estimated $\text{VO}_{2\text{max}}$ by Oja's and Leger's equations.

We found no significant difference between testing sessions in estimated $\text{VO}_{2\text{max}}$ in both tests ($p > 0.05$). However, we found significant differences between 2-km walk test and 20-m SRT in terms of total time and final stage, respectively; although we cannot translate it as a *real* statistical difference, since these results are based on minimal performance changes. For instance, the mean time to complete the 2-km walk test was of 973 seconds, while the T1 and T2 difference was of -1.48 seconds. Likewise, the mean final stage reach in the 20-m SRT was 5, while the T1 and T2 difference was of 0.04 stages. Moreover, the effect size of the mean differences was small (all, Cohen $d < 0.016$)³⁰⁶.

The reliability of the 2-km walk test and 20-m SRT were considered high based on coefficients of correlation (ICCs > 0.90) in both tests for the whole sample. Overall, the reliability of both tests did not change according to sex, age or physical activity level. Although all of them reported a high reproducibility (ICC = 0.93 to 0.99), the 20-m SRT result and its estimated $\text{VO}_{2\text{max}}$ by Leger's equation showed even higher reliability (all, ICC = 0.99) than the estimated $\text{VO}_{2\text{max}}$ by Oja's equation (ICC = 0.93 to 0.96). These results are also supported by low error values (i.e., MSE, RMSE, %Error, %SEM, %CV and SEE), indicating good data accuracy, especially for the 20-m SRT.

We have recently conducted a reliability systematic review of field-based physical fitness tests in adults³⁰⁷, and we found that the reliability of the 2-km walk test has been only previously analyzed in a sample of female and male adults aged 30–55 years³⁰². They found a T1 and T2 difference of -0.9 ± 4.4 (95% LoA = -2.7 to $4.5 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p < 0.05$) for females, and -2.2 ± 3.5 (95% LoA = -2.8 to $6.6 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, $p < 0.05$) for males, with high correlation coefficients (ICCs = 0.88 to 0.91). In our study, we found no differences when comparing by sex, and taking similar age groups (without sex differences), we observed similar results to those reported by Laukkanen et al.³⁰², Although with higher ICCs.

Finally, Bland-Altman plots support the reliability of the 2-km walk test, with differences nearly 0 and narrow LoA. Moreover, results from the heteroscedasticity analysis in the total time of the 2-km walk test, as well as in the estimated $\text{VO}_{2\text{max}}$ by Oja's equation, indicate that the variability of these measurements could be greater when the participants performed better (i.e., the fitter).

Regarding the 20-m SRT, in the aforementioned systematic review, we found that it is a reliable test for young-adults (ICCs = 0.93-0.96, SEMs < 15%)³⁰⁷. In the present study, similar results were found for the whole sample (ICC = 0.99, SEM < 1%), as well as in adults aged 18-34 years (ICC = 0.99, SEM < 2%). Moreover, we found a T1 and T2 difference of $0.03 \pm 0.04 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LoA = -1.51 to 1.58 , $p = 0.086$) for the whole sample, as well as $0.00 \pm 0.07 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (95% LoA = -1.67 to 1.67 , $p = 0.064$) in adults aged 18-34 years, being even lower than those reported in that systematic review³⁰⁷. Furthermore, results from Bland-Altman plots support the reliability of the 20-m SRT test, with differences nearly 0 and narrow LoA.

Nevertheless, some studies included in this systematic review analyzed reliability based only on the Pearson correlation coefficient which, despite being a common method to examine reliability, its use without other statistical support seems to be

inappropriate³⁰⁸. In addition, the sample of these studies mainly included participants aged <45 years, being difficult to extrapolate that results to adults over this age.

In general terms, the Leger's equation (ICCs= 0.99; RMSE= 0.62-0.84; %CV= 1.57; SEE= 0.61-0.85) seems to be slightly more reliable than Oja's equation (ICCs= 0.93-0.96; RMSE= 2.37-5.13; %CV= 7.73; SEE= 2.27-4.84), to estimate $\text{VO}_{2\text{max}}$, regardless of sex, age or physical activity level. This fact may be explained by the difficulty in developing an appropriate pace during the 2-km walk test, starting too fast, so that the participants are unable to maintain their speed throughout the test; or too slow, increasing their speed at the end of the test (which may also mean an unexpected incremented heart rate at the end of the test). On the other hand, regarding the 20-m SRT, it is possible that following an acoustic signal could be easier for self-pace regulation.

Overall, both tests can be considered **reliable**. It was not possible to compare the results of the present study in terms of measurement errors, as none were available in the current literature. Nevertheless, high ICC values and low CV and SEM values suggest high levels of reliability and reproducibility^{309, 310}, regardless of the characteristics of individuals.

Study VI: The standing long jump test.

The aim of the present study was to analyze the **criterion-related validity** and the **reliability** of the SLJ test to assess lower-body explosive muscular fitness in adult population, according to sex, age, and level of physical activity.

Our study included a homogeneously distributed sample by sex, age, and, physical activity levels, to analyse whether the criterion-related validity and reliability are dependent of these variables. The main findings suggest that the SLJ test is valid and reliable to assess lower-body explosive muscular fitness in adults aged 18-64 years. Moreover, its validity is not determined by age or physical activity levels. Regarding reliability, it seems to depend on a prior familiarization period, but not on sex, age, or physical activity levels.

Criterion-related validity

The SLJ test is proposed to evaluate lower-body explosive muscular fitness related to health in children³¹¹, and adolescents³¹², given its proven validity and reliability, while in adult population there is lack of validity to date³¹³.

The SLJ test is a multi-joint movement that involves taking off and landing with both feet in order to achieve maximum horizontal distance³¹⁴. The performance of the SLJ test is influenced by many variables, such as the maximal force and velocity of the jump at the take off^{315, 316}, as well as anthropometric variables³¹⁴. These findings are in line with our results. The SLJ distance was correlated with the force platform laboratory test, and with the isokinetic laboratory test ($r = 0.259$ to 0.826 ; all, $p < 0.001$), especially with the platform variables, such as maximal horizontal power ($r = 0.826$), maximal horizontal force ($r = 0.815$), jump take off velocity ($r = 0.768$), and horizontal jump take off velocity ($r = 0.766$). Moreover, %BF, sex, height, tricipital skinfold and sum triceps+subscapular skinfold had a moderate association with SJL ($r = -0.665$ to -0.495 ; all, $p < 0.001$).

The fact that isokinetic variables were moderately associated with SLJ test, in contrast with the stronger associations found with the force platform variables may be partially explained by the different nature of the strength performance measured by both devices. The SLJ test is performed in two scenarios, laboratory and field-based performance, being reproduced the same exercise in both conditions, with the only particularity of the absence of the force platform in the SLJ field-based test. Thus, is not surprising that strong correlations were found, since both of them are measuring lower-

body explosive muscular fitness in an open kinetic chain. On the other hand, the isokinetic dynamometer evaluates lower-body flexion and extension single-joint movement, with a closed kinetic chain, in a sitting position, isolating movements. In this sense, in a study conducted in collegiate football athletes, using a similar isokinetic 60/60 protocol that the one used in our study, found even low association between SLJ test and isokinetic variables ($r = 0.28$ to 0.36 , all $p < 0.05$)³¹⁷. Therefore, isokinetic devices may be more advantageous for evaluating closed kinetic chains, as its validity for evaluating performance in multi-joint explosive exercises, such as SLJ, seems limited. Alternatively, SLJ testing via a force platform may be more advisable.

In young population (6-17 years), the SLJ test was strongly associated with other lower-body explosive muscular fitness tests (such as vertical jump, squat jump, and countermovement jump), being considered a general index of muscular fitness in youth³¹⁸.

In children (6-12 years), Fernandez-Santos et al.³¹⁹ evaluated different lower-body explosive muscular tests (i.e., SLJ, squat jump, countermovement jump, and Abalakov test), compared with 1 repetition maximum (RM) leg extension as laboratory test. They found that the SLJ was the most closely associated test ($R^2 = 0.70$, $p < 0.001$), proposing the SLJ test as a useful tool to assess lower-body explosive muscular fitness in this population. Likewise, in children aged 7-12 years, Milliken et al.³²⁰ analysed the relationship between 1RM leg press and the vertical jump and SLJ tests. They concluded that the SLJ test would be considered a good predictor of lower-body explosive muscular fitness (better even than the vertical jump test), taking into account the BMI.

When analyzing stepwise regressions, we found that maximal horizontal power, sex, %BF, maximal horizontal force and lean mass accounted for 78% of the SLJ distance variance (all, $p < 0.05$). The SLJ follows a parabolic flight³¹⁶, hence appears reasonable that the variables related to the horizontal power path³¹⁵ were the most predictable of SLJ distance. In fact, in our regression model the maximal horizontal power explained 68% of the variance in the SLJ distance. Additionally, maximal horizontal force increased the variance in 0.5%. In this sense, the fact that maximal horizontal power strongly influenced the SLJ distance may be explained since “power” is defined as the product of force and velocity³²¹ or, in other words, as the ability of the lower limbs (mainly) to apply horizontal force with the highest speed. Therefore, the greater ability to generate power at the take off the longer jump distance.

We observed that sex explained 7% of the variance in the SLJ distance. Physiologically, sex accounts for differences in power performance between females and males³²². Factors contributing to these differences are body composition, strength (due to muscle fiber type and muscle recruitment), and neuromuscular or glycolytic performance³²². Among body composition, lean mass is one of the main components influencing muscular fitness, and it seems that even controlling by lean mass these differences still hold³²². Therefore, our results may confirm these sex differences, in terms of performance values (e.g., SLJ= 133 cm in females vs 178 cm in males, $p < 0.001$; maximal isometric peak torque= 113 Nm in females vs 180 Nm in males, $p < 0.001$), and as predicting variable for SLJ distance. In addition, our results also confirm the partial influence of body composition showed in previous studies conducted in children^{319, 320}, with %BF and lean mass together accounting for 2.6% of the explained variance in the SLJ distance. These results may partially been explained by the relationship between body composition and muscular fitness production³²³.

On the other hand, age and physical activity levels showed no significant associations with SLJ distance. In children, central nervous system maturation may play a role in muscular fitness performance, affecting the SLJ test³¹⁴, whereby age or physical activity may be involved. In adults, maturational status on SLJ performance is not relevant. As a consequence, it is possible that this lack of associations could be attributed to other factors, such as the jumping technique^{315, 316}, since most of our participants have no previous experience in performing this test. In fact, jumping skills in expert jumpers may increase SLJ performance³¹⁶.

Therefore, for all the aforementioned, we suggest that the SLJ is a **valid** test to be included as a fitness health tool, in the adult population. In addition, lower-body muscular fitness has a close relationship with health and mortality in adults³²⁴. Although more studies analyzing the criterion-related validity of this test are still needed, the SLJ test shows promise.

Reliability

The mean difference between the SLJ test T1 and T2 presented an overestimation of 12.14 cm between the second trial compared to the first trial ($p < 0.001$), with medium magnitude between these differences ($d = 0.31$). Moreover, the Bland-Altman plot showed a systematic error between trials far from 0 and large LoA. This may be due to an existing learning effect (systematic bias) among these tests when repeated measurements are

performed. These results remained the same regardless of sex, age groups or physical activity levels.

To date, a recent systematic review³⁰⁷ concluded, with moderate evidence (derived from two studies), that the SLJ test has moderate-high reliability (ICC= 0.89-0.98), with a low dispersion of T1 and T2 measurement errors (SEMs <15%; CV= 3%) in adults aged 18-45 years. Our results concur with it in terms of ICCs, showing even a lower SEM (ICC= 0.94; SEM= 8%), also including adults aged >45 years. In addition, our study shows a bigger picture regarding the reliability of the SLJ taking into account diverse reasons: first, we included a higher sample size (410 participants in our study, versus 179 and 95 participants in the others); second, we included a well-balanced sex sample distribution (i.e., 50% males in our study, versus 108 females and 166 males in the others); third, we included the full adult age range (i.e., from 18 to 64 years old in our study, versus from 18 to 45 years old in the others); fourth, in our study balanced physical active levels were included, while they did not take it into account; fifth, all these characteristics (i.e., sex, age, and physical activity levels) were analyzed to evaluate the possible influence on reliability terms. In view of this, we could conclude that the SLJ test shows high reliability in adults aged 18-64 years old, independently of sex, age, and physical activity levels.

In young population (6-18 years), España-Romero et al.³²⁵ also observed a significant T1 and T2 difference in the SLJ distance (4 cm, $p < 0.005$), suggesting learning effects on the test. Nevertheless, Fernandez-Santos et al.³¹⁹ found a high reproducibility (ICC= 0.94), without significant T1 and T2 difference in the SLJ in children (6-12 years). However, in preschoolers³²⁶, significant differences were found in a first analysis attempt (-7.31 cm, $p < 0.001$), improving its reliability after some methodological adaptations (-2 cm, $p < 0.001$).

To confirm or contrast whether a learning effect might exist, we decided to replicate the SLJ test protocol in a sub-sample of 66 adults, including a learning period. Since the systematic error was reduced substantially (i.e., from 12 to 0.5 cm; %CV from 7 to 0.5), it can be confirmed that **the SLJ test needs a familiarization period to be reliable**. Considering that the SLJ technique can be a determinant factor in the performance (e.g., coordination between arms and legs, arm swings or balance)^{315, 316, 327}, future studies should be conducted including a familiarization period with more than three trials, to observe if similar results in terms of reliability are found.

LIMITATIONS AND STRENGTHS

LIMITATIONS AND STRENGTHS

The studies included in this International Doctoral Thesis present some general and specific limitations that should be taken into account when interpreting their results. Likewise, this International Doctoral Thesis also presents general and specific strengths, which are summarized below.

General limitations

I. In the systematic reviews, meta-analyses could not be conducted owing to methodological heterogeneity such as different study designs, variability in adjustments for potential risk factors, and different statistical estimates (*studies I-IV*).

II. In the systematic reviews, only apparently healthy population at baseline was included, so it is possible that some tests may have criterion-related validity/predictive validity or reliability for some specific pathologies or particular health conditions (*studies I-IV*).

General strengths

I. In the systematic reviews, the large number of included and analyzed original studies and systematic reviews/meta-analyses, which summarize the most up-dated evidence for the exiting field-based physical fitness test in adult population, suggesting the most useful test to assess diverse components of physical fitness regarding specific characteristics, such as sex, age, weight status or physical activity/fitness levels (*studies I-IV*).

II. In the methodological studies, the relatively large sample, as well as the homogeneous distribution of the sample by sex, age, and physical activity levels (*studies V-VI*).

III. In the methodological studies, the analysis of several anthropometric variables, including height, weight, %BF, lean mass, triceps and subscapular SKF, and HC and WC to detect which may result in a more explanatory variable for the validity equations (*studies V-VI*).

IV. In the methodological studies, although ICC, Bland-Altman, SEM and CV are the most common statistic used to report reliability in sport medicine, different measurement errors to a more complete interpretation of reliability were also included (*studies V-VI*).

Specific limitations and strengths

Chapter I: Predictive Validity of Field-Based Physical Fitness Physical Test in Adults: A Systematic Review (*study I*).

Specific limitations

I. The impossibility of finding a sample of adults under 35 years of age. This fact may be explained by the fact that ageing deterioration is the main reason of multiple adverse health conditions such as falls, frailty, cognitive decline, etc.². Therefore, the clinical screening of that younger population is not yet set.

II. The results of the present systematic review should be interpreted with caution due to the variety of field-based tests used to assess motor fitness and flexibility, the heterogeneity in statistics reporting risk prediction, and protocols that make difficult to directly compare results across studies and that could somehow partially change the results presented here.

III. The heterogeneity among follow-up lengths limited direct comparison between studies as different follow-up timeframes were employed (e.g., one year vs. 25 years).

IV. The inconsistent utilization of cut-off scores for the baseline measures (in both, motor fitness and flexibility tests and health outcomes) across selected studies.

Specific strengths

I. The major strength of our study is the wide range of health outcomes introduced in the research process, including the biological and psychological spheres, providing a comprehensive overview of health, in adults and older adults.

II. The inclusion of those health outcomes with scarce literature, such as fear of falling, frailty, unplanned hospitalization, depression, or well-being.

III. The inclusion of data from near 375300 apparently healthy adults and older adults, with a mean follow-up greater than eighth years.

Chapter II. Criterion-Related Validity of Field-Based Physical Fitness Test, and Methods and Equations for Body Composition estimation in Adults: Systematic Reviews (*studies II-III*).

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

Specific limitations

I. Despite we have included different weight statuses, race-ethnicities, or activity levels, our data cannot be applied to compromised populations that may explain some of the discrepancies between our findings and those of other studies (e.g., athletes have less body fat than do non-athletes).

II. Although developing specific equations would be ideal, the more specific the population, the less general application an equation will have. Moreover, the more specific equation, the more difficult to implement that equation in epidemiologic, clinical, or field setting assessment.

III. Although an effort was made to classify the evidence based on different races-ethnicities, following data reported or study site, many people could identify with more than one race-ethnicity; therefore, the proposed categories should not be considered as absolute or viewed in isolation.

IV. It is known that direct analysis of body composition can only be performed by chemical analysis of cadavers³²⁸, and therefore, other methods serve as only indirect measurements of %BF and all techniques suffer from some type of error. In fact, estimation equations imply theoretical assumptions of body composition measurement³²⁸.

Specific strengths

I. The large sample size sorted by sex, the wide age range of the sample, the inclusion of different races-ethnics, different weight statuses, and the number of included field-based methods and derived equation studies.

II. The validity of existing body composition field-based methods was investigated using the most precise or accurate criterion methods (i.e., “gold standards”) for the populations tested.

Chapter IV. Methodological Studies of Field-Based Physical Fitness Test in Adults: Criterion-Related Validity and Reliability (*studies V-VI*).

Study V: The 2-km walk test and the 20-m shuttle run test.

Specific limitations

I. The lack of control of factors that can affect performance, such as genetics or experience and running economy.

III. Although we maintained a high level of motivation throughout the participants' test performance, psychological issues as the discomfort of strenuous effort, self-motivation, and interest span for monotonous tasks may have had some uncontrolled effect on our results.

Study VI: The standing long jump test.

Specific limitations

I. The first SLJ trial (i.e., *T1*) was performed on a force platform (laboratory test), while the second SLJ trial (i.e., *T2*) was performed on a resin sports flooring (field test), which could cause changes in the sport shoes grip during take off or landing, affecting the total distance of the test. However, anti-slip tapes were placed in the landing area to ensure a safe and slip-free landing.

Specific strengths

I. The wide number of laboratory variables analyzed, using both an isokinetic device and a force platform.

II. Since the learning effect was not initially controlled, being able to determine the results of the study, a replication study was carried out, demonstrating that a familiarization period is mandatory to provide reliability in the SLJ test.

CONCLUSIONS

[CONCLUSIONES]

CONCLUSIONS

Overall, the findings derived from the present International Doctoral Thesis suggest that:

Study I. Predictive Validity of Field-Based Physical Fitness Test in Adults: A Systematic Review.

I. Strong evidence, derived from a systematic review, indicates that:

Slower **gait speed** predicts falls, cognitive decline and impairment (including development of dementia and Alzheimer's disease), mobility disability, disability in IADL, CVD risk, all-cause mortality, and risk of institutionalization or hospitalization. Its use is suggested especially in adults over 35 years old.

Impaired **balance** predicts falls, disability in IADL or mobility disability, and all-cause mortality. Its use is suggested especially in adults over 40 years old.

Worse **TUG** performance predicts falls and fear of falling in adults over 40 years old, especially in females.

These results provide further justification to integrate the measurement of motor fitness, mainly gait speed and balance, in the prognostic assessment tools as indicators of health outcomes (such as falls risk, cognitive impairment, disability, CVD risk, all-cause mortality, and hospitalization) in both adults and older adults. Identifying these risk factors will allow for early intervention based on adequate physical fitness levels and potentially decrease morbidity and mortality over lifetime, at the same time increasing their QoL.

In spite of the limited evidence regarding **flexibility**, our results suggest that flexibility testing may help detect older adults at risk of falling, suffering from disability, or expecting all-cause mortality.

Study II: Criterion-Related Validity of Field-Based Physical Fitness Tests in Adults.

II. Strong evidence, derived from a systematic review, indicates that:

To assess **cardiorespiratory fitness**, the 20-m SRT (using Leger's equation), best assessed cardiorespiratory fitness. Alternatively, the 1.5-mile, 12-minute run/walk and YMCA step tests were other cardiorespiratory testing options. When low-level cardiorespiratory fitness existed, or if running was not possible, the 2-km (using Oja's equation) or 6-minute walk tests were appropriate alternatives.

To assess **muscular fitness**, the handgrip test with the elbow extended and with the grip span adapted to the individual's hand size (using the TKK dynamometer), offers a valid means to assess hand maximal isometric muscular fitness. The Biering-Sørensen test estimated the endurance muscular fitness of hip and back muscles. Limited evidence (only one study) supported the prone bridging and Sargent jump tests as abdominal endurance muscular fitness and lower body explosive muscular fitness surrogate markers, respectively.

To assess **flexibility**, the sit-and-reach test (and its different versions), and that the toe-to-touch tests are not valid.

Limited evidence about the criterion-related validity of **motor fitness** exists.

When there are problems of space and time, as in clinical settings, the YMCA step and the handgrip tests are good alternatives for assessing cardiorespiratory fitness and isometric muscular fitness, respectively.

Study III: Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults.

III. Strong evidence, derived from a systematic review, indicates that:

The **field-based methods**, WC, BAI, and BMI are valid indicators of body adiposity in general adult population.

Classical equations, such as Durnin/Womersley equation, Jackson/Pollock equation, Jackson, Pollock and Ward equation, and estimation equations implying SKF, alone or combined with circumferences, are most valid to estimate total body fat mass or %BF in general adult population.

These equations are valid to be “*generalized*” in the clinical or epidemiologic environments and field setting assessment.

The ease in body composition assessment may help in the rapid identification, prevention and management of the diverse health conditions related to obesity.

Study IV. Reliability of Field-Based Physical Fitness Test in Adults: A Systematic Review.

IV. Strong evidence, derived from a systematic review, indicates that:

To assess **cardiorespiratory fitness**, the 20-m SRT, the step tests and the 6-minute walk tests are reliable. The 20-m SRT is a maximal test that offers high reliability in young people, while the step tests and the 6-minute walk test emerged as possible

alternative submaximal tests to assess cardiorespiratory fitness in adults who are not very physically fit and/or find it difficult to run (e.g., obese individuals).

To assess **musculoskeletal fitness**, the handgrip test using a JAMAR dynamometer offers high test-retest reliability yet this dynamometer does not allow for adapting grip span to the individual's hand size with the same precision as the TKK dynamometer. The back-leg test, Biering-Sørensen (and its modified versions), trunk flexion sustained, and 5-reps sit-to-stand tests show high reliability. The bilateral side and prone bridge tests provide results with moderate reliability. The sit-and-reach (and its modified versions), and the toe-to-touch test show high reliability.

To assess **motor fitness**, the T-test (and its modified version) provide results demonstrating moderate reliability. The single-leg stand test (and its modified versions) is not reliable. There is inconclusive evidence regarding the reliability of balance and gait speed tests.

Study V: The 2-km walk test and the 20-m shuttle run test.

V. The 2-km walk and the 20-m SRT, as well as their corresponding Oja's and Leger's equations, are **valid** and **reliable** for estimating **cardiorespiratory fitness** in adults aged 18-64 years. However, the 20-m SRT obtained slightly greater criterion-related validity and reliability, regardless of sex, age, and physical activity level.

When existing time or space constraints, the 20-m SRT could be proposed as an ideal tool to evaluate cardiorespiratory fitness in the adult population.

Alternatively, the 2-km walk test is more suitable for adults who are unable to run.

Study VI: The standing long jump test.

VI. The SLJ test may be a **valid** tool to assess the lower-body explosive muscular fitness in the adult population, independently of age, and physical activity levels. The SLJ test may be used when controlling the possible learning effect to provide **reliability**.

The SLJ test is practical, time efficient, and low in cost and equipment requirements, being an ideal test to be used as a health evaluation tool.

VII. An evidence-based field physical fitness tests proposal can be made, in terms of **predictive validity**, **criterion-related validity** and **reliability** of the field-based physical fitness tests derived from the present International Doctoral Thesis. Therefore, a

final conclusion summarizes the evidence derived from the studies included in this International Doctoral Thesis (see below, **Figure 22**):

To assess **cardiorespiratory fitness**, the 20-m SRT, the 2-km walk, the 6-minute walk test and the YMCA step test are proposed with strong evidence.

To assess **muscular fitness**, the handgrip test and the Biering–Sørensen test are proposed with strong evidence; the prone bridging test is proposed with moderate evidence; the SLJ and the 30 s sit-to-stand test provide results with limited/inconclusive evidence.

To assess **motor fitness**, more studies are still necessary.

To assess **body composition**, WC, BAI, BMI, Durnin/Womersley equation, Jackson/Pollock equation, Jackson, Pollock and Ward equation, and estimation equations implying SKF are proposed with strong evidence.

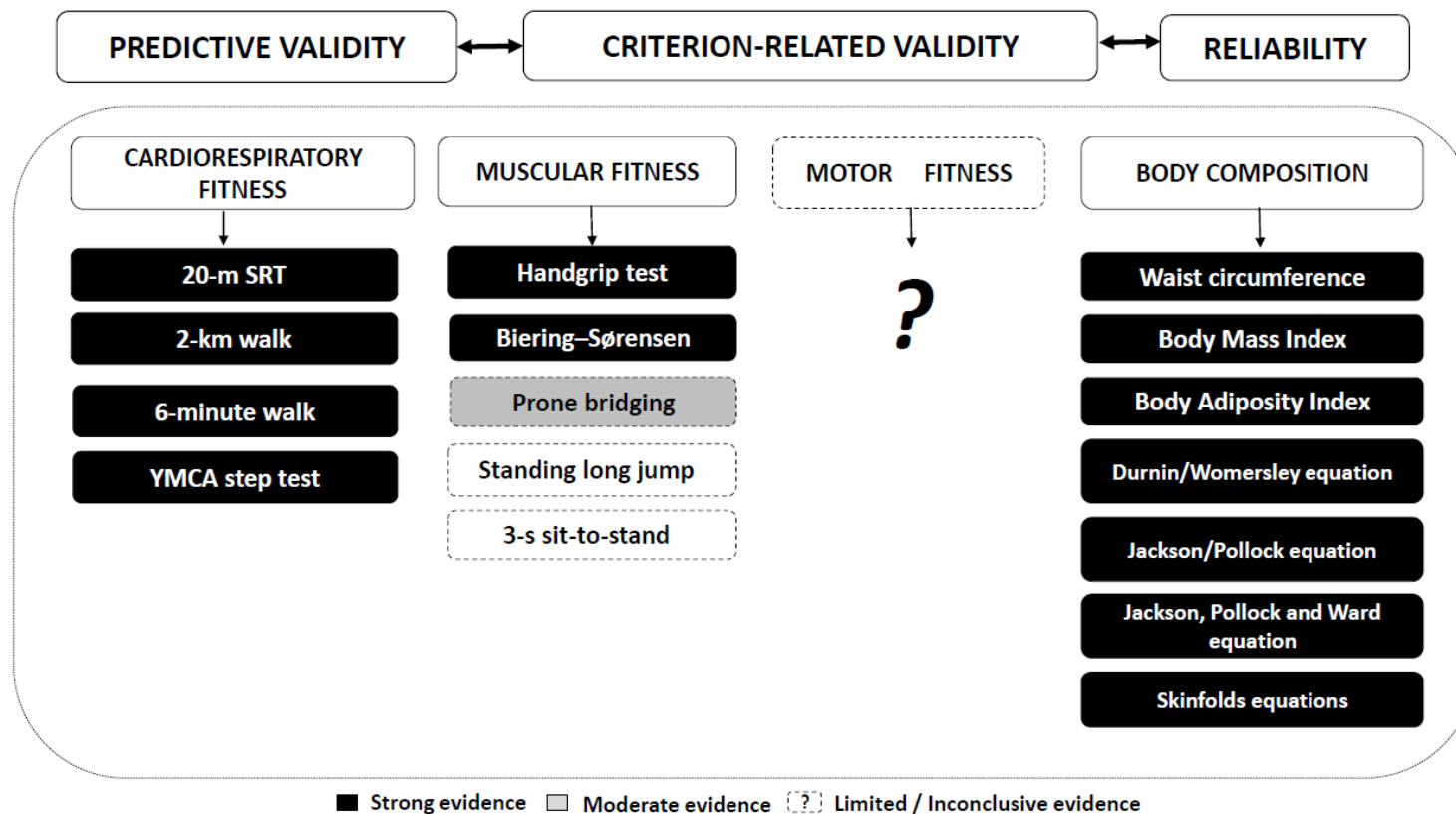


Figure 22. Evidence-based field physical fitness tests in adult population.

CONCLUSIONES

En general, los hallazgos derivados de la presente Tesis Doctoral Internacional sugieren que:

Estudio I. Validez Predictiva de Test de Campo en Adultos: Una Revisión Sistemática.

I. Fuerte evidencia, derivada de una revisión sistemática, indica que:

Una **velocidad de marcha** más lenta predice caídas, deterioro cognitivo y discapacidad (incluido el desarrollo de demencia y enfermedad de Alzheimer), movilidad reducida, discapacidad en las actividades de la vida cotidiana, riesgo de enfermedad cardiovascular, mortalidad por todas las causas y riesgo de institucionalización u hospitalización. Se sugiere su uso especialmente en adultos mayores de 35 años.

La alteración del **equilibrio** predice caídas, discapacidad en las actividades de la vida cotidiana o movilidad reducida y mortalidad por todas las causas. Se sugiere su uso especialmente en adultos mayores de 40 años.

Un peor rendimiento en el test de **levantarse y caminar (TUG)** predice caídas y miedo a caer en adultos mayores de 40 años, especialmente en mujeres.

Estos resultados brindan una justificación adicional para integrar la medición de la capacidad motora, principalmente la velocidad de la marcha y el equilibrio, en las herramientas de evaluación de pronóstico como indicadores de resultados de salud (como riesgo de caídas, deterioro cognitivo, discapacidad, riesgo de enfermedad cardiovascular, mortalidad por todas las causas y hospitalización) tanto en adultos como en adultos mayores. La identificación de estos factores de riesgo permitirá una intervención temprana basada en niveles adecuados de condición física y potencialmente disminuirá la morbilidad y la mortalidad a lo largo de la vida, al mismo tiempo que aumentará su calidad de vida.

A pesar de la evidencia limitada con respecto a la **flexibilidad**, nuestros resultados indican que las pruebas de flexibilidad pueden ayudar a detectar a los adultos mayores con riesgo de caídas, que sufren discapacidad o que esperan mortalidad por todas las causas.

Estudio II: Validez de Criterio de Test de Campo en Adultos.

II. Fuerte evidencia, derivada de una revisión sistemática, indica que:

Para evaluar la **capacidad cardiorrespiratoria**, el test de 20-m ida y vuelta (usando la ecuación de Leger), evaluó mejor la capacidad cardiorrespiratoria. Alternativamente, los test de 1,5 millas y 12-minutos corriendo/andando y el YMCA step test fueron otras opciones de test cardiorrespiratorios. Cuando existía un bajo nivel de capacidad cardiorrespiratoria, o si no era posible correr, los test de 2-km (usando la ecuación de Oja) o de 6-minutos andando fueron alternativas apropiadas.

Para evaluar la **fuerza muscular**, el test de presión manual con el codo extendido y con la amplitud de agarre adaptada al tamaño de la mano del individuo (utilizando el dinamómetro TKK), ofrece un medio válido para evaluar la fuerza muscular isométrica máxima de la mano. El test de Biering-Sørensen estimó la fuerza muscular resistencia de los músculos de la cadera y la espalda. La evidencia limitada (solo un estudio) apoyó los test de plancha prono y salto Sargent como marcadores sustitutos de la fuerza muscular resistencia abdominal y la fuerza muscular explosiva de la parte inferior del cuerpo, respectivamente.

Para evaluar la **flexibilidad**, el sit-and-reach test (y sus diferentes versiones), y el toe-to-touch test no son válidos.

Existe evidencia limitada sobre la validez de criterio relacionada con la **capacidad motora**.

Cuando existen problemas de espacio y tiempo, como en el ámbito clínico, el YMCA step test y el test de presión manual son buenas alternativas para evaluar la capacidad cardiorrespiratoria y la fuerza muscular isométrica, respectivamente.

Estudio III: Métodos y Ecuaciones para estimar la Composición Corporal en Adultos.

III. Fuerte evidencia, derivada de una revisión sistemática, indica que:

Los **métodos de evaluación de campo** de la composición corporal, como la circunferencia de cintura, el índice de adiposidad corporal y el índice de masa corporal son indicadores válidos para evaluar la composición/adiposidad corporal en población adulta.

Las **ecuaciones clásicas**, como la ecuación de Durnin/Womersley, la ecuación de Jackson/Pollock, la ecuación de Jackson, Pollock y Ward y las ecuaciones que implican pliegues corporales, solas o combinadas con circunferencias, son las más válidas para estimar la grasa corporal total o el porcentaje de grasa corporal en la población adulta.

Estas ecuaciones son válidas para ser “generalizadas” en ambientes clínicos o epidemiológicos y en la evaluación de campo.

La facilidad en la evaluación de la composición corporal puede ayudar en la rápida identificación, prevención y manejo de las diversas condiciones de salud relacionadas con la obesidad.

Estudio IV. Fiabilidad de Test de Campo en Adultos: Una Revisión Sistemática.

IV. Fuerte evidencia, derivada de una revisión sistemática, indica que:

Para evaluar la **capacidad cardiorrespiratoria**, los test de 20-m ida y vuelta, los step tests y el test 6-minutos andando son fiables. El test de 20-m ida y vuelta es un test máximo que ofrece una alta fiabilidad en jóvenes, mientras que los test sub-máximos step tests y el test 6-minutos andando surgieron como posibles pruebas alternativas para evaluar la capacidad cardiorrespiratoria en adultos con baja condición física y/o que les resulta difícil correr (por ejemplo, personas obesas).

Para evaluar la **capacidad musculoesquelética**, el test de fuerza de prensión manual usando un dinamómetro JAMAR ofrece una alta fiabilidad test-retest, pero este dinamómetro no permite adaptar el rango de agarre al tamaño de la mano del individuo con la misma precisión que el dinamómetro TKK. Los test back-leg strength, Biering-Sørensen (y sus versiones modificadas), la flexión sostenida del tronco, y 5 repeticiones de sentarse y levantarse muestran una alta fiabilidad. Los test de puente lateral y prono brindan resultados con fiabilidad moderada. El sit-and-reach test (y sus versiones modificadas) y toe-to-touch test muestran una alta fiabilidad.

Para evaluar la **capacidad motora**, el T-test (y su versión modificada) arroja resultados demostrando una fiabilidad moderada. El test de equilibrio sobre una sola pierna (y sus versiones modificadas) no es fiable. No hay evidencia concluyente con respecto a la fiabilidad de las pruebas de equilibrio y velocidad de la marcha.

Estudio V: Test de 2-km andando y test de 20-m ida y vuelta.

V. Los test de 2-km marcha (usando la ecuación de Oja) y 20-m ida y vuelta, (usando la ecuación de Leger) son **válidos** y **fiables** para estimar la capacidad cardiorrespiratoria en adultos de 18 a 64 años. Sin embargo, el test de 20-m ida y vuelta obtuvo una validez y fiabilidad ligeramente mayores, independientemente del sexo, la edad y el nivel de actividad física.

Cuando existen limitaciones de tiempo o espacio, el test de 20-m ida y vuelta podría proponerse como una herramienta ideal para evaluar la capacidad cardiorrespiratoria en la población adulta.

Alternativamente, el test de 2-km marcha es más adecuado para adultos que no pueden correr.

Estudio VI: Test de salto horizontal a pies juntos.

VI. El test de salto horizontal a pies juntos puede ser una herramienta **válida** para evaluar la fuerza muscular explosiva de la parte inferior del cuerpo en población adulta, independientemente de la edad y nivel de actividad física. El test de salto horizontal a pies juntos es **fiable** cuando se controla el posible efecto de aprendizaje.

El test de salto horizontal a pies juntos es práctico, eficiente en el tiempo y de bajo coste y requisitos de equipo, siendo un test ideal para ser utilizado como una herramienta de evaluación de la salud.

VII. Se puede realizar una propuesta de test de campo de condición física basada en la evidencia, en términos de **validez predictiva, validez de criterio y fiabilidad** de test de campo de condición física derivada de la presente Tesis Doctoral Internacional. Por tanto, una conclusión final resume la evidencia derivada de los estudios incluidos en esta Tesis Doctoral Internacional (ver más abajo, **Figura 22**):

Para evaluar la **capacidad cardiorrespiratoria** se proponen con fuerte evidencia el test de 20-m ida y vuelta, el test de 2-km andando, el test de 6-minutos andando y el YMCA step test.

Para evaluar la **fuerza muscular** se proponen con fuerte evidencia el test de prensión manual y el test de Biering-Sørensen; se propone el test de puente prono con evidencia moderada; el de salto horizontal a pies juntos y el test de sentarse y levantarse de 30 segundos brindan resultados con evidencia limitada/no concluyente.

Para evaluar la **capacidad motora**, se necesitan más estudios.

Para evaluar la **composición corporal**, la circunferencia de cintura, el índice de adiposidad corporal y el índice de masa corporal, la ecuación de Durnin/Womersley, la ecuación de Jackson/Pollock, la ecuación de Jackson, Pollock y Ward, y ecuaciones que implican pliegues corporales se proponen con fuerte evidencia.

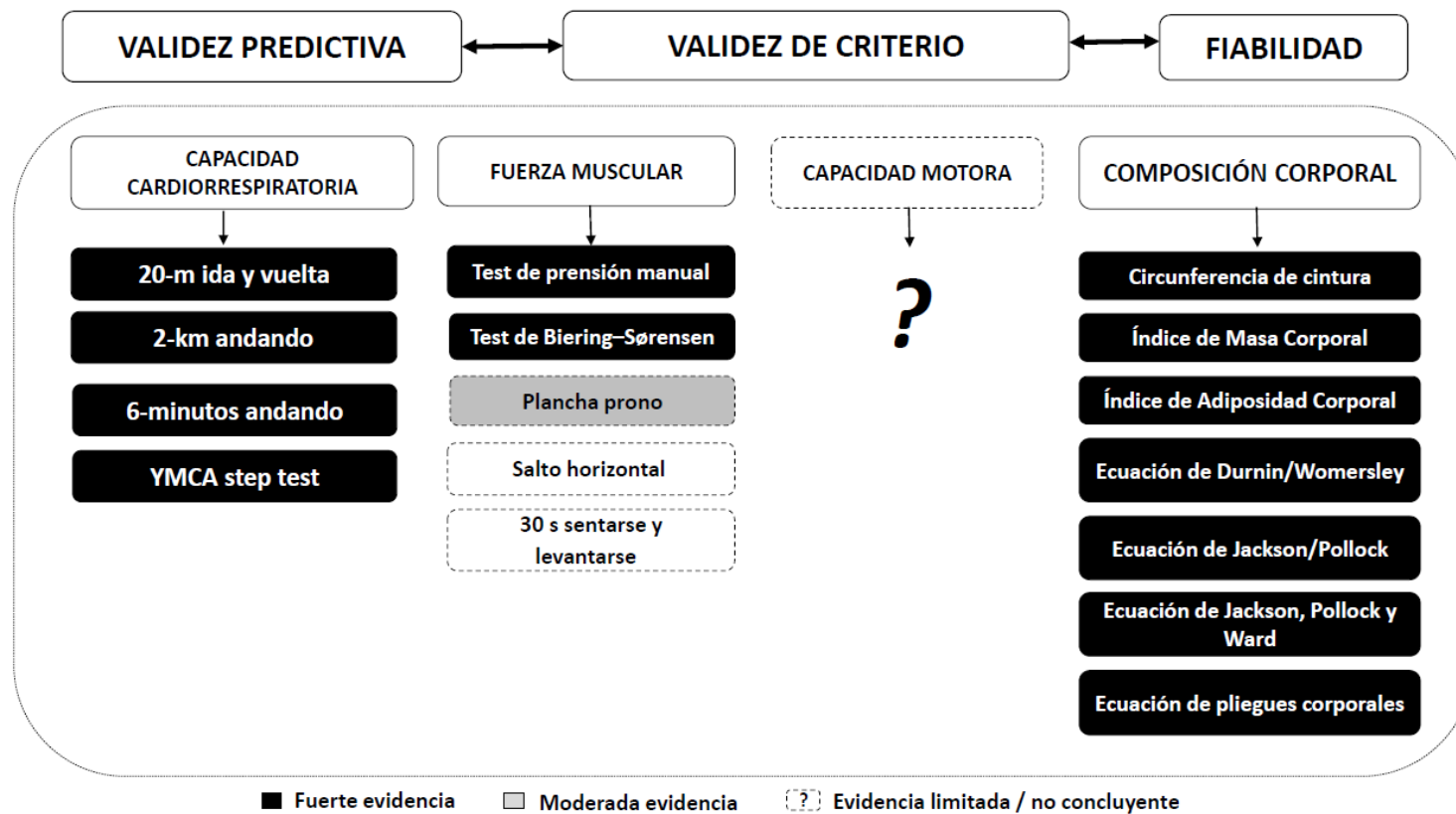


Figura 22. Test de campo de condición física basados en evidencia científica en población adulta.

FUTURE RESEARCH DIRECTIONS

FUTURE RESEARCH DIRECTIONS

The studies derived from the present International Doctoral Thesis provide evidence to propose a battery of field-based tests to assess physical fitness related to health in adults (*The ADULT-FIT health-related fitness test battery*) based on valid and reliable scientific evidence, depending on sex, age, and physical activity levels.

The ultimate aims of the ADULT-FIT project are: (i) design a valid, reliable, feasible, safe and responsiveness field-based test battery to assess physical fitness related to health in adults; (ii) homogenize the evaluation of health-related fitness tests in this population, allowing us to obtain normative reference values of physical fitness in adults; comparing physical fitness levels; and encouraging the implementation of strategies to improve the health status of adults with greater success and achieving more effective results, obtaining a significant reduction of costs in the public health monitoring system (improving the cost-effectiveness). Therefore, to achieve these aims, future methodological studies are still necessary to establish the validity and/or reliability of those field-based physical fitness tests (identified in systematic reviews as having insufficient or strong evidence) that have not been developed in this International Doctoral Thesis (e.g., 6-minute walk test, prone bridging test, sit-to-stand test, etc.). Likewise, future methodological studies are also still necessary to explore: (i) the feasibility and safety of those field-based physical fitness test identified (in the systematic reviews) as valid and reliable in adult population; (ii) the responsiveness of those field-based physical fitness test identified as valid, reliable, feasible and safe in adult population.

Finally, since there is no a field-based physical fitness test battery for older adult population that meets all the aforementioned characteristics, future research directions should be aimed at also proposing a field-based test battery to assess physical fitness related to health in older adults, that covers all the aims proposed in the ADULT-FIT project.

REFERENCES

1. Niiranen TJ, Enserro DM, Larson MG, Vasan RS. Multisystem Trajectories Over the Adult Life Course and Relations to Cardiovascular Disease and Death. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2019;74(11):1778-85.
2. Beard JR, Officer A, De Carvalho IA, Sadana R, Pot AM, Michel J-P, et al. The World report on ageing and health: a policy framework for healthy ageing. *The Lancet*. 2016;387(10033):2145-54.
3. Zucchelli A, Vetrano DL, Grande G, Calderon-Larranaga A, Fratiglioni L, Marengoni A, et al. Comparing the prognostic value of geriatric health indicators: a population-based study. *Bmc Medicine*. 2019;17(1).
4. Hoogendijk EO, Rijnhart JJM, Skoog J, Robitaille A, van den Hout A, Ferrucci L, et al. Gait speed as predictor of transition into cognitive impairment: Findings from three longitudinal studies on aging. *Experimental Gerontology*. 2020;129.
5. Cavanaugh EJ, Richardson J, McCallum CA, Wilhelm M. The Predictive Validity of Physical Performance Measures in Determining Markers of Preclinical Disability in Community-Dwelling Middle-Aged and Older Adults: A Systematic Review. *Physical Therapy*. 2018;98(12):1010-21.
6. Grande G, Triolo F, Nuara A, Welmer AK, Fratiglioni L, Vetrano DL. Measuring gait speed to better identify prodromal dementia. *Experimental Gerontology*. 2019;124.
7. Peel NM, Alapatt LJ, Jones LV, Hubbard RE. The Association Between Gait Speed and Cognitive Status in Community-Dwelling Older People: A Systematic Review and Meta-analysis. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2019;74(6):943-8.
8. Quan MH, Xun PC, Chen C, Wen J, Wang YY, Wang R, et al. Walking Pace and the Risk of Cognitive Decline and Dementia in Elderly Populations: A Meta-analysis of Prospective Cohort Studies. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*. 2017;72(2):266-70.
9. Van Kan GA, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. *Journal of Nutrition Health & Aging*. 2009;13(10):881-9.
10. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *Journal of Cachexia Sarcopenia and Muscle*. 2020;11(1):3-25.
11. Abe T, Kitamura A, Taniguchi Y, Amano H, Seino S, Yokoyama Y, et al. Pathway from gait speed to incidence of disability and mortality in older adults: A mediating role of physical activity. *Maturitas*. 2019;123:32-6.
12. Birtwhistle R, Bell NR, Thombs BD, Grad R, Dickinson JA. Periodic preventive health visits: a more appropriate approach to delivering preventive services: From the Canadian Task Force on Preventive Health Care. *Canadian Family Physician*. 2017;63(11):824-6.
13. Veronese N, Solmi M, Maggi S, Noale M, Sergi G, Manzato E, et al. Frailty and incident depression in community-dwelling older people: results from the ELSA study. *International Journal of Geriatric Psychiatry*. 2017;32:E141-E9.
14. Briggs R, Carey D, Claffey P, McNicholas T, Donoghue O, Kennelly SP, et al. Do Differences in Spatiotemporal Gait Parameters Predict the Risk of Developing Depression in Later Life? *Journal of the American Geriatrics Society*. 2019;67(5):1050-6.

15. Callisaya ML, Ayers E, Barzilai N, Ferrucci L, Guralnik JM, Lipton RB, et al. Motoric Cognitive Risk Syndrome and Falls Risk: A Multi-Center Study. *Journal of Alzheimers Disease*. 2016;53(3):1043-52.
16. Maki BE, Holliday PJ, Topper AK. A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population. *Journals of Gerontology*. 1994;49(2):M72-M84.
17. Quach L, Galica AM, Jones RN, Procter-Gray E, Manor B, Hannan MT, et al. The Nonlinear Relationship Between Gait Speed and Falls: The Maintenance of Balance, Independent Living, Intellect, and Zest in the Elderly of Boston Study. *Journal of the American Geriatrics Society*. 2011;59(6):1069-73.
18. Elbaz A, Sabia S, Brunner E, Shipley M, Marmot M, Kivimaki M, et al. Association of walking speed in late midlife with mortality: results from the Whitehall II cohort study. *Age*. 2013;35(3):943-52.
19. Sanders JB, Bremmer MA, Comijs HC, van de Ven PM, Deeg DJH, Beekman ATF. Gait Speed and Processing Speed as Clinical Markers for Geriatric Health Outcomes. *American Journal of Geriatric Psychiatry*. 2016;25:374-85.
20. Frames CW, Soangra R, Lockhart TE, Lach J, Ha DS, Roberto KA, et al. Dynamical Properties of Postural Control in Obese Community-Dwelling Older Adults. *Sensors*. 2018;18.
21. Kwan MM, Lin SI, Close JC, Lord SR. Depressive symptoms in addition to visual impairment, reduced strength and poor balance predict falls in older Taiwanese people. *Age Ageing*. 2012;41(5):606-12.
22. Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. *Journal of the American Geriatrics Society*. 1997;45:735-8.
23. Ersoy Y, MacWalter RS, Durmus B, Altay ZE, Baysal O. Predictive Effects of Different Clinical Balance Measures and the Fear of Falling on Falls in Postmenopausal Women Aged 50 Years and Over. *Gerontology*. 2009;55:660-5.
24. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society*. 1991;39(2):142-8.
25. Austin N, Devine A, Dick I, Prince R, Bruce D. Fear of falling in older women: a longitudinal study of incidence, persistence, and predictors. *J Am Geriatr Soc*. 2007;55(10):1598-603.
26. Nitz JC, Stock L, Khan A. Health-related predictors of falls and fractures in women over 40. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2013;24:613-21.
27. Mulasso A, Roppolo M, Gobbens RJ, Rabaglietti E. Mobility, balance and frailty in community-dwelling older adults: at is the best 1-year predictor of falls? *Geriatrics & Gerontology International*. 2017;17:1463-9.
28. Kang L, Han PP, Wang JZ, Ma YX, Jia LY, Fu LY, et al. Timed Up and Go Test can predict recurrent falls: a longitudinal study of the community-dwelling elderly in China. *Clinical Interventions in Aging*. 2017;12:2009-16.
29. Doi T, Hirata S, Ono R, Tsutsumimoto K, Misu S, Ando H. The harmonic ratio of trunk acceleration predicts falling among older people: results of a 1-year prospective study. *Journal of Neuroengineering and Rehabilitation*. 2013;10.
30. Clemson L, Kendig H, Mackenzie L, Browning C. Predictors of Injurious Falls and Fear of Falling Differ: An 11-Year Longitudinal Study of Incident Events in Older People. *Journal of Aging and Health*. 2015;27:239-56.

31. Abu Samah Z, Singh DKA, Murukesu RR, Shahar S, Nordin NAM, Omar MA, et al. Discriminative and Predictive Ability of Physical Performance Measures in Identifying Fall Risk among Older Adults. *Sains Malaysiana*. 2018;47:2769-76.
32. Asai T, Oshima K, Fukumoto Y, Yonezawa Y, Matsuo A, Misu S. Does dual-tasking provide additional value in timed "up and go" test for predicting the occurrence of falls? A longitudinal observation study by age group (young-older or old-older adults). *Aging Clinical and Experimental Research*. 2020.
33. Coll-Risco I, Acosta-Manzano P, Borges-Cosic M, Camiletti-Moiron D, Aranda P, Soriano-Maldonado A, et al. Body Composition Changes Following a Concurrent Exercise Intervention in Perimenopausal Women: The FLAMENCO Project Randomized Controlled Trial. *Journal of clinical medicine*. 2019;8(10):1678.
34. Luukinen H, Koski K, Laippala P, Kivela SL. Predictors for recurrent falls among the home-dwelling elderly. *Scandinavian Journal of Primary Health Care*. 1995;13(4):294-9.
35. Breton E, Beloin F, Fortin C, Martin A, Ouellet ME, Payette H, et al. Gender-specific associations between functional autonomy and physical capacities in independent older adults: Results from the NuAge study. *Archives of Gerontology and Geriatrics*. 2014;58:56-62.
36. Donoghue O, Feeney J, O'Leary N, Kenny RA. Baseline Mobility is Not Associated with Decline in Cognitive Function in Healthy Community-Dwelling Older Adults: Findings From The Irish Longitudinal Study on Ageing (TILDA). *American Journal of Geriatric Psychiatry*. 2017;26:438-48.
37. Bohannon RW. Reference values for the timed up and go test: a descriptive meta-analysis. *Journal of geriatric physical therapy*. 2006;29(2):64-8.
38. Bravell ME, Finkel D, Aslan AD, Reynolds CA, Hallgren J, Pedersen NL. Motor functioning differentially predicts mortality in men and women. *Archives of Gerontology and Geriatrics*. 2017;72:6-11.
39. Ward RE, Beauchamp MK, Latham NK, Leveille SG, Percac-Lima S, Kurlinski L, et al. Neuromuscular Impairments Contributing to Persistently Poor and Declining Lower-Extremity Mobility Among Older Adults: New Findings Informing Geriatric Rehabilitation. *Archives of Physical Medicine and Rehabilitation*. 2016;97:1316-22.
40. Bassett DR, Jr., Howley ET, Thompson DL, King GA, Strath SJ, McLaughlin JE, et al. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. *J Appl Physiol (1985)*. 2001;91(1):218-24.
41. Manttari A, Suni J, Sievanen H, Husu P, Vaha-Ypya H, Valkeinen H, et al. Six-minute walk test: a tool for predicting maximal aerobic power (VO₂ max) in healthy adults. *Clin Physiol Funct Imaging*. 2018.
42. Castro-Pinero J, Artero EG, Espana-Romero V, Ortega FB, Sjostrom M, Suni J, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med*. 2009;44(13):934-43.
43. Meredith MD, Welk GJ. FITNESSGRAM & ACTIVITYGRAM test administration manual. 4th ed. Champaign: Human Kinetics; 2010.
44. Ruiz JR, Ortega FB, Castro-Piñero J. Validity and reliability of the 1/4 mile run-walk test in physically active children and adolescents. *Nutr Hosp*. 2014;31(2):875-82.
45. Mayorga-Vega D, Bocanegra-Parrilla R, Ornelas M, Viciana J. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: a systematic review and Meta-analysis. *PLoS One*. 2016;11(3):e0151671-e.
46. McNaughton L, Hall P, Cooley D. Validation of several methods of estimating maximal oxygen uptake in young men. *Percept Mot Skills*. 1998;87(2):575-84.

47. Larsen GE, George JD, Alexander JL, Fellingham GW, Aldana SG, Parcell AC. Prediction of maximum oxygen consumption from walking, jogging, or running. *Res Q Exerc Sport*. 2002;73(1):66-72.
48. Krahenbuhl GS, Pangrazi RP, Burkett LN, Schneider MJ, Petersen G. Field estimation of VO₂ max in children eight years of age. *Med Sci Sports*. 1977;9(1):37-40.
49. McCormack WP, Cureton KJ, Bullock TA, Weyand PG. Metabolic determinants of 1-mile run/walk performance in children. *Med Sci Sports Exerc*. 1991;23(5):611-7.
50. Shephard RJ. Tests of maximum oxygen intake. A critical review. *Sports Med*. 1984;1(2):99-124.
51. Laukkanen RMT, Kukkonen-Harjula TK, Oja P, Pasanen ME, Vuori IM. Prediction of change in maximal aerobic power by the 2-km walk test after walking training in middle-aged adults. *Int J Sports Med*. 2000;21(2):113-6.
52. Di Thommazo-Luporini L, Carvalho LP, Luporini RL, Trimer R, Falasco Pantoni CB, Martinez AF, et al. Are cardiovascular and metabolic responses to field walking tests interchangeable and obesity-dependent? *Disabil Rehabil*. 2016;38(18):1820-9.
53. Laukkanen R, Oja P, Pasanen M, Vuori I. Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults. *Int J Obes Relat Metab Disord*. 1992;16(4):263-8.
54. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *Int J Sports Med*. 1991;12(4):356-62.
55. Laukkanen RMT, Oja P, Pasanen ME, Vuori IM. Criterion validity of a two-kilometer walking test for predicting the maximal oxygen uptake of moderately to highly active middle-aged adults. *Scand J Med Sci Sports*. 1993(3): 267-72.
56. Duncan MJ, Mota J, Carvalho J, Nevill AM. An Evaluation of Prediction Equations for the 6 Minute Walk Test in Healthy European Adults Aged 50-85 Years. *PLoS One*. 2015;10(9):e0139629.
57. Di Thommazo-Luporini L, Pinheiro Carvalho L, Luporini R, Trimer R, Falasco Pantoni CB, Catai AM, et al. The six-minute step test as a predictor of cardiorespiratory fitness in obese women. *Eur J Phys Rehabil Med*. 2015;51(6):793-802.
58. Burr JF, Bredin SS, Faktor MD, Warburton DE. The 6-minute walk test as a predictor of objectively measured aerobic fitness in healthy working-aged adults. *Phys Sportsmed*. 2011;39(2):133-9.
59. Kline C, Porcari JP, Hintermeister R, Freedson PS, Ward A, McCarron RF, et al. Estimation of from a one-mile track walk, gender, age and body weight. *Med Sports Exerc*. 1987;19:253-9.
60. Dolgener FA, Hensley LD, Marsh JJ, Fjelstul JK. Validation of the Rockport Fitness Walking Test in college males and females. *Res Q Exerc Sport*. 1994;65(2):152-8.
61. Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *Journal of sports sciences*. 1988;6(2):93-101.
62. Mayorga-Vega D, Aguilar-Soto P, Viciano J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med*. 2015;14(3):536-47.
63. Cooper SM, Baker JS, Tong RJ, Roberts E, Hanford M. The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *Br J Sports Med*. 2005;39(4):e19.
64. Kim J, Jung SH, Cho HC. Validity and Reliability of Shuttle-Run Test in Korean Adults. *Int J Sports Med*. 2011;32(8):580-5.

65. Flouris AD, Koutedakis Y, Nevill A, Metsios GS, Tsiotra G, Parasiris Y. Enhancing specificity in proxy-design for the assessment of bioenergetics. *J Sci Med Sport*. 2004;7(2):197-204.
66. Metsios GS, Flouris AD, Koutedakis Y, Nevill A. Criterion-related validity and test-retest reliability of the 20m square shuttle test. *J Sci Med Sport*. 2008;11(2):214-7.
67. Bennett H, Parfitt G, Davison K, Eston R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. *Sports Med*. 2016;46(5):737-50.
68. Teren A, Zachariae S, Beutner F, Ubrich R, Sandri M, Engel C, et al. Incremental value of veterans specific activity questionnaire and the ymca-step test for the assessment of cardiorespiratory fitness in population-based studies. *Eur J Prev Cardiol*. 2016;23(11):1221-7.
69. Beutner F, Ubrich R, Zachariae S, Engel C, Sandri M, Teren A, et al. Validation of a brief step-test protocol for estimation of peak oxygen uptake. *Eur J Prev Cardiol*. 2015;22(4):503-12.
70. Lee O, Lee S, Kang M, Mun J, Chung J. Prediction of maximal oxygen consumption using the Young Men's Christian Association-step test in Korean adults. *Eur J Appl Physiol*. 2019;119(5):1245-52.
71. Hong SH, Yang HI, Kim DI, Gonzales TI, Brage S, Jeon JY. Validation of submaximal step tests and the 6-min walk test for predicting maximal oxygen consumption in young and healthy participants. *Int J Environ Res Public Health*. 2019;16(23).
72. Kieu NTV, Jung SJ, Shin SW, Jung HW, Jung ES, Won YH, et al. The validity of the YMCA 3-minute step test for estimating maximal oxygen uptake in healthy Korean and Vietnamese adults. *J Lifestyle Med*. 2020;10(1):21-9.
73. Santo A, Golding LA. Predicting maximum oxygen uptake from a modified 3-minute step test. *Res Q Exer Sport*. 2003;74(1):110-5.
74. Mayhew JL, Ball TE, Ward TE, Hart CL, Arnold MD. Relationships of structural dimensions to bench press strength in college males. *J Sports Med Phys Fitness*. 1991;31(2):135-41.
75. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *Pm r*. 2011;3(5):472-9.
76. Paul DJ, Nassis GP. Testing strength and power in soccer players: the application of conventional and traditional methods of assessment. *J Strength Cond Res*. 2015;29(6):1748-58.
77. De Ste Croix M, Deighan M, Armstrong N. Assessment and interpretation of isokinetic muscle strength during growth and maturation. *Sports Med*. 2003;33(10):727-43.
78. Coorevits P, Danneels L, Cambier D, Ramon H, Vanderstraeten G. Assessment of the validity of the Biering-Sørensen test for measuring back muscle fatigue based on EMG median frequency characteristics of back and hip muscles. *J Electromyogr Kinesiol*. 2008;18(6):997-1005.
79. De Blaiser C, De Ridder R, Willems T, Danneels L, Roosen P. Reliability and validity of trunk flexor and trunk extensor strength measurements using handheld dynamometry in a healthy athletic population. *Phys Ther Sport*. 2018;34:180-6.
80. Espana-Romero V, Artero EG, Santaliestra-Pasias AM, Gutierrez A, Castillo MJ, Ruiz JR. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *J Hand Surg [Am]*. 2008;33(3):378-84.

81. Cadenas-Sanchez C, Sanchez-Delgado G, Martinez-Tellez B, Mora-Gonzalez J, Löf M, España-Romero V, et al. Reliability and validity of different models of TKK hand dynamometers. *Am J Occup Ther*. 2016;70(4):7004300010.
82. Kolimechkov S, Castro-Piñero J, Petrov A, Alexandrova A. The effect of elbow position on the handgrip strength test in children: validity and reliability of TKK 5101 and DynX dynamometers. *Pedagogy Phys Cult Sports*. 2020;24(5):240-7.
83. España-Romero V, Ortega FB, Vicente-Rodríguez G, Artero EG, Rey JP, Ruiz JR. Elbow position affects handgrip strength in adolescents: validity and reliability of Jamar, DynEx, and TKK dynamometers. *J Strength Cond Res*. 2010;24(1):272-7.
84. Balogun JA, Akomolafe CT, Amusa LO. Grip strength: effects of testing posture and elbow position. *Arch Phys Med Rehabil*. 1991;72(5):280-3.
85. NHANES. Muscle Strength Procedures Manual. National Health and Nutrition Examination Survey (NHANES). Druid Hills: CDC; 2013.
86. Ruiz-Ruiz J, Mesa JL, Gutiérrez A, Castillo MJ. Hand size influences optimal grip span in women but not in men. *J Hand Surg Am*. 2002;27(5):897-901.
87. Kim WJ, Kim KJ, Song DG, Lee JS, Park KY, Lee JW, et al. Sarcopenia and Back Muscle Degeneration as Risk Factors for Back Pain: A Comparative Study. *Asian Spine J*. 2020;14(3):364-72.
88. Mannion AF, Dolan P. Electromyographic median frequency changes during isometric contraction of the back extensors to fatigue. *Spine (Phila Pa 1976)*. 1994;19(11):1223-9.
89. Kankaanpää M, Laaksonen D, Taimela S, Kokko SM, Airaksinen O, Hänninen O. Age, sex, and body mass index as determinants of back and hip extensor fatigue in the isometric Sørensen back endurance test. *Arch Phys Med Rehabil*. 1998;79(9):1069-75.
90. Abdelraouf OR, Abdel-Aziem AA. The relationship between core endurance and back dysfunction in collegiate male athletes with and without nonspecific low back pain. *Int J Sports Phys Ther*. 2016;11(3):337-44.
91. Ozcan Kahraman B, Salik Sengul Y, Kahraman T, Kalemci O. Developing a Reliable Core Stability Assessment Battery for Patients with Nonspecific Low Back Pain. *Spine (Phila Pa 1976)*. 2016;41(14):E844-50.
92. Arab AM, Salavati M, Ebrahimi I, Ebrahim Mousavi M. Sensitivity, specificity and predictive value of the clinical trunk muscle endurance tests in low back pain. *Clin Rehabil*. 2007;21(7):640-7.
93. del Pozo-Cruz B, Mocholi MH, del Pozo-Cruz J, Parraca JA, Adsuar JC, Gusi N. Reliability and validity of lumbar and abdominal trunk muscle endurance tests in office workers with nonspecific subacute low back pain. *J Back Musculoskelet Rehabil*. 2014;27(4):399-408.
94. Ortega FB, Cadenas-Sanchez C, Sanchez-Delgado G, Mora-Gonzalez J, Martinez-Tellez B, Artero EG, et al. Systematic review and proposal of a field-based physical fitness-test battery in preschool children: the PREFIT battery. *Sports Med*. 2015;45(4):533-55.
95. Ruiz JR, Castro-Piñero J, España-Romero V, Artero EG, Ortega FB, Cuenca MM, et al. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *Br J Sports Med*. 2011;45(6):518-24.
96. Bui HT, Farinas M-I, Fortin A-M, Comtois A-S, Leone M. Comparison and analysis of three different methods to evaluate vertical jump height. *Clin Physiol Funct Imaging*. 2015;35(3):203-9.
97. Garcia-Hermoso A, Cavero-Redondo I, Ramirez-Velez R, Ruiz JR, Ortega FB, Lee DC, et al. Muscular Strength as a Predictor of All-Cause Mortality in an Apparently Healthy Population: A Systematic Review and Meta-Analysis of Data From

Approximately 2 Million Men and Women. *Arch Phys Med Rehabil*. 2018;99(10):2100-13.e5.

98. Garcia-Hermoso A, Ramirez-Velez R, Peterson MD, Lobelo F, Caverro-Redondo I, Correa-Bautista JE, et al. Handgrip and knee extension strength as predictors of cancer mortality: A systematic review and meta-analysis. *Scand J Med Sci Sports*. 2018;28(8):1852-8.

99. Leighton JR. An instrument and technic for the measurement of range of joint motion. *Arch Phys Med Rehabil*. 1955;36(9):571-8.

100. Kanbur NO, Duzgun I, Derman O, Baltaci G. Do sexual maturation stages affect flexibility in adolescent boys aged 14 years? *J Sports Med Phys Fitness*. 2005;45(1):53-7.

101. Castro-Pinero J, Chillon P, Ortega FB, Montesinos JL, Sjostrom M, Ruiz JR. Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *Int J Sports Med*. 2009;30(9):658-62.

102. Wells KF, Dillon EK. The sit-and-reach. A test of back and leg flexibility. *Res Q Exerc Sport*. 1952;23:115-8.

103. Mayorga-Vega D, Merino-Marban R, Viciano J. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: a meta-analysis. *J Sports Sci Med*. 2014;13(1):1-14.

104. Kraus H, Hirschland R. Minimum muscular fitness of the school children. *Res Q*. 1954;25:178-88.

105. Mayorga-Vega D, Viciano J, Cocca A, Merino-Marban R. Criterion-related validity of toe-touch test for estimating hamstring extensibility: A metaanalysis. *J Hum Sport Exerc*. 2014;9(1):188-200.

106. Nuzzo JL. The Case for Retiring Flexibility as a Major Component of Physical Fitness. *Sports Med*. 2020;50(5):853-70.

107. Elbaz A, Sabia S, Brunner E, Shipley M, Marmot M, Kivimaki M, et al. Association of walking speed in late midlife with mortality: results from the Whitehall II cohort study. *Age (Dordr)*. 2013;35(3):943-52.

108. Niiranen TJ, Enserro DM, Larson MG, Vasan RS. Multisystem Trajectories Over the Adult Life Course and Relations to Cardiovascular Disease and Death. *J Gerontol A Biol Sci Med Sci*. 2019;74(11):1778-85.

109. Cooper R, Strand BH, Hardy R, Patel KV, Kuh D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. *BMJ*. 2014;348:g2219.

110. Nitz JC, Stock L, Khan A. Health-related predictors of falls and fractures in women over 40. *Osteoporos Int*. 2013;24(2):613-21.

111. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *J Cachexia Sarcopenia Muscle*. 2020;11(1):3-25.

112. Briggs R, Carey D, Claffey P, McNicholas T, Donoghue O, Kennelly SP, et al. Do Differences in Spatiotemporal Gait Parameters Predict the Risk of Developing Depression in Later Life? *J Am Geriatr Soc*. 2019;67(5):1050-6.

113. Elagizi A, Kachur S, Lavie CJ, Carbone S, Pandey A, Ortega FB, et al. An overview and update on obesity and the obesity paradox in cardiovascular diseases. *Progress in cardiovascular diseases*. 2018;61(2):142-50.

114. Weir CB, Jan A. BMI classification percentile and cut off points. 2019.

115. Bergman RN, Stefanovski D, Buchanan TA, Sumner AE, Reynolds JC, Sebring NG, et al. A better index of body adiposity. *Obesity (Silver Spring)*. 2011;19(5):1083-9.

116. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nature Reviews Endocrinology*. 2020;16(3):177-89.
117. Shaw KA, Srikanth VK, Fryer JL, Blizzard L, Dwyer T, Venn AJ. Dual energy X-ray absorptiometry body composition and aging in a population-based older cohort. *Int J Obes (Lond)*. 2007;31(2):279-84.
118. Lam BC, Lim SC, Wong MT, Shum E, Ho CY, Bosco JI, et al. A method comparison study to validate a novel parameter of obesity, the body adiposity index, in Chinese subjects. *Obesity (Silver Spring)*. 2013;21(12):E634-9.
119. Gómez-Ambrosi J, Silva C, Catalán V, Rodríguez A, Galofré JC, Escalada J, et al. Clinical usefulness of a new equation for estimating body fat. *Diabetes Care*. 2012;35(2):383-8.
120. Kim SG, Ko K, Hwang IC, Suh HS, Kay S, Caterson I, et al. Relationship between indices of obesity obtained by anthropometry and dual-energy X-ray absorptiometry: The Fourth and Fifth Korea National Health and Nutrition Examination Survey (KNHANES IV and V, 2008-2011). *Obes Res Clin Pract*. 2015;9(5):487-98.
121. Castro-Piñero J, Delgado-Alfonso A, Gracia-Marco L, Gómez-Martínez S, Esteban-Cornejo I, Veiga OL, et al. Neck circumference and clustered cardiovascular risk factors in children and adolescents: cross-sectional study. *BMJ open*. 2017;7(9):e016048.
122. Organization WH. Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008. 2011.
123. Day K, Kwok A, Evans A, Mata F, Verdejo-Garcia A, Hart K, et al. Comparison of a Bioelectrical Impedance Device against the Reference Method Dual Energy X-Ray Absorptiometry and Anthropometry for the Evaluation of Body Composition in Adults. *Nutrients*. 2018;10(10).
124. Blew RM, Sardinha LB, Milliken LA, Teixeira PJ, Going SB, Ferreira DL, et al. Assessing the validity of body mass index standards in early postmenopausal women. *Obes Res*. 2002;10(8):799-808.
125. Brožek J, Keys A. The evaluation of leanness-fatness in man: norms and interrelationships. *British Journal of Nutrition*. 1951;5(2):194-206.
126. Siri WE. Body composition from fluid spaces and density: analysis of methods. 1956.
127. Brožek J, Grande F, Anderson JT, Keys A. Densitometric analysis of body composition: revision of some quantitative assumptions. *Annals of the New York Academy of Sciences*. 1963;110(1):113-40.
128. Durnin JV, Womersley J. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *British journal of nutrition*. 1974;32(1):77-97.
129. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. *British journal of nutrition*. 1978;40(3):497-504.
130. Jackson AS, Pollock ML, Ward A. Generalized equations for predicting body density of women. *Medicine and science in sports and exercise*. 1980;12(3):175-81.
131. Erselcan T, Candan F, Saruhan S, Ayca T. Comparison of body composition analysis methods in clinical routine. *Annals of Nutrition and Metabolism*. 2000;44(5-6):243-8.
132. Hodgdon JA, Beckett MB. Prediction of percent body fat for US Navy women from body circumferences and height. NAVAL HEALTH RESEARCH CENTER SAN DIEGO CA, 1984.

133. Eston RG, Rowlands AV, Charlesworth S, Davies A, Hoppitt T. Prediction of DXA-determined whole body fat from skinfolds: importance of including skinfolds from the thigh and calf in young, healthy men and women. *European Journal of Clinical Nutrition*. 2005;59(5):695-702.
134. Temple D, Denis R, Walsh MC, Dicker P, Byrne AT. Comparison of anthropometric-based equations for estimation of body fat percentage in a normal-weight and overweight female cohort: validation via air-displacement plethysmography. *Public Health Nutr*. 2015;18(3):446-52.
135. Aandstad A, Holtberget K, Hageberg R, Holme I, Anderssen SA. Validity and reliability of bioelectrical impedance analysis and skinfold thickness in predicting body fat in military personnel. *Mil Med*. 2014;179(2):208-17.
136. Aristizabal JC, Estrada-Restrepo A, García AG. Desarrollo y validación de ecuaciones antropométricas para estimar la composición corporal en mujeres adultas. *Revista Colombia Médica*. 2018;49(2):154-9.
137. Aristizabal JC, Restrepo MT, Amalia L. Validation by hydrodensitometry of skinfold thickness equations used for female body composition assessment. *Biomedica*. 2008;28(3):404-13.
138. Cui Z, Truesdale KP, Cai J, Stevens J. Evaluation of anthropometric equations to assess body fat in adults: NHANES 1999-2004. *Med Sci Sports Exerc*. 2014;46(6):1147-58.
139. Davidson LE, Wang J, Thornton JC, Kaleem Z, Silva-Palacios F, Pierson RN, et al. Predicting Fat Percent by Skinfolds in Racial Groups: Durnin and Womersley Revisited. *Medicine and Science in Sports and Exercise*. 2011;43(3):542-9.
140. Dioum A, Gartner A, Maire B, Delpeuch F, Wade S. Body composition predicted from skinfolds in African women: a cross-validation study using air-displacement plethysmography and a black-specific equation. *Br J Nutr*. 2005;93(6):973-9.
141. Friedl KE, Westphal KA, Marchitelli LJ, Patton JF, Chumlea WC, Guo SS. Evaluation of anthropometric equations to assess body-composition changes in young women. *American Journal of Clinical Nutrition*. 2001;73(2):268-75.
142. Lahav Y, Epstein Y, Kedem R, Schermann H. A novel body circumferences-based estimation of percentage body fat. *Br J Nutr*. 2018;119(6):720-5.
143. Lanham DA, Stead MA, Tsang K, Davies PSW. The prediction of body composition in Chinese Australian females. *International Journal of Obesity*. 2001;25(2):286-91.
144. Lukaski HC, Bolonchuk WW, Hall CB, Siders WA. Validation of tetrapolar bioelectrical impedance method to assess human body composition. *Journal of applied physiology*. 1986;60(4):1327-32.
145. Thomas EL, Saeed N, Hajnal JV, Brynes A, Goldstone AP, Frost G, et al. Magnetic resonance imaging of total body fat. *Journal of Applied Physiology*. 1998;85(5):1778-85.
146. Wang J, Deurenberg P. The validity of predicted body composition in Chinese adults from anthropometry and bioelectrical impedance in comparison with densitometry. *Br J Nutr*. 1996;76(2):175-82.
147. Benito PJ, Gomez-Candela C, Dolores Cabanas M, Szendrei B, Aparecida Castro E, Grp PS. COMPARISON BETWEEN DIFFERENT METHODS FOR MEASURING BODY FAT AFTER A WEIGHT LOSS PROGRAM. *Revista Brasileira De Medicina Do Esporte*. 2019;25(6):474-9.
148. Bhat DS, Yajnik CS, Sayyad MG, Raut KN, Lubree HG, Rege SS, et al. Body fat measurement in Indian men: comparison of three methods based on a two-compartment model. *Int J Obes (Lond)*. 2005;29(7):842-8.

149. Eston RG, Fu F, Fung L. VALIDITY OF CONVENTIONAL ANTHROPOMETRIC TECHNIQUES FOR PREDICTING BODY-COMPOSITION IN HEALTHY CHINESE ADULTS. *British Journal of Sports Medicine*. 1995;29(1):52-6.
150. Garcia AL, Wagner K, Hothorn T, Koebnick C, Zunft HJ, Trippo U. Improved prediction of body fat by measuring skinfold thickness, circumferences, and bone breadths. *Obes Res*. 2005;13(3):626-34.
151. Kohli S, Gao M, Lear SA. Using simple anthropometric measures to predict body fat in South Asians. *Appl Physiol Nutr Metab*. 2009;34(1):40-8.
152. Leahy S, O'Neill C, Sohun R, Toomey C, Jakeman P. Generalised equations for the prediction of percentage body fat by anthropometry in adult men and women aged 18-81 years. *Br J Nutr*. 2013;109(4):678-85.
153. Rush EC, Plank LD, Lualaba MS, Robinson SM. Prediction of percentage body fat from anthropometric measurements: comparison of New Zealand European and Polynesian young women. *The American journal of clinical nutrition*. 1997;66(1):2-7.
154. Shafer KJ, Siders WA, Johnson LK, Lukaski HC. Body density estimates from upper-body skinfold thicknesses compared to air-displacement plethysmography. *Clin Nutr*. 2010;29(2):249-54.
155. Wattanapenpaiboon N, Lukito W, Strauss BJG, Hsu-Hage BH, Wahlqvist ML, Stroud DB. Agreement of skinfold measurement and bioelectrical impedance analysis (BIA) methods with dual energy X-ray absorptiometry (DEXA) in estimating total body fat in Anglo-Celtic Australians. *International Journal of Obesity*. 1998;22(9):854-60.
156. Yao M, Roberts SB, Ma G, Pan H, McCrory MA. Field methods for body composition assessment are valid in healthy chinese adults. *J Nutr*. 2002;132(2):310-7.
157. Ball S, Swan PD, DeSimone R. Comparison of anthropometry to dual energy X-ray absorptiometry: a new prediction equation for women. *Res Q Exerc Sport*. 2004;75(3):248-58.
158. Hicks VL, Stolarczyk LM, Heyward VH, Baumgartner RN. Validation of near-infrared interactance and skinfold methods for estimating body composition of American Indian women. *Med Sci Sports Exerc*. 2000;32(2):531-9.
159. Lean M, Han TS, Deurenberg P. Predicting body composition by densitometry from simple anthropometric measurements. *The American journal of clinical nutrition*. 1996;63(1):4-14.
160. Nickerson BS, Esco MR, Bishop PA, Fedewa MV, Snarr RL, Kliszczewicz BM, et al. Validity of BMI-Based Body Fat Equations in Men and Women: A 4-Compartment Model Comparison. *J Strength Cond Res*. 2016;32(1):121-9.
161. Stout JR, Eckerson JM, Housh TJ, Johnson GO, Betts NM. Validity of percent body fat estimations in males. *Med Sci Sports Exerc*. 1994;26(5):632-6.
162. Segheto W, Coelho FA, Guimaraes da Silva DC, Hallal PC, Bouzas Marins JC, Ribeiro AQ, et al. Validity of body adiposity index in predicting body fat in Brazilians adults. *American Journal of Human Biology*. 2017;29(1).
163. Deurenberg P, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age-and sex-specific prediction formulas. *British journal of nutrition*. 1991;65(2):105-14.
164. O'Connor DP, Bray MS, McFarlin BK, Sailors MH, Ellis KJ, Jackson AS. Generalized Equations for Estimating DXA Percent Fat of Diverse Young Women and Men: The TIGER Study. *Medicine and Science in Sports and Exercise*. 2010;42(10):1959-65.
165. Gallagher D, Heymsfield SB, Heo M, Jebb SA, Murgatroyd PR, Sakamoto Y. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *The American journal of clinical nutrition*. 2000;72(3):694-701.

166. Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao D, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *International journal of obesity*. 2002;26(6):789-96.
167. Pasco JA, Nicholson GC, Brennan SL, Kotowicz MA. Prevalence of obesity and the relationship between the body mass index and body fat: cross-sectional, population-based data. *PloS one*. 2012;7(1):e29580.
168. Gallagher D, Visser M, Sepulveda D, Pierson RN, Harris T, Heymsfield SB. How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups? *American journal of epidemiology*. 1996;143(3):228-39.
169. Wang J, Thornton JC, Russell M, Burastero S, Heymsfield S, Pierson Jr RN. Asians have lower body mass index (BMI) but higher percent body fat than do whites: comparisons of anthropometric measurements. *The American journal of clinical nutrition*. 1994;60(1):23-8.
170. Goel K, Gupta N, Misra A, Poddar P, Pandey RM, Vikram NK, et al. Predictive equations for body fat and abdominal fat with DXA and MRI as reference in Asian Indians. *Obesity (Silver Spring)*. 2008;16(2):451-6.
171. Ramos-Jimenez A, Hernandez-Torres RP, Murguia-Romero M. Anthropometric equations for calculating body fat in young adults. *Archivos Latinoamericanos De Nutricion*. 2018;68(2):111-21.
172. Johnson W, Chumlea WC, Czerwinski SA, Demerath EW. Concordance of the recently published body adiposity index with measured body fat percent in European-American adults. *Obesity (Silver Spring)*. 2012;20(4):900-3.
173. Fedewa MV, Nickerson BS, Esco MR. Associations of body adiposity index, waist circumference, and body mass index in young adults. *Clinical Nutrition*. 2019;38(2):715-20.
174. Sun G, French CR, Martin GR, Younghusband B, Green RC, Xie Y-g, et al. Comparison of multifrequency bioelectrical impedance analysis with dual-energy X-ray absorptiometry for assessment of percentage body fat in a large, healthy population. *The American journal of clinical nutrition*. 2005;81(1):74-8.
175. Tucker LA, Greenwell SD. Using simple measures to estimate body fat percentage in college men. *American Journal of Health Behavior*. 2001;25(5):460-7.
176. Zanovec M, Johnson L, Marx B, Keenan M, Tuuri G. Self-reported physical activity improves prediction of body fatness in young adults. *Medicine+ Science in Sports+ Exercise*. 2009;41(2):328.
177. Lee G, Chang J, Hwang SS, Son JS, Park SM. Development and validation of prediction equations for the assessment of muscle or fat mass using anthropometric measurements, serum creatinine level, and lifestyle factors among Korean adults. *Nutrition Research and Practice*. 2021;15(1):95-105.
178. Medicine ACoS. ACSM's health-related physical fitness assessment manual: Lippincott Williams & Wilkins; 2013.
179. Sproule J, Kunalan C, McNeill M, Wright H. Validity of 20-MST for predicting VO2max of adult Singaporean athletes. *British journal of sports medicine*. 1993;27:202-4.
180. Tsigilis N, Douda H, Tokmakidis SP. Test-retest reliability of the Eurofit test battery administered to university students. *Perceptual and motor skills*. 2002;95:1295-300.
181. Cooper S-M, Baker JS, Tong RJ, Roberts E, Hanford M. The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *British journal of sports medicine*. 2005;39:e19.

182. Lamb KL, Rogers L. A re-appraisal of the reliability of the 20 m multi-stage shuttle run test. *European journal of applied physiology*. 2007;100:287-92.
183. Aandstad A, Holme I, Berntsen S, Anderssen SA. Validity and reliability of the 20 meter shuttle run test in military personnel. *Military medicine*. 2011;176:513-8.
184. Kim J, Jung SH, Cho HC. Validity and Reliability of Shuttle-Run Test in Korean Adults. *International Journal of Sports Medicine*. 2011;32:580-5.
185. Metsios GS, Flouris AD, Koutedakis Y, Nevill A. Criterion-related validity and test-retest reliability of the 20m square shuttle test. *Journal of science and medicine in sport*. 2008;11:214-7.
186. Penry JT, Wilcox AR, Yun J. Validity and reliability analysis of Cooper's 12-minute run and the multistage shuttle run in healthy adults. *Journal of strength and conditioning research*. 2011;25:597-605.
187. Flouris AD, Metsios GS, Famisis K, Geladas N, Koutedakis Y. Prediction of VO₂max from a new field test based on portable indirect calorimetry. *Journal of Science and Medicine in Sport*. 2010;13:70-3.
188. Mikawa K, Yano Y, Senju H. Development of a field test for evaluating aerobic fitness. *International journal of sports medicine*. 2012;33:346-50.
189. Taylor S, Frost H, Taylor A, Barker K. Reliability and responsiveness of the shuttle walking test in patients with chronic low back pain. *Physiotherapy research international : the journal for researchers and clinicians in physical therapy*. 2001;6:170-8.
190. Jurgensen SP, Trimer R, Dourado VZ, Di Thommazo-Luporini L, Bonjorno-Junior JC, Oliveira CR, et al. Shuttle walking test in obese women: test-retest reliability and concurrent validity with peak oxygen uptake. *Clinical physiology and functional imaging*. 2015;35:120-6.
191. Curb JD, Ceria-Ulep CD, Rodriguez BL, Grove J, Guralnik J, Willcox BJ, et al. Performance-based measures of physical function for high-function populations. *Journal of the American Geriatrics Society*. 2006;54:737-42.
192. Reuter SE, Massy-Westropp N, Evans AM. Reliability and validity of indices of hand-grip strength and endurance. *Australian occupational therapy journal*. 2011;58:82-7.
193. Wilken JM, Darter BJ, Goffar SL, Ellwein JC, Snell RM, Tomalis EA, et al. Physical performance assessment in military service members. *The Journal of the American Academy of Orthopaedic Surgeons*. 2012;20 Suppl 1:S42-7.
194. Larsson UE, Reynisdottir S. The six-minute walk test in outpatients with obesity: reproducibility and known group validity. *Physiotherapy research international : the journal for researchers and clinicians in physical therapy*. 2008;13:84-93.
195. Simmonds MJ, Olson SL, Jones S, Hussein T, Lee CE, Novy D, et al. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine*. 1998;23:2412-21.
196. Laukkanen RMT, Kukkonen-Harjula TK, Oja P, Pasanen ME, Vuori IM. Prediction of change in maximal aerobic power by the 2-km walk test after walking training in middle-aged adults. *International Journal of Sports Medicine*. 2000;21:113-6.
197. Gabriel KKP, Rankin RL, Lee C, Charlton ME, Swan PD, Ainsworth BE, et al. Test-Retest Reliability and Validity of the 400-Meter Walk Test in Healthy, Middle-Aged Women. *Journal of Physical Activity & Health*. 2010;7:649-57.
198. Larsen GE, George JD, Alexander JL, Fellingham GW, Aldana SG, Parcell AC. Prediction of maximum oxygen consumption from walking, jogging, or running. *Research quarterly for exercise and sport*. 2002;73:66-72.

199. Léger L, Bouchet R. An indirect continuous running multistage field test: the Université de Montréal Track Test. *Canadian Journal of Applied Sport sciences*. 1980;5:77-84.
200. Beriault K, Carpentier AC, Gagnon C, Menard J, Baillargeon J-P, Ardilouze J-L, et al. Reproducibility of the 6-minute walk test in obese adults. *International journal of sports medicine*. 2009;30:725-7.
201. Carvalho LP, Di Thommazo-Luporini L, Aubertin-Leheudre M, Bonjorno Junior JC, de Oliveira CR, Luporini RL, et al. Prediction of Cardiorespiratory Fitness by the Six-Minute Step Test and Its Association with Muscle Strength and Power in Sedentary Obese and Lean Young Women: A Cross-Sectional Study. *Plos One*. 2015;10.
202. Arcuri JF, Borghi-Silva A, Labadessa IG, Sentanin AC, Candolo C, Pires Di Lorenzo VA. Validity and Reliability of the 6-Minute Step Test in Healthy Individuals: A Cross-sectional Study. *Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine*. 2016;26:69-75.
203. Giacomantonio N, Morrison P, Rasmussen R, MacKay-Lyons MJ. Reliability and Validity of the 6-Minute Step Test for Clinical Assessment of Cardiorespiratory Fitness in People at Risk of Cardiovascular Disease. *Journal of strength and conditioning research*. 2018.
204. Buckley JP, Sim J, Eston RG, Hession R, Fox R. Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. *British journal of sports medicine*. 2004;38:197-205.
205. Sykes K, Roberts A. The Chester step test—a simple yet effective tool for the prediction of aerobic capacity. *Physiotherapy*. 2004;90:183-8.
206. Ritchie C, Trost SG, Brown W, Armit C. Reliability and validity of physical fitness field tests for adults aged 55 to 70 years. *Journal of science and medicine in sport*. 2005;8:61-70.
207. Santa Maria DL, Kinnear GR, Kearney JT, Martin PT. The Objectivity, Reliability, and Validity of the OSU Step Test for College Males. *Research Quarterly American Alliance for Health, Physical Education and Recreation*. 1976;47:445-52.
208. Gerodimos V. Reliability of Handgrip Strength Test in Basketball Players. *Journal of Human Kinetics*. 2012;31:25-36.
209. Savva C, Karagiannis C, Rushton A. Test-retest reliability of grip strength measurement in full elbow extension to evaluate maximum grip strength. *The Journal of hand surgery, European volume*. 2013;38:183-6.
210. Bohannon RW. Test-retest reliability of the five-repetition sit-to-stand test: a systematic review of the literature involving adults. *Journal of strength and conditioning research*. 2011;25:3205-7.
211. van Meeteren J, van Rijn RM, Selles RW, Roebroek ME, Stam HJ. Grip strength parameters and functional activities in young adults with unilateral cerebral palsy compared with healthy subjects. *Journal of rehabilitation medicine*. 2007;39:598-604.
212. Boissy P, Bourbonnais D, Carlotti MM, Gravel D, Arsenault BA. Maximal grip force in chronic stroke subjects and its relationship to global upper extremity function. *Clinical rehabilitation*. 1999;13:354-62.
213. Chkeir A, Jaber R, Hewson DJ, Duchene J. Reliability and validity of the Grip-Ball dynamometer for grip-strength measurement. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference*. 2012;2012:1996-9.
214. Lusardi M BR. Hand Grip Strength: Comparability of Measurements Obtained with a Jamar Dynamometer and a Modified Sphygmomanometerle. *Journal of Hand Therapy*. 1991;4:117-22.

215. Hamilton GF, McDonald C, Chenier TC. Measurement of grip strength: validity and reliability of the sphygmomanometer and jamar grip dynamometer. *The Journal of orthopaedic and sports physical therapy*. 1992;16:215-9.
216. Cadenas-Sanchez C, Sanchez-Delgado G, Martinez-Tellez B, Mora-Gonzalez J, Lof M, Espana-Romero V, et al. Reliability and Validity of Different Models of TKK Hand Dynamometers. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association*. 2016;70:7004300010.
217. Bohannon RW. Test-retest reliability of the MicroFET 4 hand-grip dynamometer. *Physiotherapy theory and practice*. 2006;22:219-21.
218. Reijnierse EM, de Jong N, Trappenburg MC, Blauw GJ, Butler-Browne G, Gapeyeva H, et al. Assessment of maximal handgrip strength: how many attempts are needed? *Journal of cachexia, sarcopenia and muscle*. 2017;8:466-74.
219. Peolsson A, Hedlund R, Oberg B. Intra- and inter-tester reliability and reference values for hand strength. *Journal of rehabilitation medicine*. 2001;33:36-41.
220. Plant CE, Parsons NR, Edwards AT, Rice H, Denninson K, Costa ML. A comparison of electronic and manual dynamometry and goniometry in patients with fracture of the distal radius and healthy participants. *Journal of hand therapy : official journal of the American Society of Hand Therapists*. 2016;29:73-80; quiz
221. Hamilton A, Balnave R, Adams R. Grip strength testing reliability. *Journal of hand therapy : official journal of the American Society of Hand Therapists*. 1994;7:163-70.
222. Shechtman O, MacKinnon L, Locklear C. Using the BTE Primus to measure grip and wrist flexion strength in physically active wheelchair users: an exploratory study. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association*. 2001;55:393-400.
223. Shechtman O, Davenport R, Malcolm M, Nabavi D. Reliability and validity of the BTE-Primus grip tool. *Journal of hand therapy : official journal of the American Society of Hand Therapists*. 2003;16:36-42.
224. Shechtman O, Gestewitz L, Kimble C. Reliability and validity of the DynEx dynamometer. *Journal of hand therapy : official journal of the American Society of Hand Therapists*. 2005;18:339-47.
225. Coldham F, Lewis J, Lee H. The reliability of one vs. three grip trials in symptomatic and asymptomatic subjects. *Journal of hand therapy : official journal of the American Society of Hand Therapists*. 2006;19:318-26; quiz 27.
226. Svantesson U, Norde M, Svensson S, Brodin E. A comparative study of the Jamar (R) and the Grippit (R) for measuring handgrip strength in clinical practice. *Isokinetics and Exercise Science*. 2009;17:85-91.
227. Mathiowetz V. Comparison of Rolyan and Jamar dynamometers for measuring grip strength. *Occupational therapy international*. 2002;9:201-9.
228. Shechtman O. The coefficient of variation as a measure of sincerity of effort of grip strength, Part I: the statistical principle. *Journal of hand therapy : official journal of the American Society of Hand Therapists*. 2001;14:180-7.
229. Espana-Romero V, Artero EG, Santaliestra-Pasias AM, Gutierrez A, Castillo MJ, Ruiz JR. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *Journal of Hand Surgery*. 2008;33:378-84.
230. Ruiz JR, Espana-Romero V, Ortega FB, Sjöström M, Castillo MJ, Gutierrez A. Hand span influences optimal grip span in male and female teenagers. *Journal of Hand Surgery*. 2006;31:1367-72.
231. Ruiz-Ruiz J, Mesa JL, Gutiérrez A, Castillo MJ. Hand size influences optimal grip span in women but not in men. *Journal of Hand Surgery, American*. 2002;27:897-901.

232. Sanchez-Delgado G, Adenas-Sanchez C, Mora-Gonzalez J, Martinez-Tellez B, Chillón P, Löf M, et al. Assessment of handgrip strength in preschool children aged 3 to 5 years. *Journal of Hand Surgery, European*. 2015;40:966-72.
233. Ten Hoor GA, Musch K, Meijer K, Plasqui G. Test-retest reproducibility and validity of the back-leg-chest strength measurements. *Isokinetics and Exercise Science*. 2016;24:209-16.
234. Coldwells A, Atkinson G, Reilly T. Sources of variation in back and leg dynamometry. *Ergonomics*. 1994;37:79-86.
235. Whitehead PN, Schilling BK, Peterson DD, Weiss LW. Possible new modalities for the Navy physical readiness test. *Military medicine*. 2012;177:1417-25.
236. Keller A, Hellesnes J, Brox JI. Reliability of the isokinetic trunk extensor test, Biering-Sorensen test, and Astrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. *Spine*. 2001;26:771-7.
237. Ljungquist T, Harms-Ringdahl K, Nygren A, Jensen I. Intra- and inter-rater reliability of an 11-test package for assessing dysfunction due to back or neck pain. *Physiotherapy research international : the journal for researchers and clinicians in physical therapy*. 1999;4:214-32.
238. Ito T, Shirado O, Suzuki H, Takahashi M, Kaneda K, Strax TE. Lumbar trunk muscle endurance testing: an inexpensive alternative to a machine for evaluation. *Archives of physical medicine and rehabilitation*. 1996;77:75-9.
239. del Pozo-Cruz B, Mocholi MH, del Pozo-Cruz J, Parraca JA, Adsuar JC, Gusi N. Reliability and validity of lumbar and abdominal trunk muscle endurance tests in office workers with nonspecific subacute low back pain. *Journal of back and musculoskeletal rehabilitation*. 2014;27:399-408.
240. Reiman MP, Krier AD, Nelson JA, Rogers MA, Stuke ZO, Smith BS. Reliability of alternative trunk endurance testing procedures using clinician stabilization vs. traditional methods. *Journal of strength and conditioning research*. 2010;24:730-6.
241. Kahraman BO, Sengul YS, Kahraman T, Kalemci O, Ozcan Kahraman B, Salik Sengul Y, et al. Developing a Reliable Core Stability Assessment Battery for Patients With Nonspecific Low Back Pain. *Spine*. 2016;41:E844-E50.
242. Hyytiäinen K, Salminen J, Suviö T, Wickström G, Pentti J. Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. *Scan J Rehab Med*. 1991;23:3-10.
243. Larsson H, Tegern M, Monnier A, Skoglund J, Helander C, Persson E, et al. Content Validity Index and Intra- and Inter-Rater Reliability of a New Muscle Strength/Endurance Test Battery for Swedish Soldiers. *PloS one*. 2015;10:e0132185.
244. Latimer J, Maher CG, Refshauge K, Colaco I. The reliability and validity of the Biering-Sorensen test in asymptomatic subjects and subjects reporting current or previous nonspecific low back pain. *Spine*. 1999;24:2085-9; discussion 90.
245. Gruther W, Wick F, Paul B, Leitner C, Posch M, Matzner M, et al. Diagnostic accuracy and reliability of muscle strength and endurance measurements in patients with chronic low back pain. *Journal of rehabilitation medicine*. 2009;41:613-9.
246. Evans K, Refshauge KM, Adams R. Trunk muscle endurance tests: reliability, and gender differences in athletes. *Journal of science and medicine in sport*. 2007;10:447-55.
247. Greene PF, Durall CJ, Kernozek TW. Intersession Reliability and Concurrent Validity of Isometric Endurance Tests for the Lateral Trunk Muscles. *Journal of Sport Rehabilitation*. 2012;21:161-6.

248. Durall CJ, Greene PF, Kernozek TW. A comparison of two isometric tests of trunk flexor endurance. *Journal of strength and conditioning research*. 2012;26:1939-44.
249. De Blaiser C, De Ridder R, Willems T, Danneels L, Vanden Bossche L, Palmans T, et al. Evaluating abdominal core muscle fatigue: Assessment of the validity and reliability of the prone bridging test. *Scandinavian journal of medicine & science in sports*. 2018;28:391-9.
250. Diener M, Golding L, Diener D. Validity and reliability of a one-minute half sit-up test of abdominal strength and endurance. *Sports Medicine, Training and Rehabilitation: An International Journal*. 1995;6:105-19.
251. Moreland J, Finch E, Stratford P, Balsor B, Gill C. Interrater reliability of six tests of trunk muscle function and endurance. *The Journal of orthopaedic and sports physical therapy*. 1997;26:200-8.
252. ROBERTSON LD, MAGNUSDOTTIR N. Evaluation of Criteria Associated with Abdominal Fitness Testing. *RESEARCH QUARTERLY FOR EXERCISE AND SPORT*. 1987;58:355-9.
253. Suni JH, Oja P, Laukkanen RT, Miilunpalo SI, Pasanen ME, Vuori IM, et al. Health-related fitness test battery for adults: aspects of reliability. *Archives of physical medicine and rehabilitation*. 1996;77:399-405.
254. Vincent WJ, Britten SD. Evaluation Of The Curl Up asubstitute for the bent knee sit up. *Journal of Physical Education and Recreation*. 1980;51:74-5.
255. Brotons-Gil E, Garcia-Vaquero MP, Peco-Gonzalez N, Vera-Garcia FJ. Flexion-rotation trunk test to assess abdominal muscle endurance: reliability, learning effect, and sex differences. *Journal of strength and conditioning research*. 2013;27:1602-8.
256. Juan-Recio C, Lopez-Plaza D, Barbado Murillo D, Pilar Garcia-Vaquero M, Vera-Garcia FJ. Reliability assessment and correlation analysis of 3 protocols to measure trunk muscle strength and endurance. *JOURNAL OF SPORTS SCIENCES*. 2018;36:357-64.
257. Wolinsky FD, Miller DK, Andresen EM, Malmstrom TK, Miller JP. Reproducibility of physical performance and physiologic assessments. *Journal of aging and health*. 2005;17:111-24.
258. Bohannon RW, Bubela DJ, Magasi SR, Gershon RC. Relative reliability of three objective tests of limb muscle strength. *Isokinetics and Exercise Science*. 2011;19:77-81.
259. Kahraman T, Ozcan Kahraman B, Salik Sengul Y, Kalemci O. Assessment of sit-to-stand movement in nonspecific low back pain: a comparison study for psychometric properties of field-based and laboratory-based methods. *International journal of rehabilitation research Internationale Zeitschrift fur Rehabilitationsforschung Revue internationale de recherches de readaptation*. 2016;39:165-70.
260. LeBrasseur NK, Bhasin S, Miciek R, Storer TW. Tests of muscle strength and physical function: reliability and discrimination of performance in younger and older men and older men with mobility limitations. *Journal of the American Geriatrics Society*. 2008;56:2118-23.
261. Fielitz L, Coelho J, Horne T, Brechue W. Inter-Rater Reliability and Intra-Rater Reliability of Assessing the 2-Minute Push-Up Test. *Military medicine*. 2016;181:167-72.
262. Negrete RJ, Hanney WJ, Kolber MJ, Davies GJ, Ansley MK, McBride AB, et al. Reliability, minimal detectable change, and normative values for tests of upper extremity function and power. *Journal of strength and conditioning research*. 2010;24:3318-25.
263. Clemons JM. Construct validity of a modification of the flexed arm hang test. *Journal of strength and conditioning research*. 2014;28:3523-30.

264. Haitz K, Shultz R, Hodgins M, Matheson GO. Test-retest and interrater reliability of the functional lower extremity evaluation. *The Journal of orthopaedic and sports physical therapy*. 2014;44:947-54.
265. Munro AG, Herrington LC. Between-session reliability of four hop tests and the agility T-test. *Journal of strength and conditioning research*. 2011;25:1470-7.
266. Moir G, Shastri P, Connaboy C. Intersession reliability of vertical jump height in women and men. *Journal of strength and conditioning research*. 2008;22:1779-84.
267. Ortiz A, Olson SL, Roddey TS, Morales J. Reliability of selected physical performance tests in young adult women. *Journal of strength and conditioning research*. 2005;19:39-44.
268. Clemons JM, Campbell B, Jeanson C. VALIDITY AND RELIABILITY OF A NEW TEST OF UPPER BODY POWER. *Journal of Strength and Conditioning Research*. 2010;24:1559-65.
269. Cowley PM, Swensen TC. Development and reliability of two core stability field tests. *Journal of strength and conditioning research*. 2008;22:619-24.
270. Bozic PR, Pazin NR, Berjan BB, Planic NM, Cuk ID. EVALUATION OF THE FIELD TESTS OF FLEXIBILITY OF THE LOWER EXTREMITY: RELIABILITY AND THE CONCURRENT AND FACTORIAL VALIDITY. *Journal of Strength and Conditioning Research*. 2010;24:2523-31.
271. Minkler S, Patterson P. The validity of the modified sit-and-reach test in college-age students. *Research quarterly for exercise and sport*. 1994;65:189-92.
272. Wear. Relationship of Flexibility Measurements to Length of Body Segments Title. *The Research Quarterly*. 1963;Vol. 34, N.
273. Liemohn WP, Sharpe GL, Wasserman JF. Lumbosacral movement in the sit-and-reach and in Cailliet's protective-hamstring stretch. *Spine*. 1994;19:2127-30.
274. Hui SC, Yuen PY, Morrow JRJ, Jackson AW. Comparison of the criterion-related validity of sit-and-reach tests with and without limb length adjustment in Asian adults. *Research quarterly for exercise and sport*. 1999;70:401-6.
275. Shephard RJ, Berridge M, Montelpare W. On the generality of the "sit and reach" test: an analysis of flexibility data for an aging population. *Research quarterly for exercise and sport*. 1990;61:326-30.
276. Ayala F, de Baranda RS, De Ste Croix M, Santonja F. Reproducibility and criterion-related validity of the sit and reach test and toe touch test for estimating hamstring flexibility in recreationally active young adults. *Physical Therapy in Sport*. 2012;13:219-26.
277. Hui SS, Yuen PY. Validity of the modified back-saver sit-and-reach test: a comparison with other protocols. *Medicine and science in sports and exercise*. 2000;32:1655-9.
278. Lopez Minarro PA, de Baranda Andujar PS, Rodriguez Garcia PL, Ortega Toro E. A comparison of the spine posture among several sit-and-reach test protocols. *Journal of Science and Medicine in Sport*. 2007;10:456-62.
279. Leard JS, Crane BA, Ball KA. INTRARATER AND INTERRATER RELIABILITY OF 22 CLINICAL MEASURES ASSOCIATED WITH LOWER QUARTER MALALIGNMENT. *Journal of Manipulative and Physiological Therapeutics*. 2009;32:270-6.
280. Atamaz F, Ozcaldiran B, Ozdedeli S, Capaci K, Durmaz B. Interobserver and intraobserver reliability in lower-limb flexibility measurements. *The Journal of sports medicine and physical fitness*. 2011;51:689-94.

281. Perret C, Poiraudéau S, Fermanian J, Colau MML, Benhamou MAM, Revel M. Validity, reliability, and responsiveness of the fingertip-to-floor test. *Archives of Physical Medicine and Rehabilitation*. 2001;82:1566-70.
282. KIPPERS V, PARKER aAW. Toe-Touch Test A Measure of Its Validity. *PHYSICAL THERAPY*. 1987;67 (11).
283. Gauvin MG, Riddle DL, Rothstein JM. Reliability of Clinical Measurements of Forward Bending Using the Modified Fingertip-to-Floor Method. *Physical Therap*. 1990;70 (7).
284. Dijkstra PU, de Bont LG, van der Weele LT, Boering G. Joint mobility measurements: reliability of a standardized method. *Cranio : the journal of craniomandibular practice*. 1994;12:52-7.
285. Jackson AW, Morrow JRJ, Jensen RL, Jones NA, Schultes SS. Reliability of the prudential FITNESSGRAM trunk lift test in young adults. *Research quarterly for exercise and sport*. 1996;67:115-7.
286. van den Dolder PA, Ferreira PH, Refshauge K. Intra- and inter-rater reliability of a modified measure of hand behind back range of motion. *Manual therapy*. 2014;19:72-6.
287. Kent M. Oxford dictionary of sports science and medicine: OUP Oxford; 2006.
288. Medicine ACoS. ACSM's guidelines for exercise testing and prescription: Lippincott williams & wilkins; 2013.
289. Rinne MB, Pasanen ME, Miilunpalo SI, Oja P. Test-retest reproducibility and inter-rater reliability of a motor skill test battery for adults. *International journal of sports medicine*. 2001;22:192-200.
290. Kinzey SJ, Armstrong CW. The reliability of the star-excursion test in assessing dynamic balance. *The Journal of orthopaedic and sports physical therapy*. 1998;27:356-60.
291. Gribble PA, Kelly SE, Refshauge KM, Hiller CE. Interrater reliability of the star excursion balance test. *Journal of athletic training*. 2013;48:621-6.
292. Punakallio A. Trial-to-trial reproducibility and test-retest stability of two dynamic balance tests among male firefighters. *International journal of sports medicine*. 2004;25:163-9.
293. López-Plaza D, Juan-Recio C, Barbado D, Ruiz-Pérez I, Vera-Garcia FJ. Reliability of the star excursion balance test and two new similar protocols to measure trunk postural control. *PM&R*. 2018;10(12):1344-52.
294. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*. 1991;39:142-8.
295. Spagnuolo DL, Jurgensen SP, Iwama AM, Dourado VZ. Walking for the assessment of balance in healthy subjects older than 40 years. *Gerontology*. 2010;56:467-73.
296. Raya MA, Gailey RS, Gaunard IA, Jayne DM, Campbell SM, Gagne E, et al. Comparison of three agility tests with male servicemembers: Edgren Side Step Test, T-Test, and Illinois Agility Test. *Journal of rehabilitation research and development*. 2013;50:951-60.
297. Sassi RH, Dardouri W, Yahmed MH, Gmada N, Mahfoudhi ME, Gharbi Z. RELATIVE AND ABSOLUTE RELIABILITY OF A MODIFIED AGILITY T-TEST AND ITS RELATIONSHIP WITH VERTICAL JUMP AND STRAIGHT SPRINT. *Journal of Strength and Conditioning Research*. 2009;23:1644-51.

298. Beekhuizen KS, Davis MD, Kolber MJ, Cheng M-SS. Test-retest reliability and minimal detectable change of the hexagon agility test. *Journal of strength and conditioning research*. 2009;23:2167-71.
299. Manderoos SA, Vaara ME, Maki PJ, Malkia EA, Aunola SK, Karppi S-L. A New Agility Test for Adults: Its Test-Retest Reliability and Minimal Detectable Change in Untrained Women and Men Aged 28-55. *Journal of strength and conditioning research*. 2016;30:2226-34.
300. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *International journal of sports medicine*. 1991;12(04):356-62.
301. Laukkanen R, Oja P, Pasanen M, Vuori I. Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*. 1992;16(4):263-8.
302. Laukkanen R, Kukkonen-Harjula T, Oja P, Pasanen M, Vuori I. Prediction of change in maximal aerobic power by the 2-km walk test after walking training in middle-aged adults. *International journal of sports medicine*. 2000;21(02):113-6.
303. Laukkanen R, Oja R, Pasanen M, Vuori I. Criterion validity of a two-kilometer walking test for predicting the maximal oxygen uptake of moderately to highly active middle-aged adults. *Scandinavian journal of medicine & science in sports*. 1993;3(4):267-72.
304. Mayorga-Vega D, Aguilar-Soto P, Viciano J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *Journal of sports science & medicine*. 2015;14(3):536.
305. Mayorga-Vega D, Bocanegra-Parrilla R, Ornelas M, Viciano J. Criterion-related validity of the distance-and time-based walk/run field tests for estimating cardiorespiratory fitness: a systematic review and meta-analysis. *PloS one*. 2016;11(3):e0151671.
306. Cohen J. Statistical power analysis for the behavioral sciences: Routledge; 2013.
307. Cuenca-Garcia M, Marin-Jimenez N, Perez-Bey A, Sánchez-Oliva D, Camiletti-Moiron D, Alvarez-Gallardo IC, et al. Reliability of Field-Based Fitness Tests in Adults: A Systematic Review. *Sports Medicine*. 2022:1-20.
308. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine*. 1998;26(4):217-38.
309. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Manual therapy*. 1999;4(2):87-93.
310. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research*. 2005;19:231-40.
311. Ortega FB, Cadenas-Sánchez C, Sánchez-Delgado G, Mora-González J, Martínez-Téllez B, Artero EG, et al. Systematic review and proposal of a field-based physical fitness-test battery in preschool children: the PREFIT battery. *Sports medicine*. 2015;45(4):533-55.
312. Ruiz JR, Castro-Piñero J, España-Romero V, Artero EG, Ortega FB, Cuenca MM, et al. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *British journal of sports medicine*. 2011;45(6):518-24.
313. Castro-Piñero J, Marin-Jimenez N, Fernandez-Santos JR, Martin-Acosta F, Segura-Jimenez V, Izquierdo-Gomez R, et al. Criterion-Related Validity of Field-Based

- Fitness Tests in Adults: A Systematic Review. *Journal of clinical medicine*. 2021;10(16):3743.
314. Fernandez-Santos JR, Gonzalez-Montesinos JL, Ruiz JR, Jiménez-Pavón D, Castro-Piñero J. Kinematic analysis of the standing long jump in children 6-to 12-years-old. *Measurement in Physical Education and Exercise Science*. 2018;22(1):70-8.
 315. Mackala K, Stodółka J, Siemiński A, Coh M. Biomechanical analysis of standing long jump from varying starting positions. *The Journal of Strength & Conditioning Research*. 2013;27(10):2674-84.
 316. Grosprêtre S, Ufland P, Jecker D. The adaptation to standing long jump distance in parkour is performed by the modulation of specific variables prior and during take-off. *Movement & Sport Sciences-Science & Motricité*. 2018(100):27-37.
 317. Smothermon JR. Do isokinetic variables generated correlate to vertical jump and standing long jump measures in collegiate football athletes? 2016.
 318. Castro-Piñero J, Ortega FB, Artero EG, Girela-Rejón MJ, Mora J, Sjöström M, et al. Assessing muscular strength in youth: usefulness of standing long jump as a general index of muscular fitness. *The Journal of Strength & Conditioning Research*. 2010;24(7):1810-7.
 319. Fernandez-Santos JR, Ruiz JR, Cohen DD, Gonzalez-Montesinos JL, Castro-Piñero J. Reliability and validity of tests to assess lower-body muscular power in children. *The Journal of Strength & Conditioning Research*. 2015;29(8):2277-85.
 320. Milliken LA, Faigenbaum AD, Loud RL, Westcott WL. Correlates of upper and lower body muscular strength in children. *The Journal of Strength & Conditioning Research*. 2008;22(4):1339-46.
 321. Sale D, Norman R. Testing strength and power. *Physiological testing of the high-performance athlete*. 1991:21-106.
 322. Mayhew J, Salm PC. Gender differences in anaerobic power tests. *European journal of applied physiology and occupational physiology*. 1990;60(2):133-8.
 323. Wolfe RR. The underappreciated role of muscle in health and disease. *The American journal of clinical nutrition*. 2006;84(3):475-82.
 324. García-Hermoso A, Cavero-Redondo I, Ramírez-Vélez R, Ruiz JR, Ortega FB, Lee D-C, et al. Muscular strength as a predictor of all-cause mortality in an apparently healthy population: a systematic review and meta-analysis of data from approximately 2 million men and women. *Archives of physical medicine and rehabilitation*. 2018;99(10):2100-13. e5.
 325. España-Romero V, Artero E, Jimenez-Pavón D, Cuenca-Garcia M, Ortega F, Castro-Piñero J, et al. Assessing health-related fitness tests in the school setting: reliability, feasibility and safety; the ALPHA Study. *International journal of sports medicine*. 2010;31(07):490-7.
 326. Cadenas-Sanchez C, Martinez-Tellez B, Sanchez-Delgado G, Mora-Gonzalez J, Castro-Piñero J, Löf M, et al. Assessing physical fitness in preschool children: Feasibility, reliability and practical recommendations for the PREFIT battery. *Journal of science and medicine in sport*. 2016;19(11):910-5.
 327. Wu W-L, Wu J-H, Lin H-T, Wang G-J. Biomechanical analysis of the standing long jump. *Biomedical Engineering: Applications, Basis and Communications*. 2003;15(05):186-92.
 328. Wells J, Fewtrell M. Measuring body composition. *Archives of disease in childhood*. 2006;91(7):612-7.

ANNEXES



SHORT CV

CURRICULUM VITAE [SHORT CV]

Personal Data

Name: Nuria Marín Jiménez

E-mail: nuria.marin@uca.es

Birth Date: 27/04/1992

Birth Place: Jaén

Nationality: Spanish

ORCID profile: [0000-0002-0687-7554](https://orcid.org/0000-0002-0687-7554)

Research Gate: Nuria Marin-Jimenez

Education and training

2019-2022 PhD candidate in Health Sciences. University of Cádiz, Spain.

2017-.... PhD candidate in Biomedicine. University of Granada, Spain.

2017-2018 Master's degree in Research on Physical activity and Health Specializing in Physical Activity and Health. University of Granada, Spain.

2016-2017 Master's Degree in Teaching Compulsory and Pre-University Secondary Education, Vocational Training and Language Teaching. Specializing in Physical Education. University of Granada, Spain.

2011-2016 Bachelor's Degree in Sport Sciences. University of Granada, Spain.

Erasmus Student Internship - 9 months at University of Tallin, Estonia (2015-2016).

National fellowships

2019-2022 Regional Government of Andalusia and University of Cadiz: Research and Knowledge Transfer Fund (PPIT-FPI19).

National internships

2021 Instituto Mixto Universitario Deporte y Salud, University of Granada. Professor Virginia A. Aparicio-Molina (2 months).

2019 Faculty of Sport Sciences, Polytechnic University of Madrid. Professor Rocio Cupeiro Coto (3 months).

International internships

- 2022 Funding: Erasmus+ Programme (K107) for teaching. Center: Tbilisi State University, Georgia (1 week).
- 2021 Funding: University of Cadiz: Research and Knowledge Transfer Fund. Center: Karolinska Institutet, Sweden. Professor Marie Löf (3 months).
- 2019 Funding: Erasmus+ Programme (K103) for teaching. Center: University of Tallin, Estonia (1 week).

Participation in research projects

Researcher 2019–...

Project Title & Design: Health-related physical fitness test battery design for the adult population: The ADULT-FIT project.

Principal Investigator: José Catro Piñero and Magdalena Cuenca García

Department of Physical Education and Sport, University of Cadiz, Spain

Funding: Spanish Ministry of Economy, Industry, and Competitiveness (112.000€)

Researcher 2020–...

Project Title & Design: Influencia de un programa de ejercicio físico concurrente durante el embarazo sobre la composición corporal, condición física y desarrollo motor.
Acronym: GESTAFITOS

Principal Investigator: Virginia A Aparicio

Department of Physical Education and Sport, University of Granada, Spain

Funding: I+D+I projects in the framework of FEDER Andalucía, 2014-2020 (6.400€)

Researcher 2017–2019

Project Title & Design: Effects of a supervised physical exercise intervention during pregnancy on telomere length and gene expression markers related to adiposity in the mother and neonate. A randomized controlled trial.

Principal Investigator: Virginia A Aparicio

Department of Physical Education and Sport, University of Granada, Spain

Funding: Spanish Ministry & European Funding (*MSCA-COFUND* €56000)

Researcher 2015–2017

Project Title & Design: Influence of physical activity levels, physical fitness and nutritional habits of the pregnant women on diverse foetal and maternal health markers”

Acronym: GESTAFIT.

Principal Investigator: VirginiaA Aparicio

Department of Physical Education and Sport, University of Granada, Spain

Funding: Spanish Ministry & European Funding €100000

Student Assistant-Collaborator 2015–2018

Project Title & Design: Cost-effectiveness of an exercise intervention program in perimenopausal women. The Fitness League Against MENopause COst (FLAMENCO) randomized controlled trial.

Principal Investigator: Virginia A Aparicio

Department of Physical Education and Sport, University of Granada, Spain

Funding: Spanish Ministry €38500

Scientific and organizing committee member activities

2019 XII Simposio Internacional de Fuerza y Proyecto IRONFEMME / INSCA Spain National Conference, Madrid, Spain. Organization coordinator (volunteer) and mediator.

2019 **24th Annual Congress of the European College of Sports Sciences**, Prague, Czech Republic. Volunteer.

Reviewer activity

Nuria Marín Jiménez has contributed as reviewer of indexed journals, such as *Climacteric*, and *RETOS*.

Publications (not included as part of this Doctoral Thesis)

1. **Marín-Jiménez, N**, Flor-Alemany, M, Baena-García, L, Coll-Risco, I, Castro-Piñero, J, Aparicio, V. A. Physical fitness and maternal body composition indices during pregnancy and postpartum: the GESTAFIT project. *European Journal of Sport Science*, 2022; 1-11.
2. Baena-García L, Flor-Alemany M, **Marín-Jiménez N**, Aranda P, Aparicio VA. A 16-week multicomponent exercise training program improves menopause-related symptoms in middle-aged women. The FLAMENCO project randomized control trial. *Menopause*, 2022; 29(5), 537-544.
3. Ruiz-Montero P, **Marín-Jiménez N**, Borges-Cosic M, Aparicio V. Association of objectively measured physical fitness with health-related quality of life of mid-life women: the FLAMENCO project. *Climacteric*. 2021; 24(3):282-288.
4. Romero-Parra N, Maestre-Cascales C, **Marín-Jiménez N**, et al. Exercise-Induced Muscle Damage in Postmenopausal Well-Trained Women. *Sports Health*. 2021; 13(6):613-621.
5. **Marín-Jiménez N**, Castro-Piñero J, Rodríguez-Ayllón M, Marchán-Rubio A, Delgado-Fernández M, Aparicio VA. The favourable association of self-reported physical fitness with depression and anxiety during pregnancy. The GESTAFIT project. *European Journal of Sport Science*. 2021:1-9.
6. **Marín-Jiménez N**, Borges-Cosic M, Ocón-Hernández O, et al. Association of Self-Reported Physical Fitness with Pregnancy Related Symptoms the GESTAFIT Project. *International Journal of Environmental Research and Public Health*. 2021; 18(7):3345.
7. Flor-Alemany M, Acosta P, **Marín-Jiménez N**, Baena-García L, Aranda P, Aparicio VA. Influence of the degree of adherence to the mediterranean diet and its components on cardiometabolic risk during pregnancy. The GESTAFIT project. *Nutrition, Metabolism and Cardiovascular Diseases*. 2021; 31(8):2311-2318.
8. Baena-García L, **Marín-Jiménez N**, Romero-Gallardo L, et al. Association of Self-Reported Physical Fitness during Late Pregnancy with Birth Outcomes and Oxytocin

Administration during Labour—The GESTAFIT Project. *International Journal of Environmental Research and Public Health*. 2021; 18(15):8201.

9. Baena-García L, Acosta-Manzano P, Ocón-Hernández O, Borges-Cosic M, Romero-Gallardo L, **Marín-Jiménez N**, Aparicio, VA. Objectively measured sedentary time and physical activity levels in Spanish pregnant women. Factors affecting the compliance with physical activity guidelines. *Women & Health*. 2021; 61(1):27-37.

10. Aparicio VA, Flor-Alemany M, **Marín-Jiménez N**, Coll-Risco I, Aranda P. A 16-week concurrent exercise program improves emotional well-being and emotional distress in middle-aged women: the FLAMENCO project randomized controlled trial. *Menopause*. 2021; 28(7):764-771.

11. **Marín-Jiménez N**, Ruiz-Montero PJ, De la Flor-Alemany M, Aranda P, Aparicio VA. Association of objectively measured sedentary behavior and physical activity levels with health-related quality of life in middle-aged women: The FLAMENCO project. *Menopause*. 2020; 27(4):437-443.

12. Flor-Alemany M, Nestares T, Alemany-Arrebola I, **Marín-Jiménez N**, Borges-Cosic M, Aparicio VA. Influence of Dietary Habits and Mediterranean Diet Adherence on Sleep Quality during Pregnancy. The GESTAFIT Project. *Nutrients*. 2020; 12(11):3569.

13. Flor-Alemany M, **Marín-Jiménez N**, Nestares T, Borges-Cosic M, Aranda P, Aparicio V. Mediterranean diet, tobacco consumption and body composition during perimenopause. the FLAMENCO project. *Maturitas*. 2020; 137:30-36.

14. Flor-Alemany M, **Marín-Jiménez N**, Coll-Risco I, Aranda P, Aparicio VA. Influence of dietary habits and Mediterranean diet adherence on menopausal symptoms. The FLAMENCO project. *Menopause*. 2020; 27(9):1015-1021.

15. **Marín-Jiménez N**, Acosta-Manzano P, Borges-Cosic M, et al. Association of self-reported physical fitness with pain during pregnancy: The GESTAFIT Project. *Scandinavian journal of medicine & science in sports*. 2019; 29(7):1022-1030.

16. Aparicio VA, **Marín-Jiménez N**, Coll-Risco I, et al. Doctor, ask your perimenopausal patient about her physical fitness; association of self-reported physical fitness with cardiometabolic and mental health in perimenopausal women: the FLAMENCO project. *Menopause*. 2019; 26(10):1146-1153.

Teaching experience

Lecturer at University of Cadiz (200 h)

- Valoración de la Condición Física relacionada con la Salud
- Fundamentos de las habilidades motrices
- Fisiología del ejercicio 2
- Dinamización y recreación en piscinas y playas
- Expresión corporal
- Fundamentos de los deportes individuales 1
- Actividad Física para Grupos de Población Específica
- Nuevas tendencias del Fitness y el Wellness
- Actividad Física y Salud
- Trabajo Fin de Grado
- Aula Universitaria de Mayores

Lecturer at Karolinska Institutet (10 h)

- Diet, Physical Activity and Fitness - assessment and evaluation (Master's Program in Nutrition Science)

Lecturer at University of Granada (2 h)

- Máster propio en alimentación, ejercicio y deporte para la salud (Food & Fit)

Other merits

Twenty-seven communications or posters both in national and international conferences.

LIST OF STUDIES INCLUDED

and quality indexes

LIST OF STUDIES INCLUDED AND QUALITY INDEXES

- I. **Marín-Jiménez, N.**, Cruz-León, C., Pérez-Bey, A., Conde-Caveda, J., Grao-Cruces, A., Aparicio, V. A., Castro-Piñero, J. & Cuenca-García, M. (2022). Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review. *Journal of Clinical Medicine*, 11(2), 328.
Journal Impact Factor (JCR): 4.964 (Q1).
- II. Castro-Piñero, J., **Marín-Jiménez, N.**, Fernández-Santos, J. R., Martín-Acosta, F., Segura-Jiménez, V., Izquierdo-Gómez, R., R. Ruiz, J. & Cuenca-Garcia, M. (2021). Criterion-Related Validity of Field-Based Fitness Tests in Adults: A Systematic Review. *Journal of clinical medicine*, 10(16), 3743.
Journal Impact Factor (JCR): 4.964 (Q1).
- III. **Marín-Jiménez, N.**, Cruz-León, C., Sánchez-Oliva, D., Jiménez-Iglesias, J., Caraballo, I., Padilla-Moledo, C., Cadenas-Sánchez, C., Cuenca-García, M. & Castro-Piñero, J. (2022). Criterion-Related Validity of Field-Based Methods and Equations for Body Composition estimation in Adults: A Systematic Review. *Current Obesity Reports*, Accepted, In production.
Journal Impact Factor (JCR): 8.023 (Q1).
- IV. Cuenca-García, M., **Marín-Jiménez, N.**, Pérez-Bey, A., Sánchez-Oliva, D., Camiletti-Moirón, D., Álvarez-Gallardo, I. C., Ortega, F. B. & Castro-Piñero, J. (2022). Reliability of Field-Based Fitness Tests in Adults: A Systematic Review. *Sports Medicine*, 1-20.
Journal Impact Factor (JCR): 11.928 (Q1).
- V. **Marín-Jiménez, N.**, Sánchez-Parente, S., Expósito-Carrillo, P., Jiménez-Iglesias, J., C. Álvarez-Gallardo, I., Cuenca-García, M. & Castro-Piñero, J. Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level. *Submitted**.

- VI. **Marín-Jiménez, N.,** Pérez-Bey, A., Cruz-León, C., Conde-Caveda, J., Segura-Jiménez, V., Castro-Piñero, J. & Cuenca-García, M. Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project. *Submitted***.

*Study V is currently submitted for publication in a high-impact science journal. Its current status is *under review*. Journal of Science and Medicine in Sport (JCR: 4.319, Q1).

**Study VI is currently submitted for publication in a high-impact science journal. Its current status is *submitted*. Journal of Strength and Conditioning Research (JCR: 4.415, Q1).

AGREEMENT OF CO-AUTHORS

AGREEMENT OF CO-AUTHORS AND CONTRIBUTIONS

The following pages contain the approval of the co-authors for the presentation of the included studies as part of the current Doctoral Thesis, and their resignation to present these studies as part of another Doctoral Thesis at anyother University.

Nuria Marín Jiménez states her contribution in the included studies as follows:

- Studies I, IV, V, and VI: Nuria Marín Jiménez contributed in the acquisition of the data, the creation of the databases, performed the data analyses, proceeded with the presentation of the results, and wrote the manuscript, being the first author.
- Studies II and III: Nuria Marín Jiménez contributed in the acquisition of the data, the creation of the databases, the elaboration of the manuscripts' figures, gave critical discussion of the manuscript, and contributed in the final manuscript as accepted, being the second author and the corresponding author.

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. Inmaculada C. Álvarez Gallardo con DNI 28788284N, coautora de las publicaciones que se identifican a continuación:

- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.
- Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

ALVAREZ
GALLARDO
INMACULADA
CONCEPCION
- 28788284N

Firmado digitalmente
por ALVAREZ
GALLARDO
INMACULADA
CONCEPCION -
28788284N
Fecha: 2022.10.06
11:48:58 +02'00'

Fdo.: Inmaculada C. Álvarez Gallardo

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. Virginia A. Aparicio García-Molina con DNI 53265170E, coautora de la publicación que se identifica a continuación:

- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Granada, a 5 de octubre de 2022

Fdo.:



Tesis por compendio de publicaciones

Documento de conformidad y renuncia de coautores

Dña. Cristina Cadenas Sánchez con DNI o Pasaporte: 78982081N, coautora de la publicación que se identifican a continuación:

- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada “VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT”.

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Granada, a 5 de octubre de 2022

CADENAS
SANCHEZ
CRISTINA -
78982081N

Firmado
digitalmente por
CADENAS SANCHEZ
CRISTINA -
78982081N
Fecha: 2022.10.05
23:42:55 +02'00'

Fdo.: Cristina Cadenas Sánchez

Tesis por compendio de publicaciones

Documento de conformidad y renuncia de coautores

D. Daniel Camiletti Moirón con DNI o Pasaporte: 08836816D, coautor de la publicación que se identifica a continuación:

- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

Firmado por CAMILETTI MOIRON DANIEL -
08836816D el día 05/10/2022 con un
certificado emitido por AC FNMT Usuarios

Fdo.:

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Israel Caraballo Vidal con DNI o Pasaporte: 48889518-M, coautor de las publicaciones que se identifican a continuación:

- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

**CARABALLO
VIDAL ISRAEL -
48889518M**

Fdo.:

Firmado digitalmente por
CARABALLO VIDAL ISRAEL -
48889518M
Fecha: 2022.10.06 12:43:13
+02'00'

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. José Castro Piñero con DNI o Pasaporte: 31656395-T, coautor de las publicaciones que se identifican a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.
- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.
- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.
- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.
- Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level.
- Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

CASTRO PIÑERO
Fdo.: JOSE - 31656395T

Firmado digitalmente por CASTRO
PIÑERO JOSE - 31656395T
Fecha: 2022.11.04 10:08:30 +01'00'

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Julio Conde Caveda con DNI o Pasaporte: 00820007-B, coautor de las publicaciones que se identifican a continuación:

- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.
- Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022



Fdo.: Julio Conde Caveda

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. Carolina Cruz León con DNI o Pasaporte: Y7499243S, coautora de las publicaciones que se identifican a continuación:

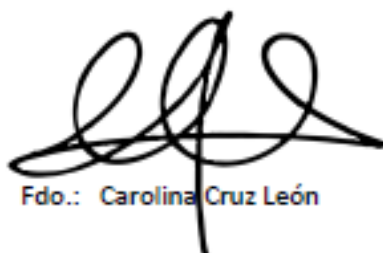
- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.
- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.
- Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022



Fdo.: Carolina Cruz León

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. **Magdalena Cuenca García** con DNI o Pasaporte: 74728962-F, coautora de las publicaciones que se identifican a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.
- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.
- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.
- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.
- Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level.
- Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. **Nuria Marín Jiménez**, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

Fdo.: CUENCA GARCIA MARIA MAGDALENA - 74728962F  Firmado digitalmente por CUENCA GARCIA MARIA MAGDALENA - 74728962F
Fecha: 2022.11.03 12:50:54 +01'00'

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Pablo Expósito Carrillo con DNI o Pasaporte 26977873T, coautor de la publicación que se identifica a continuación:

- Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

Fdo.:



Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Jorge del Rosario Fernández Santos con DNI o Pasaporte: 75778284-T, coautor de la publicación que se identifica a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

Fdo.:

**FERNANDEZ
SANTOS JORGE
DEL ROSARIO -
75778284T**

Firmado digitalmente
por FERNANDEZ
SANTOS JORGE DEL
ROSARIO -
75778284T
Fecha: 2022.10.11
15:50:19 +02'00'

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Alberto Grao Cruces con DNI o Pasaporte: 14619249N, coautor de la publicación que se identifica a continuación:

- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 6 de octubre de 2022

Fdo.:

GRAO CRUCES ALBERTO -
14619249N

Firmado digitalmente por GRAO CRUCES ALBERTO, 14619249N.
Nombre de correo electrónico: GRAO CRUCES ALBERTO, 14619249N.
DNI: 14619249N. en GRAO CRUCES ALBERTO, 14619249N.
Fecha: 2022.10.06 11:08:46 +0200

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. José Jiménez Iglesias con DNI o Pasaporte: 76648742-T, coautor de las publicaciones que se identifican a continuación:

- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.
- Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

JIMENEZ
IGLESIAS JOSE
- 76648742T

Firmado digitalmente
por JIMENEZ IGLESIAS
JOSE - 76648742T
Fecha: 2022.10.11
11:36:07 +02'00'

Fdo.:

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. **Rocío Izquierdo Gómez** con DNI o Pasaporte: 49017224S, coautora de la publicación que se identifica a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. **Nuria Marín Jiménez**, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

Fdo.:

IZQUIERDO
GOMEZ
ROCIO -
49017224S

Firmado
digitalmente por
IZQUIERDO GOMEZ
ROCIO - 49017224S
Fecha: 2022.10.06
09:51:11 +11'00'

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. **Fátima Martín-Acosta** con DNI o Pasaporte: 74735341S, coautora de la publicación que se identifica a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. **Nuria Marín Jiménez**, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022



Fdo.:

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Francisco B. Ortega Porcel con DNI o Pasaporte: 75228381-G, coautor de la publicación que se identifica a continuación:

- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT-FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Granada, a 5 de octubre de 2022



Fdo.: Francisco B. Ortega

Tesis por compendio de publicaciones Documento de conformidad y renuncia de coautores

Dña. **Carmen Padilla Moledo** con DNI o Pasaporte: 31243940-G, coautora de la publicación que se identifican a continuación:

- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. **Nuria Marín Jiménez**, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

PADILLA MOLEDO MARIA
DEL CARMEN - 31243940G
Firmado digitalmente por
PADILLA MOLEDO MARIA
DEL CARMEN - 31243940G
Fecha: 2022.10.05 20:14:52
+02'00'

En Puerto Real, a 5 de octubre de 2022

Fdo.:

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. **Alejandro Pérez Pérez** con DNI o Pasaporte: 49074232Y, coautor de las publicaciones que se identifican a continuación:

- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.
- Predictive Validity of Motor Fitness and Flexibility Tests in Adults and Older Adults: A Systematic Review.
- Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. **Nuria Marín Jiménez**, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar las citadas publicaciones como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 6 de octubre de 2022

Fdo.:

**PEREZ
PEREZ
ALEJANDRO**
- 49074232Y

Firmado digitalmente
por PEREZ PEREZ
ALEJANDRO -
49074232Y
Fecha: 2022.10.06
18:36:54 +02'00'

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Jonatan Ruiz Ruiz con DNI o Pasaporte: 46779494R, coautor de la publicación que se identifica a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT-FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Granada, a 5 de octubre de 2022



Fdo.: Jonatan Ruiz Ruiz

Tesis por compendio de publicaciones

Documento de conformidad y renuncia de coautores

D. David Sánchez Oliva con DNI o Pasaporte: 76129318-P, coautor de las publicaciones que se identifican a continuación:

- Reliability of Field-Based Fitness Tests in Adults: A Systematic Review.
- Criterion-related validity of field-based methods and equations for body composition estimation in adults: a systematic review.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de las citadas publicaciones como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

**SANCHEZ
OLIVA DAVID**
- 76129318P

Firmado digitalmente
por SANCHEZ OLIVA
DAVID - 76129318P
Fecha: 2022.10.06
09:14:18 +02'00'

Fdo.:

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

Dña. **Sandra Sánchez Parente** con DNI o Pasaporte: 15458799Q, coautora de la publicación que se identifica a continuación:

- Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: the role of sex, age and physical activity level.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT-FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Puerto Real, a 5 de octubre de 2022

Fdo.:

Firmado por SANCHEZ PARENTE SANDRA -
***5879** (el día 06/10/2022) con un
certificado electrónico emitido por AC
FNMT Usuarios

Tesis por compendio de publicaciones
Documento de conformidad y renuncia de coautores

D. Víctor Segura Jiménez con DNI o Pasaporte: 74934073G, coautor de la publicación que se identifica a continuación:

- Criterion-related validity of field-based fitness tests in adults: A systematic review.
- Criterion-related validity and reliability of the standing long jump test in adult population: The Adult-Fit project.

de acuerdo con lo establecido en el Artículo 23.4 del Reglamento UCA/CG06/2012, de 27 de junio de 2012, por el que se regula la ordenación de los estudios de doctorado en la Universidad de Cádiz (BUOCA nº 208).

Manifiesta su conformidad para la presentación de la citada publicación como parte de la tesis doctoral de Dña. Nuria Marín Jiménez, titulada "VALIDEZ Y FIABILIDAD DE TEST DE CAMPO PARA EVALUAR LA CONDICIÓN FÍSICA RELACIONADA CON LA SALUD EN ADULTOS. PROYECTO ADULT- FIT".

Y expresa su renuncia a presentar la citada publicación como parte de otra tesis doctoral en cualquier otra universidad.

En Granada, a 5 de octubre de 2022

**SEGURA
JIMENEZ
VICTOR -
74934073G**

Firmado digitalmente por SEGURA
JIMENEZ VICTOR - 74934073G
Nombre de reconocimiento (DN):
c=ES,
serialNumber=IDCES-74934073G,
givenName=VICTOR, sn=SEGURA
JIMENEZ, cn=SEGURA JIMENEZ
VICTOR - 74934073G
Fecha: 2022.10.06 09:42:50 +02'00'

