International Doctoral Thesis / Tesis Doctoral Internacional

PHYSICAL FITNESS, ACADEMIC ACHIEVEMENT AND BRAIN IN CHILDREN

CONDICIÓN FÍSICA, RENDIMIENTO ACADÉMICO Y CEREBRO EN NIÑOS

PROGRAMA OFICIAL DE DOCTORADO EN BIOMEDICINA

DEPARTAMENTO DE EDUCACIÓN FÍSICA Y DEPORTIVA
FACULTAD DE CIENCIAS DEL DEPORTE
UNIVERSIDAD DE GRANADA

CRISTINA CADENAS SÁNCHEZ

2018

Editor: Universidad de Granada. Tesis Doctorales
Autor: Cristina Cadenas Sánchez
ISBN: 978-84-9163-814-8
URI: http://hdl.handle.net/10481/50138

A todos vosotros que me hicisteis, me seguís haciendo y me haréis sonreír en cada momento

PHYSICAL FITNESS, ACADEMIC ACHIEVEMENT AND BRAIN IN CHILDREN

CONDICIÓN FÍSICA, RENDIMIENTO ACADÉMICO Y CEREBRO EN NIÑOS

CRISTINA CADENAS SÁNCHEZ

Doctoral Thesis Supervisors [Directores de la Tesis Doctoral]

Francisco B. Ortega Porcel
PhD
Profesor Contratado Doctor
Previamente Investigador
Ramón y Cajal
Universidad de Granada

Andrés Catena Martínez
PhD
Catedrático de Universidad
Universidad de Granada

Doctoral Thesis Committee [Miembros del Tribunal]

Manuel Delgado Fernández
PhD
Catedrático de Universidad
Universidad de Granada

Vicente Martínez Vizcaíno
PhD
Catedrático de Universidad
Centro de Estudios Socio-Sanitarios
Universidad de Castilla-La Mancha

Kirk I. Erickson
PhD
Professor
Center for the Neural Basis of Cognition
University of Pittsburgh

Magdalena Cuenca García
PhD
Profesora Ayudante Doctora
Universidad de Cádiz

Virginia A. Aparicio García-Molina
PhD
Profesora Contratada Doctora
Universidad de Granada

Granada, 9 de Marzo de 2018

Prof. Dr. D. Francisco B. Ortega Porcel
Profesor Contratado Doctor
Previamente Investigador Ramón y Cajal
--
Departamento de Educación Física y Deportiva
Facultad de Ciencias del Deporte
Universidad de Granada

FRANCISCO B. ORTEGA PORCEL, PROFESOR CONTRATADO DOCTOR DE LA UNIVERSIDAD DE GRANADA

CERTIFICA:

Que la Tesis Doctoral titulada "Physical fitness, academic achievement and brain in children" que presenta Dña. **Cristina Cadenas Sánchez** al superior juicio del Tribunal que designe la Universidad de Granada, ha sido realizada bajo mi dirección durante los años 2014-2018, siendo expresión de la capacidad técnica e interpretativa de su autora en condiciones tan aventajadas que la hacen merecedora del Título de Doctora por la Universidad de Granada, siempre y cuando así lo considere el citado Tribunal.

Fdo. Francisco B. Ortega Porcel

En Granada, 9 de Marzo de 2018

Prof. Dr. D. Andrés Catena Martínez
Catedrático de Universidad
--
Departamento de Psicología Experimental
Centro de Investigación Mente, Cerebro y Comportamiento
Universidad de Granada

ANDRÉS CATENA MARTÍNEZ, CATEDRÁTICO DE LA UNIVERSIDAD DE GRANADA

CERTIFICA:

Que la Tesis Doctoral titulada "Physical fitness, academic achievement and brain in children" que presenta Dña. **Cristina Cadenas Sánchez** al superior juicio del Tribunal que designe la Universidad de Granada, ha sido realizada bajo mi dirección durante los años 2014-2018, siendo expresión de la capacidad técnica e interpretativa de su autora en condiciones tan aventajadas que la hacen merecedora del Título de Doctora por la Universidad de Granada, siempre y cuando así lo considere el citado Tribunal.

Fdo. Andrés Catena Martínez

En Granada, 9 de Marzo de 2018

La doctoranda Dña. CRISTINA CADENAS SÁNCHEZ y los directores de la tesis D. FRANCISCO B. ORTEGA PORCEL Y ANDRÉS CATENA MARTÍNEZ:

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por la doctoranda bajo la dirección de los directores de tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

En Granada, 9 de Marzo de 2018.

Directores de la Tesis

Francisco B. Ortega Porcel

Andrés Catena Martínez

Doctoranda

Cristina Cadenas Sánchez

La doctoranda Dña. **Cristina Cadenas Sánchez** ha realizado la presente Tesis Doctoral Internacional como beneficiaria de una beca-contrato con cargo al programa de Contratos predoctorales para la formación de doctores (FPI) del Ministerio de Economía y Competitividad (Código: BES-2014-068829), por resolución de 22 de Abril de 2015 de la Secretaria de Estado de Investigación, Desarrollo e Innovación (BOE-A-2015-4778, publicado el 30 de Abril de 2015).

CONTENTS

List of Tables

Annexes-Tables

Table S6. Generalized Additive Model for Location, Scale and Shape (GAMLSS) models used to calculate physical fitness and anthropometric reference standards.

Table S7. Stages, speed (km/h) and laps of the original 20m shuttle run test.

Table S8. Comparison of the reliability of fitness tests between preschoolers (PREFIT study) and adolescents (HELENA study).

List of Figures

Annexes-Figures

RESEARCH PROJECTS AND FUNDING

The present International Doctoral Thesis was performed mainly as a result of the following research projects:

- The **PREFIT Project**, "Assessing physical FITness in PREeschoolers" (Evaluación del FITness en PREescolares). Grant for opening new research lines based on Ramon y Cajal fellowship program (RYC-2011-09011).
  Principal Investigator: Francisco B. Ortega Porcel.
  Duration: 1st May 2014 to 31st December 2015.
  Funding: 15.000 €.

- The **ActiveBrains Project**, "Effects of an exercise-based randomized controlled trial on cognition, brain structure and brain function in overweight preadolescent children". Spanish Ministry of Economy and Competitiveness (i+D+I program), reference DEP2013-47540-R.
  Principal Investigator: Francisco B. Ortega Porcel.
  Duration: 8th October 2014 to 28th February 2017.
  Funding: 120.000 €.

  Additional funding of the ActiveBrains project was obtained from:

  - The author of the present International Doctoral Thesis was funded by a grant from the Spanish Ministry of Economy and Competitiveness (BES-2014-068829) associated to ActiveBrains Project.

  - Other grants associated with ActiveBrains project are: FPU14/06837, FPU15/02645, FPU16/02760 and FJCI-2014-19563.

Another project related to the present International Doctoral Thesis:

- The **MINISTOP Project**, "Mobile-based INtervention Intended to Stop Obesity in Preschoolers". Swedish Research Council, the Swedish Research Council for Health, Working life and Welfare, Karolinska Institutet, and Bo and Vera Axson Johnssons' foundation.
  Principal Investigator: Marie Löf.
  Duration: 2013 – 2016.
  Funding: 62.391 €.

RESUMEN

Los objetivos principales de esta Tesis Doctoral Internacional fueron: aportar métodos nuevos para la evaluación de la condición física en niños de edad preescolar, y proporcionar valores de referencia para poder interpretarla (*Sección 1*); y examinar las asociaciones existentes entre la condición física y el rendimiento , así como explorar el rol del fenotipo metabólicamente sano con sobrepeso/obesidad en el cerebro y su posible efecto sobre el rendimiento académico en niños preadolescentes (*Sección 2*). Para abordar estos objetivos, se llevaron a cabo nueve estudios en el contexto de dos proyectos.

El proyecto PREFIT (*Sección 1*, *Estudios I al VII*) es un estudio transversal que consta de una primera fase metodológica llevada a cabo en Granada (España) y una segunda fase multicéntrica llevada a cabo en más de 3000 niños preescolares procedentes de 10 ciudades españoles para un mejor entendimiento de la evaluación de la condición física en este grupo de edad. El proyecto ActiveBrains (*Sección 2*, *Estudios VIII al IX*) es un estudio aleatorizado controlado cuyo objetivo es examinar el efecto de un programa de ejercicio sobre la salud física y mental en niños con sobrepeso/obesidad.En la presente tesis se basa en análisis transversales centrados en los datos de la evaluación inicial.

Los resultados principales, hallazgos y conclusiones que se desprenden de los nueve estudios incluidos en esta Tesis son: **I)** La batería PREFIT propuesta se basa en el resultado de una revisión sistemática en niños en edad preescolar; **II)** La fiabilidad intra-instrumento fue excelente para todos los dinamómetros utilizados para la medición de la fuerza manual en la batería PREFIT; **III)** El test de 20m de ida y vuelta PREFIT es viable, máximo y fiable en niños preescolares; **IV)** La batería PREFIT es una herramienta viable y fiable para evaluar la condición física en niños de edad preescolar, sin embargo el test de salto horizontal con pies juntos ha mostrado resultados inconsistentes. El test de equilibrio a una pierna mostró una baja fiabilidad en nuestro estudio, no apoyando por tanto su uso en niños de 3 a 5 años; **V y VI)** Proporcionamos valores de referencia específicos de sexo y edad de la condición física y antropometría de una muestra relativamente grande de niños preescolares geográficamente distribuidos en España; y **VII)** Los preescolares españoles presentaron mayor prevalencia de sobrepeso/obesidad en comparación con los preescolares suecos, mientras que las diferencias en la condición física fueron inconsistentes.

En relación con la *Sección 2*, los principales resultados fueron: **VIII)** No sólo la capacidad cardiorrespiratoria, sino también la fuerza muscular y velocidad-agilidad se asociaron positivamente con el rendimiento académico en niños con sobrepeso/obesidad. Por otro lado, la actividad física no demostró ninguna asociación con ninguno de los indicadores de rendimiento académico estudiados; y **IX)** Los niños con sobrepeso/obesidad y un fenotipo metabólicamente sano parecen tener una mayor cantidad de materia gris en el cerebro y mayor volumen cerebral total en comparación con aquellos niños con sobrepeso/obesidad y fenotipo metabólicamente no sano, lo que a su vez se relacionó con un mejor rendimiento académico; sin embargo, tales asociaciones desaparecieron o fueron atenuadas tras ajustar por capacidad cardiorrespiratoria.

Los resultados de esta Tesis Doctoral Internacional incrementan nuestro conocimiento sobre cómo evaluar e interpretar la condición física en niños con edad preescolar, y sobre el papel de la condición física en el rendimiento académico de los niños y el rol del fenotipo metabólicamente sobre el cerebro y el rendimiento académico en niños preadolescentes. Estos resultados darán lugar a futuros estudios prospectivos y de intervención centrados en la salud física y cerebral en la infancia y en el futuro.

ABSTRACT

The major aims of the present International Doctoral Thesis were: provide new methods for physical fitness assessment in preschool children, and provide reference standards for interpreting fitness assessment (*Section 1*); and to examine the associations between physical fitness and academic achievement in preadolescent children, as well as to explore the rol of metabolic healthy overweight/obesity phenotype in the brain and its associations with academic achievement (*Section 2*). To address these aims, nine studies were conducted in the context of two projects.

The PREFIT project (*Section 1*, *Studies I to VII*) is a cross-sectional study that takes place 2 stages: first, a methodological stage conducted in Granada and a second a multi-center stage carried out in more than 3000 Spanish pre-schoolers from 10 Spanish cities, for a better understanding of physical fitness assessment in this age group. The ActiveBrains project (*Section 2, Studies VIII to IX*) is a randomized controlled trial that aimed to examine the effect of physical exercise programme on physical and mental health in overweight/obese children.In the present thesis focused on cross-sectional analyses using the baseline data.

The main findings and conclusions derived from the nine studies included in this thesis were: **I)** The PREFIT battery hereby proposed is based on the output of the current systematic review in preschool children; **II)** The intra-instrument test–retest reliability was excellent for all of the dynamometers used for handgrip strength assessment in the PREFIT battery; **III)** The PREFIT 20m shuttle run test is feasible, maximum and reliable in preschool children; **IV)** The PREFIT battery is a feasible and reliable tool to assess physical fitness in preschool children yet standing long jump has shown mixed findings. The one-leg stance test showed poor reliability in our study, not supporting thus its use in 3 to 5 years-old; **V and VI)** We provide reference standards for physical fitness and anthropometry by sex and age from a relatively large sample of preschool children geographically distributed across Spain; and **VII)** Higher prevalence of overweight/obesity in Spain compared with Sweden is present already at early childhood, while differences in physical fitness components showed mixed findings.

In regards to the *Section 2*, the main findings were: **VIII)** Not only cardiorespiratory fitness but also muscular strength and speed-agility were positively associated with academic achievement in overweight/obese children. Physical activity did not demonstrate an association with any of the academic outcomes studied; and **IX)** Metabolically healthy overweight/obesity related to higher gray matter volume and total brain volume compared to metabolically unhealthy overweight/obese children, which in turn related to better academic achievement, although such associations disappeared or were attenuated after adjusting for cardiorespiratory fitness.

The results of this International Doctoral Thesis enhance our understanding about how to assess and interpret physical fitness and fatness in preschool children; and also how physical fitness relate with academic achievement and the metabolically healthy overweight/obesity phenotype with brain in preadolescent children. These results will lead to future prospective and intervention investigations on the physical and brain health at childhood and later in life.

ABBREVIATIONS

ANCOVA = Analysis of covariance.

ANOVA = Analysis of variance.

b = Beta unstandardized coefficients.

β = Beta standardized coefficients.

BCCG = Box-Cox Cole and Green.

BCPE = Box-Cox power exponential.

BCT = Box-Cox t distribution.

BDNF = Brain Derived Neurotrophic Factor.

BMI = Body Mass Index.

BOT-2 = Bruininks-Oseretsky Test of Motor Proficiency, second edition.

CFI = Comparative Fit Index.

CI = Confidence interval.

Const. = Constant.

Cs = Cubic splines.

CV = Coefficient of variation.

DARTEL = Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra.

ENMO = Euclidean Norm Minus One g.

GAMLSS = Generalized Additive Model for Location, Scale and Shape.

GDBT = Ghent Developmental Balance Test.

GMF = Gleichgewichts-Koordintations-System.

ICC = Intraclass Correlation Coefficient.

IGF = Insulin-like Growth Factor.

MABC-2 = Movement Assessment Battery for Children-second edition.

METs = Metabolic Equivalents measured as ml/kg/min.

MHO = Metabolically Healthy Obesity.

MHR = Maximum Heart Rate.

MMT = Maastricht's Motor Test.

MNI = Montreal Neurological Institute.

MRI = Magnetic Resonance Imaging.

MUO = Metabolically Unhealthy Obesity.

MVPA = Moderate-to-Vigorous Physical Activity.

NIH = National Institutes of Health.

NS = No significant difference.

PA = Physical Activity.

RM = Repetition Maximum.

SD = Standard Deviation.

PACER = Progressive Aerobic Cardiovascular Endurance Run.

PDMS-2 = Peabody Developmental Motor Scales, second edition.

PGMQ = Preschooler Gross Motor Quality Scale.

RIHM = Rotterdam Intrinsic Hand Myometer.

RMSEA = Root Mean Square Error of Approximation.

S1= Study 1.

S2 = Study 2.

SDD = Smallest Detectable Difference.

SEM = Standard Error of Measurement.

r_c = Lin's concordance correlation coefficient.

VEGF = Vascular Endothelial Growth Factor.

VO_2max = Maximum oxygen consumption.

VO_2peak = Peak oxygen consumption.

WHO = World Health Organization.

WOF = World Obesity Federation.

INTRODUCTION

INTRODUCTION

Physical fitness is a set of attributes related to a person's ability to perform physical activities that require aerobic fitness, endurance, strength or flexibility and is determined by a combination of regular activity and genetically inherited ability[1]. A growing body of evidence shows that physical fitness, particularly cardiorespiratory fitness and muscular strength, is a powerful marker of health in children and adolescents. For instance, low fitness level during childhood has been associated with a high risk of developing cardiovascular diseases, overweight/obesity, mental disorders, and skeletal problems later in life[2,3]. Existent evidence refers primarily to children above 6 years of age and adolescents with nearly no information about fitness and health in preschoolers. However, there is no reason to believe that fitness is less important in preschoolers than in older children.

Physical activity is defined as any bodily movement produced by skeletal muscles that result in energy expenditure above the basal levels (i.e., 1.5 metabolic equivalents)[1]. This broad term means that physical activity includes almost everything that a person does. *Physical inactivity* is defined as a level of activity considered insufficient, which in practical terms is defined relating to the International Physical Activity Guidelines for children and adults specifically, e.g., 60min/day of moderate-to-vigorous physical activity (MVPA) at least 5 days of the week for children and adolescents. Consequently, a child who accumulates a total of 50min of MVPA 3 times per week will be considered inactive. The concept of inactivity should not be mixed with sedentarism. The Sedentary Behaviour Research Network defines *sedentary behaviour* as any waking behavior characterized by an energy expenditure ≤1.5 metabolic equivalents (METs, ml/kg/min) while in a sitting or reclining posture[4]. Based on these definitions, sedentarism and inactivity are different constructs and consequently a person can be at the same time active (if meeting the minimum physical activity guidelines of activity per week) and sedentary (if accumulating much time in sedentary behaviours). Another related but not interchangeable term is *physical exercise*, defined as a subset of physical activity that is planned, structured and systematic.

Recently, it has been stated the concern about the low levels of physical activity and the marked high levels of sedentary time among youths[5]. In line with this assumption, it is important to consider the health problems related to the lack of activity and high amounts of time spent in sedentary behaviours, i.e., less favourable cardiometabolic health markers, lower academic achievement, etc.[6].

Overweight and *obesity* are defined as an abnormal or excessive fat accumulation that presents a risk to health. Childhood obesity is currently recognized as a worldwide epidemic problem and is associated with an array of adverse consequences across the lifespan[7]. Beyond the total amount of fat, its distribution in the body is also associated with a poor health status[8]. *Central* or *abdominal body fatness* is associated with metabolic risk factors and evidence from adults showed that has been more closely related to inflammation than the total adiposity[9].

Academic achievement refers to the performance of the person that indicates how they have accomplished in specific goals in educational setting. Recently, it has been published a position stand by the American College of Sport Medicine which states that physical fitness and physical activity could benefit executive function and academic achievement in youth[10,11].

Brain structure comprises gray matter, white matter and cerebrospinal fluid (all together is known as *total brain volume*) that can be easily discriminated by magnetic resonance imaging (MRI). *Gray matter* (named gray for the pinkish-gray color) contains the cell bodies, dendrites and non-myelinated axons. Its function is focused on processing the information, and also linked to reasoning functioning. Otherwise, *white matter* contains myelinated axons connecting different parts of the gray matter to each other and carrying nerve impulses among neurons. Its function is related to processing speed, since this is the tissue through which messages pass between different areas of gray matter. Information available about brain and its relation with physical health outcomes is less known[10]. However, a growing body of literature is showing a relationship of brain structure (gray and white matter) with physical activity and physical fitness which, in turn, could be associated with executive function and academic achievement[12-16].

The present International Doctoral Thesis is structured in two sections. *Section 1* is related to physical fitness and fatness in preschool children (aged 3 to 5 years). *Section 2* comprises the topic related to physical fitness, academic achievement and brain in preadolescent children.

Section 1: Physical fitness and fatness in preschool children

Physical fitness assessment in preschool children: field-fitness tests and metric properties

As we have stated in the section above, the assessment of physical fitness in young people is relevant from a public health point of view. In this context, the European Commission funded the ALPHA project (Assessing Levels of Physical Activity and Fitness, http://www.thealphaprojet.net) which primarily aimed to improve existing methods for assessing physical activity and physical fitness. Under the framework of the ALPHA project, our group conducted three systematic reviews regarding the reliability[17], criterion-related validity[18], and relationship with health[2] of existing field-based physical fitness tests in children and adolescents (aged 6-18 years). The main findings of these systematic reviews, together with the outcome of several methodological studies were summarized and an evidence-based ALPHA fitness-test battery was proposed[19]. This fitness test battery includes the following tests: the 20m shuttle run test, the handgrip strength test, the standing long jump test and the 4x10m shuttle run test for assessing cardiorespiratory fitness, upper and lower limbs muscular strength and speed-agility fitness, respectively. Recently, a Committee composed of members of councils from the Institute of Medicine and National Academy of Sciences in the US, among others, has revisited the literature on health-related physical fitness in youth and concurred with the ALPHA project's recommendations for the field-based physical fitness tests[20,21].

Therefore, the two major projects mentioned above provide a useful and scientific-based platform for simple and inexpensive health-related physical fitness assessment in children and adolescents. However, the recommended tests were proposed for children and adolescents aged 6-18 years but it was unknown which are the most reliable, valid and health-related physical fitness tests to be used in preschool children (aged 3-5 years) in field-based conditions. It is reasonable to think that physical fitness might be as important for health in preschool children as it has shown to be in their older peers, as well as to quantify changes in physical fitness as a consequence of physical activity-based interventions in preschoolers. Thus, studies providing field-based fitness tests for preschool children, as well as addressing its feasibility, reliability and validity are needed. This information will contribute to better inform the researchers and practitioners about which physical fitness tests should be used in preschoolers based on their metric properties (e.g., feasible, reliable) which, in turn, could result in a substantial improvement in the body of literature related to physical fitness in preschoolers.

Physical fitness and anthropometric reference standards in preschool children

Reference values are necessary for classifying children, for health screening, and for early prevention of non-communicable diseases which are aggravated during growth and development. Fitness reference data have been reported in children and adolescents (6-18 years old) from different countries using standardized measures[22-26]. However, literature addressing reference data in preschool children is rather scarce. In fact, as far as we know, only one study provided reference values for one specific test (i.e., standing long jump) in preschoolers from one province in the south of Spain (Jaén)[27]. Thus, studies providing reference values in preschool children for all fitness components with harmonized measurements are warranted.

Sex- and age-specific reference standards are needed to accurately assess anthropometric parameters and to correctly interpret prevalence and health-related consequences of different weight statuses in preschool children. Previous studies have shown anthropometric reference standards in children and adolescents for specific populations[28-31]. However, international cut-points for markers of abdominal adiposity in children is still undefined. For instance, for the definition of metabolic syndrome in youth, central obesity is defined as a waist circumference above the sex- and age-specific 95th percentile[32]. Therefore, from a public health point of view, these reference standards would help health care professionals to detect and/or prevent health problems related to low and high levels of total and central fatness later in life. Nevertheless, most of the examined studies provided reference values for children at the age of ≥6[30,31]. To the best of our knowledge, there is no study providing reference values in Spanish preschool children.

Prevalence of the different weight status categories in preschool children

Recent evidence shows that the rates of overweight/obesity have reached epidemic proportions also in children under the age of 5[33]. The World Health Organization (WHO) reported that more than 41 million children of this age group were classified as overweight or obese in 2014. Likewise, it is estimated that by 2020 the number of overweight/obese children between the ages of 3 and 5 will reach up to 60 million. At the European level, data prevalence of overweight/obese children (between 2 and 9.9 years old) based on the World Obesity Federation (WOF, formerly known as International Obesity Task Force) criteria ranged from 21 to 42% in the south of Europe and remained under 11% in the north, being higher in girls than in boys (21 % vs. 19%)[34]. Further studies in preschool children with larger sample sizes and geographically distributed across multiple countries may provide a more real description of the global problem. Furthermore, previous studies in older children and adolescents observed that the increase of overweight/obesity prevalence has slowed down or even decreased[35,36]. Nevertheless, the prevalence of severe and morbid obesity rates in childhood seems to increase in several countries[37-39]. To our knowledge, there is no previous information about the prevalence of severe/morbid obesity in preschool children. Therefore, it is relevant from a clinical and public health perspective to assess the prevalence of different grades of obesity (i.e., type I, II and III, or mild, severe, and morbid respectively) in preschoolers, using the new cut-points linked to adult standard body mass index (BMI) cut-offs[40].

Underweight, which can be found to the opposite end of the BMI spectrum, has received much less attention, even though the health consequences of its different degrees could affect growth and development, being also of public health concern[41]. A trend analysis conducted by de Onis et al.[42] showed that the worldwide prevalence of childhood underweight was projected to decline from 26.5% in 1990 to 17.6% in 2015. Particularly, in developed countries the prevalence was estimated to decrease from 1.6% to 0.9%. However, the percentage of underweight children changes from 1% to 23% depending on the characteristics of the study sample, the year of study, the country, and the definition of underweight[43,44]. Thus, classification of different degrees of obesity and underweight and different weight statuses according to different cut-offs criteria such as WOF or WHO are needed to approach the inconsistency in the prevalence findings about underweight children.

Physical fitness levels and prevalence of overweight/obesity in preschool children from north compared with the south of Europe

Considering the role of cardiorespiratory fitness and muscular strength over health outcomes and examining the differences between physical fitness levels in preschoolers from different parts of Europe are matters of interest. Previous evidence showed that adolescents from central-northern Europe are fitter in terms of cardiorespiratory fitness, muscular strength, and speed-agility compared than those from the south[45]. Lang et al.[46] concluded that children from northern Europe were better performers in the 20m shuttles run test (i.e., cardiorespiratory fitness) than their southern peers in a systematic review which included more than 1 million children and youth. In line with these findings, Ortega et al.[47] observed that the level of cardiorespiratory fitness (measured by 20m shuttle run test) of Spanish adolescents was worse than that of adolescents from other countries (e.g., Netherlands, Belgium, Denmark, etc.). To the best of our knowledge, there is no study comparing fitness levels in different parts of Europe in preschool children, however it would be interesting to see whether the differences observed in children and adolescents are present earlier in life at preschool age.

In regards to weight status prevalence, Spain, Malta and Sicily (i.e., southern Europe) lead the ranking of the number of overweight and obese in children aged 5-10 years following the last report of the WOF[48]. In relation with differences between north and south of Europe, recently Ortega et al.[45] showed that the prevalence of overweight/obesity and waist circumference in adolescents from southern Europe was substantially higher compared with central-northern European adolescents (31% vs. 21% for prevalences, and 73.8 vs. 71.3 cm for waist circumference). However, information about whether south-north gradient regarding the prevalence of overweight/obesity observed in preschool children is scarce. In this context, it is relevant to note that van Stralen et al.[49] observed the highest rates of overweight/obesity among 4-7 year-olds in southern European countries when comparing six European countries. Furthermore, information about central obesity or abdominal adiposity as measured by waist circumference in preschool children from different parts of Europe is insufficient. Thus, comparison about prevalence of overweight/obesity and central obesity in preschool children from north and south of Europe are needed.

Section 2: Physical fitness, academic achievement and brain in preadolescent children

Physical fitness, physical activity and academic achievement in overweight/obese children

In childhood, cognition is an important predictor of physical and mental health across the lifespan[50]. One indicator of cognition in children is academic achievement, and interest in this outcome has grown rapidly in recent years due to its relevance to school settings and educational policies[10]. Schools have received mounting pressure to increase the time devoted to core subjects (e.g., language, math and science) and consequently, decreased time has been allotted to other subjects such as physical education, arts, and music. However, a recent position stand from the American College of Sport Medicine[10] suggests that this strategy is ill-suited, since health-related factors, such as physical fitness and physical activity inherent to the physical education curriculum, are positively associated with cognitive health and academic achievement.

As we have indicated before, fitness is considered a powerful marker of health in children and adolescents[3]. Previous studies have mainly examined the relationship of cardiorespiratory fitness with academic outcomes, concluding that a higher fitness level is associated with a higher academic achievement in children[10,51-54]. However, far less investigated are the associations between muscular and speed-agility fitness with academic achievement[10]. In this regard, Kao et al.[55] highlighted the importance of research focusing on muscular fitness and its association with cognitive health. In consonance, our group recently examined the role of muscular and speed-agility fitness on brain structure, concluding that speed-agility, but not muscular, might be another component of fitness (along with cardiorespiratory fitness) with a strong potential for improving brain development, with implications for better academic performance in overweight/obese children[12]. Thus, a greater understanding of the relationship between muscular and speed-agility fitness with academic achievement is needed.

Physical activity, an independent but strongly related to fitness construct, is associated with a wide-range of benefits for children's physical and mental health[56]. Therefore, time spent in different physical activity intensities may also be associated with academic achievement. Previous systematic reviews and meta-analyses have shown a positive influence of physical activity on academic performance[10,57-59], with MVPA being the intensity most studied in this relationship. However, such a conclusion remains controversial, given that a number of reports appear in the literature demonstrating positive[60,61], negative[62] and non-significant results[60,63-65] between physical activity and academic outcomes.

Further, previous studies examining physical fitness and physical activity in relation to academic achievement have mainly focused on healthy weight children, while there is a lack of information in overweight/obese children (for exception, see Davis et al.[66]). Therefore, studies examining physical fitness and physical activity in relation with academic achievement in overweight/obese children are warranted.

Metabolic healthy phenotype and brain structure in overweight and obese children

Childhood overweight/obesity is one of the major health concerns of this century as it has reached epidemic proportions worldwide[39]. These results are alarming since the excess of weight has well-known comorbidities including insulin resistance, hypertension, and type-2 diabetes[67,68] as well as mental health problems, and measurable changes in brain and academic achievement[69-72]. However, there is a subset of the population, referred to as metabolically healthy overweight/obesity (hereinafter abbreviated as MHO to refer to both overweight and obese children with a healthy metabolic profile), who do not present these metabolic abnormalities. MHO phenotype is defined as an excessive body weight based on international BMI cut-points and meet none of the metabolic syndrome criteria (i.e., triglycerides, glucose, high-density lipoprotein and systolic and diastolic blood pressure)[67], which in children can be done using the age- and sex-specific cut-points proposed by Jolliffe and Janssen[73].

Recent systematic reviews and meta-analyses have concluded that MHO individuals are at a higher risk of cardiovascular mortality and morbidity than those who are metabolically healthy normal weight[74-78], however, it has been shown that cardiorespiratory fitness could at least partially explain these differences[67,79,80]. When the focus is on the same weight status, i.e., overweight/obesity, MHO individuals are characterized by having lower amounts of visceral adipose tissue and adipose cell size, and higher cardiorespiratory fitness and physical activity levels than metabolically unhealthy overweight/obesity (MUO)[67,79,81,82], yet it remains unexplored whether metabolic differences between MHO and MUO may also extend to brain and academic achievement in young individuals.

There is only one study in obese adolescents showing that metabolic syndrome was related to a reduction of hippocampal volume and academic and cognitive measures[83]. Clearly, there is a need to explore differences between MHO and MUO using a whole brain approach. In addition, given the major role of cardiorespiratory fitness in the characterization and prognosis of the MHO phenotype[67,79,80,84], and the link between fitness and gray matter shown in our previous study[12], it is interesting to examine role of cardiorespiratory fitness also in the relation of the MHO phenotype with brain and academic achievement.

OBJETIVOS/AIMS

OBJETIVOS

Los objetivos principales de esta Tesis Doctoral Internacional fueron: aportar métodos nuevos para la evaluación de la condición física en niños de edad preescolar, y proporcionar valores de referencia para poder interpretarla (*Sección 1*); y examinar las asociaciones existentes entre la condición física y rendimiento académico , así como explorar el rol del fenotipo metabólicamente sano con sobrepeso/obesidad en el cerebro y sus asociaciones con el rendimiento académico en niños preadolescentes (*Sección 2*).

Los resultados de esta Tesis Doctoral International están organizados en nueve estudios, basándose en los siguientes objetivos específicos:

Sección 1

1. Estudio I. Revisar de forma sistemática los estudios realizados en niños de edad preescolar usando test de campo de condición física, y estudiar su fiabilidad, validez y relación con la salud. Nuestro objetivo final fue proponer una batería de test de campo de condición física que se pueda usar en preescolares a partir de los estudios revisados.

2. Estudio II. Examinar la fiabilidad y validez de los dinamómetros analógico y digital TKK para la medición de la fuerza de prensión manual en la batería PREFIT usando pesos conocidos calibrados.

3. Estudio III. Describir la adaptación del test de 20m de ida y vuelta, su viabilidad, maximidad y fiabilidad en niños de 3 a 5 años.

4. Estudio IV. Determinar la viabilidad de todos los tests de la batería PREFIT en niños de 3 a 5 años, así como su fiabilidad para toda la muestra y separado por grupo de edad (es decir, 3, 4 y 5 años); y proporcionar recomendaciones prácticas para la realización de batería PREFIT.

5. Estudio V. Proporcionar valores de referencia de condición física específicos para sexo y edad en niños españoles de edad preescolar; y estudiar las diferencias en condición física entre sexos, así como caracterizar los niveles de condición física a lo largo de la etapa preescolar.

6. Estudio VI. Determinar la prevalencia de bajo peso, normo peso, sobrepeso y diferentes grados de obesidad (media, severa y mórbida) basándose en los criterios de la Federación Mundial de la Obesidad y de la Organización Mundial de la Salud; y proporcionar valores de referencia específicos de sexo y edad de antropometría en una muestra relativamente grande (i.e., ~3000 participantes) y geográficamente distribuida de niños preescolares de España (3.0 – 6.25 años).

7. Estudio VII. Comparar la prevalencia de sobrepeso/obesidad, perímetro de cintura, y niveles de condición física en niños preescolares de 4 años de Suecia (norte de Europa) y España (sur de Europa).

Sección 2

1. Estudio VIII. Examinar las asociaciones de los diferentes componentes de la condición física (capacidad cardiorrespiratoria, fuerza muscular y velocidad-agilidad), además de la actividad física, con el rendimiento académico en niños con sobrepeso/obesidad.

2. Estudio IX. Examinar si los niños con sobrepeso/obesidad con un fenotipo metabólicamente sano tienen un mayor volumen global y regional de materia gris comparado con aquellos niños metabólicamente no sanos. También examinamos la asociación entre el volumen de materia gris y el rendimiento académico, y por último, estudiamos el rol de la capacidad cardiorrespiratoria en todas estas asociaciones.

AIMS

The major aims of the present International Doctoral Thesis were: provide new methods for physical fitness assessment in preschool children, and provide reference standards for interpreting it (Section 1); and to examine the associations between physical fitness and academic achievement, as well as to explore the rol of metabolic healthy overweight/obesity phenotype in the brain and its associations with academic achievement in preadolescent children (Section 2).

The findings of this International Doctoral Thesis are organized in nine studies, based on the following specific aims:

Section 1

1. Study I. To systematically review the studies conducted in preschool children using field-based fitness tests, and to examine their reliability, validity, and relationship with health outcomes. Our ultimate goal was to propose a field-based physical fitness-test battery to be used in preschool children.

2. Study II. To examine the reliability and validity of the analog and digital models of TKK handgrip dynamometers for the measurement of muscular fitness in the PREFIT battery using calibrated known weights.

3. Study III. To describe the adaptation of the original 20m shuttle run test, its feasibility, maximality and reliability in children from 3 to 5 years.

4. Study IV. To determine the feasibility of all the tests of the PREFIT battery, as well as its reliability in 3–5 years old children for the whole sample and separated by age groups (i.e., 3, 4 and 5 years-old); and to provide practical recommendations for the PREFIT battery.

5. Study V. To provide sex- and age-specific physical fitness reference standards for Spanish preschool children; to study sex differences across this age period and to characterize fitness performance throughout the preschool period.

6. Study VI. To determine the prevalence of underweight, normal weight, overweight, and different grades of obesity (mild, severe, and morbid) based on World Obesity Federation and World Health Organization criteria; and to provide sex- and age-specific anthropometric reference standards obtained in a relatively large (i.e., ~3000 participants) and geographically distributed sample of Spanish preschool children aged 3.0-6.25.

7. Study VII. To compare the prevalence of overweight/obesity, waist circumference and physical fitness levels in preschool children aged 4 years from Sweden (north of Europe) and Spain (south of Europe).

Section 2

1. Study VIII. To examine the associations of the different components of physical fitness (cardiorespiratory, muscular and speed-agility), as well as physical activity, with academic achievement in overweight/obese children.

2. Study IX. To examine whether metabolically healthy overweight/obese children have higher global and regional gray matter volumes than their metabolically unhealthy peers using a whole-brain analytical approach. Also, we examine the association between gray matter volume and academic achievement, and finally we examined the role of cardiorespiratory fitness in all these associations among overweight/obese children.

MATERIAL AND METHODS

MATERIAL AND METHODS

Section 1: Physical fitness and fatness in preschool children

Study I. Systematic review of physical fitness in preschoolers

Search strategy

We selected two major electronic databases to conduct the search -PubMed and Web of Science. We screened for studies conducted in preschool children that included at least one field-based physical fitness test. When searching in PubMed, we used Medical Subject Heading (MeSH) terms in our search. MeSH is the National Library of Medicine's controlled vocabulary thesaurus used for indexing articles for PubMed. This is a powerful method to enhance the search. As an example, 'physical fitness' is a MeSH term introduced in the MeSH database in 1996, and a number of related terms are linked to it (see Annexes, Table S1). When 'physical fitness' [MeSH] is entered in the search browser, the search engine will look for papers that include 'physical fitness' or any of the related terms included in the MeSH database. The search terms not available in the MeSH database (e.g. agility) were entered in the browser in the normal way, i.e. without the command MeSH attached at the end.

Since we wanted to focus the search on preschool children, all the key search terms concerning physical fitness were combined (using the connector 'AND') with this term, as shown in Table S1. Search example 1: 'child, preschool' [MeSH] AND 'physical fitness' [MeSH]; search example 2: 'child, preschool' [MeSH] AND 'muscle strength' [MeSH]. In addition, we searched whether a set of fitness-test batteries (e.g. EUROFIT) commonly used in children and adolescents, identified in a previous review[18], have been used in preschool children. Finally, we combined all the searches using the connector 'OR' (e.g. search example 1 OR search example 2; see these examples above), so that we obtained the final number of articles (duplicates excluded by using the connector 'OR' when combining searches) that met all the search criteria. The same search strategy and combination of terms was repeated in Web of Science but without using MeSH terms or equivalent since a similar option does not exist in Web of Science (Table S2).

We first checked the review articles identified in our search (n = 122) in order to find out whether there were any previous reviews on reliability, validity and/or relationship with health of field-based physical fitness tests in preschool children. Since we did not find any, no starting year limit was set. Therefore, our review includes all articles published on this topic in this age group to 1 April 2014.

Eligibility criteria

Studies were considered eligible for their inclusion if: 1) the age of participants was from 3 to 5 years (i.e. preschool children), 2) participants were healthy (i.e. not presenting any disease or pathology), 3) fitness tests were measured in field condition, 4) we had access to the full text, and 5) report information on reliability, validity or the relationship with health outcomes.

Two independent reviewers for each database (four in total) examined the articles retrieved for inclusion in this review.

Study II. Reliability and validity of handgrip dynamometers

Instruments

A total of six TKK handgrip dynamometers, three digital and three analog, were used to assess the reliability and criterion related validity. The digital handgrip dynamometers (TKK Model 5401; Takei, Tokyo, Japan) had a range of measure from 5.0 kg to 100.0 kg, whereas the analog handgrip dynamometers (TKK Model 5001; Takei, Tokyo, Japan) had a range of measure from 0.0 kg to 100.0 kg (Figure S1). Four dynamometers were new (i.e., bought for this study), and two (one digital and one analog) were old (i.e., had been used for >6 years in population studies), allowing us to compare the reliability and validity between new and old TKK dynamometers. The dynamometers, weights, and scale were calibrated by the manufacturer at purchase. The verification of all weights was performed by means of a new (bought for this study) high-precision SECA scale (Model 769; SECA, Hamburg, Germany). We assumed that the SECA scale was perfectly calibrated because we could not test its validity against a gold standard. However, we assessed its reliability using known weights from 1 to 70 kg with increments of 1 kg up to 20 kg and increments of 5 kg thereafter. We observed very high reliability; that is, the mean difference between the known weights and the SECA scale measures was 0.004 kg (standard deviation, SD = 0.02 kg; p= 0.300).

Measurement

Known weights were used to analyse the criterion-related validity (known weights vs. dynamometers) and reliability (intra-instrument, inter-instrument, and inter-model measures) of the six dynamometers (Figure 1).

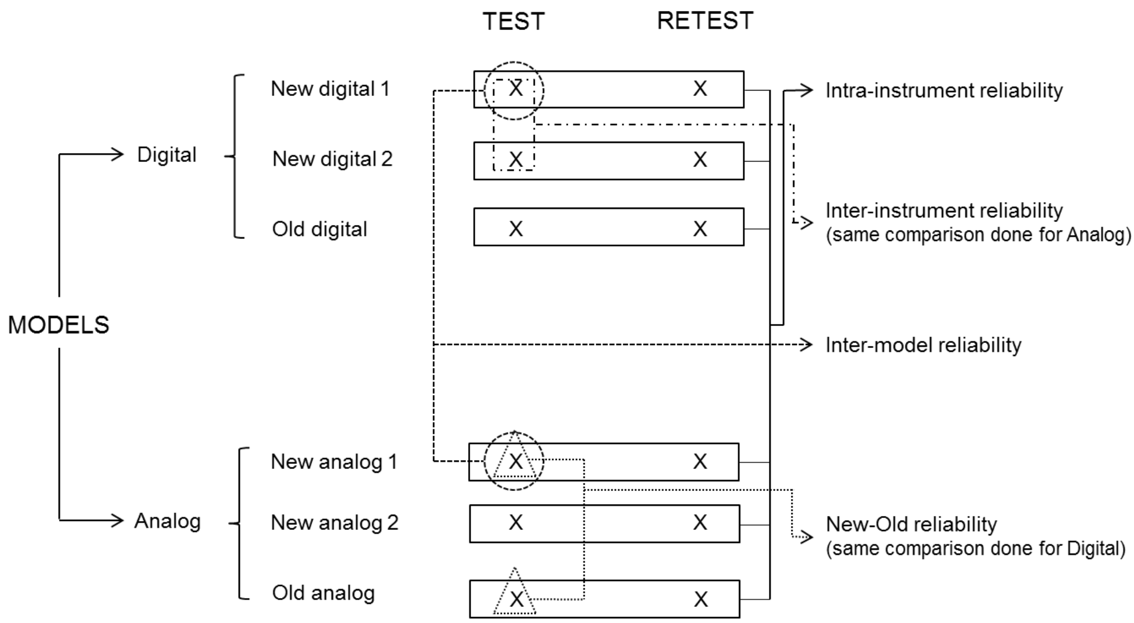


Figure 1. Graphic illustration of the study design for reliability analyses using TKK dynamometers.
Numbering of dynamometers (i.e., 1 or 2) was done randomly.
Note. Inter-instrument and inter-model reliability and new–old reliability were done for both TKK models, digital and analog, but only one of them is marked in the figure to avoid too much overlapping.

The dynamometers were positioned between two wooden supports with the handle fixed. The known weights were suspended with a loading belt from the center of the dynamometer's handle; the weights ranged from 1 kg to 70 kg, and increments of 1 kg were added up to 20 kg and increments of 5 kg thereafter. The dynamometer's handle was marked for consistent placement of the loading belt with the known weight. The weights were added in a randomized order, and each weight measure was repeated twice (test–retest). The order of testing of the dynamometers was also randomized. As commonly described in the literature, a 5.0-cm grip span was used (roughly corresponding to Position 3 of a Jamar grip span). The time between trials was approximately 50–60 s.

Studies III to VII. Fitness and fatness in preschoolers: methodological studies, reference values and comparison between Spanish and Swedish.

Study design

This study was conducted under the PREFIT project framework (http://profith.ugr.es/prefit). The main objective of this study was to assess physical fitness and anthropometric characteristics in preschool children. The PREFIT project started in 2014 with methodological studies in Granada and in 2015 we extend the physical fitness and fatness assessment to others 9 different cities/towns in Spain (i.e.

Almería, Cádiz, Castellón de la Plana, Cuenca, Las Palmas de Gran Canaria, Madrid, Palma de Mallorca, Vitoria-Gasteiz, and Zaragoza). The study protocol was approved by the local Review Committee for Research Involving Human Subjects, in accordance with the Declaration of Helsinki 1961 (revision of Edinburgh 2013).

Participants

A total of 4338 preschoolers and their parents were invited to participate in the PREFIT project. The teaching staff from each school delivered an information sheet and an informed consent to parents and/or guardians. These included the purpose of the study and brief explanations concerning the tests to be used. Finally, 3198 parents agreed to participate in the study (participation rate: 73.7%). Among them, 19 children were excluded after the assessments (i.e. they presented any motor or cerebral disease that limited the test performance reported by the school teachers, they cried during most tests, they had cough and mucus, or they did not understand correctly the instructions of the tests). Thereby, a total of 3179 (4.6 ± 0.9 years old, 1678 boys, 52.8%) preschool children participated in the PREFIT Project (Figure 2).

We measured children attending first, second, and third grades of preschool centers. In Spain, this mainly includes children aged 3, 4, and 5 years old. However, these groups included some children younger than 3 (n=44, 1.4%) and some older than 5.9 (n=112, 3.5%).

To note, *Study III*, *Study IV*, and *Study VII* were performed with a subsample of one of the centers involved in PREFIT Project, (i.e. Granada), and thus, the sample size decrease up to 161 as maximum.

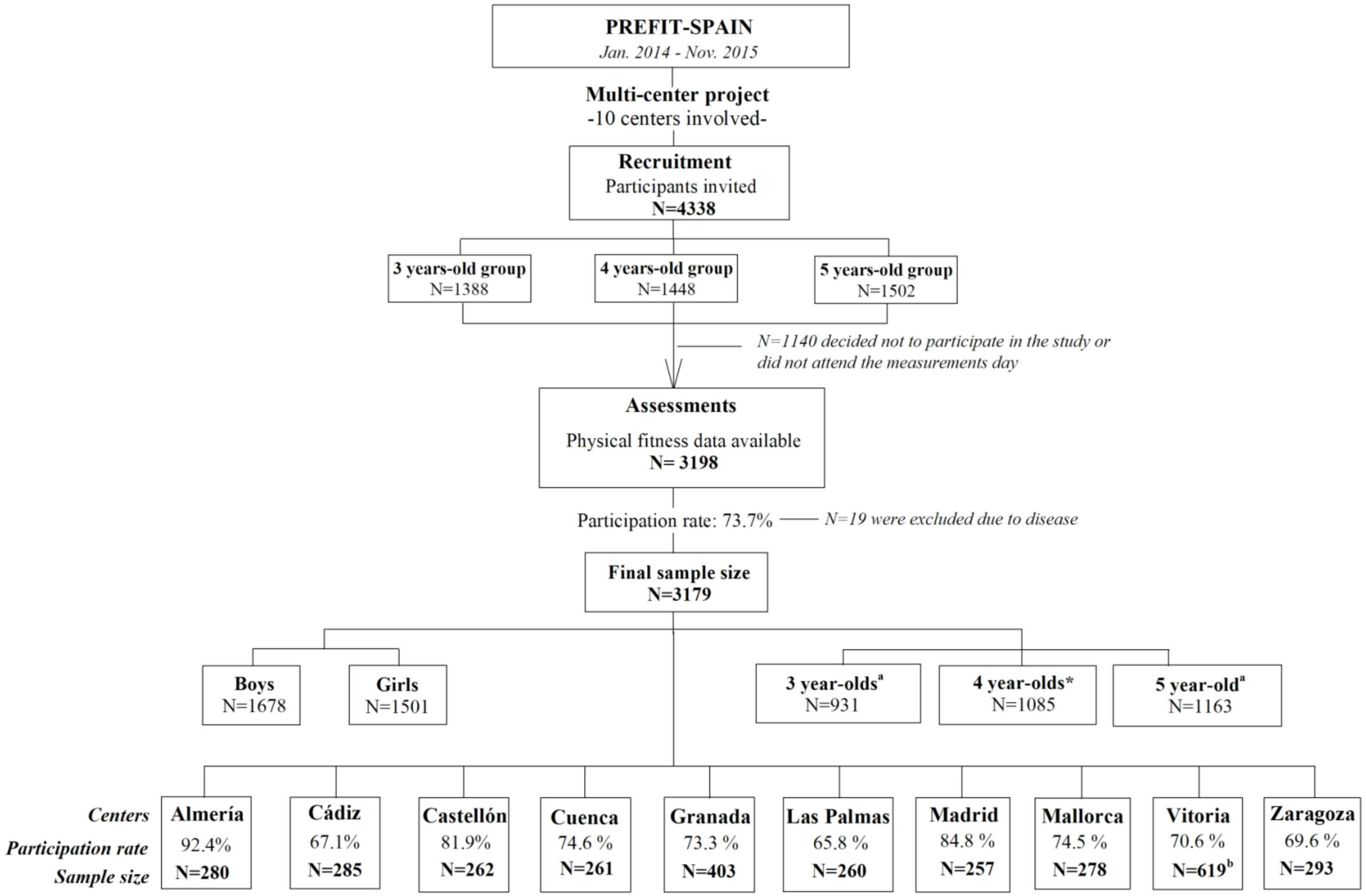


Figure 2. Flow-chart of the population involved in the PREFIT Project.

a We measured first, second, and third grade of preschool centers, which mainly consists in 3, 4, and 5-year-olds in Spain. However, these natural groups included some children younger than 3 (n=44, 1.4%) and some older than 5.9 (n=112, 3.5%).

b The sample size in this center in the north of Spain was over dimensioned in order to compensate the sample size in the south of Spain, Andalucía, where 3 centers participated in the study.

Measurements

Anthropometry

Weight (kg) and height (cm) were assessed without shoes and in light clothing with a balance scale (SECA 213, Hamburg, Germany) and stadiometer (SECA 213, Hamburg, Germany), respectively. Thereafter, we calculated BMI (kg/m^2). The cut-off points used to determine weight status categories and, thus, prevalence were those established by the WOF[40,85] and by the WHO[86]. The WOF cut-offs were developed for youth aged from 2 to 18 years old and provided weight status categories for each month of age from underweight to severe obesity[85]. Additionally, morbid obesity was just recently defined, which provides new opportunities to study the most pathological obesity from very young ages[40]. The WHO child growth standards provided cut-offs based on sex-and age-specific z-scores[87].

We assessed waist circumference (cm) with a non-elastic tape (SECA 200, Hamburg, Germany) at the level of the umbilicus zone. The measures were taken at the end of the normal expiration without the non-elastic tape compressing the skin. Thereafter, we calculated the waist-to-height ratio.

All the anthropometric measurements were performed twice, and we calculated the mean of both measurements, which were then used for the analyses. Likewise, a strict protocol was performed to ensure standardization across all the involved centres.

Physical fitness

Physical fitness (i.e. cardiorespiratory fitness, muscular strength, speed-agility, and balance) was assessed with the PREFIT battery (**Figure 3**). Just before the measurements, we told a motivating fairy tale based on *"Cofito and his adventures on the Lipid Island"or "Cofita and her adventures in the banana space"* with the aim to encourage children and make the tests more attractive. Briefly, participants had to help "Cofito" (boy) and "Cofita" (girl) to overcome sme adventures in imaginary places, such as "Lipid Island" or in the "Banana space" (a visual example of one of the fairy tale is presented in Figure S2). Cardiorespiratory fitness was assessed with the PREFIT 20m shuttle run test. Briefly, the test consisted in running back and forth between two lines (20m apart) following an audio signal. From the original version proposed by Leger et al.[88] two adaptations were introduced for preschool children: 1) the test started at 6.5 km/h with an increment of 0.5 km/h every minute, and 2) one evaluator runs in front of the preschoolers and another behind them (e.g. 4-8 preschoolers of the same age) in order to help them to maintain the pace. The test finished for each child when they could not reach the line with the audio signal on two consecutive occasions or when they stopped due to fatigue. One of the evaluators was the person responsible for taking out the children of the test when they finished. This was the last test to be done and it was only done on one occasion. The result was registered in laps.

Upper-limb muscular strength was assessed with the handgrip strength test. This test consisted in squeezing as much as possible for 2-3 seconds. The analog version of TKK dynamometer (TKK 5001, Grip-A, Takei, Tokyo) was used and grip span 4.0 cm was fixed [89]. The elbow had to be extended without being in contact with anything except for the hand touching the dynamometer. Preschoolers performed two non-consecutive attempts with each hand. We chose the best result and registered the average of both hands in kg.

Lower-limb muscular strength was assessed with the standing long jump test. This test consisted in jumping forward as far as possible, with the feet separated at the shoulders' width, landing upright. We recorded the distance between the starting line and the location of the foot closest to the starting line. We registered in cm the best of three attempts. In *Study IV*, in order to help and guide the preschoolers to jump, first we put a stick in the take-off line where they had to place their feet nearer it (hereinafter called standing long jump test S1). We included the stick in the last assessments of the test and in all the assessments of the tests due to when jumping forward many children step over the starting line, so that the attempt was considered not valid and it was needed to repeat it. However, after see the systematic worse performance in the retest, we thought that this could have made the children to jump more vertically than needed in order to reach as far as possible. In the replication study (n=106 preschoolers, 64 boys), a slight modification was applied (standing long jump test S2). Instead of the stick as an interference object to jump, we decided to draw footprints to help children to detect the take-off line and start jump. When referring to this test during the *Study IV*, we will use S1 and S2, for Study 1 main study and Study 2 replication study, respectively. For the rest of the studies included, (i.e., *Study V* and *Study VII)*, we drew footprints on the floor to guide the preschoolers towards the starting line to jump.

Speed-agility was assessed with the PREFIT 4x10m shuttle run test. In this test, children had to run four times between two lines (10m apart) as fast as possible. Two evaluators stood at each line and preschoolers had to touch the evaluator's hand and return to the starting line as fast as possible. The best of two attempts were manually registered by an experienced evaluator (lowest duration in seconds)[90].

Static balance was assessed with the one-leg stance test. The test consisted in standing on one-leg still and bending the other leg at approximately 90º. The beginning of the test starts when one of the legs is no longer in contact with the floor. Children had to maintain the balance position for as long as they could. The test finished when the child could not continue in the required position. Children had one attempt with each leg, and the average time was registered in seconds.

Figure 3. Physical fitness tests in preschool children: the PREFIT battery.

Familiarization trials and explanations providing examples of how to perform the tests were very important to ensure that children had understood correctly the process. More information about practical recommendations and how we approached several situations during the assessments can be found in Annexes (See Table S3). The manual of operations, audio of the PREFIT 20m shuttle run test, and videos showing how to perform and score the fitness tests are in the Thesis CD and freely available in Spanish and English at: http://profith.ugr.es/recursos/prefit.

Section 2: Physical fitness, academic achievement and brain in preadolescent children

Study VIII and Study IX. Physical fitness, physical activity and academic achievement; and differences in brain in metabolically healthy vs. unhealthy: associations with academic achievement and role of cardiorespiratory fitness

Note: The information about the methods of the ActiveBrains project included in this section *(Study VIII and IX)* is extracted from methodological paper of the project, published by Cadenas-Sanchez et al. (see short curriculum vitae), which is also one of the studies included in this thesis listed as Study 0 of the Thesis, since it appears in the thesis book earlier (i.e. Methods section) than the Studies I to IX that are presented in the Results and Discussion sections.

Study design

The ActiveBrains project is an individual randomized controlled trial (1:1) that aimed to examine the effects of a 20-weeks physical exercise programme on brain structure and function, cognitive performance, academic achievement and physical and mental health in overweight/obese children. For this Thesis we used only baseline data of the project. Participants with overweight/obesity, meeting the eligibility criteria, were included in our study. The ActiveBrains project has been approved by the Review Committee for Research Involving Human Subjects at the University of Granada (Reference: 848, February 2014), and registered in the ClinicalTrials.gov (Identifier: NCT02295072).

Participants

A total of 110 overweight/obese children (categorized based on WOF cut-off points[85]) aged 8-11 years were recruited from Granada (Spain). Parents or legal guardians were informed of the aims of the project and signed the informed consent to participate in the study.

Particularly, for the *Study VIII,* 106 (10.0±1.1 years, 61 boys) had valid data for fitness, physical activity and academic achievement outcomes, and were included in the analyses. On the other hand, for the *Study IX*, a total of 97 children (10.0±1.2 years, 60 boys; metabolically healthy n=52) with valid data were included in this study.

Measurements

Anthropometry (*Study VIII and IX*)

Weight, height, BMI, waist circumference and triceps and subscapular skinfolds thickness are evaluated. In addition, we assess peak height velocity as an accurate and discriminant measure of maturational status[91]. For this purpose we additionally measured sitting height and leg length; using this information we can calculated peak height velocity using the following equation: −9.236 + 0.0002708·Leg Length and Sitting Height interaction −0.001663·Age and Leg Length interaction + 0.007216·Age and Sitting Height interaction + 0.02292·Weight by Height ratio (for boys) and -9.376 + 0.0001882·Leg Length and Sitting Height interaction + 0.0022·Age and Leg Length interaction + 0.005841·Age and Sitting Height interaction − 0.002658·Age and Weight interaction + 0.07693·Weight by Height ratio (for girls)[91].

Physical fitness: Field (*Study VIII and IX*) *and laboratory (Study VIII) tests*

Physical fitness was assessed by the ALPHA battery[19]. Collectively, these fitness tests have been shown to be feasible, reliable and valid for this age group[2,17,18]. Briefly, cardiorespiratory fitness was assessed by a 20m shuttle-run test, which consisted of running back and forth between two lines (20m apart) following an audio signal. The test started at 8.5 km/h and the speed progressively increased 0.5 km/h per minute. The test finished when the participant stopped because of exhaustion or when they did not reach the end lines concurrent with the audio signal on two consecutive laps. The number of completed laps was recorded for the analyses. Muscular strength was assessed by handgrip strength and standing long jump test. Handgrip strength test measures the maximal strength of the upper-limb using a digital dynamometer (TKK 5401, Grip-D, Takei, Tokyo, Japan). Children squeezed as much as possible for 2-3 seconds. Each hand was measured twice and the best result of each hand was retained and averaged for the analyses. The standing long jump test assesses the explosive strength of the lower-limbs. This test required the participant to jump as far as possible and remaining upright. The distance jumped from the take-off line to the point where the back of the heel nearest to the take-off line landed on the ground was recorded. This test was performed three times and the best result was used for the analyses. Speed-agility fitness was assessed by the 4x10m shuttle-run test. This test consisted of running and changing direction as fast as possible between 2 lines (10 m apart). Children had to exchange sponges at each line (4 times x 10 meters). For this test, the fastest completion time was recorded in seconds (sec). This test was performed twice, and the best result (the lowest completion time) was used. For analyses purposes, we inverted this variable by multiplying test completion time (sec) by −1. Thus, higher scores indicated better performance.

Additionally, we assessed cardiorespiratory and muscular strength fitness in laboratory conditions. Cardiorespiratory fitness was assessed by a maximal incremental treadmill test recommended by the American College of Sport Medicine for poorly fit children[92]. The test consisted of walking on a treadmill at a constant speed (4.8 km/h) starting at a 6% slope with grade increments of 1% per minute until volitional exhaustion. The criteria for reaching maximal oxygen uptake were: volitional fatigue (>8 points in the OMNI scale), a plateau in maximum oxygen consumption (VO_2max) during the last two stages of exercise (<2.0 mL·kg−1·min−1), achieved >85% of age-predicted maximum heart rate (MHR), and/or a respiratory exchange ratio ≥1.0.

Muscular strength was assessed by one repetition maximum (1-RM) exercises. Children performed both bench press and leg press tests to evaluate upper- and lower-limb muscular strength, respectively. A familiarization phase was performed with the aim of ensuring that children knew the technique and could perform the exercise correctly (i.e. control movement and proper breathing). The protocol for 1-RM tests in children has been published previously[93]. All tests were performed once and on different days.

Physical activity (*Study VIII*)

Physical activity was assessed by accelerometers. A tri-axial accelerometer (Actisleep GT3x, Pensacola, FL, USA) is used to assess physical activity and sedentary time over seven consecutive days. Participants are instructed to wear two devices: one attached using an elastic band to the right hip and one attached to the non-dominant wrist (which in all cases is the left one). Children wear the accelerometers during 7 consecutive days (24hrs) and remove it only while bathing or swimming. Also, children have a log in order to record the time when they go to bed, wake up, and remove the device.

Raw accelerations collected at a sampling frequency of 100 Hz were processed to derive the Euclidean norm minus one g (ENMO) metric in R (v. 3.1.2, https://www.cran.r-project.org/) using the GGIR package (v. 1.5-12, https://cran.r-project.org/web/packages/GGIR/)[94]. Detailed information about the data processing is shown in Annexes (Supplemental Information 1)**.** Physical activity (i.e. light, moderate, vigorous and MVPA) was classified based on Hildebrand et al.[95,96] cut-off points. To thoroughly investigate the independent associations of MVPA with academic achievement, we also identified bout durations lasting at least 1, 5, and 10 minutes with a drop tolerance of the 20% of the time.

Academic achievement: Woodcock-Muñoz (Study VIII, and IX) and school grades (Study VIII)

Woodcock-Muñoz test battery (the Spanish adaptation of the Woodcock-Johnson III) is a standardized academic achievement test, which measures components of language, mathematics and sciences[97]. This battery is divided in two parts: standard and extended. In our study, we used the standard tests, which contain: three tests of reading, two tests of oral language, three tests of written language, and three tests of mathematics. Further, since science is considered a core academic subject in school, we additionally evaluated one test of the extended battery, which was based on the biological sciences, social sciences, and humanities. All tests were performed individually by a trained evaluator and the testing time range from 100 to 120 min.

Data registered for each child was independently checked by two trained evaluators. All data were recorded in the Compuscore and Profile software version 3.1 (Riverside Publishing Company, Itasca, IL, USA) and grouped into components (i.e. total achievement, reading, oral language, writing, written expression, mathematics, calculation skills, and sciences). Detailed information about the interpretation of each of these components is provided in Annexes (Table S4). The standard score punctuation was selected for the analyses.

As a second indicator of academic achievement, school grades were collected from the official school records in 80% of the study sample (n=83). Since school records were provided qualitatively (i.e. insufficient, sufficient, good, very good, outstanding), we registered the grades based on a scale from 1 to 5. Mean grade of the previous academic course per subject was used for the analyses. Grade-point average, language, mathematics, natural and social sciences, and foreign language (i.e. English) were recorded.

Blood analysis (Study IX)

Cardiometabolic risk factors such as serum triglycerides concentration, glucose, and high-density lipoprotein cholesterol were measured from fasting blood samples collected at the hospital between 8:00 and 10:30h. Systolic and diastolic blood pressures were measured in a sitting position from the left arm with an automatic sphygmomanometer (Omron M6, the Netherlands). For blood pressure, two measurements were collected on different days and the minimum value was registered for the analyses.

Based on previous reviews conducted by our group[67,79], children were characterized as MHO if they did not present altered values for any of the following four risk factors (excluding waist circumference): triglycerides, glucose, high-density lipoprotein and systolic and/or diastolic blood pressure. On the contrary, MUO individuals were classified if children present one or more values indicative of metabolic disturbances (i.e., one or more altered values for the risk factors mentioned above). In this study, when we refer to MHO and MUO we are including both overweight and obese children due to the relatively small sample size (n=97).

Cut-offs for metabolically healthy and unhealthy categorization were based on sex- and age-specific (from 12 to 18 years) for adolescent population provided by Jolliffe and Janssen[73]. Due to the age range of our sample was from 8 t o11 years, we used the cut-off provided for boys and girls aged 12 (see Table S5). The strengths of these cut-offs are the link to the International Diabetes Federation and Adult Treatment Panel, and the adaptation by age-and sex-specific based on growth curves.

Brain (Study IX)

Brain volume was assessed by magnetic resonance imaging (Siemens Magnetom Tim Trio, 3T, Germany) using a whole-brain analytical approach. A high resolution sagittal three dimensional T1-weighted image was collected using magnetization-prepared rapid gradient-echo sequence (MPRAGE)[12]. Total acquisition time for the T1 sequence was 7 min and 31s for every child (Total time= 40-45min, including rest between MRI sequences).

Structural images were pre-processed using the Statistical Parametric Mapping software (SPM, version 12, Wellcome Department of Cognitive Neurology, London, United Kingdom). First, we checked the T1-images for each participant to detect any artefacts/movement and we aligned the image to the anterior and posterior commissure. After this first screening, we segmented T1-weighted structural images into gray matter, white matter, and cerebrospinal fluid[98]. Second, we created a template using Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL)[99] based on gray and white matter tissues images. After template creation, DARTEL estimates the best set of smooth deformations for every child's tissue to their common average, applies the deformations to create a new average, and finally reiterates the process until convergence. The resultant images were spatially normalized to Montreal Neurological Institute (MNI) space with affine transformation to create the

DARTEL template. Each participant's segmented images were normalized to the DARTEL template. To perform a volume change correction, the normalized gray matter images were modulated (each voxel) with the Jacobian determinants derived from spatial normalization[100]. Lastly, the volumetric images were smoothed by convolution with an isotropic Gaussian kernel of 8mm full-width at half-maximum (FWHM). The calculation of the global brain volume was derived by the non-normalized segmented images.

Table 1. Summary table of the methods used in this International Doctoral Thesis.

Project	Study	Design	Participants	Main variables	Methods
Section 1					
PREFIT Project	I. Systematic review of physical fitness in preschoolers.	Systematic review	-	Studies conducted in preschool children that included at least one field-based physical fitness test.	Two major electronic databases to conduct the search (i.e., PubMed and Web of Science).
PREFIT Project	II. Reliability and validity of handgrip dynamometers.	Cross-sectional	-	Different models of dynamometers (digital and analog) and years (new and old).	The dynamometers were positioned between two wooden supports with the handle fixed. Known weights were suspended with a loading belt from the center of the dynamometer's handle.
PREFIT Project	III. Adaptation of the original 20m shuttle run test in preschool children.	Cross-sectional	N = 130	Cardiorespiratory fitness and MHR.	Incremental field PREFIT 20m shuttle run test with MHR recording each 5 seconds.
PREFIT Project	IV. Feasibility, reliability and practical recommendations for the PREFIT battery.	Cross-sectional	N = 161	Anthropometry and physical fitness.	Weight, height, waist circumference, PREFIT 20m shuttle run, handgrip strength, standing long jump, 4x10m shuttle run and one-leg stance tests.
PREFIT Project	V. Physical fitness reference standards for Spanish preschool children.	Cross-sectional	N = 3179	Physical fitness.	PREFIT 20m shuttle run, handgrip strength, standing long jump, 4x10m shuttle run and one-leg stance tests.
PREFIT Project	VI. Weight status and fatness reference standards for Spanish preschool children.	Cross-sectional	N = 3178	Weight status categories and anthropometry.	Prevalence of underweight, normal weight, overweight, obesity type I, obesity type II and obesity type III based on WOF and WHO, weight, height, BMI, waist

					circumference and waist to height ratio.
PREFIT and MINISTOP Projects	VII. Comparison of fatness and physical fitness between Swedish and Spanish preschool children.	Cross-sectional	PREFIT, N = 128 MINISTOP, N = 315	Anthropometry and physical fitness.	Weight, height, waist circumference, waist to height ratio, overweight/obesity prevalence, original 20m shuttle run, PREFIT 20m shuttle run, handgrip strength, standing long jump, 4x10m shuttle run and one-leg stance tests.
Section 2					
ActiveBrains Project	VIII. Physical fitness, physical activity and academic achievement preadolescent children.	Cross-sectional	N = 106	Anthropometry, physical fitness, physical activity and academic achievement.	Weight, height, BMI, 20m shuttle run test, handgrip strength test, standing long jump test, 4x10m shuttle run test, incremental maximal treadmill test, 1-Repetition maximum bench press and leg press, light physical activity, moderate PA, Vigorous PA, MVPA, 1-min bouts MVPA, 5-min bouts MVPA, and 10-min bouts MVPA, Woodcock Muñoz test, and school grades.
ActiveBrains Project	IX. Differences in brain in metabolic healthy vs. unhealthy: associations with academic achievement and role of cardiorespiratory fitness.	Cross-sectional	N = 97	Brain volume, academic achievement and cardiorespiratory fitness.	Global and regional gray matter volume, total academic achievement and 20m shuttle run test.

BMI = Body mass index, MHR = maximum heart rate, MVPA = Moderate-to-vigorous physical activity, PA = Physical activity, WOF = World Obesity Federation, WHO = World Health Organization. Studies that for their methodological characteristics cannot meet any of the criteria of the table has been marked with an hyphen (i.e., -).

Statistical analyses

Descriptive statistics are shown as means and SDs or frequencies. Percentages were used when necessary. In order to decide whether to analyse the data for whole sample together or separated by sex, we overall examined sex interactions. Confounders were selected after examining their influence on the estimates analysed in each study. All the statistical approaches in the different studies were performed by using the Statistical Package for the Social Sciences software (version 20.0, IBM Corporation) and the level of significance was set at $p < 0.05$, unless otherwise indicated.

The statistical approach undertaken to accomplish the aims of this International Doctoral Thesis is presented below and is summarized in **Table 2.**

Section 1: Physical fitness and fatness in preschool children

Study I. Systematic review of physical fitness in preschoolers

No statistical analyses were performed due to the study design (i.e., systematic review).

Study II. Reliability and validity of handgrip dynamometers

The agreement among intra-instrument, inter-instrument, and inter-model trials (i.e., reliability) and the agreement between the known weights and dynamometer measures (i.e., criterion-related validity) were assessed using the Bland and Altman's method[101]. Mean difference (error) and the 95% limits of agreement (error ± 1.96 of the standard error of the difference) were calculated. Results were graphically examined by plotting the differences against their mean[101]. One-sample t-test was used to test whether the mean difference (i.e., systematic error) was significantly different from zero (reference).

Study III. Adaptation of the original 20m shuttle run test in preschool children

In order to test the maximum level of the test in preschool children, two-way ANOVA was performed, having sex and age groups as fixed factor and theoretical MHR and the percentage of the MHR as dependent variable. Due to there is no equation for estimating the MHR in preschoolers, we decided to use the equation 208 – 0.7 (age) following the results obtained by Mahon et al.[102] in >7 years old children. Age groups were created according the following categorization: 3 years (from 3.0 to 3.9 years), 4 years (from 4.0 to 4.9 years) and 5 years (from 5.0 onwards).

Test-retest reliability of the PREFIT 20m shuttle run test was assessed following the Bland and Altman's method[101]. We calculated the mean difference (systematic error) and 95% limits of agreement (error ± 1.96 of the difference). Results were graphically examined by plotting the differences against their mean[101]. To know whether the mean difference (test-retest) of the PREFIT 20m shuttle run test was significant different between sex and age, we performed two-way ANOVA, having sex and age as fixed factor and the mean difference of the test as dependent variable. Likewise, with the aim to examine the concordance-correlation between test-retest measurements of the

PREFIT 20m shuttle run and MHR, we applied the Lin's concordance correlation coefficient[103].

For reliability purposes, the heteroscedasticity analysis is important since shows whether the error increases or decreases as the magnitude of the measure changes. To calculate it, negative values from the mean difference were changed to positive (multiplying by -1 the negative values), using this variable as dependent in a one-way ANOVA analysis. As fixed factor, cardiorespiratory fitness level was grouped as quartiles, being quartile 1 categorized as low level of fitness and quartile 4 as high level of fitness.

Study IV. Feasibility, reliability and practical recommendations for the PREFIT battery

The agreement between test-retest trials of all tests was assessed following the Bland-Altman's method[101]. One sample t-test was used in order to test whether error was significantly different from zero (reference). To know whether the mean differences of the tests studied were significantly different between sex and age, we used two-way ANOVA. To test whether the reliability differs across age groups, we used a polynomic contrast with linear trend. To determine the correlation-concordance between test-retest measures, we used the Lin's concordance correlation coefficient[103]. Cohen's effect size was also calculated[104].

Heteroscedasticity analyses showed whether the error increases or decreases as the magnitude of the measure changes. To calculate it, the negative values of the difference were transformed to positive and fitted into a one-way ANOVA model. Fitness groups were represented by quartiles based on the sex and age group, i.e. quartile 1 corresponds with a low level of fitness and quartile 4 corresponds to a high level of fitness, except in the 4x10m shuttle run test, where the quartile 1 corresponds to the fastest participants and the quartile 4 to the slowest participants.

Studies V to VI. Physical fitness and fatness reference standards in preschool children

To obtain percentile curves for preschool children, we applied the Generalized Additive Model for Location, Scale and Shape (GAMLSS)[105]. We used the GAMLSS package (version 4.4-0) of the statistical software R (version 3.3.1). GAMLSS is able to model up to four parameters of different distributions: μ accounts for the location, σ for the scale, υ for the skewness, and τ for the kurtosis. The Box-Cox Cole and Green, Box-Cox t, and Box-Cox power exponential distribution were fitted to the observed data. Furthermore, the influence of age on the distribution parameters was modelled constantly, linearly, or as a cubic spline function. We assessed the goodness of fit applying the Bayesian information criterion and worm plots[106]. More information about the procedure has been described elsewhere[107]. Percentile curves for the 1^{st}, 3^{rd}, 5^{th}, 10^{th}, 15^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, 85^{th}, 90^{th}, 95^{th}, 97^{th}, and 99^{th} percentiles were calculated based on the model that showed the best goodness of fit (Table S6). We provide reference standards with a precision of 0.25 years of age (each trimester) as the main outcome of this article and also every 0.025 years of age (equivalent to 9 days).

For *Study V*, to test sex differences in fitness across the whole age period and percentiles 25^{th}, 50^{th}, and 75^{th}, we performed a t-student test for independent samples. We used these percentiles to test sex differences in the fit preschoolers (percentile 75^{th}), in averagely fit preschoolers (percentile 50^{th}), and in unfit preschoolers (percentile 25^{th}). We had previously calculated and depicted mean differences for every 0.025 years of age (i.e., 9 days) and percentiles.

Study VII. Comparison of fatness and physical fitness between Swedish and Spanish preschool children

Pearson's chi-square was calculated to test differences between countries in overweight/obesity prevalence. Differences in anthropometry and physical fitness between preschool children from Sweden and Spain were analysed by analysis of covariance (ANCOVA). To test whether differences in anthropometry and fitness between preschoolers from Sweden and Spain might be due to age, sex, height or BMI, we tested the differences using ANCOVA with and without these confounders. Anthropometric tests were adjusted for age and sex, except waist circumference that was additionally adjusted for height. This additional adjustment was needed as taller individuals have larger bones, including larger ribs and pelvis, therefore, having a larger waist circumference does not necessary indicate more abdominal fat[108]. To test whether differences in fitness could be explained by BMI, we first adjusted all the fitness tests for age and sex (model 1), and secondly, we additionally adjusted for BMI (model 2). To make different variables visually comparable above/below the mean and on the same scale, they were all transformed into sex-specific z-scores. There were significant sex interactions in 20m shuttle run test and thus these data were presented for boys and girls separately.

Section 2: Physical fitness, academic achievement and brain in preadolescent children

Study VIII. Physical fitness, physical activity and academic achievement preadolescent children

Linear regression analyses were performed, which included only those outcomes that were strongly correlated with the independent measures ($p<0.01$) to better describe the pattern of the association of physical fitness components and physical activity with academic achievement. Basic confounders such as age, sex and maternal education were introduced in the analyses (model 1). Additional analyses were performed by adjusting for BMI and cardiorespiratory or speed-agility fitness (model 2).

Physical activity estimates were presented based on those obtained from the wrist-worn accelerometer and calculated with ENMO metric using the Hildebrand et al.[95] cut-points. Nevertheless, given that different placements and processing criteria result in different estimations of physical activity[109,110], we also examined whether the methods used for physical activity estimations, i.e., ENMO for hip and activity counts for hip and wrist, change the results in its association with academic achievement. This was done since none of the methods have thus far demonstrated to out-perform the rest. Thus, until a harmonization is reached, it is highly relevant to evaluate whether investigated associations between accelerometer variables and investigated outcomes differ when using different placements and processing criteria[109].

Study IX. Differences in brain in metabolic healthy vs. unhealthy: associations with academic achievement and role of cardiorespiratory fitness

To test whether differences existed between MHO and MUO in global gray matter, we performed ANCOVA models having MHO/MUO as a fixed factor, and global volume as the dependent variables. Confounders (i.e., sex, peak height velocity, parental education university level and BMI) were selected after examining their influence on the estimates. Thus, data were presented without any adjustment, adjusted for basic confounders, and additionally for cardiorespiratory fitness. As exploratory analyses (presented in Supplementary material), we also examined the differences between MHO and MUO in global white matter and total brain volume. All statistical procedures (t-test, chi-square, and ANCOVA) were performed using IBM SPSS Statistical Software (version 20 for Windows) with an alpha level of 0.05.

For regional gray matter volume, we performed general linear model analyses using the SPM12 software (Wellcome Department of Cognitive Neurology, London, United Kingdom). The difference in gray matter between MHO and MUO was analysed using factor models, adjusted for sex, peak height velocity, parental education university level, BMI, and total brain volume (model 1). We additionally adjusted for cardiorespiratory fitness (model 2). For those brain regions that showed statistical significance, we extracted the eigenvalues of each significant cluster (k). Effect size (Cohen's d and its 95% confidence intervals) was calculated for each cluster extracted. The statistical threshold in the regional gray matter analyses was calculated with AlphaSim in Resting-State functional magnetic resonance imaging Data Analysis Toolkit toolbox plus v1.2 (RESTplus)[111]. Parameters were defined as follows: cluster connection radius (rmm) = 5 mm and the actual smoothness of the data after model estimation, incorporating a gray mask volume of 12819 voxels. The voxel-level alpha significance (threshold, $p < 0.001$ uncorrected) along with the appropriate cluster size for controlling for multiple comparisons in each analysis were indicated in the results. The resulting cluster extents were further adjusted for Hayasaka et al.[112] to account for the non-isotropic smoothness of structural images. As sensitivity analyses, we further analysed how these differences differ between MHO and MUO in only obese sample.

Lastly, to examine the associations between global gray matter volume and academic achievement, we performed a linear regression analysis, where academic achievement was used as dependent variable, and gray matter volume as an independent variable. Model 1 was adjusted for basic confounders (i.e., sex, age, parental education university level and BMI) and Model 2 additionally adjusted for cardiorespiratory fitness. Likewise, as exploratory analysis and in order to know the inter-relationships among main exposures, outcomes and confounders studied (i.e., sex, age, parental education university level, BMI, metabolic phenotype, gray matter and academic achievement), we applied Structural Equation Modeling methods using SPSS Amos statistical software. The model fit was examined using the comparative fit index (CFI) and the root mean square error of approximation (RMSEA)[113]. CFI is considered good when is higher than 0.9. RMSEA was considered acceptable when is lower than 0.05 and excellent when is lower than 0.01[113].

Table 2. Summary table of the statistical approach of each study included in the present International Doctoral Thesis.

Study	Statistical analysis
Section 1	
I. Systematic review of physical fitness in preschoolers.	-
II. Reliability and validity of handgrip dynamometers.	-Reliability and validity: Bland and Altman's method and one-sample t test.
III. Adaptation of the original 20m shuttle run test in preschool children.	-Feasibility: - -Maximality: two-way ANOVA (fixed factor: sex and age; dependent variable: theoretical MHR and the percentage of the MHR). -Reliability: Bland and Altman's method and two-way ANOVA (fixed factor: sex and age; dependent variable: mean difference). Heteroscedasticity: one-way ANOVA (fixed factor: cardiorespiratory fitness level categorized in quartiles; dependent variable: mean difference in absolute values). Correlation-concordance: Lin's concordance correlation coefficient.
IV. Feasibility, reliability and practical recommendations for the PREFIT battery.	-Feasibility: - -Reliability: Bland and Altman's method. One-sample t test and two-way ANOVA (fixed factor: sex and age; dependent variable: mean difference of the tests studied). Heteroscedasticity: one-way ANOVA (fixed factor: fitness level categorized in quartiles; dependent variable: mean difference in absolute values). Correlation-concordance: Lin's concordance correlation coefficient. -Practical recommendations: -
V. Physical fitness reference standards for Spanish preschool children.	-Reference standards: GAMLSS method. -Sex differences across the whole age period and percentiles: Independent t-test.
VI. Weight status and fatness reference standards for Spanish preschool children.	-Weight status prevalence: Frequencies. -Reference standards: GAMLSS method.

VII. Comparison of fatness and physical fitness between Swedish and Spanish preschool children.	-Differences between countries in weight status: Pearson chi-square. -Differences in anthropometric characteristics: ANCOVA (adjusted for age and sex. Waist circumference was additionally adjusted for height). -Differences in fitness characteristics: ANCOVA (model 1: adjusted for age and sex, model 2: additionally adjusted for BMI).
Section 2	
VIII. Physical fitness, physical activity and academic achievement preadolescent children.	-Association of fitness and physical activity on academic achievement: linear regression model (independent variable: fitness/physical activity and dependent variable: academic achievement; model 1: adjusted for age, sex, and maternal education; model 2: additionally adjusted for BMI and cardiorespiratory fitness or speed-agility fitness).
IX. Differences in brain in metabolic healthy vs. unhealthy: associations with academic achievement and role of cardiorespiratory fitness.	-Differences between metabolic phenotypes in global brain: ANCOVA (fixed factor: metabolically healthy/unhealthy; dependent variable: gray matter volume; model 1: unadjusted, model 2: adjusted for sex, peak height velocity, parental education and BMI, model 3: additionally adjusted for cardiorespiratory fitness). -Differences between metabolic phenotypes in regional gray matter volume: linear regression analyses (factor models; model 1: adjusted for sex, peak height velocity, parental education, BMI, and total brain volume; model 2: additionally adjusted for cardiorespiratory fitness). -Associations between gray matter and academic achievement: Linear regression model (independent variable: global gray matter volume dependent variable: academic achievement; model 1: adjusted for sex, age, parental education and BMI; model 2: additionally adjusted for cardiorespiratory fitness).

ANOVA= analysis of variance. ANCOVA= analysis of covariance. BMI= Body mass index. MHR= maximum heart rate, GAMLSS= Generalized Additive Model for Location, Scale and Shape.

RESULTS

RESULTS

The results of each individual study comprising the present International Doctoral Thesis are presented below.

Section 1: Physical fitness and fatness in preschool children

Study I. Systematic review of physical fitness in preschoolers

We carried out a two-step selection procedure with the 2,109 articles initially identified: (1) screening based on title and abstract; (2) search of the full text of the articles selected in the previous step for comprehensive reading and final inclusion/exclusion. **Figure 4** graphically shows the flow of the search process.

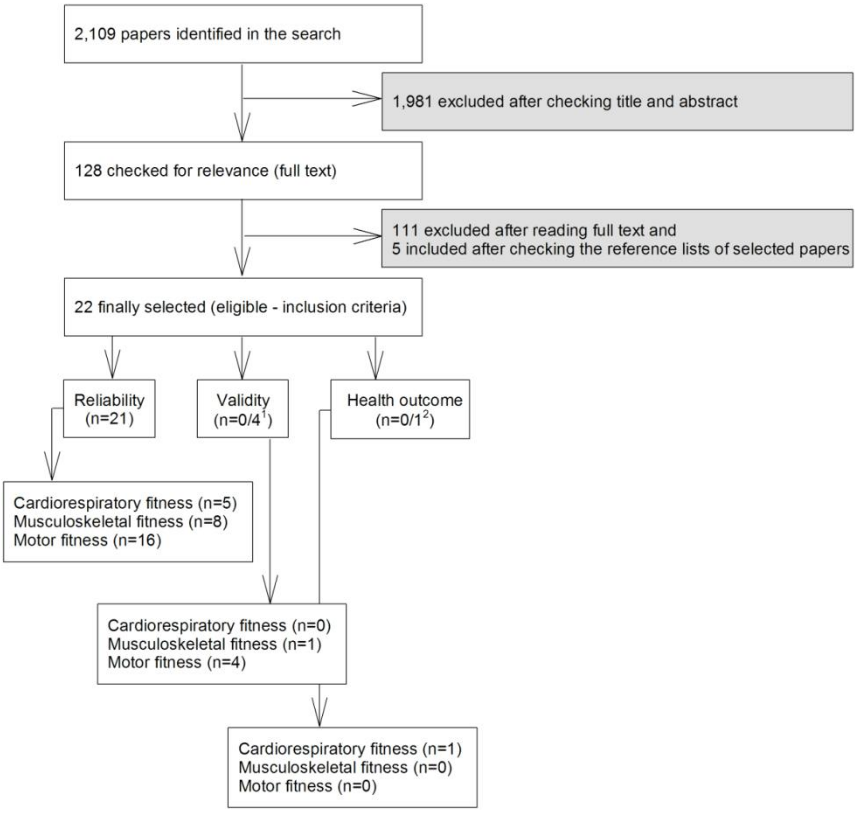


Figure 4. Flow chart of the literature search and paper selection process.
[1] Zero studies examining criterion-related validity / 4 studies examining construct (convergent) validity.
[2] Zero longitudinal studies / 1 cross-sectional study.

We found a limited number of studies examining the reliability, validity or relationship with health outcomes of field-based fitness tests in preschool children (See **Tables 3 to 5**): 21 (96%) articles examined reliability, 0 examined criterion-related validity and 4 (5%) examined construct (convergent) validity. Otherwise, 0 longitudinal and 1 (5%) cross-sectional studies examined the relationship with health outcomes. These findings show that the majority of the currently available information about fitness in preschool age is concerning reliability of field-based physical fitness tests, and that nearly no information is available about validity and relationship with health outcomes in preschool children.

Noteworthy is that the number of studies found was markedly smaller than those found in our previous systematic reviews on reliability[17], criterion-related validity[18] and longitudinal relationship with health outcomes[2] of field-based fitness tests in children and adolescents aged 6 to 18 years, even though these reviews retrieved studies published from 1990 to 2008/9 while the present review included all studies published until April 1st 2014 (no starting date). These findings illustrate the lack of information about which physical fitness tests should be used in preschool children, supporting the need to further research on this topic and population. Further, contrasting what was found in older children and adolescents[2,17,18], cardiorespiratory fitness was the least studied component of physical fitness in preschool children, being motor fitness and particularly balance, the most studied component. Figure 4 shows the exact number of studies examining the reliability, validity and relationship with health outcomes of different physical fitness components.

Cardiorespiratory fitness

Reliability

A total of 5 (23%) studies examined the reliability of field-based tests assessing cardiorespiratory fitness (Table 3). The 20m shuttle run[114,115], 1/2-mile walk/run[115,116] and distance run in 3min[117,118] tests were examined for reliability twice; whereas the 1-mile walk/run[116] and 3/4-mile walk/run[116] tests were examined only once. It is important to highlight that all these studies were conducted in children aged 4 to 6 years, so that reliability of these tests in preschool children aged 3 years is not known. Rikli et al.[116] did the most comprehensive reliability study in a sample of 180 preschoolers aged 5 years and concluded that the 1/2-mile walk-run test is more reliable and recommendable to be used in preschool children (r>0.73) than the other two walk-run tests (ranged from 0.34 to 0.63). Niederer et al.[114] reported a good reliability for the 20m shuttle run test in Swiss preschool children aged 4 to 6 years (r=0.84). It is not possible to test if the 20m shuttle run test is more or less reliable than the 1/2-mile walk-run test, since different statistics were used in different studies (i.e. correlation coefficients, intra-class correlation coefficient (ICC) and paired t-tests for repeated measures)[114-116]. Reeves et al.[115] studied both of these tests, but statistics for each test were not provided and it was only reported that no significant test-retest differences were found for any test. Finally, the distance run in 3 min also showed acceptable reliability in children aged 4 to 6 years, as reported by Bénéfice et al.[117] (r=0.83 for girls and 0.82 for boys) and Oja et al.[118] (ICCs ranged from 0.60 to 0.80).

Validity

Unfortunately, no criterion-related validity study or construct (convergent) validity study of field-based cardiorespiratory fitness tests in preschool children was found (Table 4). We cannot therefore know which of the field-based tests available is more valid in preschool children.

Relationship with health outcomes

Likewise, no longitudinal study was found examining the association between performance in field-based cardiorespiratory fitness tests in preschool children and future health outcomes (Table 5). Only one (5%) cross-sectional study was found and concluded that the 1/2-mile walk-run test was negatively correlated with body weight (r=-0.49), BMI (r=-0.52) and percent body fat (r=-0.54)[115]. This study also included the 20m shuttle run test, but no information was provided about its association with these anthropometric markers.

Musculoskeletal fitness: muscular strength and flexibility

Reliability

The reliability of handgrip strength test in preschool children was analyzed in 2 (9%) studies (Table 3). Molenaar et al.[119], examined the test-retest reliability of two handgrip strength dynamometers in 30 preschoolers aged 4 to 6 years, and concluded that the Lode dynamometer was more reliable than the Martin vigorimeter (ICC=0.91 compared with 0.76 respectively, for dominant hand and similar difference for non-dominant hand). Bénéfice et al.[117] studied also the reliability of the Martin vigorimeter in 84 preschool children aged 4 to 6 years and observed correlation coefficients of 0.84 and 0.71 for girls and boys respectively.

The standing long jump test was also studied in 3 (14%) of the reliability papers selected (Table 3). Krombholz[120] observed a correlation coefficient of 0.68 for standing long jump when assessed twice 8 months apart in 3 to 7 year-old children. It is important to note that test-retest intervals are usually 1-2 weeks, and the 8 months used by Krombholz might reflect developmental changes that take place relatively quickly at these early ages. Bénéfice et al.[117] observed similar correlation coefficients for standing long jump as those reported for handgrip, i.e., 0.85 and 0.71 for girls and boys respectively. Likewise, Oja et al.[118] observed acceptable reliability statistics for the standing long jump test in 4 and 5 year-old children, i.e., ICC=0.65 and 0.89 for girls and boys respectively. Reeves et al.[115] reported no significant differences between the test and retest for all the fitness tests of the muscular strength subset of the BOTMP test-battery, which includes the standing long jump test. Other muscular strength tests, such as curl-ups[115], sit-ups (ICCs ranged from 0.61 to 0.78)[118], trunk lifts[115], bent arm hang[115], straight arm hang (r from 0.61 to 0.76)[120], throws with dominant hand (ICCs ranged from 0.70 to 0.80)[118], muscle strength of the fingers and thumb (using the

Rotterdam Intrinsic Hand Myometer) (ICCs ranged from 0.61 to 0.98)[121], as well as strength of the foot and ankle (ICCs ranged between 0.85 and 0.96)[122], were examined (1 study per test) and reported acceptable reliability for all of them. Likewise, reliability of the sit-and-reach rest[119,123] and its later version the back-saver sit and reach test[115], were studied (1 study per test) and acceptable reliability was also reported (ICC for sit and reach test= from 0.74 to 0.94).

Validity

Our search retrieved no criterion-related validity study of musculoskeletal fitness. We found only the article by Deitz et al.[124] in which the authors review the construct (convergent) validity of the Bruininks-Oseretsky Test of Motor Proficiency (BOT-2) test battery in 4 to 5 year-old children (Table 4). The authors reported a moderate correlation (r=0.75) between the muscular strength subset of the BOT-2 battery and the Peabody Developmental Motor Scales, second edition (PDMS-2) Gross Motor Quotient (which is designed to measure similar constructs). There is however no information about which test included in this subset is more valid, since only statistics for the overall score of the subset were provided. No validity study on flexibility was found.

Relationship with health outcomes

We did not find any longitudinal or cross-sectional study relating musculoskeletal fitness field-based tests and health outcomes in preschool children (Table 5).

Motor fitness: speed-agility and balance

Reliability

Our search identified 6 (27%) reliability studies of speed-agility tests: 4 (18%) studies for different versions of the shuttle run test (4x10m shuttle without sponges[117,118] 4x4m shuttle run with wooden blocks[120] and 100 feet (32.8m) shuttle run from the BOTMP test-battery[115]) , one (5%) for a 25m sprint test[125] and one (5%) for an obstacle course test[114](Table 3). The results reported in these studies suggest that all these tests provide a reliable measure of speed-agility in preschool children.

Most of the reliability studies selected in this review were focused on balance, n=12 (55%) studies[114,115,126-136]. The tests used to assess balance were many and very heterogeneous aiming mainly to assess static and dynamic balance. In most studies, the authors reported statistics for overall scores of a balance subsets belonging to test batteries, e.g. BOTMP, BOT-2, Movement Assessment Battery for Children-second edition (MABC-2), Preschooler Gross Motor Quality Scale (PGMQ) and Maastricht's Motor Test (MMT), rather than on individual tests. Since the purpose of our review is to select "the best" tests to assess health-related fitness components, studies reporting reliability statistics for single test are more valuable for the purpose of the current review. For more information about motor and neuro-developmental test-batteries in preschool children see the review by Cools et al.[136].

Some of the batteries include many items, for instance, the Ghent Developmental Balance Test (GDBT) include 35 items (tests) to assess static and dynamic balance. Likewise, the Maastricht's Motor Test (MMT) includes a quantity and a quality scores for a set of static balance tests (14 items) and dynamic balance tests (20 items). After analyzing all the studies selected, the single-leg stance tests seems to be the most commonly used test to assess balance and provide reliable results in preschool children (ICC between 0.73 to 0.99, r=0.84-0.97)[114,115,120,126,129-131,134]. The test has mainly 4 versions, standing on the floor (on a drawn line or without line) or on a beam, and with open or close eyes. The scoring protocols also differ largely among tests, being the most common way to record the time (seconds) that the person is able to keep the requested balance position or the number of times ("errors") that the person fail to maintain the requested position.

Validity

While no criterion-related validity study was found, four (18%) construct (convergent) validity studies of motor fitness were identified in our search[124,126,128,131] (Table 4). Deitz et al.[124] reported that the correlation between the running speed-agility subset of the BOT-2 battery and the PDMS-2 Gross Motor Quotient in 4 to 5 year-old children was moderate (r=0.75); there was however no information about which test included in this subset was more valid, since only statistics for the overall score of the subset were provided.

Similarly, 4 studies examined associations between balance subsets from different batteries (i.e. MABC-2, GDBT, PGMQ, BOT-2 and the Gross Motor Quotient or Stationary subset of the PDMS-2)[124,126,128,131]. Except for the study by Hua et al.[126], that observed a weak correlation (r=0.07), the others reported moderate correlations (all r>0.50) between the mentioned batteries, supporting the construct (convergent) validity of the subsets. However, there is almost no information about which individual test included in these subsets is the most valid and describes better the balance construct.

Relationship with health outcomes

We did not find any longitudinal or cross-sectional study relating motor fitness field-based tests and health outcomes in preschool children (Table 5).

Table 3. Overview of reliability studies on field-based fitness tests in preschool children.

Reference	Age	Participants	Fitness test	Statistical method	Design (time interval)	Main outcomes and conclusions
Cardiorespiratory fitness						
Niederer et al.[137]	4-6 yr	n=20	20m shuttle run	Correlation (not reported if Pearson or Spearman)	Test-retest (interval between trials not reported)	r=0.84
Reeves et al.[138]	5-6 yr	n=17	20m shuttle run (PACER) 1/2-mile walk/run	Paired t test for differences between repeated measures	Test-retest (1 week)	Statistics for individual tests were not reported. No significant differences were found between tests conducted on day 1 versus day 2.
Bénéfice et al.[139]	4-6.5 yr	n= 84 (36 girls, 48 boys)	Distance run in 3 min	Correlation (not reported if Pearson or Spearman)	Test-retest (1 day)	Girls, r=0.83 Boys, r=0.82
Oja and Jürimäe[140]	4-5 yr	n=124 (60 girls and 64 boys)	Distance run in 3 min	ICC, Cronbach's alpha, and CV (%)	Test-retest (1 week)	The test showed an acceptable reliability; the test was more reliable in 4 year-old girls than in 5 year-old girls. ICC, Cronbach's alpha coefficients and CV are presented below (in this order) by sex and age: Girls, 4yr: 0.795, 0.882 and 7.5 Girls, 5yr: 0.601, 0.643 and 9.1

						Boys, 4yr: 0.749, 0.800 and 4.0 Boys, 5yr: 0.808, 0.903 and 4.3
Rikli et al.[116]	5 yr	n=180 (82 girls, 98 boys)	1-mile walk/run 3/4-mile walk/run 1/2-mile walk/run	ICC using one-way ANOVA	Test-retest reliability (1 week apart) in the fall (October) and the spring (May)	1-mile walk/run: boys: ICC=0.53 in fall and 0.44 in spring; girls: ICC=0.39 in fall and 0.34 in spring 3/4-mile walk/run: boys: ICC=0.48 in fall and 0.58 in spring and girls: ICC=0.58 in fall, 0.63 in spring 1/2-mile walk/run: boys: R=0.77 in fall and 0.88 in spring; girls: R=0.73 in fall and 0.74 in spring. There was no marked difference for reliability estimates between fall and spring. The median of the fall reliability estimates was 0.74, and the spring median reliability was 0.73. Authors' conclusion: the only distance meeting minimal reliability standards is the 1/2 mile for both boys and girls. There was unanimous agreement among the teachers that the 1/2-mile distance is preferable for children in kindergarten.
				Percent of agreement relative to classification categories	Criterion-referenced reliability, i.e. consistency of classification	The 3/4-mile walk/run and 1/2 –mile walk/run tests, but not the 1 mile walk/run test, were acceptably reliable for classifying students with respect to meeting the standards.
Musculoskeletal fitness (muscle strength and flexibility)						
Krombholz[120]	3-7 yr	n=1871	Standing long jump	Pearson	Inter-trials	r=0.88

		(912 girls, 959 boys)	Straight arm hang	correlation	variability (between 2 attempts)	r=0.76
		n=428	Standing long jump Straight arm hang	Pearson correlation	Test-retest reliability (8 month)	r=0.68 r=0.61 The authors concluded that the reliability estimates for these tests were adequate.
Molenaar et al.[121]	4-6 yr	n=16 (8 girls, 8 boys)	Muscle strength of single fingers and thumb (RIHM): Thumb palmar abduction Thumb opposition Thumb flexion at the metacarpal-phalangeal joint Index finger abduction Little finger abduction	ICC, SDD and Bland-Altman plots	Test-retest reliability (26 days)	The authors concluded that the RIHM was reliable for use in preschool children. ICCs ranged between 0.61 and 0.98, and the normalized SDD ranged from 9% to 31%. Bland-Altman plots showed an even distribution around zero.
Molenaar et al.[141]	4-6 yr	n=30 (18 girls, 12 boys)	Handgrip strength measured by 2 methods: Lode dynamometer Martin vigorimeter	ICC and SDD	Test-retest reliability (29 days)	Both the Lode dynamometer and the Martin vigorimeter were reliable instruments to measure the grip strength in preschool children; however, the Lode dynamometer showed better test-retest reliability. Furthermore, comparison of the SDD showed the Lode dynamometer to be a more accurate instrument.

						ICC for Lode-dominant hand=0.91 ICC for Lode-non-dominant hand=0.73 ICC for Martin-dominant hand=0.76 ICC for Martin-non-dominant hand=0.49 Normalized SDD for Lode-dominant hand=27.6% Normalized SDD for Lode-non-dominant hand=35.5% Normalized SDD for Martin-dominant hand=34.9% Normalized SDD for Martin-non-dominant hand=33.6%
Rose et al.[122]	2-4 yr	n=60	Foot and ankle strength measured by a hand-held dynamometer (Citec model)	ICC and SEM	Intra-rater test-retest reliability (later on the same day) and inter-rater reliability (same time)	For all muscle groups reliability was as follows: Intra-rater reliability, ICC=0.85- 0.94, SEM=1.5–4.7N Inter-rater, ICC=0.88–0.96, SEM=1.2–4.6N Reliability was high in preschool children. Reliability was generally highest in 3- and 4-year-old children and lowest in 2-year-old children. Hand-held dynamometry can reliably measure foot and ankle strength in preschool children
Reeves et al.[138]	5-6 yr	n=17	FITNESSGRAM battery: Curl-ups Trunk lift Bent arm hang Back-saver sit-and-reach Shoulder stretch BOTMP battery: [a]	Paired t test for differences between repeated measures	Test-retest (1 week)	Statistics for single tests were not reported. No significant differences were found between tests conducted on day 1 versus day 2.

			Muscular strength subset			
Oja and Jürimäe[140]	4-5 yr	n=124 (60 girls and 64 boys)	Muscular strength: Standing long jump Sit-ups in 30sec Throws, dominant Flexibility: Sit-and-reach	ICC, Cronbach's alpha, and CV (%)	Test-retest (1 week)	Overall, the test showed an acceptable reliability. ICC, Cronbach's alpha coefficients and CV are presented below (in this order) by sex and age: Standing long jump: Girls, 4yr: 0.727, 0.839 and 7.5 Girls, 5yr: 0.572, 0.686 and 9.1 Boys, 4yr: 0.888, 0.945 and 5.2 Boys, 5yr: 0.898, 0.928 and 4.3 Sit-ups in 30sec: Girls, 4yr: 0.629, 0.760 and 25.0 Girls, 5yr: 0.902, 0.946 and 13.4 Boys, 4yr: 0.911, 0.953 and 17.7 Boys, 5yr: 0.608, 0.623 and 23.5 Throws, dominant hand: Girls, 4yr: 0.699, 0.717 and 22.7 Girls, 5yr: 0.782, 0.872 and 8.2 Boys, 4yr: 0.739, 0.830 and 14.2 Boys, 5yr: 0.800, 0.886 and 11.6 Sit-and-reach: Girls, 4 yr: 0.877, 0.924 and 7.6 Girls, 5yr: 0.933, 0.963 and 4.4 Boys, 4yr: 0.847, 0.900 and 6.5 Boys, 5yr: 0.744, 0.786 and 9.5
Bénéfice et al.[139]	4-6.5	n= 84	Handgrip strength	Correlation	Test-retest	Handgrip strength:

	yr	(36 girls, 48 boys)	(Martin vigorimeter) Standing long jump	(not reported if Pearson or Spearman)	(1 day)	Girls, r=0.84 Boys, r=0.71 Standing long jump: Girls, r=0.85 Boys, r=0.71
Koslow[123]	3-4 yr 5-6 yr	n=100 (50 girls, 50 boys) n=100 (50 girls, 50 boys)	Sit and reach	ICC across trials in the same session and in two different sessions	2 measuring sessions two weeks apart; 3 trials per session	ICC across trials: 3-4 yr=0.85 5-6 yr=0.88 ICC across sessions: 3-4 yr=0.80 5-6 yr=0.84 These values were interpreted as an acceptable reliability.
Motor fitness (speed, agility and balance)						
Hua et al.[142]	3-6 yr	n=184 (92 girls, 92 boys)	MABC-2 battery, balance subset: One-leg balance (preferred leg) One-leg balance (non-preferred leg) Walking heels raised Jumping on mats	ICC	Inter-rater (same day) and intra-rater test-retest (2 weeks)	Authors concluded that reliability was good for all the balance tests: Inter-rater reliability: One-leg balance (preferred leg), ICC=0.997 One-leg balance (non-preferred leg), ICC=0.998 Walking heels raised, ICC=0.895 Jumping on mats, ICC=0.993 Intra-rater test-retest reliability: One-leg balance (preferred leg), ICC=0.970 One-leg balance (non-preferred leg), ICC=0.985 Walking heels raised, ICC=0.832

						Jumping on mats, ICC=0.936
Kakabeeke et al.[143]	3-5 yr	n=20	Static balance: standing on one leg [b]	Weighted Cohen's kappa with quadratic weights in ranked (not raw) observations (when this test is used in continuous variables is equivalent to ICC)	Test-retest (within a week)	Dominant leg: Kappa=0.45 Non-dominant leg: Kappa=0.59 According to Landis and Koch's benchmarks[144], the test-retest reliability for this test would be fair.
Kasuga et al.[127]	4-6 yr	n=342	Dynamic balance: Walk on a pathway 2m (length) x 10cm (width) with: - No obstacle - Low obstacle (5cm) - High obstacle (10cm)	ICC	Trial-to-trial reliability (interval between trials not reported)	The ICC of the walking time for each condition was 0.62 to 0.76 in 4-year-old boys, 0.71 to 0.81 in 4-year-old girls, 0.64 to 0.83 in 5-year-old boys, 0.69 to 0.82 in 5-year-old girls, 0.66 to 0.78 in 6-year-old boys and 0.60 to 0.72 in 6-year-old girls.
De Kegel et al.[145]	1.5-5 yr	n=144	GDBT, which include 35 items (tests) assessing static and dynamic balance [c]	ICC, SEM and SDD for the total score, and weighted-Kappa statistics for	Inter-rater reliability (same day, 3 raters, n=22) Test-retest	Test-retest and inter-rater reliability were excellent for the GDBT total scores, for the whole group and for different age groups, with ICC of 0.99 and 0.98, SEM values of 0.21 and 0.78, and SDD of 0.58 and 2.08, respectively. The weighted-Kappa statistics for each

				single tests	intra-rater reliability (within a week, n=144)	item (test) showed a very good agreement, with kappa values all higher than 0.73.
Niederer et al.[137]	4-6 yr	n between 14 and 40 depending on the test	Agility: Obstacle course Dynamic balance: Balance beam	Correlation (not reported if Pearson or Spearman)	Inter-rater and intra-rater test-retest reliability (interval between trials not reported)	Obstacle course: The inter-rater correlation and intra-rater test-retest reliability (n=14) were r=0.99 and r=0.82 respectively. Balance beam: The inter-rater and intra-rater test-retest reliability between the two better attempts (n=15) were r=0.97 (p<0.01) and r=0.84 (p<0.01) respectively. Note: this study included also a test to assess static balance, using a balance platform with 4 sensors (GKS 1000, IMM). Our review is however focused on simple and inexpensive field-based fitness tests so that more complex systems and laboratory tests are excluded, as indicated in the inclusion criteria.
Krombholz[120]	3-7 yr	n=1871 (912 girls, 959 boys)	Speed-agility: 4x4m shuttle run (with wooden blocks)	Pearson correlation	Inter-trials variability (between 2 attempts)	r=0.82
		n=429	Speed-agility: 4x4m shuttle run (with wooden blocks)	Pearson correlation	Test-retest reliability (8 month)	r=0.57 The authors concluded that the reliability estimates for this tests were adequate.
Ellinoudis et al.[146]	3-5 yr	n=60 (30 girls, 30 boys)	MABC-2 battery, balance subset: One-leg balance	ICC	Test-retest (1 week)	Test-retest reliability was good for all the balance tests: One-leg balance (preferred leg), ICC=0.73 One-leg balance (non-preferred leg), ICC=0.85

			(preferred leg) One-leg balance (non-preferred leg) Walking heels raised Jumping on mats			Walking heels raised, ICC=0.96 Jumping on mats, ICC=0.87
Smits-Engelsman et al.[147]	3 yr	n=28	MABC-2 battery, balance subset: One-leg balance (preferred leg) One-leg balance (non-preferred leg) Walking heels raised Jumping on mats	ICC	Intra-rater test-retest reliability (1-2 weeks)	Statistics were not reported for each test. Intra-tester test-retest for balance subset: ICC=0.75 No significant differences between the number of failed items were found between younger (mean 3.3, SD 3.6) and older children (mean 2.2, SD 3.4) or between boys (mean 2.9, SD 3.6) and girls (mean 2.6, SD 3.5).
Nguyen et al.[148]	3-5 yr	n=32 (12 girls, 20 boys)	25-m sprint	ICC CV One-way ANOVA	Test-retest (1 week)	Authors concluded that time to complete the 25-m sprint was highly reliable, i.e., ICC=0.91 and CV=3.7%. One-way ANOVA revealed no differences in CVs between age groups.
Sun et al.[149]	3-6 yr	n=39	Preschooler Gross Motor Quality Scale [d] Balance subset	ICC	Inter-rater and intra-rater test-retest reliability (interval between trials not reported)	No statistics were reported for single tests. The ICCs for inter-rater reliability ranged from 0.41 to 1.00 and for intra-rater reliability ranged from 0.60 to 1.00.
Kroes et al.[132]	5-6 yr	n=27-43	Maastricht's Motor Test, which includes a quantity and a	ICC	Inter-rater (same time), intra-rater	Statistics for single tests were not reported. ICC for inter-rater, intra-rater and test-retest were

			quality score [e] Static balance subset Dynamic balance subset		(watch the same video of the test 1 month later) and test-retest (tested twice 1 month later)	(respectively): Quality score-static balance score: 0.87, 0.92 and 0.43 Quality score-dynamic balance score: 0.87, 0.85 and 0.48 Quantity score-static balance score: 0.95, 0.98 and 0.48 Quantity score-dynamic balance score: 0.97, 0.90 and 0.77
Reeves et al.[138]	5-6 yr	n=17	BOTMP battery: [a] Running speed and agility subset Balance subset	Paired t test for differences between repeated measures	Test-retest (1 week)	Statistics for single tests were not reported. No significant differences were found between tests conducted on day 1 versus day 2.
Bénéfice et al.[139]	4-6.5 yr	n= 84 (36 girls, 48 boys)	Speed-agility: 4x10m shuttle run (without sponges)	Correlation (not reported if Pearson or Spearman)	Test-retest (1 day)	Girls, r=0.50 Boys, r=0.58
Oja and Jürimäe[140]	4-5 yr	n=124 (60 girls and 64 boys)	Speed-agility: 4x10m shuttle run (without sponges)	ICC, Cronbach's alpha, and CV (%)	Test-retest (1 week)	Overall, the test showed an acceptable reliability; the test was more reliable in 4 year-old than in 5 year-old. ICC, Cronbach's alpha coefficients and CV are presented below (in this order) by sex and age: Girls, 4 yr: 0.921, 0.954 and 2.9 Girls, 5yr: 0.515, 0.669 and 4.1 Boys, 4yr: 0.841, 0.890 and 3.3 Boys, 5yr: 0.683, 0.736 and 3.9

Broadstone et al.[133]	4-6.5 yr	n=18	Two-legs stand on a tiltboard with eyes open and closed	ICC	Test-retest (1 week)	ICC for test-retest reliability ranged from 0.49 to 0.54 in typically developing preschool children. The authors concluded that these tiltboard tests do not give stable and reliable measurements across test sessions.
Greenspan[134]	3-6 yr	n=10	Unilateral stance test (duration or up to 60s) on each foot	Pearson correlation	Intra-rater and inter-rater reliability (videotaping 10 children, later in the day)	Intra-rater reliability was estimated as 1.00, and inter-rater reliability was 0.9.

Abbreviations for test batteries and devices mentioned in the table: GMF, Gleichgewichts-Koordintations-System; MABC-2, Movement Assessment Battery for Children-second edition; PACER, Progressive Aerobic Cardiovascular Endurance Run; PDMS-2, Peabody Developmental Motor Scales, second edition; RIHM, Rotterdam Intrinsic Hand Myometer.

Abbreviations for statistical indices: CV indicates coefficient of variation; ICC, intraclass correlation coefficient; SEM, standard error of measurement; SDD, smallest detectable difference; SD, standard deviation; ANOVA, analysis of variance.

[a] The BOTMP is composed of a large number of items (tests). These tests are grouped in composites and subsets. Part of this battery aim to assess fine manual control and dexterity, but for the purpose of the present review, we focused only on the fitness components included in the battery, which are 3 subsets: muscular strength, running speed and agility, and balance. Each of these subsets is composed of several fitness tests.

[b] Other fitness tests (e.g. dynamic balance) were included in this study. However, they are not reported here, since they were scored in a scale from 0 to 4, and the aim for the PREFIT battery is to propose a set of fitness tests measured in continuous variables so that they can be discriminant enough to detect small changes, for example, as a result of an exercise-based intervention.

[c] Ghent Developmental Balance Test (GDBT), which include 35 items (tests) assessing static and dynamic balance: stepping across a thin mat, bipedal standing, bipedal standing on a balance pad, stepping across a soft balance beam, stepping across a balance pad, stepping across a soft balance beam that rests on a thin mat, kneeling and turning the head, kicking a ball 1, semi-tandem stance, semi-tandem stance with eyes closed, stepping with one foot on the line, standing on a line 1, stepping on tiptoes, stepping on a line 1, kicking a ball 2, stepping on tiptoes on a line, standing on one leg (preferred foot) 1, stepping on a line, standing on a line 2, stepping on a line 3, run and stop, standing on one leg (preferred foot) 2, stepping backward on a line 1, jumping forward on one leg across a line, stepping backward on a line 2, hopping, standing on one leg (preferred foot) 3, standing on a line with eyes closed, standing on tiptoes with arms up, Standing on a line with the heel touching the toes (preferred foot in front), standing on a line with the heel touching the toes (non-preferred foot in front), standing on one leg (preferred foot) 4, standing on one leg (non-preferred foot) 5, standing on a line with the heel touching the toes (preferred foot in front) and with eyes closed, standing on a line with the heel touching the toes (non-preferred foot in front) and with eyes closed.

[d] Preschooler Gross Motor Quality Scale (PGMQ), balance subset includes the following tests: single leg standing, tandem standing, walking line forward, walking line backward.

[e] Maastricht's Motor Test (MMT), which includes a quantity and a quality scores for a set of static balance tests (14 items) and dynamic balance tests (20 items).

Table 4. Overview of validity studies on field-based fitness tests in preschool children.

Reference	Age	Participants	Fitness test	Statistical tests used	Gold standard/construct	Main outcomes and conclusions
Criterion-related validity studies						
None found						
Construct (convergent) validity studies						
Cardiorespiratory fitness						
None found						
Musculoskeletal fitness (muscle strength and flexibility)						
Deitz et al.[124]	4-5 yr	n=38	BOT-2 battery: [a] Muscular strength subset	Pearson correlation	PDMS-2 Gross Motor Quotient, (is another assessment tool theoretically measuring similar constructs)	Moderate correlation (r=0.75) No statistics were reported for single tests.

Motor fitness (speed, agility and balance)						
Hua et al.[142]	3-6 yr	n=184 (92 girls, 92 boys)	MABC-2 battery, balance subset	Spearman correlation	PDMS-2 Gross Motor Quotient	A weak correlation was observed between these subsets. r=0.066 No statistics were reported for single tests.
De Kegel et al.[145]	1.5-5 yr	n=94 (74 with typical development and 20 diagnosed with motor retardation)	Static and dynamic balance: GDBT, which include 35 items (tests)	Pearson correlation	BOT-2, MABC-2 and PDMS-2.	Pearson correlations between the *z* scores on GDBT and the standardized scores of specific balance subscales of the BOT-2, MABC-2 and PDMS-2 were moderate to high (i.e. r=0.50 to 0.89), whereas correlations with subscales measuring constructs other than balance were low (r<0.50).
Sun et al.[149]	3-6 yr	n=39	Preschooler Gross Motor Quality Scale Balance subset: Single leg standing Tandem standing Walking line forward Walking line backward	Pearson correlation	PDMS-2 Stationary subset	No statistics were reported for single tests. Moderate correlation was observed between the subsets. r=0.605

Deitz et al.[124]	4-5 yr	n=38 girls and boys	BOT-2 battery: [a] Running speed and agility subset Balance subset	Pearson correlation coefficients	PDMS-2 Gross Motor Quotient PDMS-2 Gross Motor Quotient	There was moderate correlation (r=0.75) There was moderate correlation (r=0.65)

Abbreviations for test batteries: BOT-2, Bruininks-Oseretsky Test of Motor Proficiency, second edition; MABC-2, Movement Assessment Battery for Children-second edition; PDMS-2, Peabody Developmental Motor Scales, second edition; GDBT, Ghent Developmental Balance Test .

[a] The BOT-2 battery is composed of a large number of items (tests), e.g. the BOT-2 includes 53 tests, while a short-form version of the BOT-2 includes 12 tests. These tests are grouped in composites and subsets. Part of this battery aim to assess fine manual control and dexterity, but for the purpose of the present review, we focused only on the fitness components included in the battery, which are 3 subsets: muscular strength, running speed and agility, and balance. Each of these subsets is composed of several fitness tests, which for BOT-2 are as following. *Muscular strength subset* is composed of *5 tests*: standing long jump, knee (or full) push-ups, sit-ups, wall sit and v-up. *Running speed and agility subset* is composed of *5 tests*: shuttle run, stepping sideways over a balance beam, one-legged stationary hop, one-legged side hop and two-legged side hop. *Balance subset* is composed of *9 tests*: standing with feet apart on a line-eyes open, walking forward on a line, standing on one leg on a line-eyes open, standing with feet apart on a line-eyes closed, walking forward heel-to-toe on a line, standing on one leg on a line-eyes closed, standing on one leg on a balance beam-eyes open, standing heel-to-toe on a balance beam and standing on one leg on a balance beam-eyes closed.

Table 5. Overview of studies examining the relationship between field-based fitness tests and health outcomes in preschool children.

Reference	Age	Participants	Fitness test	Statistical tests used	Health outcomes	Main outcomes and conclusions
Longitudinal studies						
None found.						
Cross-sectional studies						
Cardiorespiratory fitness						
Reeves et al.[138]	5-6 yr	N=51 (29 girls, 22 boys)	1/2 mile walk/run 20m shuttle run	Pearson correlation	Body weight, BMI, and percent body fat (estimated from skinfolds, Slaughter)	The 1/2 mile walk/run performance was negatively correlated with body weight (r=-0.49), BMI (r=-0.52) and percent body fat (r=-0.54). Correlations between 20m shuttle run and anthopometric markers were not reported.
Musculoskeletal fitness (muscle strength and flexibility)						
None found.						

Motor fitness (speed, agility and balance)
None found.

Abbreviation for health outcomes: BMI, body mass index.

Study II. Reliability and validity of handgrip dynamometers

Reliability

Table 6 presents the mean differences and the SD of differences between repeated measures with the same instrument (intra-instrument reliability); different instrument but same model (inter-instrument reliability); different model (i.e., digital vs. analog, inter-model reliability) and old compared to new dynamometers. When the measure is done twice with the same instrument (intra-instrument reliability), the mean difference ranged from |0.04 to 0.25 kg| for the digital dynamometer, and from |0.09 to 0.33 kg| for the analog dynamometer. The systematic error between different instruments within the same models (e.g. digital 2 vs. digital 1, inter-instrument reliability) was |0.6 kg| for both the digital and analog dynamometer. Systematic error ranged from |0.24 to 0.35 kg| for the comparison between different new models (i.e. digital vs. analog). The systematic error between old and new digital dynamometers was |1.08 kg| for the digital and |0.78 kg| for the analog dynamometers.

Reliability analyses are graphically shown using Bland-Altman plots (**Figures 5 and 6**). Plots are shown only for one dynamometer per model; similar plots were obtained for the other dynamometers (data not shown).

Table 6. Mean differences (kg) and standard deviation (SD) of differences of TKK dynamometers reliability.

Dynamometers	Mean ± SD	p*
Comparison between the same instruments (Intra-instrument reliability: Retest minus Test)		
New digital 1	0.04 ± 0.60	0.742
New digital 2	-0.06 ± 0.55	0.613
Old digital	-0.25 ± 0.65	0.063
New analog 1	0.09 ± 0.65	0.437
New analog 2	-0.22 ± 0.53	**0.034**
Old analog	-0.33 ± 0.69	**0.013**
Comparison between different instruments, same models (Inter-instrument reliability)		
New digital 2 *minus* new digital 1	-0.62 ± 2.03	0.140
New analog 2 *minus* new analog 1	-0.64 ± 1.33	**0.013**
New digital 1 *minus* new analog 1	-0.35 ± 1.51	0.260
New digital 2 *minus* new analog 2	-0.25 ± 1.72	0.499
Comparison old vs. new dynamometers		
Old digital *minus* new digital 1	1.08 ± 2.02	**0.014**
Old analog *minus* new analog 1	-0.78 ± 1.32	**0.003**

* One-sample T- test. The inter-trials difference was entered as a dependent variable. P value refers to whether the mean differences is significantly different from 0 for all measures (from 1kg to 70kg).

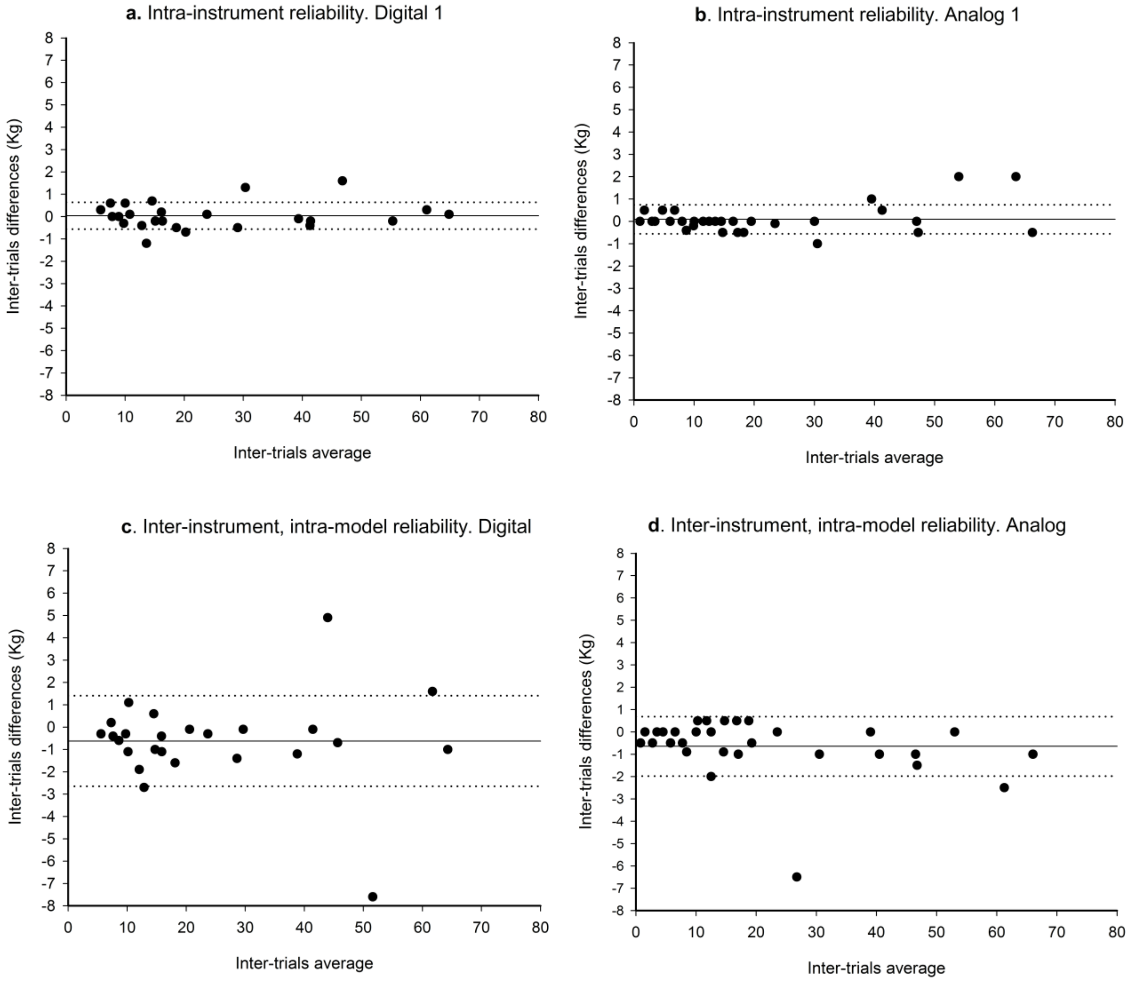


Figure 5. Reliability of the weight's trials in TKK dynamometer respectively by Bland and Altman plots [101].

The central line represents the mean difference (error) between known weights and different trials. Upper and lower broken lines represent the 95% limits of agreement (mean difference ± 1.96 of the differences). **Figure 5a**, intra-instrument reliability refers to the comparison within the same digital 1 dynamometer and model. **Figure 5b**, intra-instrument reliability refers to the comparison within the same analog 1 dynamometer and model. Figure 5a and 5b calculations were done as retest minus test, so that positive values indicate that the retest showed higher values than the test. **Figure 5c**, inter-instrument reliability means the comparison between different digital dynamometers of the same model. **Figure 5d**, inter-instrument reliability means the comparison between different analog dynamometers and same model. Figure 5c and 5d calculations were done as new digital/analog 2 minus new digital/analog 1 dynamometer.

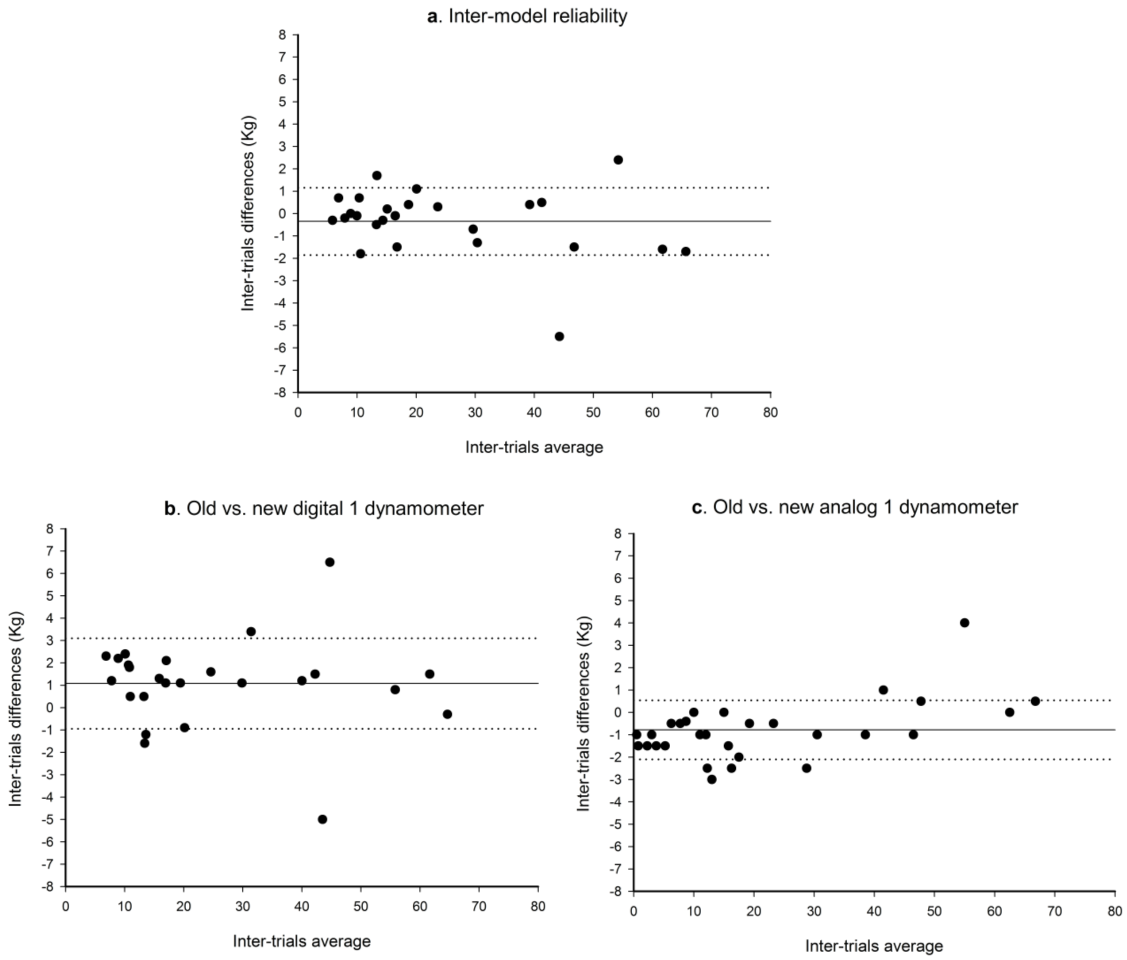


Figure 6. Reliability of the weight's trials in TKK dynamometer respectively by Bland and Altman plots[101].

The central line represents the mean difference (error) between known weights and different trials. The upper and lower broken lines represent the 95% limits of agreement (mean difference ± 1.96 of differences). **Figure 6a**, inter-model reliability refers to the comparison between the digital model and the analog model. Calculations were done as digital 1 minus analog 1, so that positive values denote that the digital 1 showed higher values than the analog 1 dynamometer. **Figure 6b**, old vs. new digital 1 refers to the comparison between the old digital and the new digital 1 dynamometer. **Figure 6c**, old vs. new analog 1 refers to the comparison between the old analog and the new analog 1 dynamometer. Figure 6b and 6c calculations were done as old digital/analog minus new digital/analog 1.

Validity

Criterion related-validity for the dynamometers against known weights is provided in **Table 7.** A negative systematic error (underestimation) was observed for the new digital (range -2.64 kg to -2.02 kg) and new analog (range -2.15 kg to -1.51 kg) dynamometers; as well as for the old digital and analog dynamometers (-0.94 kg and -2.29 kg, respectively).

Table 7. Mean differences (kg) and standard deviation (SD) of differences of TKK dynamometers vs. known weights.

Dynamometers	Mean ± SD	p*
New digital 1	-2.02 ± 2.60	**0.001**
New digital 2	-2.64 ± 2.73	**0.001**
Old digital	-0.94 ± 3.05	0.135
New analog 1	-1.51 ± 1.73	**0.001**
New analog 2	-2.15 ± 2.20	**0.001**
Old analog	-2.29 ± 1.27	**0.001**

*One-sample T- test. The inter-trials difference was entered as a dependent variable. P value refers whether the mean differences is significantly different from 0 for all measures (from 1kg to 70kg).

Study III. Adaptation of the original 20m shuttle run test in preschool children

Descriptive characteristics of the study sample are shown in **Table 8.** Results are presented for the whole sample, and divided by sex and age group. Table S7 shows the comparison of the 20m shuttle run test and PREFIT 20m shuttle run test.

Table 8. Descriptive characteristics of the study sample (N=137).

		Sex		Age group		
	All (N=137)	Boys (N=59)	Girls (N=78)	3 years-old (N=31)	4 years-old (N=32)	5 years-old (N=74)
Age (years)	4.9 ± 0.8	4.9 ± 0.8	4.8 ± 0.8	3.6 ± 0.2	4.5 ± 0.3	5.6 ± 0.3
Weight (Kg)	19.7 ± 3.5	19.8 ± 3.4	19.5 ± 3.7	16.6 ± 1.9	18.2 ± 1.9	21.7 ± 3.4
Height (cm)	109.2 ± 7.6	101.1 ± 7.6	107.9 ± 7.4	99.9 ± 3.6	106.2 ± 3.9	114.5 ± 5.1
Body Mass Index (Kg/m^2)	16.4 ± 1.6	16.2 ± 1.4	16.7 ± 1.8	16.6 ± 1.2	16.1 ± 1.2	16.4 ± 1.9
Waist circumference (cm)	53.2 ± 4.1	53.1 ± 3.7	53.5 ± 4.6	51.4 ± 3.2	52.7 ± 3.0	54.2 ± 4.6

Data are shown as mean ± standard deviation.

Feasibility

Time preparation/explication of the tests was 3 minutes approximately and the duration of the test depends to their cardiorespiratory fitness level. The participants completed a minimum of 3 laps and maximum of 71, being 25 the mean of the laps reached in the test. We observed that the performance of the test in optimal conditions were when preschool children were grouped in 4-8, yet, in 3 years-old preschool children we do not recommend perform the test with more than 4-5 children. Preschool children understood correctly the instruction of the test and showed a high motivation for performing it.

Maximality

Figure 7 shows an example of the heart rate during the test. In regards to the maximality, we did not observed differences between sexes in the percentage of the theoretical MHR in the test (97.5% both in girls and in boys, p=0.972) neither in the retest (97.1% for girls and 96.1% for boys, p=0.373). Differences between age groups in percentage of the theoretical MHR were marginally significant (test, p=0.067). Particularly, we observed that 3 years-old preschool children reached a percentage of theoretical MHR (94.8%) lower than the 4 and 5 years-old preschool children (97.6 and 98.4%, respectively). Similar findings were observed in the retest (p=0.067). Lin's concordance correlation of the whole sample was 0.311.

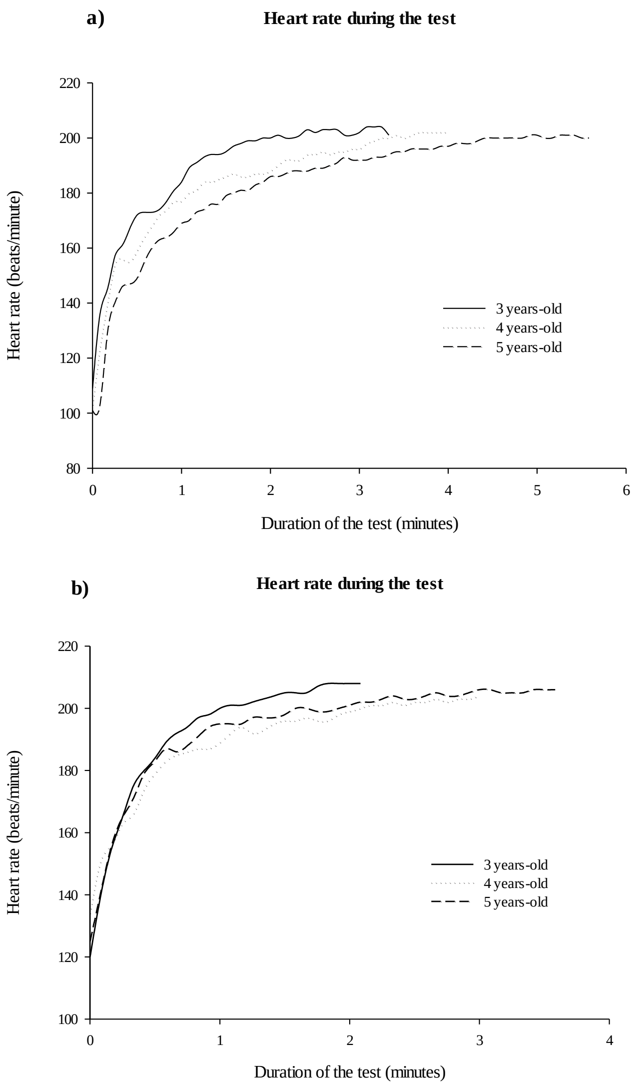


Figure 7. Heart rate recording during the performance of the PREFIT 20m shuttle run test.

Figure 7a depicts the heart rate in boys. Figure 7b shows the heart rate in girls.

Reliability

Table 9 shows the number of laps done for preschool children in the test, retest, and the mean difference. Boys showed higher number of laps than girls and a mean difference of 1 lap while this value increased up to 3.5 laps in girls. In regards to the age group, 5 years-old boys were the ones who reached the maximum number of laps in the PREFIT

20m shuttle run test. Nevertheless, the mean difference between test-retest was not sex- and age- significant (p>0.05).

Heteroscedasticity analysis for the PREFIT 20m shuttle run test showed no significant difference (p>0.05), concluding the no lineal relation between the magnitude (cardiorespiratory fitness level) and the test-retest variability (reliability). Lin's concordance correlation showed a coefficient of 0.827.

Table 9. Test, retest and mean difference of the number of laps done in PREFIT 20m shuttle run test.

PREFIT 20m shuttle run test		N	Test	N	Retest	r_c	N	Mean difference (Retest-Test)	p of the difference
All		137	25.00 ± 13.4	134	27.12 ± 13.2	0.838	130	2.00 ± 7.6	**0.002**
Sex	Girls	59	20.15 ± 9.7	53	23.83 ± 10.5	0.769	53	3.41 ± 6.9	0.360
	Boys	78	28.67 ± 14.7	81	29.27 ± 14.4	0.852	77	1.03 ± 8.0	
Age	3 years	31	12.90 ± 5.9	29	15.21 ± 6.2	0.643	29	1.76 ± 5.0	0.804
	4 years	32	21.09 ± 7.4	32	23.97 ± 6.5	0.615	29	3.34 ± 6.1	
	5 years	74	31.76 ± 13.5	73	33.23 ± 13.7	0.789	72	1.56 ± 8.9	

Data are shown as mean ± standard deviation. r_c = Lin's concordance correlation coefficient.

Study IV. Feasibility, reliability and practical recommendations for the PREFIT battery

Feasibility

Overall, fitness testing in preschoolers using the PREFIT battery was feasible, without any major problem detected when was implemented. The time required to assess each of the PREFIT battery tests is described in **Table 10**. In total, the duration of the PREFIT battery in a group of 20 preschoolers is around 1 h 40 minutes with 5 evaluators and 2 h 30 minutes with 2 evaluators.

Table 10. Time required to assess the PREFIT battery tests.

	Duration
Explain each test	3-5 min
Anthropometric assessment	3 min
Weight	40 s
Height	50 s
Waist circumference	90 s
Muscular strength	
Handgrip strength	3 min
Standing long jump S1 and S2	4 min
Cardiorespiratory fitness	
PREFIT 20m shuttle run test	From 24 s to 10 min*
Static Balance	
One-leg stance test	From 1 s to 145 s*
Total time PREFIT battery	
With 5 evaluators	1 h 40 min
With 2 evaluators	2 h 30 min
Total time PREFIT battery excluding one-leg stance test	
With 5 evaluators	1h 25 min
With 2 evaluators	2 h 10 min

S1= Study 1, S2= Study 2.
*The duration of this test depends on the fitness level of each participant. We observed large variability depending on sex, age and developmental stage.

Reliability

Mean differences and SD of differences between trials of each test are presented in **Table 11**. Anthropometric measures, PREFIT 20m shuttle run test, handgrip strength, standing long jump S2 and 4×10m shuttle run tests showed a low mean difference between the test and retest for the whole sample. The largest systematic errors were found for the standing long jump S1 and one-leg stance tests (|7.31cm| and |8.01s|, respectively). Cohen's effect size was estimated to allow comparisons between tests with different units of measure (Table 11). All effect sizes were small ($d \leq 0.37$). Systematic and random variability are also graphically shown using Bland-Altman plots (**Figure 8**).

Table 11. Test, retest, mean differences, concordance correlation coefficient and effect size of PREFIT physical fitness tests.

	N	Test	Retest	Mean differences *(retest - test)*	r_c	Effect size	p*
Age (years)	161	4.87 ± 0.83	4.91 ± 0.83	-	-	-	-
Weight (kg)	144	19.60 ± 3.44	19.80 ± 3.46	0.04 ± 1.09	0.948	0.058	0.628
Height (cm)	156	109.02 ± 7.33	109.30 ± 7.47	0.22 ± 1.75	0.971	0.027	0.109
Body mass index (kg/m^2)	144	16.39 ± 1.55	16.42 ± 1.81	-0.00 ± 0.86	0.868	0.018	0.964
Waist circumference (cm)	156	53.04 ± 3.95	53.00 ± 4.25	-0.08 ± 1.71	0.913	0.010	0.542
PREFIT 20m shuttle run (laps)	130	25.16 ± 13.23	26.48 ± 12.90	2.00 ± 7.61	0.833	0.101	**0.003**
Handgrip strength (kg)	156	7.90 ± 2.44	7.69 ± 2.45	-0.24 ± 1.28	0.859	0.086	**0.017**
Standing long jump-S1 (cm)	153	78.23 ± 20.03	70.57 ± 21.28	-7.31 ± 14.40	0.710	0.371	**<0.001**
Standing long jump-S2 (cm)	106	67.12 ± 19.09	64.43 ± 16.18	-2.33 ± 13.09	0.702	0.152	0.069
4×10m shuttle run (s)	153	15.74 ± 1.83	15.85 ± 2.05	0.12 ± 1.02	0.861	0.057	0.124
One-leg stance (s)	156	19.31 ± 20.60	27.63 ± 27.84	8.01 ± 20.84	0.604	0.341	**<0.001**

r_c= Lin's concordance correlation coefficient, S1= Study 1, S2= Study 2.
*One-sample t-test. P value refers whether a mean difference is significantly different from 0 for all measures.
Standing long jump-S2 results corresponds to the replication study performed for this test in different preschoolers.

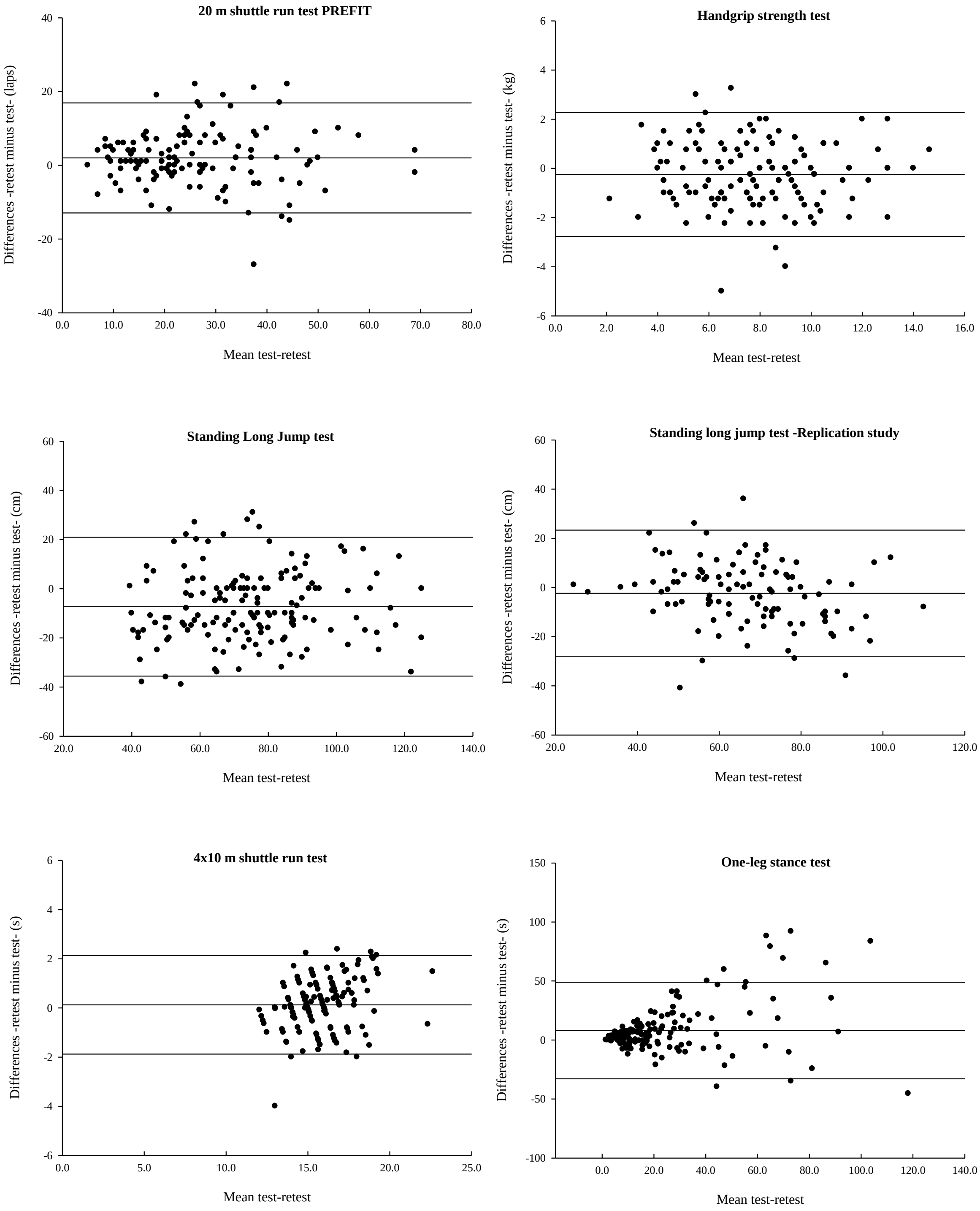


Figure 8. Bland and Altman plots[12] of the PREFIT 20m shuttle run test, handgrip strength test, standing long jump tests, 4×10m shuttle run test and one-leg stance test according fitness levels in preschoolers.

Central line represents mean difference (systematic error) between retest and test. Upper and lower lines represent 95% limits of agreement (mean difference ± 1.96 standard deviation of differences).

Table 12 shows the reliability statistics by sex. Overall, there were no significant differences in test-retest mean differences between boys and girls (p>0.05) except in BMI (p=0.023). Mean, SD and mean differences of all tests of the PREFIT battery by age group are provided in **Table 13.** Mean differences of anthropometric measures, PREFIT 20m shuttle run test, handgrip strength test and 4×10m shuttle run test were similar in 3, 4 and 5 years group. The standing long jump test S1 showed a systematic error significantly higher for 3 years group (|12.42cm|) than for 4 or 5 years group (|6.03cm| and |6.25cm|, respectively). On the contrary, its second study, the highest systematic error was observed in 5 years group (|8.19cm|) compared with 3 or 4 years group (|1.17cm| and |1.55cm|, respectively). The systematic error of the one-leg stance test increased exponentially with age (p=0.020). We observed heteroscedasticity in the 4×10m shuttle run test (p=0.038) and one-leg stance test (p<0.001). No significant heteroscedasticity was observed for the rest of tests (p>0.050). Although non-significant, we observed that for the standing long jump S2 children with the highest performance in this test tended to systematically perform worse in the second assessment (mean difference=-6.56cm), whereas this difference was smaller in children with lower performance in this test (-2.99cm). This difference was not altered after adjustment for age. Table S3 shows practical consideration when using the tests proposed in the PREFIT battery with preschoolers.

Table 12. Test, retest and mean differences of PREFIT physical fitness tests divided by sex.

	Boys		Girls		Mean differences *(retest minus test)*				p_{sex}†
	Test	Retest	Test	Retest	Boys	p*	Girls	p*	
Age (years)	4.90 ± 0.86	4.94 ± 0.86	4.82 ± 0.79	4.87 ± 0.79	-	-	-	-	-
Weight (kg)	19.66 ± 3.34	19.76 ± 3.43	19.51 ± 3.60	19.86 ± 3.54	-0.01 ± 1.21	0.919	0.12 ± 0.90	0.281	0.181
Height (cm)	109.71 ± 7.43	109.98 ± 7.51	108.11 ± 7.14	108.32 ± 7.37	0.27 ± 2.09	0.219	0.16 ± 1.12	0.248	0.674
Body mass index (kg/m^2)	16.23 ± 1.34	16.12 ± 1.52	16.60 ± 1.78	16.86 ± 2.10	-0.08 ± 0.68	0.239	0.11 ± 1.07	0.401	**0.023**
Waist circumference (cm)	52.92 ± 3.65	52.73 ± 4.15	53.20 ± 4.36	53.38 ± 4.40	-0.19 ± 1.78	0.305	0.07 ± 1.60	0.722	0.294
PREFIT 20m shuttle run (laps)	28.47 ± 14.66	28.99 ± 14.00	20.68 ± 9.42	22.86 ± 10.19	1.02 ± 7.98	0.263	3.41 ± 6.87	**0.001**	0.360
Handgrip strength (kg)	8.40 ± 2.40	8.02 ± 2.56	7.24 ± 2.34	7.22 ± 2.21	-0.38 ± 1.32	**0.007**	-0.05 ± 1.21	0.701	0.059
Standing long jump-S1 (cm)	82.11 ± 21.05	74.51 ± 22.40	72.97 ± 17.37	65.03 ± 18.37	-7.51 ± 14.34	**<0.001**	-7.03 ± 14.59	**<0.001**	0.875
Standing long jump-S2 (cm)	71.55 ± 19.11	66.00 ±17.72	61.55 ± 17.70	62.38 ± 13.82	-4.10 ±12.99	**0.017**	-0.02 ±13.00	0.991	0.187
4×10m shuttle run (s)	15.49 ± 1.65	15.56 ± 1.84	16.09 ± 2.00	16.28 ± 2.28	0.07 ± 0.92	0.455	0.21 ± 1.15	0.159	0.199
One-leg stance (s)	19.35 ± 22.24	28.27 ± 27.48	19.25 ± 18.34	26.72 ± 28.54	8.91 ± 20.40	**<0.001**	6.71 ± 21.56	**0.015**	0.484

S1= Study 1, S2= Study 2.
*One-sample t-test. P value refers whether a mean difference is significantly different from 0 for all measures.
† Two-way ANOVA. Mean differences were entered as dependent variables and sex (i.e. boys and girls) and group of ages as fixed factors.
Standing long jump-S2 results corresponds to the replication study performed for this test in different preschoolers.

Table 13. Test, retest and mean differences of PREFIT physical fitness tests divided by groups of age.

	3 years-old		4 years-old		5 years-old		Mean differences *(retest minus test)*						p_{age}
	Test	Retest	Test	Retest	Test	Retest	3 years-old	p*	4 years-old	p*	5 years-old	p*	
Age (years)	3.64 ± 0.22	3.68 ± 0.21	4.49 ± 0.28	4.53 ± 0.28	5.63 ± 0.35	5.67 ± 0.35	-	-	-	-	-	-	-
Weight (kg)	16.61 ± 1.93	17.02 ± 2.03	18.51 ± 2.25	18.48 ± 2.36	21.58 ± 3.46	21.70 ± 3.37	0.41 ± 1.09	0.049	0.02 ± 1.68	0.914	-0.09 ± 0.45	0.077	0.051
Height (cm)	100.10 ± 3.62	100.14 ± 3.77	106.48 ± 4.91	106.54 ± 4.81	114.40 ± 5.25	114.86 ± 5.15	0.03 ± 0.66	0.752	0.19 ± 2.62	0.587	0.32 ± 1.21	0.025	0.674
Body mass index (kg/m^2)	16.53 ± 1.12	16.97 ± 1.77	16.29 ± 1.22	16.20 ± 1.52	16.41 ± 1.89	16.34 ± 1.96	0.43 ± 1.37	0.096	-0.05 ± 0.88	0.708	-0.15 ± 0.45	0.006	**0.001**
Waist circumference (cm)	51.20 ± 3.16	50.94 ± 3.05	52.57 ± 2.78	51.87 ± 3.21	54.11 ± 4.64	54.61 ± 4.72	0.26 ± 1.16	0.230	-0.66 ± 1.95	**0.016**	0.40 ± 1.58	0.032	**0.002**
PREFIT 20m shuttle run (laps)	13.45 ± 5.70	15.21 ± 6.23	21.09 ± 7.44	23.57 ± 7.82	31.59 ± 13.51	32.91 ± 13.92	1.75 ± 5.06	0.072	3.34 ± 6.14	**0.007**	1.55 ± 8.90	0.143	0.638
Handgrip strength (kg)	5.62 ± 1.71	5.44 ± 1.49	7.34 ± 1.67	6.94 ± 1.84	9.22 ± 2.37	9.11 ± 2.24	-0.17 ± 1.20	0.447	-0.35 ± 1.39	0.067	-0.19 ± 1.24	0.174	0.658
Standing long jump-S1 (cm)	62.21 ± 10.57	49.58 ± 13.01	73.94 ± 15.57	66.96 ± 17.39	87.46 ± 20.98	81.28 ± 19.41	-12.42 ± 12.83	**<0.001**	-6.03 ± 15.76	**0.009**	-6.25 ± 13.72	<0.001	0.083
Standing long jump-S2 (cm)	59.10 ± 15.94	61.50 ± 13.04	62.82 ± 18.65	58.68 ± 15.92	82.63 ± 14.13	74.78 ±16.34	1.17 ± 12.62	0.535	-1.55 ± 12.00	0.485	-8.19 ± 13.13	0.002	**0.027**
4×10m shuttle run (s)	17.71 ± 1.97	18.10 ± 1.99	15.96 ± 1.81	16.11 ± 1.25	14.80 ± 1.50	14.81 ± 1.79	0.55 ± 1.05	**0.010**	0.07 ± 1.08	0.609	0.00 ± 0.93	1.000	**0.022**
One-leg stance (s)	6.03 ± 4.78	8.47 ± 5.27	16.77 ± 15.67	21.72 ± 20.75	26.37 ± 24.47	39.38 ± 31.90	2.43 ± 4.75	**0.010**	4.44 ± 17.91	0.077	12.76 ± 25.31	<0.001	**0.018**

S1= Study 1, S2= Study 2.

*One-sample t-test. P value refers whether the mean differences is significantly different from 0 for all measures.

† Two-way analysis of variance (ANOVA). Mean differences were entered as dependent variables and sex and groups of age (i.e. 3, 4 and 5) as fixed factors.

Standing long jump-S2 results corresponds to the replication study performed for this test in different preschoolers.

Study V. Physical fitness reference standards for Spanish preschool children

Anthropometric and physical fitness characteristics of the study sample (whole sample and separated by sex and age) are shown in **Table 14**. Boys were heavier and taller than girls ($p \leq 0.001$). Older children were heavier, taller, and presented higher BMI than younger groups ($p \leq 0.016$). In regards to physical fitness, boys performed better than girls in PREFIT 20m shuttle run, handgrip strength, standing long jump, and 4x10m shuttle run tests (all $p<0.001$). However, girls performed better than boys in one-leg stance test (balance) ($p<0.001$). Older children had better performance in all fitness tests than their younger counterparts (all $p<0.001$).

Table 14. Anthropometric and physical fitness characteristics in preschool children (N=3179).

	All	Boys	Girls	p_{sex}*	3-year-olds	4-year-olds	5-year-olds	p_{age}*
	Mean ± SD	Mean ± SD	Mean ± SD		Mean ± SD	Mean ± SD	Mean ± SD	
Age (years)	4.59 ± 0.87	4.60 ± 0.87	4.59 ± 0.88	0.806	3.52 ± 0.30	4.51 ± 0.29	5.53 ± 0.33	**<0.001**
Anthropometry								
Weight (kg)	18.99 ± 3.75	19.17 ± 3.85	18.77 ± 3.63	**0.001**	16.33 ± 2.26	18.66 ± 2.94	21.40 ± 3.84	**<0.001**
Height (cm)	106.93 ± 7.51	107.44 ± 7.56	106.37 ± 7.42	**<0.001**	99.33 ± 4.63	106.49 ± 4.77	113.43 ± 5.22	**<0.001**
Body mass index (kg/m^2)	16.5 ± 1.8	16.5 ± 1.8	16.5 ± 1.8	0.964	16.50 ± 1.42	16.38 ± 1.71	16.54 ± 2.02	0.093
Physical fitness								
PREFIT 20m shuttle run (laps)	20.0 ± 11.6	21.6 ± 12.3	18.3 ± 10.5	**<0.001**	11.91 ± 7.13	19.91 ± 8.98	26.53 ± 12.54	**<0.001**
Handgrip strength (kg)	7.03 ± 2.48	7.36 ± 2.57	6.66 ± 2.33	**<0.001**	4.94 ± 1.66	6.88 ± 1.84	8.84 ± 2.16	**<0.001**
Standing long jump (cm)	73.63 ± 22.25	76.98 ± 22.05	69.88 ± 21.87	**<0.001**	54.31 ± 17.74	74.48 ± 16.90	88.16 ± 18.03	**<0.001**
4×10m shuttle run (s)	16.81 ± 2.52	16.51 ± 2.43	17.15 ± 2.57	**<0.001**	19.27 ± 2.45	16.53 ± 1.69	15.13 ± 1.48	**<0.001**
One-leg stance (s)	13.87 ± 17.25	12.89 ± 16.80	14.95 ± 17.67	**<0.001**	5.17 ± 4.86	11.98 ± 10.55	22.54 ± 23.45	**<0.001**

SD= Standard deviation.
* Two-way analysis of variance (ANOVA). Descriptive characteristics were entered as dependent variables and sex and age groups as fixed factors.

Sex- and age-specific reference standards

Reference standards for the 1st, 3rd, 5th, 10th, 15th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 85th, 90th, 95th, 97th, and 99th percentiles and for every 0.025 years of age (i.e. 9 days) are provided in the Thesis CD (Tables in xlsx format) and will be freely available at http://profith.ugr.es/recursos/prefit. A summary of these reference standards (i.e. percentiles: 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, 99th) for each trimester (i.e. 0.25 years of age, 3 months) is provided in **Tables 15** and **16**. **Figure 9** and **10** depict sex- and age-specific fitness reference data according the 1st, 5th, 15th, 25th, 50th, 75th, 85th, 95th, and 99th percentiles.

We found higher values in boys compared to girls in the entire fitness tests battery except for the one-leg stance test, where girls showed better performance in all the percentiles analyzed. Also, along the percentiles analyzed, the performance increased with age. Larger differences between P95 and P99 were found in older preschool children than in their younger counterparts in PREFIT 20m shuttle run test (for girls), standing long jump test (for girls), and one-leg stance test (for boys and girls) (Figure 9-10). In the 4x10m shuttle run test, younger children showed larger differences (for boys and girls) between these P95 and P99 percentiles (Figure 10).

Sex differences across preschool period

Figure 11 shows sex differences across the preschool age and percentiles 25th, 50th, and 75th. Mean differences between boys and girls increased with age in most of the fitness components and percentiles studied (all $p<0.001$, data not shown), except in the standing long jump and speed-agility tests, where the differences remained similar for boys and girls.

Table 15. Reference standards of cardiorespiratory fitness and muscular strength tests calculated with Generalized Additive Model for Location, Scale and Shape (GAMLSS) in preschool children.

Boys											Girls										
Cardiorespiratory fitness: PREFIT 20m shuttle run test (laps)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	4	5	6	7	8	10	12	14	18	32	**3.00-3.24**	4	5	6	7	8	9	10	11	14	21
3.25-3.49	5	6	7	9	10	12	14	17	21	34	**3.25-3.49**	5	6	7	8	9	10	12	14	17	25
3.50-3.74	6	7	9	11	12	14	17	20	24	37	**3.50-3.74**	6	7	8	10	11	12	14	16	19	30
3.75-3.99	7	9	11	12	14	17	19	22	27	40	**3.75-3.99**	6	8	10	11	12	14	16	18	22	34
4.00-4.24	8	10	12	14	16	19	21	25	30	43	**4.00-4.24**	7	9	11	12	14	16	18	21	25	39
4.25-4.49	9	12	14	16	18	21	24	27	32	46	**4.25-4.49**	8	10	12	14	16	18	20	23	28	43
4.50-4.74	10	13	15	18	20	23	26	30	35	49	**4.50-4.74**	9	11	13	15	17	19	22	25	31	47
4.75-4.99	11	14	17	20	22	25	29	32	38	52	**4.75-4.99**	10	12	14	16	19	21	24	27	33	51
5.00-5.24	12	16	19	21	24	27	31	35	41	56	**5.00-5.24**	10	13	15	18	20	23	26	30	36	55
5.25-5.49	13	17	20	23	26	30	33	38	44	59	**5.25-5.49**	11	14	16	19	21	24	27	32	38	59
5.50-5.74	14	18	22	25	28	32	36	40	47	64	**5.50-5.74**	12	15	18	20	23	26	29	33	41	63
5.75-5.99	16	20	23	27	30	34	38	43	50	68	**5.75-5.99**	12	16	19	21	24	27	31	35	43	66
6.00-6.25	17	21	25	29	32	36	41	46	53	72	**6.00-6.25**	13	17	20	22	25	29	32	37	45	70
Upper-limb muscular strength: Handgrip strength test (kg)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	2.6	3.2	3.6	4.0	4.3	4.7	5.0	5.5	6.2	8.0	**3.00-3.24**	2.3	2.9	3.3	3.6	3.9	4.2	4.5	5.0	5.6	7.7
3.25-3.49	3.0	3.6	4.1	4.5	4.8	5.2	5.6	6.1	6.8	8.8	**3.25-3.49**	2.7	3.3	3.7	4.0	4.3	4.7	5.0	5.5	6.2	8.2
3.50-3.74	3.4	4.1	4.5	4.9	5.3	5.7	6.2	6.7	7.4	9.5	**3.50-3.74**	3.1	3.7	4.1	4.5	4.8	5.2	5.6	6.0	6.7	8.9
3.75-3.99	3.8	4.5	5.0	5.4	5.8	6.3	6.7	7.3	8.1	10.2	**3.75-3.99**	3.4	4.1	4.5	4.9	5.3	5.6	6.1	6.6	7.3	9.4
4.00-4.24	4.2	4.9	5.5	5.9	6.4	6.8	7.3	7.8	8.7	10.9	**4.00-4.24**	3.8	4.5	4.9	5.3	5.7	6.1	6.6	7.1	7.8	9.9
4.25-4.49	4.6	5.4	5.9	6.4	6.9	7.3	7.8	8.4	9.2	11.5	**4.25-4.49**	4.2	4.9	5.4	5.8	6.2	6.6	7.0	7.6	8.3	10.3

4.50-4.74	5.1	5.9	6.4	6.9	7.4	7.8	8.3	9.0	9.8	12.2	**4.50-4.74**	4.6	5.3	5.8	6.2	6.7	7.1	7.5	8.1	8.9	10.8
4.75-4.99	5.5	6.3	6.9	7.4	7.9	8.4	8.9	9.5	10.4	12.8	**4.75-4.99**	5.0	5.7	6.2	6.7	7.1	7.6	8.0	8.6	9.4	11.4
5.00-5.24	6.0	6.8	7.4	7.9	8.4	8.9	9.4	10.1	11.0	13.4	**5.00-5.24**	5.4	6.1	6.7	7.1	7.6	8.0	8.5	9.1	9.9	11.9
5.25-5.49	6.4	7.3	7.9	8.4	8.9	9.4	9.9	10.6	11.5	14.0	**5.25-5.49**	5.8	6.6	7.1	7.6	8.0	8.5	9.0	9.6	10.4	12.5
5.50-5.74	6.9	7.8	8.4	8.9	9.4	9.9	10.5	11.1	12.1	14.5	**5.50-5.74**	6.2	7.0	7.6	8.0	8.5	9.0	9.5	10.1	11.0	13.1
5.75-5.99	7.4	8.2	8.9	9.4	9.9	10.4	11.0	11.7	12.6	15.1	**5.75-5.99**	6.6	7.4	8.0	8.5	9.0	9.5	10.0	10.6	11.5	13.7
6.00-6.25	7.8	8.7	9.4	9.9	10.4	11.0	11.5	12.2	13.2	15.6	**6.00-6.25**	7.0	7.8	8.4	8.9	9.4	9.9	10.5	11.1	12.0	14.3

Lower-limb muscular strength: Standing long jump (cm)

Age (years)	P10	P20	P30	P40	P50	P60	P70	P80	P90	P99	Age (years)	P10	P20	P30	P40	P50	P60	P70	P80	P90	P99
3.00-3.24	36.4	42.0	45.8	49.0	51.9	54.7	57.8	61.4	66.6	79.4	**3.00-3.24**	30.4	35.7	39.4	42.5	45.2	48.0	51.0	54.5	59.4	71.5
3.25-3.49	39.8	46.0	50.2	53.6	56.7	59.8	63.2	67.2	72.9	86.9	**3.25-3.49**	33.7	39.6	43.7	47.0	50.0	53.0	56.3	60.1	65.7	79.3
3.50-3.74	43.3	50.0	54.6	58.3	61.7	65.0	68.6	72.9	79.1	94.6	**3.50-3.74**	36.9	43.5	47.9	51.5	54.8	58.0	61.5	65.7	71.8	87.1
3.75-3.99	46.8	54.1	59.0	63.0	66.5	70.1	74.0	78.7	85.4	102.2	**3.75-3.99**	40.1	47.2	52.0	55.9	59.3	62.7	66.4	71.0	77.6	94.5
4.00-4.24	50.2	58.0	63.2	67.5	71.3	75.0	79.2	84.2	91.4	109.6	**4.00-4.24**	43.1	50.8	55.9	60.0	63.6	67.2	71.1	76.0	83.2	101.8
4.25-4.49	53.2	61.5	67.1	71.5	75.5	79.5	83.9	89.2	96.8	116.3	**4.25-4.49**	46.1	54.3	59.7	64.0	67.7	71.4	75.6	80.8	88.5	108.8
4.50-4.74	56.0	64.7	70.5	75.2	79.3	83.5	88.0	93.6	101.7	122.3	**4.50-4.74**	48.9	57.7	63.4	67.9	71.7	75.6	79.9	85.4	93.6	115.6
4.75-4.99	58.4	67.5	73.6	78.4	82.7	87.0	91.7	97.5	105.9	127.7	**4.75-4.99**	51.7	61.0	66.9	71.6	75.5	79.5	84.0	89.7	98.4	122.1
5.00-5.24	60.6	70.0	76.3	81.3	85.7	90.1	95.0	101.0	109.7	132.5	**5.00-5.24**	54.2	64.0	70.2	75.0	79.0	83.0	87.7	93.6	102.8	128.1
5.25-5.49	62.6	72.4	78.9	84.0	88.5	93.0	98.0	104.2	113.3	137.0	**5.25-5.49**	56.6	66.8	73.2	78.0	82.1	86.2	91.0	97.1	106.7	133.7
5.50-5.74	64.5	74.6	81.3	86.5	91.1	95.7	100.8	107.2	116.6	141.3	**5.50-5.74**	58.7	69.3	75.8	80.8	84.9	89.0	93.8	100.1	110.1	138.6
5.75-5.99	66.4	76.8	83.6	89.0	93.6	98.3	103.5	110.0	119.7	145.3	**5.75-5.99**	60.6	71.5	78.2	83.2	87.3	91.4	96.3	102.8	113.0	143.1
6.00-6.25	68.2	78.9	85.8	91.3	96.1	100.8	106.1	112.8	122.8	149.3	**6.00-6.25**	62.4	73.6	80.4	85.5	89.5	93.6	98.6	105.2	115.7	147.3

Data are presented for every 0.25 years of age, which correspond to one trimester (i.e. every 3 months of age). Age at the midpoint of each interval was selected to provide percentiles. For instance, for the interval 3.00-3.24, data presented were those corresponding to an exact age of 3.125 years old.
P10: 10^{th} percentile; other percentiles are abbreviated accordingly.

Table 16. Reference standards of speed-agility and balance tests calculated with Generalized Additive Model for Location, Scale and Shape (GAMLSS) in preschool children.

Boys											Girls										
Speed-agility: 4x10m shuttle run test (seconds)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	17.2	18.0	18.6	19.1	19.7	20.2	20.9	21.7	23.0	27.2	**3.00-3.24**	17.8	18.8	19.4	20.0	20.5	21.0	21.7	22.5	24.0	29.0
3.25-3.49	16.6	17.4	18.0	18.5	19.0	19.5	20.1	20.8	22.0	25.9	**3.25-3.49**	17.3	18.2	18.8	19.3	19.7	20.2	20.8	21.6	22.8	27.2
3.50-3.74	16.1	16.8	17.3	17.8	18.3	18.8	19.3	20.0	21.1	24.6	**3.50-3.74**	16.8	17.6	18.1	18.6	19.0	19.4	19.9	20.6	21.8	25.6
3.75-3.99	15.6	16.3	16.8	17.2	17.6	18.1	18.6	19.2	20.2	23.4	**3.75-3.99**	16.3	17.0	17.5	17.9	18.3	18.6	19.1	19.7	20.8	24.1
4.00-4.24	15.1	15.7	16.2	16.6	17.0	17.4	17.9	18.5	19.4	22.3	**4.00-4.24**	15.7	16.4	16.8	17.2	17.6	17.9	18.4	19.0	19.9	22.8
4.25-4.49	14.7	15.3	15.7	16.1	16.5	16.9	17.3	17.8	18.7	21.4	**4.25-4.49**	15.2	15.8	16.3	16.6	17.0	17.4	17.8	18.3	19.2	21.7
4.50-4.74	14.3	14.9	15.3	15.7	16.0	16.4	16.8	17.3	18.1	20.6	**4.50-4.74**	14.8	15.3	15.8	16.1	16.5	16.9	17.3	17.8	18.6	20.9
4.75-4.99	14.0	14.5	14.9	15.3	15.6	15.9	16.3	16.8	17.5	19.9	**4.75-4.99**	14.4	15.0	15.4	15.7	16.1	16.5	16.9	17.4	18.2	20.3
5.00-5.24	13.7	14.2	14.6	14.9	15.2	15.5	15.9	16.4	17.1	19.4	**5.00-5.24**	14.1	14.7	15.1	15.4	15.8	16.1	16.6	17.1	17.8	19.9
5.25-5.49	13.4	13.9	14.3	14.6	14.9	15.2	15.6	16.0	16.7	18.9	**5.25-5.49**	13.9	14.4	14.8	15.1	15.5	15.8	16.2	16.7	17.5	19.5
5.50-5.74	13.1	13.6	14.0	14.3	14.6	14.9	15.3	15.7	16.4	18.5	**5.50-5.74**	13.6	14.1	14.5	14.9	15.2	15.5	15.9	16.4	17.1	19.1
5.75-5.99	12.9	13.4	13.7	14.0	14.3	14.6	15.0	15.4	16.1	18.1	**5.75-5.99**	13.3	13.8	14.2	14.6	14.9	15.2	15.6	16.1	16.8	18.7
6.00-6.25	12.7	13.2	13.5	13.8	14.1	14.4	14.7	15.1	15.8	17.8	**6.00-6.25**	13.1	13.5	13.9	14.2	14.6	14.9	15.3	15.8	16.4	18.2
Balance: One-leg stance test (seconds)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	1.1	1.5	1.8	2.2	2.7	3.2	3.9	4.9	6.8	16.2	**3.00-3.24**	1.0	1.5	1.8	2.3	2.7	3.3	4.0	5.1	7.0	15.1
3.25-3.49	1.4	1.9	2.3	2.8	3.3	4.0	4.9	6.1	8.6	20.3	**3.25-3.49**	1.4	2.0	2.6	3.1	3.8	4.6	5.6	7.0	9.7	21.0
3.50-3.74	1.7	2.3	2.8	3.4	4.1	4.9	6.0	7.5	10.5	25.0	**3.50-3.74**	1.9	2.6	3.4	4.1	5.0	6.0	7.3	9.2	12.7	27.5
3.75-3.99	2.1	2.8	3.5	4.2	5.0	6.0	7.3	9.2	12.9	30.6	**3.75-3.99**	2.4	3.3	4.2	5.2	6.3	7.5	9.2	11.6	16.0	34.7
4.00-4.24	2.5	3.4	4.2	5.1	6.1	7.3	8.9	11.2	15.7	37.2	**4.00-4.24**	2.9	4.1	5.2	6.4	7.7	9.3	11.3	14.3	19.7	42.6
4.25-4.49	3.0	4.1	5.1	6.2	7.4	8.8	10.7	13.5	18.9	44.8	**4.25-4.49**	3.5	4.9	6.2	7.6	9.2	11.1	13.5	17.1	23.6	51.0

4.50-4.74	3.6	4.9	6.1	7.3	8.7	10.5	12.7	16.0	22.4	53.2	**4.50-4.74**	4.1	5.7	7.3	9.0	10.8	13.0	15.9	20.1	27.7	59.9
4.75-4.99	4.2	5.7	7.1	8.6	10.2	12.2	14.9	18.8	26.2	62.3	**4.75-4.99**	4.7	6.6	8.4	10.3	12.5	15.1	18.4	23.2	32.0	69.2
5.00-5.24	4.9	6.6	8.2	9.9	11.8	14.2	17.2	21.7	30.3	72.0	**5.00-5.24**	5.4	7.6	9.6	11.8	14.2	17.1	20.9	26.4	36.4	78.8
5.25-5.49	5.6	7.5	9.4	11.3	13.5	16.1	19.6	24.7	34.5	82.0	**5.25-5.49**	6.0	8.5	10.8	13.2	16.0	19.3	23.5	29.7	40.9	88.5
5.50-5.74	6.2	8.4	10.5	12.7	15.1	18.1	22.0	27.8	38.8	92.2	**5.50-5.74**	6.7	9.4	12.0	14.7	17.7	21.4	26.1	32.9	45.4	98.3
5.75-5.99	6.9	9.4	11.7	14.1	16.8	20.1	24.5	30.9	43.1	102.4	**5.75-5.99**	7.4	10.4	13.2	16.1	19.5	23.5	28.7	36.2	50.0	108.1
6.00-6.25	7.6	10.3	12.9	15.5	18.5	22.2	26.9	34.0	47.5	112.7	**6.00-6.25**	8.0	11.3	14.4	17.6	21.2	25.6	31.3	39.5	54.4	117.8

Data are presented for every 0.25 years of age, which correspond to one trimester (i.e. every 3 months of age). Age at the midpoint of each interval was selected to provide percentiles. For instance, for the interval 3.00-3.24, data presented were those corresponding to an exact age of 3.125 years old.

P10: 10^{th} percentile; other percentiles are abbreviated accordingly.

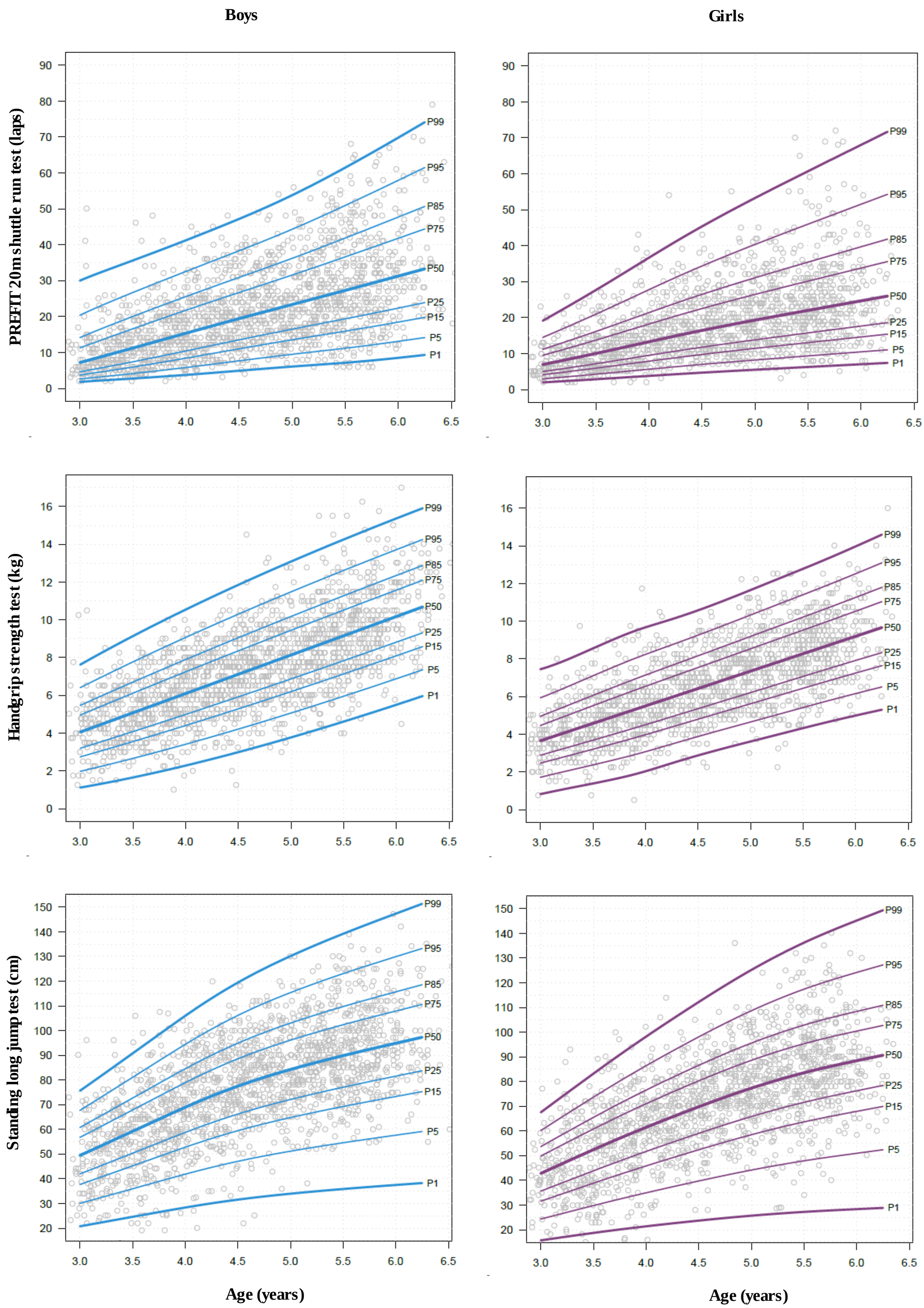


Figure 9. Percentile curves of the PREFIT 20m shuttle run, handgrip strength, and standing long jump tests in preschool children from 3 to 6.25 years old.

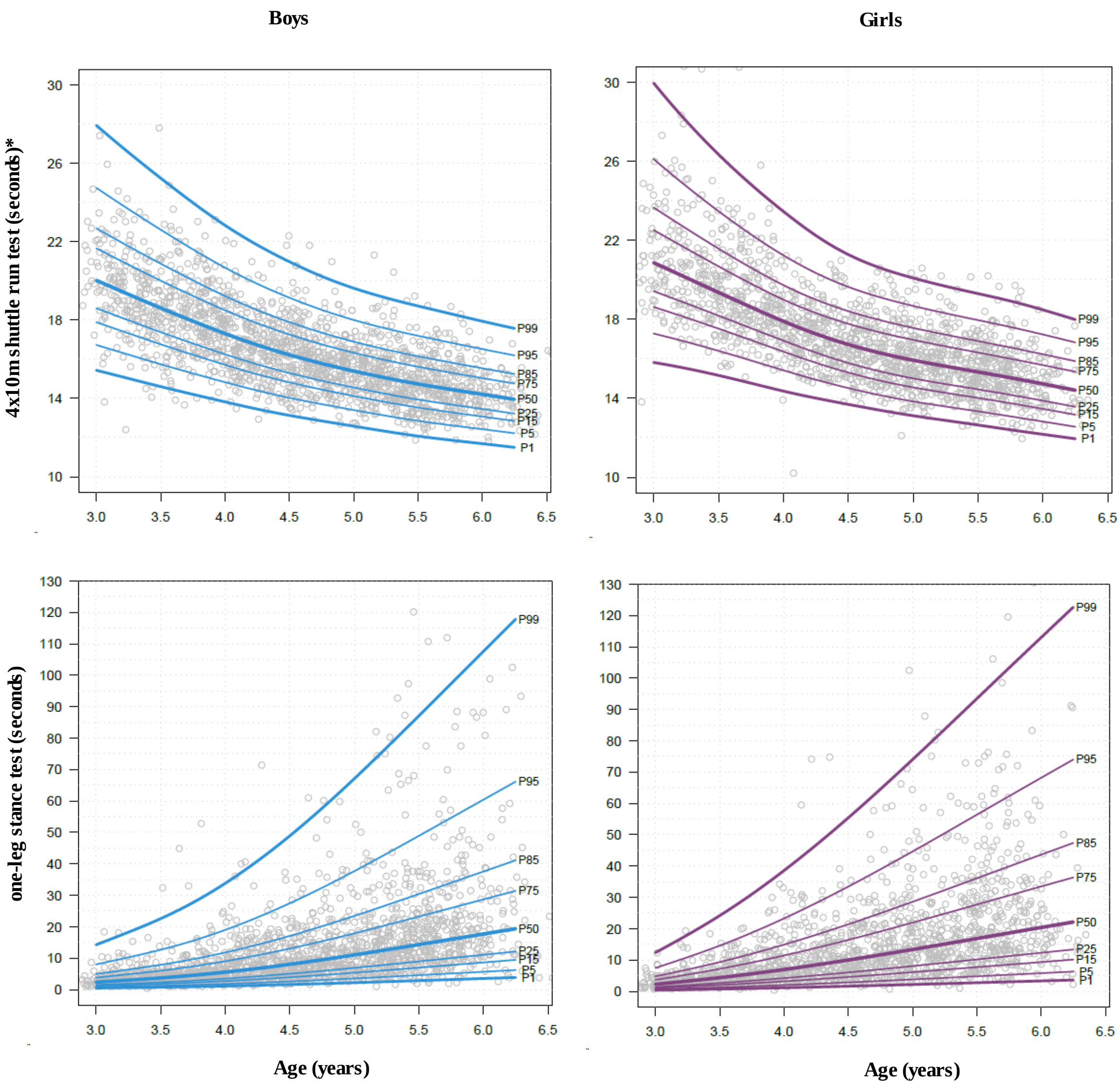


Figure 10. Percentile curves of the 4x10m shuttle run and one-leg stance tests in preschool children from 3 to 6.25 years old.

*In the 4x10m shuttle run test, lower scores (less seconds in running the fixed distance) indicate better performance (children are faster and more agile).

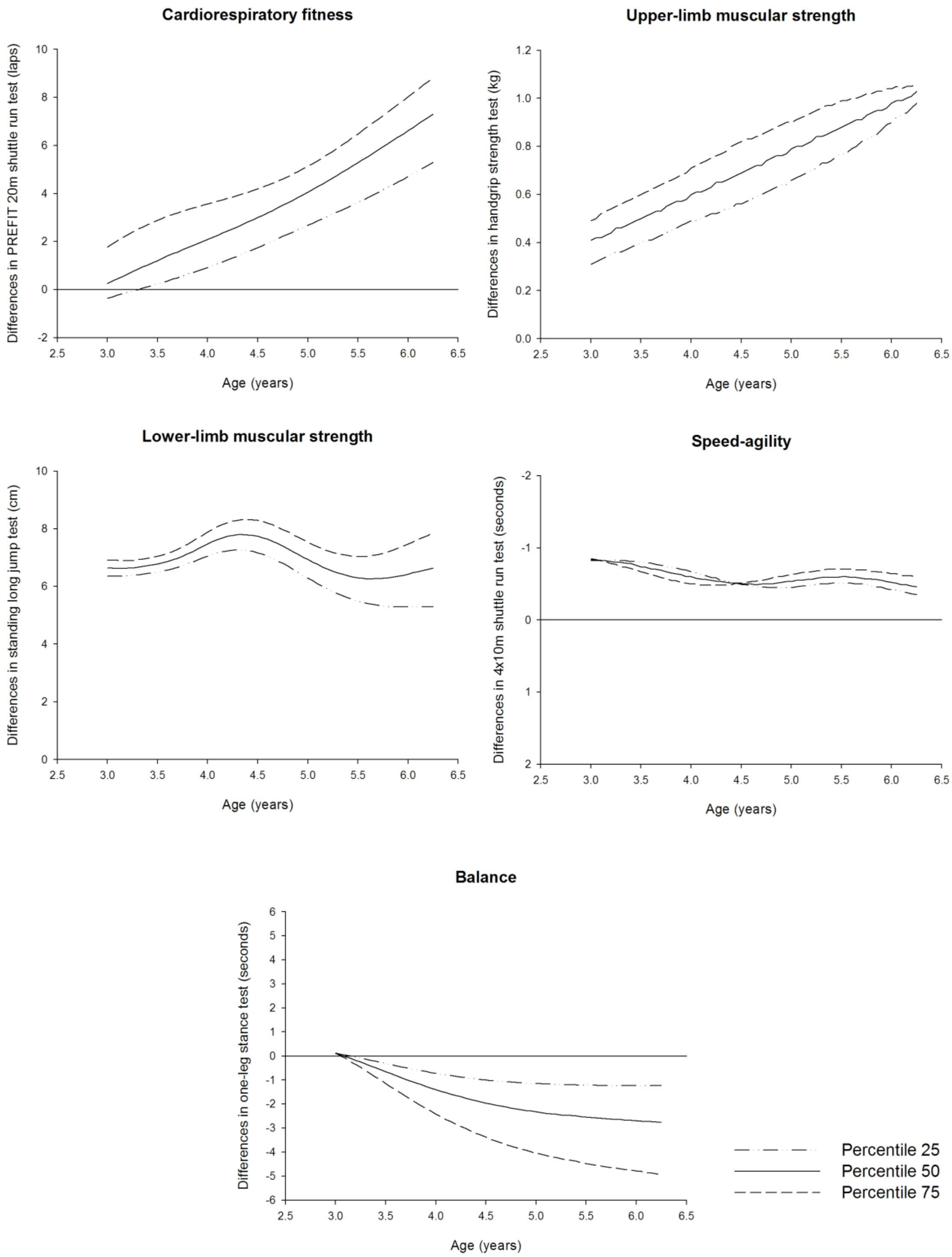


Figure 11. Sex differences over the whole age period and across different percentiles (25th, 50th, and 75th).

Differences are calculated as percentiles of boys minus age-corresponding percentiles of girls.

Study VI. Weight status and fatness reference standards for Spanish preschool children

Table 17 shows anthropometric characteristics of the study sample. Briefly, boys were slightly heavier and taller compared to girls (all $p<0.001$). In contrast, girls showed higher waist circumference and waist-to-height ratio compared to boys (all $p\leq0.010$). As expected, in regards to the age group, older children were heavier, taller, and presented higher waist circumference compared to younger children (all $p<0.001$). Waist-to-height ratio decreased as the age increased ($p<0.001$).

Table 17. Descriptive characteristics of the study sample (N=3178).

	All		Boys		Girls		p_{sex}*	3 years		4 years		5 years		p_{age}*
	N	Mean ± SD	N	Mean ± SE	N	Mean ± SE		N	Mean ± SE	N	Mean ± SE	N	Mean ± SE	
Age (years)	3178	4.59 ± 0.87	1677	4.52 ± 0.01	1501	4.52 ± 0.01	0.807	930	3.52 ± 0.01	1085	4.51 ± 0.01	1163	5.54 ± 0.01	**<0.001**
Weight (kg)	3176	18.99 ± 3.75	1675	18.99 ± 0.08	1501	18.60 ± 0.08	**<0.001**	929	16.33 ± 0.10	1084	18.65 ± 0.10	1163	21.39 ± 0.09	**<0.001**
Height (cm)	3176	106.93 ± 7.51	1675	106.91 ± 0.12	1501	105.86 ± 0.13	**<0.001**	929	99.30 ± 0.16	1084	106.46 ± 0.15	1163	113.40 ± 0.14	**<0.001**
Body mass index (kg/m^2)	3176	16.48 ± 1.76	1675	16.48 ± 0.43	1501	16.48 ± 0.46	0.974	929	16.51 ± 0.06	1084	16.39 ± 0.05	1163	16.54 ± 0.05	0.093
Waist circumference (cm)	3174	53.16 ± 5.02	1676	52.80 ± 0.12	1500	53.24 ± 0.12	**0.010**	928	51.06 ± 0.16	1085	52.94 ± 0.14	1163	55.07 ± 0.14	**<0.001**
Waist-to-height ratio	2775	0.50 ± 0.04	1674	0.49 ± 0.01	1500	0.50 ± 0.01	**<0.001**	927	0.51 ± 0.01	1084	0.50 ± 0.01	1163	0.49 ± 0.01	**<0.001**

SD: Standard deviation. SEM: Standard error of measurement.

* Two-way analysis of variance (ANOVA). Descriptive characteristics were entered as dependent variables and sex (i.e. boys and girls) and age groups (3 to 5 year-olds) as fixed factors. Sex and age groups data were presented as adjusted means.

Prevalence of weight status categories

Prevalence of weight status categories using both the WOF and the WHO definitions in preschool children is shown in **Figure 12**. Overall, following the WOF cut-off points, overweight prevalence in preschool children was 15.4%, while the prevalence of mild (obesity type I), severe (obesity type II) and morbid obesity (obesity type III) were 3.5%, 1.2% and 1.3%, respectively, which sum up to a total prevalence of obesity of 6%. The prevalences of the different degrees of underweight were 0.2, 0.6 and 4.1% for underweight type III, II and I, respectively. Normal weight prevalence was 73.7%. On the other hand, using the WHO references, the prevalence of overweight and obesity were 29.0% and 5.8%, respectively. Underweight prevalence was 0.1, 0.3 and 2.8% for underweight type III, II and I respectively. A percentage of 62.1% of the sample was classified as normal weight.

It was observed that overweight/obesity percentage were higher in girls than in boys using either the WOF (23.9% *vs*. 19.6%, p=0.002, for girls and boys, respectively) or the WHO (37.9% *vs*. 32% p<0.001, for girls and boys, respectively). For underweight sex-comparison, different patterns were observed for WOF and WHO, showing that girls have lower percentages than boys following the WOF but not following the WHO criteria (WOF: 5.1 *vs.* 4.6; WHO: 1.9 *vs*. 4.5 for boys and girls, respectively). Main differences between WOF and WHO prevalence were observed for normal weight (73.7% with WOF *vs.* 62.1% with WHO) and overweight (15.4% with WOF *vs.* 29.0% with WHO), while underweight and obesity prevalence were much closer (underweight: 4.8% with WOF *vs*. 3.1% with WHO; obesity: 6% with both WOF and WHO).

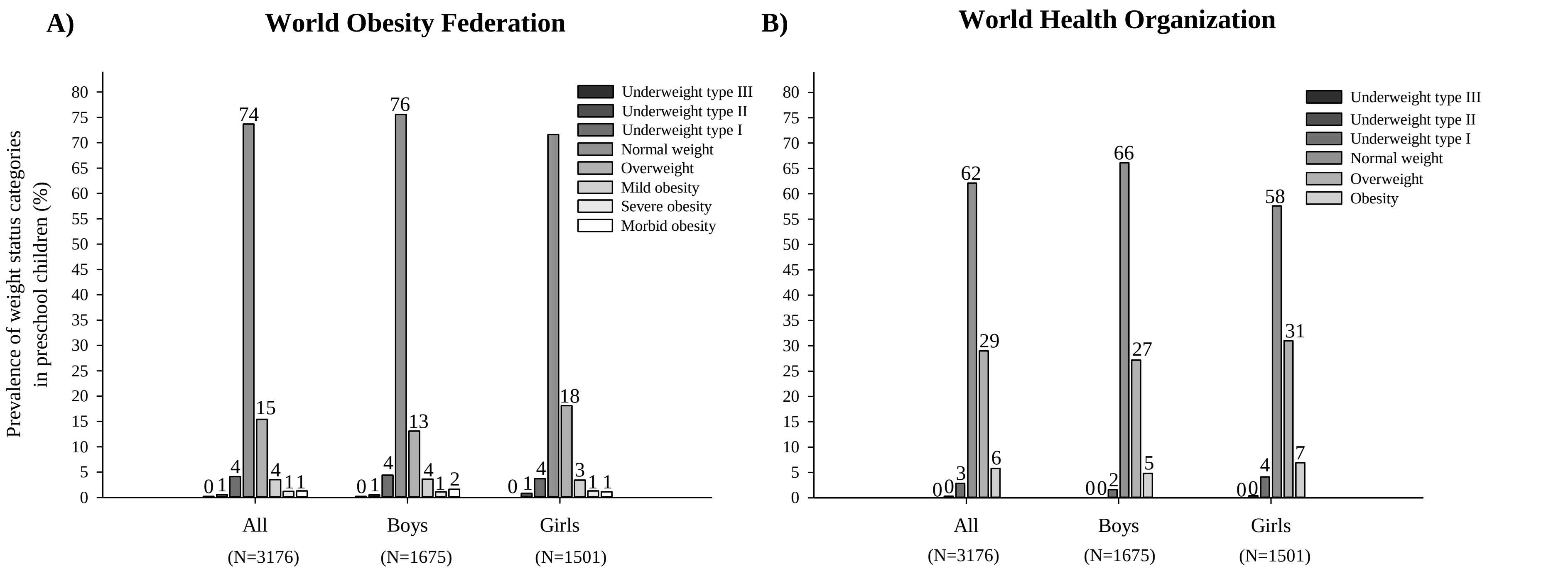


Figure 12. Prevalence of weight status categories based on World Obesity Federation (Figure 12a) and World Health Organization (Figure 12b) in preschool children.

Sex- and age-specific anthropometric reference standards

Reference standards (i.e. 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, 99th percentiles) for weight, height, BMI, waist circumference and waist-to-height ratio for each 0.25 years of age (trimester of year) from 3 to 6.25 years are shown in **Tables 18** and **19**. Even more detailed reference standards for the 1st, 3rd, 5th, 10th, 15th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 85th, 90th, 95th, 97th and 99th percentiles and for each 0.025 years of age are provided in the Thesis CD (Tables in xlsx format) and will be freely available at http://profith.ugr.es/recursos/prefit. The 1st, 5th, 15th, 25th, 50th, 75th, 85th, 95th and 99th percentiles curves for each anthropometric parameter are shown in **Figure 13** and **14.**

Differences between the 95th and 99th percentile curves were larger in weight, BMI, waist circumference and waist-to-height ratio than between other shown percentiles in boys and girls and in all age groups, although these differences were more marked in older children. The 50th percentile curve of BMI showed a plateau in boys from 4 to 5.5 years and in girls across the whole age range. In boys but not in girls, the 95th and 99th percentile curves in BMI for boys have a J-curve which is in particular the case for the 99th percentile curve.

Table 18. Reference standards of weight, height and body mass index calculated with Generalized Additive Model for Location, Scale and Shape (GAMLSS) in preschool children.

Boys											Girls										
Weight (kg)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	13.3	14.0	14.5	14.9	15.4	15.9	16.4	17.1	18.3	22.8	**3.00-3.24**	13.0	13.6	14.1	14.6	15.0	15.5	16.1	16.8	17.8	21.0
3.25-3.49	13.8	14.5	15.0	15.4	15.9	16.4	16.9	17.6	18.8	23.4	**3.25-3.49**	13.4	14.1	14.6	15.1	15.6	16.1	16.7	17.4	18.5	21.9
3.50-3.74	14.3	15.0	15.5	15.9	16.4	16.9	17.5	18.2	19.4	24.1	**3.50-3.74**	13.9	14.6	15.1	15.6	16.1	16.7	17.3	18.1	19.3	22.9
3.75-3.99	14.7	15.4	16.0	16.5	16.9	17.4	18.0	18.8	20.1	24.9	**3.75-3.99**	14.3	15.0	15.6	16.1	16.7	17.2	17.9	18.7	20.0	23.9
4.00-4.24	15.2	15.9	16.5	17.0	17.5	18.0	18.6	19.4	20.8	25.9	**4.00-4.24**	14.7	15.5	16.1	16.7	17.2	17.8	18.5	19.4	20.8	25.0
4.25-4.49	15.6	16.4	17.0	17.5	18.0	18.6	19.3	20.1	21.5	27.1	**4.25-4.49**	15.1	16.0	16.6	17.2	17.8	18.4	19.2	20.1	21.5	26.1
4.50-4.74	16.0	16.8	17.5	18.1	18.6	19.2	19.9	20.9	22.4	28.6	**4.50-4.74**	15.6	16.5	17.1	17.8	18.4	19.0	19.8	20.8	22.3	27.2
4.75-4.99	16.4	17.3	18.0	18.6	19.2	19.9	20.7	21.7	23.4	30.3	**4.75-4.99**	16.1	17.0	17.7	18.3	19.0	19.7	20.5	21.5	23.1	28.4
5.00-5.24	16.9	17.8	18.5	19.2	19.9	20.6	21.4	22.5	24.4	32.1	**5.00-5.24**	16.6	17.5	18.3	18.9	19.6	20.3	21.2	22.2	23.9	29.6
5.25-5.49	17.3	18.3	19.1	19.8	20.5	21.3	22.2	23.4	25.4	34.1	**5.25-5.49**	17.1	18.1	18.9	19.6	20.3	21.0	21.9	23.0	24.8	30.9
5.50-5.74	17.8	18.9	19.7	20.5	21.2	22.1	23.1	24.4	26.6	36.5	**5.50-5.74**	17.6	18.6	19.4	20.2	20.9	21.7	22.6	23.8	25.7	32.4
5.75-5.99	18.3	19.4	20.3	21.2	22.0	22.9	24.0	25.4	27.9	39.1	**5.75-5.99**	18.1	19.2	20.0	20.8	21.5	22.4	23.3	24.6	26.6	34.4
6.00-6.25	18.8	20.0	21.0	21.8	22.7	23.7	24.9	26.4	29.2	42.0	**6.00-6.25**	18.5	19.7	20.6	21.4	22.2	23.0	24.0	25.4	27.6	37.4
Height (cm)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	92.0	93.6	94.7	95.7	96.6	97.6	98.6	99.9	101.7	106.4	**3.00-3.24**	91.1	92.7	93.9	94.9	95.9	96.9	97.9	99.1	100.9	105.0
3.25-3.49	93.8	95.4	96.5	97.5	98.5	99.5	100.6	101.9	103.7	108.4	**3.25-3.49**	92.9	94.5	95.7	96.8	97.7	98.7	99.8	101.0	102.8	107.0
3.50-3.74	95.5	97.2	98.4	99.4	100.4	101.4	102.5	103.9	105.7	110.5	**3.50-3.74**	94.6	96.3	97.5	98.6	99.5	100.6	101.6	102.9	104.7	109.0
3.75-3.99	97.2	98.9	100.2	101.2	102.3	103.3	104.5	105.8	107.7	112.5	**3.75-3.99**	96.3	98.0	99.3	100.3	101.3	102.4	103.5	104.8	106.6	111.0
4.00-4.24	98.9	100.6	101.9	103.1	104.1	105.2	106.3	107.7	109.7	114.5	**4.00-4.24**	98.0	99.7	101.0	102.1	103.1	104.2	105.3	106.6	108.4	112.9
4.25-4.49	100.5	102.3	103.7	104.8	105.9	107.0	108.2	109.6	111.6	116.4	**4.25-4.49**	99.7	101.4	102.7	103.8	104.9	105.9	107.1	108.4	110.3	114.9

4.50-4.74	102.1	104.0	105.4	106.6	107.7	108.8	110.0	111.5	113.5	118.4	**4.50-4.74**	101.3	103.1	104.4	105.6	106.6	107.7	108.9	110.2	112.1	116.8
4.75-4.99	103.7	105.7	107.1	108.3	109.5	110.6	111.9	113.3	115.4	120.3	**4.75-4.99**	103.0	104.8	106.1	107.3	108.4	109.5	110.6	112.0	114.0	118.7
5.00-5.24	105.3	107.3	108.8	110.0	111.2	112.4	113.6	115.1	117.2	122.1	**5.00-5.24**	104.6	106.5	107.8	109.0	110.1	111.2	112.4	113.8	115.8	120.6
5.25-5.49	106.8	108.9	110.4	111.7	112.9	114.1	115.4	116.9	119.0	124.0	**5.25-5.49**	106.2	108.1	109.5	110.7	111.8	112.9	114.1	115.5	117.5	122.4
5.50-5.74	108.3	110.5	112.1	113.4	114.7	115.9	117.2	118.8	120.9	125.9	**5.50-5.74**	107.8	109.7	111.1	112.3	113.4	114.6	115.8	117.3	119.3	124.2
5.75-5.99	109.8	112.1	113.7	115.1	116.4	117.6	119.0	120.5	122.7	127.7	**5.75-5.99**	109.4	111.3	112.7	113.9	115.1	116.3	117.5	119.0	121.0	126.1
6.00-6.25	111.3	113.7	115.3	116.8	118.1	119.4	120.7	122.3	124.5	129.5	**6.00-6.25**	110.9	112.9	114.3	115.6	116.7	117.9	119.2	120.7	122.8	127.9

Body mass index (kg/m²)

Age (years)	P10	P20	P30	P40	P50	P60	P70	P80	P90	P99	Age (years)	P10	P20	P30	P40	P50	P60	P70	P80	P90	P99
3.00-3.24	15.2	15.6	15.9	16.2	16.5	16.8	17.2	17.7	18.5	22.0	**3.00-3.24**	14.9	15.3	15.7	16.0	16.2	16.5	16.9	17.4	18.1	20.5
3.25-3.49	15.0	15.5	15.8	16.1	16.4	16.7	17.0	17.5	18.3	21.6	**3.25-3.49**	14.8	15.3	15.6	15.9	16.2	16.6	16.9	17.4	18.2	20.7
3.50-3.74	14.9	15.4	15.7	16.0	16.3	16.6	16.9	17.4	18.2	21.5	**3.50-3.74**	14.8	15.3	15.6	15.9	16.2	16.6	16.9	17.4	18.3	20.9
3.75-3.99	14.8	15.3	15.6	15.9	16.2	16.5	16.8	17.3	18.1	21.4	**3.75-3.99**	14.7	15.2	15.6	15.9	16.2	16.6	17.0	17.5	18.3	21.2
4.00-4.24	14.8	15.2	15.5	15.8	16.1	16.4	16.8	17.3	18.1	21.5	**4.00-4.24**	14.7	15.2	15.6	15.9	16.2	16.6	17.0	17.6	18.5	21.5
4.25-4.49	14.7	15.1	15.5	15.8	16.1	16.4	16.8	17.3	18.1	21.9	**4.25-4.49**	14.6	15.1	15.5	15.9	16.2	16.6	17.0	17.6	18.6	21.8
4.50-4.74	14.6	15.1	15.4	15.8	16.1	16.4	16.8	17.3	18.3	22.4	**4.50-4.74**	14.6	15.1	15.5	15.9	16.2	16.6	17.1	17.7	18.7	22.2
4.75-4.99	14.6	15.1	15.4	15.8	16.1	16.4	16.9	17.4	18.4	23.0	**4.75-4.99**	14.5	15.1	15.5	15.9	16.2	16.6	17.1	17.7	18.7	22.5
5.00-5.24	14.5	15.0	15.4	15.8	16.1	16.5	16.9	17.5	18.6	23.6	**5.00-5.24**	14.5	15.1	15.5	15.9	16.2	16.6	17.1	17.8	18.8	22.7
5.25-5.49	14.5	15.0	15.4	15.8	16.1	16.5	17.0	17.6	18.7	24.2	**5.25-5.49**	14.5	15.0	15.5	15.9	16.2	16.7	17.1	17.8	18.9	23.0
5.50-5.74	14.5	15.0	15.5	15.8	16.2	16.6	17.1	17.8	18.9	24.9	**5.50-5.74**	14.4	15.0	15.4	15.8	16.2	16.7	17.2	17.9	19.0	23.3
5.75-5.99	14.5	15.1	15.5	15.9	16.3	16.7	17.2	17.9	19.1	25.5	**5.75-5.99**	14.4	15.0	15.4	15.8	16.2	16.7	17.2	17.9	19.1	23.6
6.00-6.25	14.6	15.2	15.6	16.0	16.4	16.8	17.3	18.1	19.4	26.2	**6.00-6.25**	14.3	14.9	15.4	15.8	16.2	16.7	17.2	17.9	19.2	23.9

Data are presented by 0.25 years steps which correspond to one trimester (i.e. each 3 months of year). Age at the midpoint of each interval was selected for providing percentiles. For instance, for the interval 3.00-3.24, data presented were those corresponding to an exact age of 3.125 years.

Table 19. Reference standards of waist circumference and waist-to-height ratio calculated with Generalized Additive Model for Location, Scale and Shape (GAMLSS) in preschool children.

Boys											Girls										
Waist circumference (cm)																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	46.2	47.3	48.2	49.0	49.7	50.6	51.5	52.8	54.8	61.6	**3.00-3.24**	46.2	47.4	48.4	49.3	50.1	51.0	52.1	53.4	55.5	62.0
3.25-3.49	46.5	47.7	48.5	49.3	50.1	50.9	51.9	53.2	55.2	62.4	**3.25-3.49**	46.5	47.8	48.8	49.7	50.6	51.5	52.6	54.0	56.1	62.8
3.50-3.74	46.8	48.0	48.9	49.7	50.5	51.3	52.3	53.6	55.7	63.3	**3.50-3.74**	46.9	48.2	49.2	50.1	51.0	52.0	53.1	54.5	56.7	63.7
3.75-3.99	47.2	48.4	49.3	50.1	50.9	51.7	52.7	54.0	56.2	64.1	**3.75-3.99**	47.2	48.5	49.6	50.5	51.5	52.4	53.6	55.1	57.4	64.6
4.00-4.24	47.5	48.7	49.7	50.5	51.3	52.2	53.2	54.5	56.8	65.0	**4.00-4.24**	47.5	48.9	50.0	50.9	51.9	52.9	54.1	55.6	58.0	65.6
4.25-4.49	47.9	49.1	50.1	50.9	51.7	52.6	53.7	55.1	57.4	66.1	**4.25-4.49**	47.8	49.3	50.4	51.4	52.3	53.4	54.6	56.1	58.6	66.7
4.50-4.74	48.2	49.5	50.5	51.3	52.2	53.1	54.2	55.7	58.1	67.3	**4.50-4.74**	48.2	49.6	50.8	51.8	52.8	53.8	55.1	56.7	59.2	67.8
4.75-4.99	48.5	49.8	50.9	51.8	52.7	53.6	54.8	56.3	58.9	68.7	**4.75-4.99**	48.5	50.0	51.2	52.2	53.2	54.3	55.6	57.2	59.9	69.0
5.00-5.24	48.8	50.2	51.3	52.2	53.1	54.1	55.3	57.0	59.7	70.2	**5.00-5.24**	48.8	50.4	51.6	52.7	53.7	54.8	56.1	57.8	60.6	70.3
5.25-5.49	49.2	50.6	51.7	52.7	53.6	54.7	55.9	57.6	60.5	72.0	**5.25-5.49**	49.2	50.8	52.0	53.1	54.1	55.3	56.6	58.4	61.3	71.8
5.50-5.74	49.5	51.0	52.1	53.2	54.2	55.2	56.6	58.3	61.4	73.9	**5.50-5.74**	49.5	51.2	52.4	53.5	54.6	55.7	57.1	58.9	62.0	73.5
5.75-5.99	49.9	51.4	52.6	53.7	54.7	55.8	57.2	59.1	62.3	76.1	**5.75-5.99**	49.8	51.6	52.9	54.0	55.1	56.2	57.6	59.5	62.7	75.4
6.00-6.25	50.2	51.8	53.1	54.2	55.2	56.4	57.8	59.8	63.3	78.5	**6.00-6.25**	50.2	52.0	53.3	54.5	55.5	56.6	58.0	60.0	63.4	77.6
Waist-to-height ratio																					
Age (years)	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**	**Age (years)**	**P10**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**	**P90**	**P99**
3.00-3.24	0.48	0.49	0.50	0.51	0.51	0.52	0.53	0.54	0.56	0.62	**3.00-3.24**	0.49	0.50	0.51	0.52	0.52	0.53	0.54	0.55	0.57	0.63
3.25-3.49	0.48	0.49	0.49	0.50	0.51	0.52	0.53	0.54	0.55	0.61	**3.25-3.49**	0.48	0.49	0.50	0.51	0.52	0.53	0.54	0.55	0.57	0.63
3.50-3.74	0.47	0.48	0.49	0.50	0.50	0.51	0.52	0.53	0.55	0.61	**3.50-3.74**	0.48	0.49	0.50	0.51	0.51	0.52	0.53	0.54	0.56	0.62
3.75-3.99	0.47	0.48	0.48	0.49	0.50	0.51	0.51	0.53	0.54	0.60	**3.75-3.99**	0.47	0.48	0.49	0.50	0.51	0.52	0.53	0.54	0.56	0.62
4.00-4.24	0.46	0.47	0.48	0.49	0.49	0.50	0.51	0.52	0.54	0.60	**4.00-4.24**	0.47	0.48	0.49	0.50	0.50	0.51	0.52	0.54	0.56	0.62
4.25-4.49	0.46	0.47	0.47	0.48	0.49	0.50	0.51	0.52	0.54	0.60	**4.25-4.49**	0.46	0.47	0.48	0.49	0.50	0.51	0.52	0.53	0.55	0.62

4.50-4.74	0.45	0.46	0.47	0.48	0.49	0.49	0.50	0.52	0.53	0.60	**4.50-4.74**	0.46	0.47	0.48	0.49	0.50	0.51	0.52	0.53	0.55	0.62
4.75-4.99	0.45	0.46	0.47	0.48	0.48	0.49	0.50	0.51	0.53	0.60	**4.75-4.99**	0.45	0.47	0.48	0.48	0.49	0.50	0.51	0.53	0.55	0.62
5.00-5.24	0.45	0.46	0.46	0.47	0.48	0.49	0.50	0.51	0.53	0.60	**5.00-5.24**	0.45	0.46	0.47	0.48	0.49	0.50	0.51	0.52	0.54	0.62
5.25-5.49	0.44	0.45	0.46	0.47	0.48	0.49	0.50	0.51	0.53	0.60	**5.25-5.49**	0.44	0.46	0.47	0.48	0.49	0.50	0.51	0.52	0.54	0.62
5.50-5.74	0.44	0.45	0.46	0.47	0.48	0.48	0.49	0.51	0.53	0.61	**5.50-5.74**	0.44	0.45	0.46	0.47	0.48	0.49	0.50	0.52	0.54	0.62
5.75-5.99	0.44	0.45	0.46	0.47	0.47	0.48	0.49	0.51	0.53	0.61	**5.75-5.99**	0.44	0.45	0.46	0.47	0.48	0.49	0.50	0.52	0.54	0.62
6.00-6.25	0.43	0.45	0.45	0.46	0.47	0.48	0.49	0.51	0.53	0.61	**6.00-6.25**	0.43	0.45	0.46	0.47	0.48	0.49	0.50	0.51	0.54	0.63

Data are presented by 0.25 years steps which correspond to one trimester (i.e. each 3 months of year). Age at the midpoint of each interval was selected for providing percentiles. For instance, for the interval 3.00-3.24, data presented were those corresponding to an exact age of 3.125 years.

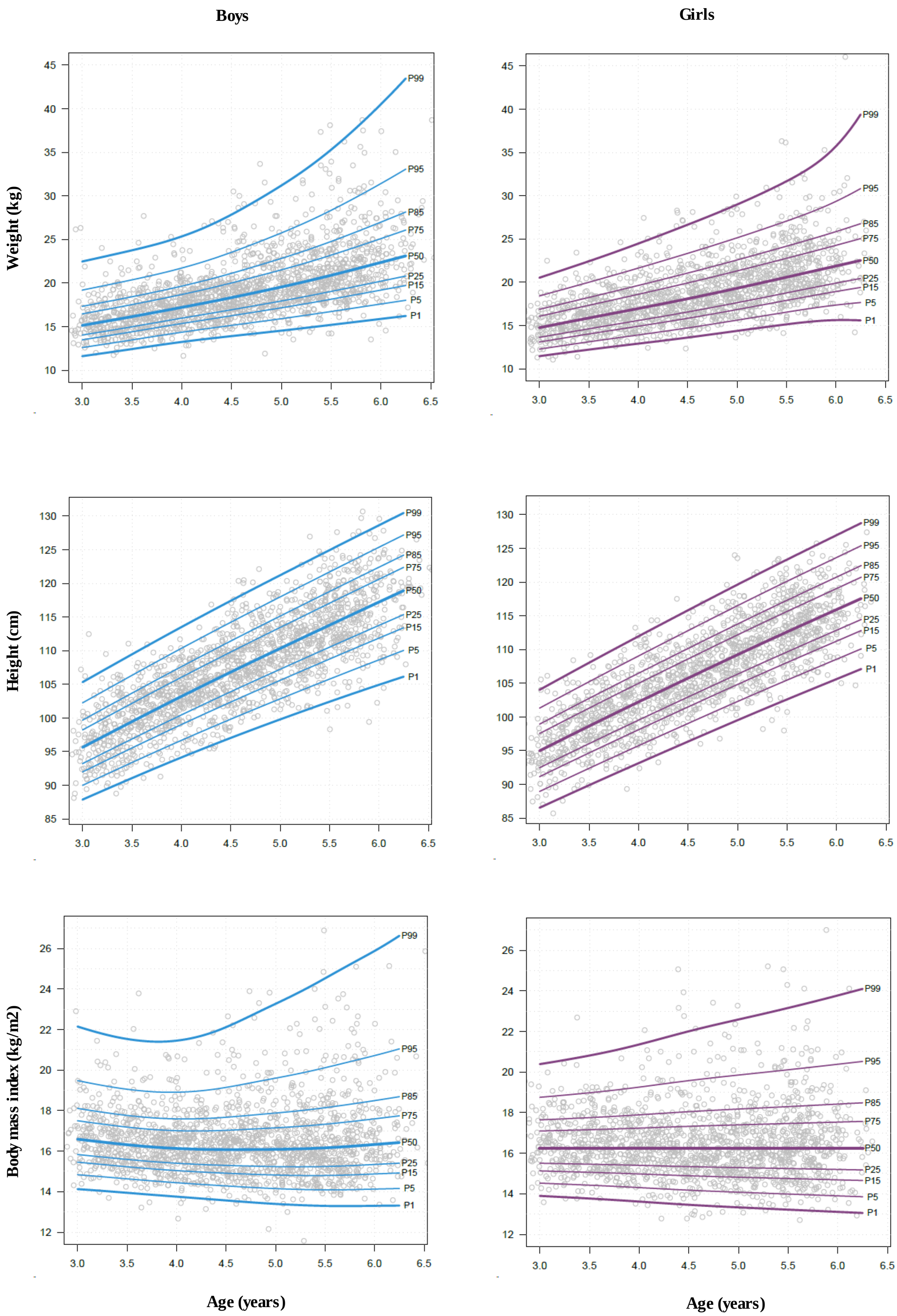


Figure 13. Sex-and age-specific percentile curves of weight, height and body mass index in preschool children.

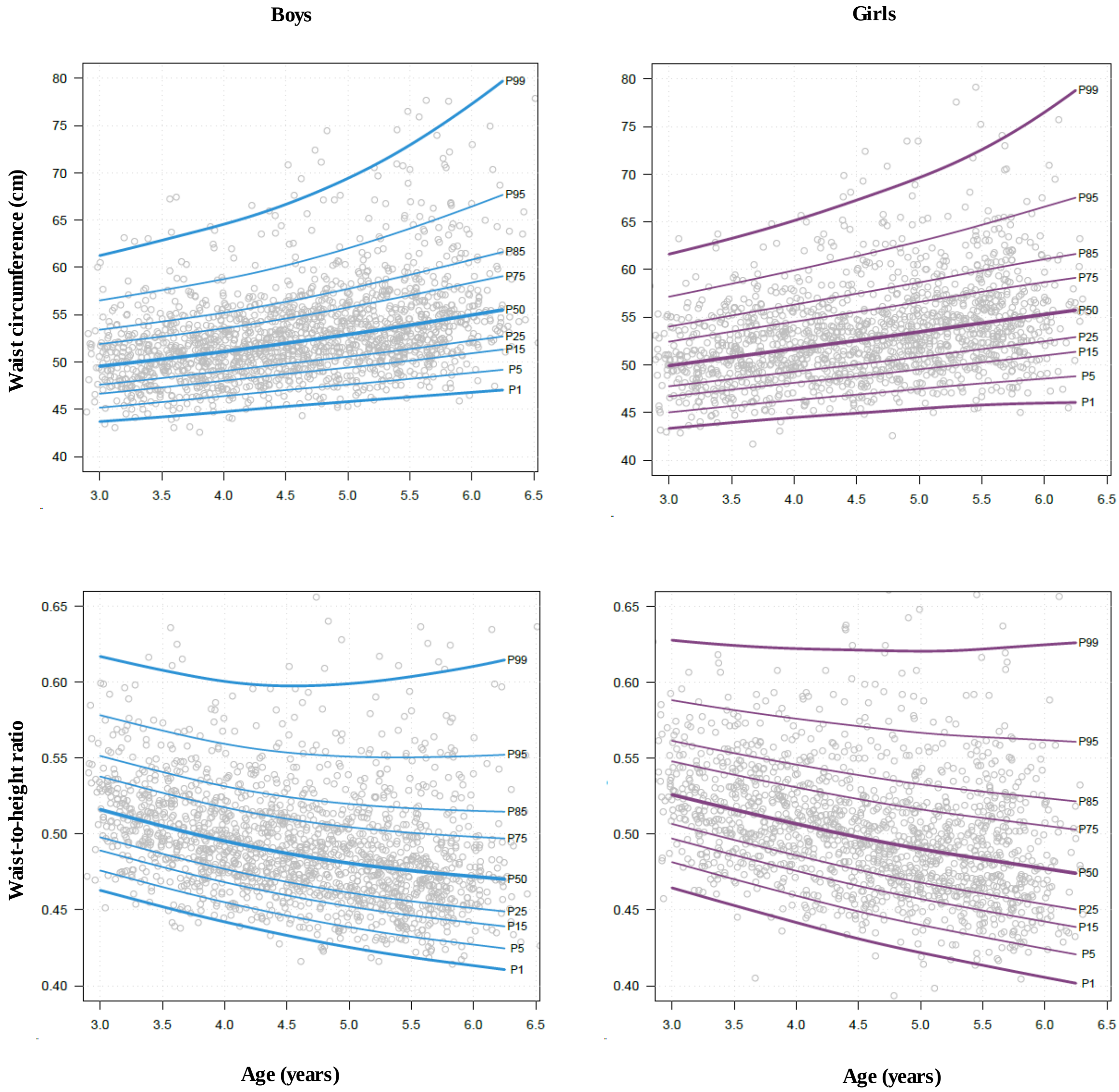


Figure 14. Sex-and age-specific percentile curves of waist circumference and waist-to-height ratio in preschool children.

Study VII. Comparison of fatness and physical fitness between Swedish and Spanish preschool children

Preschool children from Sweden had lower values for BMI (mean difference=-0.6 kg/m^2; p=0.001) and were taller (mean difference = +1.3 cm, p=0.003) than those from Spain (**Table 20**). We did not observe significant differences in weight, waist circumference and waist to height ratio (mean difference ranged from 0 to 0.2; p>0.050).

Table 20. Differences in anthropometrics measures in preschool children from Sweden and Spain.

	Swedish preschoolers			Spanish preschoolers			*p**
	n	Mean	SEM	*n*	Mean	SEM	
Age (years)	315	4.48	0.01	128	4.45	0.02	0.188
Anthropometrics measures							
Weight (kg)	315	18.3	0.15	128	18.5	0.23	0.388
Height (cm)	315	107.5	0.23	128	106.2	0.36	**0.003**
Total adiposity: BMI	315	15.7	0.09	128	16.3	0.13	**0.001**
Abdominal adiposity:							
Waist circumference (cm)	312	53.3	0.20	128	53.3	0.31	0.852
Waist to height ratio	312	0.5	0.02	128	0.5	0.03	0.835
Overweight/obesity prevalences	Frequency (%)			Frequency (%)			*p*
World Obesity federation:							
All	27 (8.6%)			22 (17.2%)			**0.010**
Boys	14 (8.1%)			13 (17.1%)			**0.037**
Girls	13 (9.1%)			9 (17.3%)			0.109
World Health Organization:							
All	63 (20.0%)			40 (31.3%)			**0.011**
Boys	39 (22.7%)			26 (34.2%)			**0.057**
Girls	24 (16.8%)			14 (26.9%)			0.114

* ANCOVA models adjusted for age and sex. Waist circumference was additionally adjusted for height. The means presented are adjusted for the covariate indicated above. SEM= Standard error of measurement.

Figure 15 depicts the prevalence of overweight and obese preschoolers from Sweden and Spain. For the whole sample, the prevalence of overweight/obesity was roughly 10% higher in Spanish preschoolers than Swedish preschoolers using WOF and WHO definitions (WOF=17% vs. 9%, p=0.010 and WHO=31% vs. 20%, p=0.011). For boys, similar results were observed between countries (WOF=17% vs. 8%, p=0.037 and WHO=34% vs. 23%, p=0.057 for Spanish and Swedish preschoolers, respectively). Correspondingly, the prevalence of overweight and obesity were higher for Spanish girls compared to Swedish girls, however, no significant differences were observed (all p>0.109).

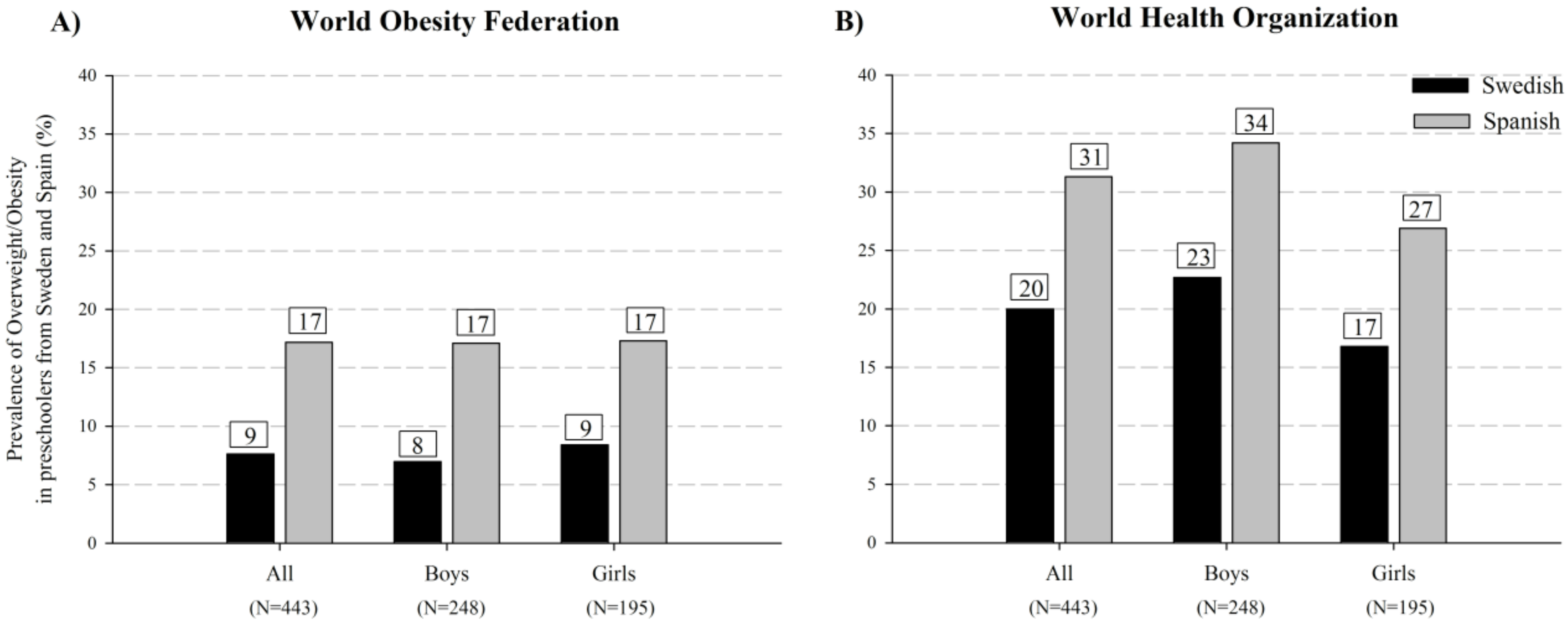


Figure 15. Prevalence of overweight/obesity in preschool children from Sweden and Spain following World Obesity Federation (Figure 15a) and the World Health Organization (Figure 15b) classifications.

Differences in physical fitness

Significant interaction in cardiorespiratory fitness test according to sex was found. Thus, the data from the main analyses were presented for boys and girls together except the original and PREFIT 20m shuttle run test.

Concerning fitness when we adjusted for age and sex (model 1, **Table 21**), preschoolers from Spain performed better in the cardiorespiratory fitness (+6.6 laps for boys and +2.3 laps for girls, respectively, all p<0.001), handgrip strength (mean difference= +0.4kg, p=0.010), 4x10m shuttle run test (mean difference=- 1.9s, p=0.001), and one-leg stance test (mean difference= +4.0s, p=0.001) and had worse results in standing long jump (mean difference= +7.1cm, p<0.001) than those from Sweden. However, the differences in upper-body muscular strength disappeared after adjustment for BMI (p reduced from 0.010 to 0.118) (model 2, **Table 21**).

Table 21. Differences in physical fitness tests in preschool children from Sweden and Spain.

	Model 1							Model 2						
	Swedish preschoolers			Spanish preschoolers			*p*	Swedish preschoolers			Spanish preschoolers			*p*
	n	Mean	SEM	*n*	Mean	SEM		*n*	Mean	SEM	*n*	Mean	SEM	
Original 20m shuttle run/ PREFIT 20m shuttle run (laps)*:														
Boys	168	16.9	0.4	71	23.4	0.6	**0.001**	168	16.7	0.4	71	23.7	0.6	**0.001**
Girls	142	18.1	0.3	50	20.4	0.6	**0.001**	142	18.1	0.3	50	20.4	0.6	**0.001**
Handgrip strength (kg)	313	6.8	0.1	128	7.2	0.1	**0.010**	313	6.9	0.1	128	7.1	0.1	0.118
Standing long jump (cm)	314	71.4	0.9	128	64.3	1.4	**0.001**	314	71.4	0.9	128	64.2	1.4	**0.001**
4x10m shuttle run (s)**	315	18.1	0.1	125	16.2	0.2	**0.001**	315	18.1	0.1	125	16.2	0.2	**0.001**
One-leg stance test (s)	314	10.6	0.6	128	14.6	0.9	**0.001**	314	10.5	0.6	128	14.8	0.9	**0.001**

Model 1: ANCOVA models were adjusted for age and sex.

Model 2: Model 1 additionally adjusted for BMI. The means presented are adjusted for the covariate indicated above. SEM= Standard error of measurement.

* The original 20m shuttle run was used in the MINISTOP project (i.e. preschoolers from Sweden) and the PREFIT 20m shuttle run was used in the PREFIT project (i.e. preschoolers from Spain). We applied correction factors to the scores of preschool children from Sweden in order to compare both tests. There were significant sex-interactions according to sex and thus, this data was presented for boys and girls separately.

** In this test the lower the score (in seconds) the higher the performance.

Section 2: Physical fitness, academic achievement and brain in preadolescent children

Study VIII. Physical fitness, physical activity and academic achievement preadolescent children

The demographic characteristics of the study sample are shown in **Table 24**. Briefly, the weight status analysis showed that 24.5% of the study sample was overweight, 44.3% was categorized as obesity type 1, 19.8% presented obesity type 2, and the remaining 11.3% was grouped as obesity type 3. Boys showed greater fitness performance in all field and laboratory tests compared to girls (all p>0.01), except for lower-limb muscular strength (i.e.1-RM bench press) (p=0.45). Objective physical activity outcomes were also higher in boys compared to girls (all p≤0.01), with the exception of light physical activity, which did not differ between sexes. Academic achievement characteristics did not show differences between sexes (all p≥0.1, **Table 25**).

Table 24. Descriptive characteristics of the study sample (N=106).

	All	Boys	Girls	p_{sex}
	Mean ± SD	*Mean ± SD*	*Mean ± SD*	
Anthropometric and sociodemographic characteristics (n=106; 61 boys):				
Age (years)	10.0 ± 1.1	10.1 ± 1.1	9.9 ± 1.1	0.473
Weight (kg)	56.1 ± 10.9	56.5 ± 10.5	55.5 ± 11.7	0.624
Height (cm)	144.2 ± 8.3	144.5 ± 7.3	143.7 ± 9.6	0.655
Body mass index (kg/m^2)	26.8 ± 3.6	26.9 ± 3.7	26.6 ± 3.4	0.634
Weight status (n, %):*				
Overweight	26, 24.5%	15, 24.6%	11, 24.4%	0.721
Obesity type 1	47, 44.3%	28, 45.9%	19, 42.2%	
Obesity type 2	21, 19.8%	10, 16.4%	11, 24.4%	
Obesity type 3	12, 11.3%	8, 13.1%	4, 8.9%	
Maternal education (n, %):				
Primary	28, 26.4%	16, 26.2%	12, 26.7%	0.088
Secondary	51, 48.1%	34, 55.7%	17, 37.8%	
University	27, 25.5%	11, 18.0%	16, 35.6%	
Physical fitness:				
Field-based test, physical fitness (n=106; 61 boys):				
20m shuttle-run (laps)	16.0 ± 7.8	17.0 ± 8.2	14.6 ± 7.0	0.127
Handgrip strength (kg)	16.8 ± 4.1	17.2 ± 3.8	16.3 ± 4.4	0.284
Standing long jump (cm)	105.0 ± 18.4	106.3 ± 17.4	103.2 ± 19.7	0.406
4x10m shuttle-run (sec)	15.1 ± 1.6	14.9 ± 1.7	15.4 ± 1.5	0.128
Laboratory-based test, Cardiorespiratory fitness, (n=104; 60 boys):				
Treadmill test (VO_2peak, ml/kg/min)	36.9 ± 5.0	37.2 ± 5.0	36.4 ± 5.1	0.420
Laboratory-based test, Muscular strength, (n=88; 55 boys):				
1-RM bench press (kg)	21.5 ± 4.3	22.4 ± 4.4	20.0 ± 3.8	**0.009**
1-RM leg press (kg)	135.4 ± 26.3	137.5± 28.5	133.1 ± 22.5	0.456

Physical activity (min/day; n=100, 58 boys)‡:				
Light PA	260.0 ± 37.4	254.2 ± 62.3	268.1 ± 38.4	0.068
Moderate PA	44.3 ± 16.4	50.2 ± 17.5	36.2 ± 10.2	**<0.001**
Vigorous PA	7.03 ± 4.3	8.8 ± 4.5	4.7 ± 2.4	**<0.001**
MVPA	51.4 ± 19.8	58.9 ± 21.0	40.9 ± 11.7	**<0.001**
1-min bouts MVPA	12.2 ± 5.9	14.8 ± 5.7	8.7 ± 4.1	**<0.001**
5-min bouts MVPA	2.9 ± 3.1	4.0 ± 3.4	1.6 ± 1.8	**<0.001**
10-min bouts MVPA	3.1 ± 5.2	4.5 ± 6.1	1.3 ± 2.5	**0.002**

SD= Standard deviation. VO_2max= maximum oxygen volume VO_2peak= peak oxygen volume. RM= Repetition maximum. PA= physical activity. MVPA= moderate-to-vigorous physical activity.

Weight status and maternal education were presented as frequency and percentage. The rest of the sections were presented as mean ± standard deviation. For continuous variables, p value was obtained by independent t-test in order to show whether the mean is the same/different for boys compared to girls. For categorical variables, p value was obtained by Chi-square test.

*Classified according to Cole et al.[85] and Bervoets et al.[40]

†Classified according to Hildebrand et al.[95,96] cut-off points for wrist.

Table 25. Academic achievement measured by Woodcock-Muñoz standardized test and school grades characteristics of the study sample.

	All	Boys	Girls	p_{sex}
	Mean ± *SD*	*Mean* ± *SD*	*Mean* ± *SD*	
Woodcock-Muñoz Test (n=106; 61 boys):				
Total achievement	108.9 ± 12.1	108.2 ± 11.6	109.8 ± 12.9	0.493
Reading	107.9 ± 13.3	107.7 ± 12.1	108.2 ± 14.8	0.852
Oral language	90.2 ± 13.7	89.6 ± 13.3	91.1 ± 14.2	0.595
Writing	113.5 ± 13.0	111.7 ± 12.2	115.9 ± 13.7	0.098
Written expression	103.4 ± 8.9	102.5 ± 9.0	104.7 ± 8.8	0.210
Mathematics	101.5 ± 10.8	101.7 ± 11.7	101.2 ± 9.5	0.833
Calculation skills	102.9 ± 11.9	101.8 ± 12.9	104.4 ± 10.3	0.272
Sciences	95.8 ± 13.2	96.3 ± 12.0	95.3 ± 14.7	0.702
School grades, scale from 1 to 5 (n=83; 49 boys):				
Grade point average	3.7 ± 0.9	3.6 ± 0.9	3.8 ± 0.9	0.397
Language	3.7 ± 1.0	3.6 ± 0.9	3.9 ± 1.0	0.243
Mathematics	3.7 ± 1.0	3.7 ± 1.0	3.6 ± 1.0	0.659
Natural and Social Sciences	3.7 ± 1.0	3.6 ± 1.1	3.8 ± 0.9	0.382
Foreign language	3.6 ± 1.1	3.5 ± 1.2	3.9 ± 1.1	0.157

SD= Standard deviation.
Data were presented as mean ± standard deviation. P value was obtained by independent t-test in order to test whether the mean differed between boys and girls.

Fitness and academic achievement

The partial correlation analyses (adjusted for basic confounders) between fitness and academic achievement measured by the Woodcock-Muñoz standardized test and school grades are shown in **Table 26**. Briefly, in field tests, cardiorespiratory fitness (i.e. the 20m shuttle-run test) was associated with writing skills measured by Woodcock-Muñoz (all $r \geq 0.273$, all $p<0.01$); the upper-limbs muscular strength (i.e. handgrip strength) was associated with school grades including grade point average, natural and social sciences, and foreign language (all $r \geq 0.282$, all $p<0.01$); speed-agility fitness (i.e.4x10m shuttle-run test) was associated with writing skills measured by Woodcock-Muñoz (all $r \geq 0.318$, all $p<0.01$). The laboratory fitness tests only showed significant associations between the lower-limbs muscular strength (i.e.1-RM leg press) and mathematics skills measured by Woodcock-Muñoz (all $r \geq 0.264$, all $p<0.01$).

Table 26. Partial correlations between physical fitness and academic achievement measured by Woodcock-Muñoz and school grades adjusted for basic confounders.

	20m shuttle-run (laps)	Handgrip strength (kg)	Standing long jump (cm)	4x10m shuttle-run test (sec)†	Treadmill test (VO_2peak)	1-RM bench press	1-RM leg press
Woodcock-Muñoz:							
Total achievement	**0.206***	0.128	**0.229***	**0.235***	0.086	0.085	**0.240***
Reading	0.093	0.038	0.180	0.120	0.021	0.100	0.196
Oral language	0.060	**0.209***	0.118	0.006	0.046	0.037	-0.041
Writing	**0.273****	0.108	**0.196***	**0.318****	0.078	0.029	0.185
Written expression	**0.348****	0.140	**0.212***	**0.377****	**0.207***	0.092	**0.218***
Mathematics	0.172	0.179	**0.204***	0.170	0.096	0.086	**0.264****
Calculation skills	0.054	0.156	0.106	0.172	0.062	0.062	**0.316****
Sciences	-0.105	0.101	0.055	0.074	-0.010	0.048	0.141
School grades:							
Grade point average	0.123	**0.300****	0.103	0.024	-0.016	0.014	0.147
Language	0.098	0.206	0.051	0.048	-0.015	0.078	0.159
Mathematics	0.073	**0.262***	0.008	-0.055	-0.086	0.032	0.128
Natural and Social Sciences	0.179	**0.282****	0.126	0.045	0.069	0.004	0.077
Foreign Language	0.088	**0.307****	0.170	0.046	-0.023	-0.056	0.164

RM= Repetition maximum. VO_2max= maximum oxygen volume.
Data were adjusted for basic confounders (age, sex, and maternal education).
*p≤0.05
**p≤0.01
†Values were multiplied by -1 before analyses so that higher values indicate better performance.

After additional adjustment of BMI and cardiorespiratory fitness or speed-agility (model 2), the results observed for the field and laboratory tests were attenuated, and significant associations disappeared except for those in the lower-limbs muscular strength (1-RM leg press) (**Table 27**).

Table 27. Partial correlations between physical fitness and academic achievement measured by Woodcock-Muñoz and school grades adjusted for basic confounders plus body mass index and fitness.

	20m shuttle-run (laps)	Handgrip strength (kg)	Standing long jump (cm)	4x10m shuttle-run test (sec)†	Treadmill test (VO_2peak)	1-RM bench press	1-RM leg press
Woodcock-Muñoz:							
Total achievement	0.078	0.153	0.212*	0.139	-0.040	0.033	**0.276****
Reading	0.030	0.049	0.192	0.084	-0.038	0.128	**0.266***
Oral language	0.058	**0.233****	0.108	0.054	0.017	0.062	0.007
Writing	0.093	0.160	0.065	0.193	-0.124	-0.103	0.169
Written expression	0.146	0.161	0.037	**0.212***	0.016	-0.076	**0.219***
Mathematics	0.094	0.184	0.141	0.083	0.028	0.038	**0.290****
Calculation skills	-0.050	0.137	0.097	0.197	0.036	0.042	**0.342****
Sciences	-0.080	0.094	0.129	0.010	0.040	0.086	0.152
School-grades:							
Grade point average	0.170	**0.260***	0.062	0.055	0.001	-0.083	0.035
Language	0.115	0.166	0.014	0.000	0.006	-0.002	0.073
Mathematics	0.182	0.213	-0.018	0.110	-0.027	-0.065	-0.017
Natural and Social Sciences	0.198	**0.274***	0.047	0.109	0.025	-0.105	-0.004
Foreign Language	0.170	**0.265***	0.161	-0.018	0.011	-0.117	0.071

RM= Repetition maximum. VO_2max= maximum oxygen volume.
Data were adjusted for basic confounders (age, sex, and maternal education), body mass index, and cardiorespiratory fitness, except for cardiorespiratory fitness, which was adjusted for speed-agility fitness.
*p≤0.05
**p≤0.01
†Values were multiplied by -1 before analyses so that higher values indicate better performance.

Figure 16 shows linear regression analyses for the components that showed a strong association (p<0.01) in the correlation analyses between fitness and academic achievement in both models. In the field tests, after adjusting for basic confounders, cardiorespiratory fitness (i.e. the 20m shuttle-run test) was significantly associated with writing and written expression, according to the Woodcock-Muñoz test (standardized ß ranging from 0.281 to 0.365, p<0.01). Positive associations were found for the upper-limbs muscular strength (i.e. handgrip strength test) and the following school grades: grade point average, natural and social sciences, and foreign language (standardized ß ranging from 0.303 to 0.378, p<0.01). Speed-agility fitness (i.e. 4x10m shuttle-run test) was positively associated with writing skills in the Woodcock-Muñoz test (standardized ß ranging from 0.325 to 0.393, p<0.01).

For the laboratory-based measures, the lower-limbs muscular strength (i.e.1-RM leg press) was positively associated with mathematics skills, measured by the Woodcock-Muñoz test (standardized ß from 0.244 to 0.300, p<0.01). No significant associations were found for the remaining fitness components (data not shown).

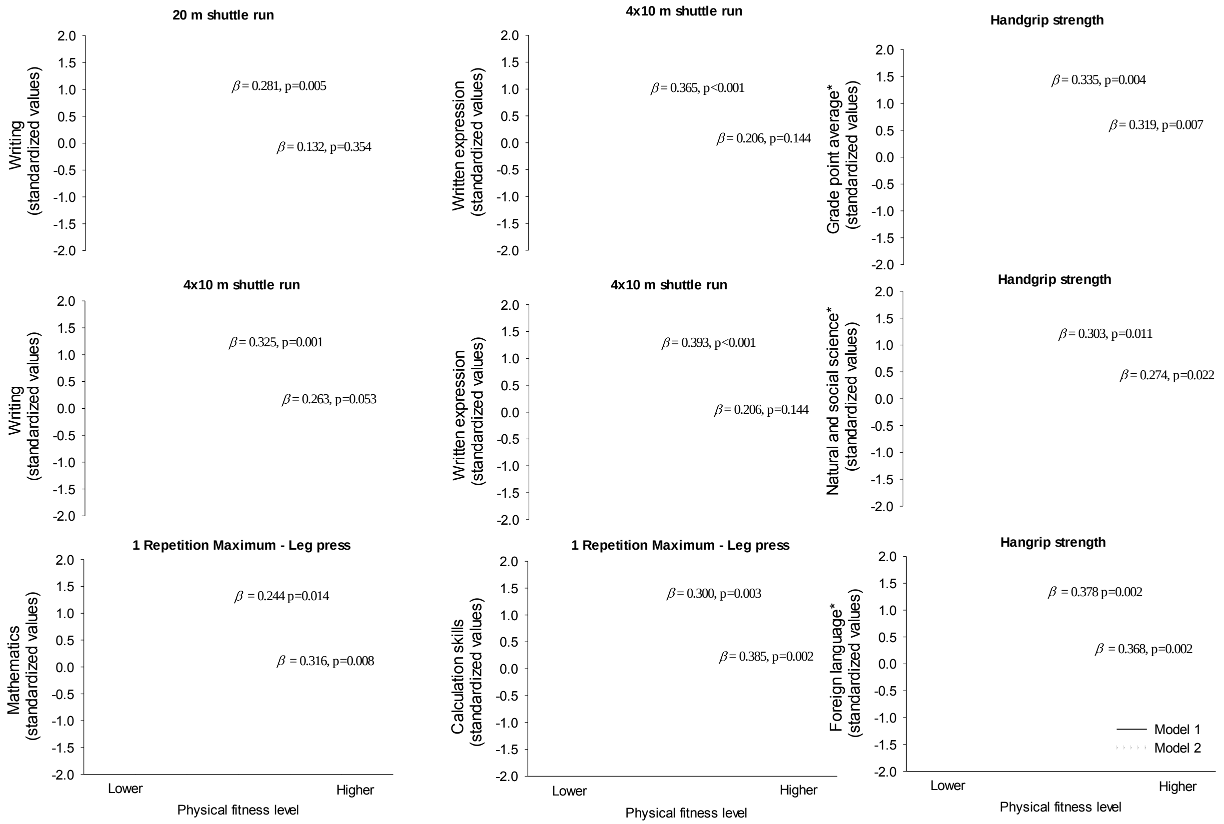


Figure 16. Linear regression analyses of the strongest associations between physical fitness components and academic achievement measured by Woodcock-Muñoz and school grades.

ß= beta standardized coefficients.

Data presented are the main associations (p<0.01) observed in the correlation analyses (see Table 26).

Model 1 was adjusted for basic confounders (age, sex, and maternal education).

Model 2 was additionally adjusted for body mass index and cardiorespiratory fitness, except for cardiorespiratory fitness which was adjusted for speed-agility fitness.

*These analyses used the information from school grades, while the rest used the Woodcock-Muñoz test.

Physical activity and academic achievement

The partial correlations between physical activity and academic achievement using the Woodcock-Muñoz test and school grades are shown in **Table 28**. Overall, the results showed that physical activity was not significantly associated with academic achievement. For this reason, we did not perform linear regression analyses with physical activity and academic achievement outcomes. In sensitivity analyses, we examined the relationships between physical activity obtained from different accelerometer metrics (i.e., ENMO and activity counts) and placements (i.e., right hip and non-dominant wrist), and the findings were consistent overall, showing no association (data not shown).

Table 28. Partial correlations between physical activity and academic achievement measured by Woodcock-Muñoz and school-grades.

	Light PA	Moderate PA	Vigorous PA	MVPA	1-min bouts MVPA	5-min bouts MVPA	10-min bouts MVPA
Woodcock-Muñoz:							
Total achievement	0.134	0.091	0.036	0.084	0.064	0.048	0.045
Reading	0.114	0.031	-0.001	0.026	0.017	-0.026	-0.043
Oral language	0.141	0.132	0.121	0.138	0.156	0.089	0.102
Writing	0.105	0.060	0.039	0.059	0.040	0.010	0.081
Written expression	0.123	0.085	0.111	0.096	0.052	-0.037	0.107
Mathematics	0.108	0.140	0.050	0.129	0.123	0.135	0.090
Calculation skills	0.110	0.056	-0.028	0.041	0.008	0.048	0.059
Sciences	-0.004	0.017	0.003	0.015	0.086	0.056	0.029
School grades:							
Grade point average	0.015	0.073	0.138	0.092	0.096	0.091	0.015
Language	-0.001	-0.011	0.050	0.002	0.026	0.013	-0.051
Mathematics	-0.009	0.080	0.145	0.099	0.129	0.123	0.077
Natural and Social Sciences	0.053	0.086	0.181	0.112	0.112	0.109	-0.022
Foreign Language	0.008	0.096	0.108	0.105	0.070	0.075	0.043

PA= Physical activity. MVPA= Moderate-to-vigorous physical activity.
Data were adjusted for basic confounders (age, sex, and maternal education).
All p>0.05
† Classified according to the Hildebrand et al.[95,96] cut-off points.

Study IX. Differences in brain in metabolic healthy vs. unhealthy: associations with academic achievement and role of cardiorespiratory fitness

Descriptive characteristics of the study sample (all and separated by sex and metabolic phenotype) are shown in **Table 29**.

Table 29. Descriptive characteristics of the study sample (Metabolically healthy=52, metabolically unhealthy=45).

	All (n=97)	Boys	Girls	Metabolically healthy (n=52)	Metabolically unhealthy (n=45)	$p_{phenotype*}$
	Mean ± SD	*Mean ± SD*	*Mean ± SD*	*Mean ± SD*	*Mean ± SD*	
Descriptive characteristics:						
Age (years)	10.0 ± 1.2	10.2 ± 1.1	9.8 ± 1.2	9.9 ± 1.1	10.1 ± 1.2	0.473
Peak height velocity (years)	-2.0 ± 1.0	-2.5 ± 0.7	-1.1 ± 0.8	-2.2 ± 1.0	-1.7 ± 0.9	**0.031**
Weight (kg)	56.0 ± 11.1	57.0 ± 10.7	54.3 ± 11.6	53.4 ± 11.2	59.0 ± 10.1	**0.011**
Height (cm)	144.1 ± 8.3	144.8 ± 7.4	142.9 ± 9.6	143.3 ± 8.4	144.9 ± 8.2	0.325
Body mass index (kg/m^2)	26.8 ± 3.7	27.0 ± 3.8	26.3 ± 3.5	25.7 ± 3.5	28.0 ± 3.5	**0.003**
*Weight status (n, (%))**:*						**0.016**
Overweight	24 (24.7)	15 (25.4)	9 (25.0)	19 (36.5)	5 (11.1)	
Obesity type 1	42 (43.3)	26 (44.1)	15 (41.7)	21 (40.4)	21 (46.7)	
Obesity type 2	19 (19.6)	11 (18.6)	8 (22.2)	6 (11.5)	13 (28.9)	
Obesity type 3	12 (12.4)	7 (11.9)	4 (11.1)	6 (11.5)	6 (13.3)	
Parental education (n (%)):						0.248
None with university studies	65 (67.0)	42 (71.2)	21 (58.3)	31 (59.6)	34 (75.6)	
Only one with university studies	17 (17.5)	10 (16.9)	7 (19.4)	11 (21.2)	6 (13.3)	
Both of them with university studies	15 (15.5)	7 (11.9)	8 (22.2)	10 (19.2)	5 (11.1)	
Cardiorespiratory fitness:						
Cardiorespiratory fitness (VO_2max)†	40.8 ± 2.7	40.8 ± 2.7	40.7 ± 2.9	41.6 ± 2.6	39.8 ± 2.6	**0.001**
Metabolic risk factors:						
Triglycerides (mg/dL)	98.6 ± 57.7	89.2 ± 51.7	114.8 ± 67.2	73.3 ± 23.7	128.6 ± 72.8	**<0.001**
Glucose (mg/dL)	86.3 ± 6.6	87.6 ± 6.4	84.5 ± 6.5	86.4 ± 5.8	86.4 ± 7.5	0.987

High Density Lipoprotein (mg/dL)	50.3 ± 11.2	51.3 ± 11.4	48.4 ± 12.7	56.7 ± 9.7	42.6 ± 9.6	**<0.001**
Systolic blood pressure (mmHg)	99.6 ± 12.9	100.4 ± 12.1	98.9 ± 14.3	98.9 ± 10.7	100.9 ± 15.2	0.458
Diastolic blood pressure (mmHg)	56.0 ± 12.3	57.0 ± 13.1	55.0 ± 10.8	55.2 ± 10.4	57.4 ± 14.1	0.383
Academic achievement:						
Total achievement	108.8 ± 12.4	108.6 ± 11.5	109.3 ± 13.8	111.38 ± 12.9	105.9 ± 11.2	**0.028**

SD: Standard deviation.
Data are presented as mean and standard deviations unless indicated otherwise. Statistically significant values are shown in bold.
*Independent t-test and chi-square test (for continuous and categorical variables, respectively). Descriptive characteristics were entered as dependent variables and metabolic phenotypes (i.e., metabolically healthy obesity and metabolically unhealthy obesity) as grouping variable.
**Classified according to Cole et al.[85] and Bervoets et al.[40]
†Measured by the 20-m shuttle run test, estimated following Leger et al.[88] equation.
Academic achievement was measured by Woodcock-Muñoz test.

Global brain volume

Figure 17 depicts differences between MHO and MUO in global gray matter volume. In an unadjusted model, MHO showed higher gray matter volumes than MUO (P=0.001). After adjusting for basic confounders (i.e. sex, peak height velocity, parental education university level and BMI), the differences became marginally significant (p=0.056). Finally, when cardiorespiratory fitness was included in the model, no significant differences were found in gray matter volume between MHO and MUO (p=0.135). Exploratory analyses revealed a similar pattern for global white matter (Figure S3a). MHO showed a significantly higher total brain volume compared to MUO after adjustment for basic confounders (p=0.035), yet this difference became non-significant after additional adjustment for cardiorespiratory fitness (p=0.100) (Figure S3b).

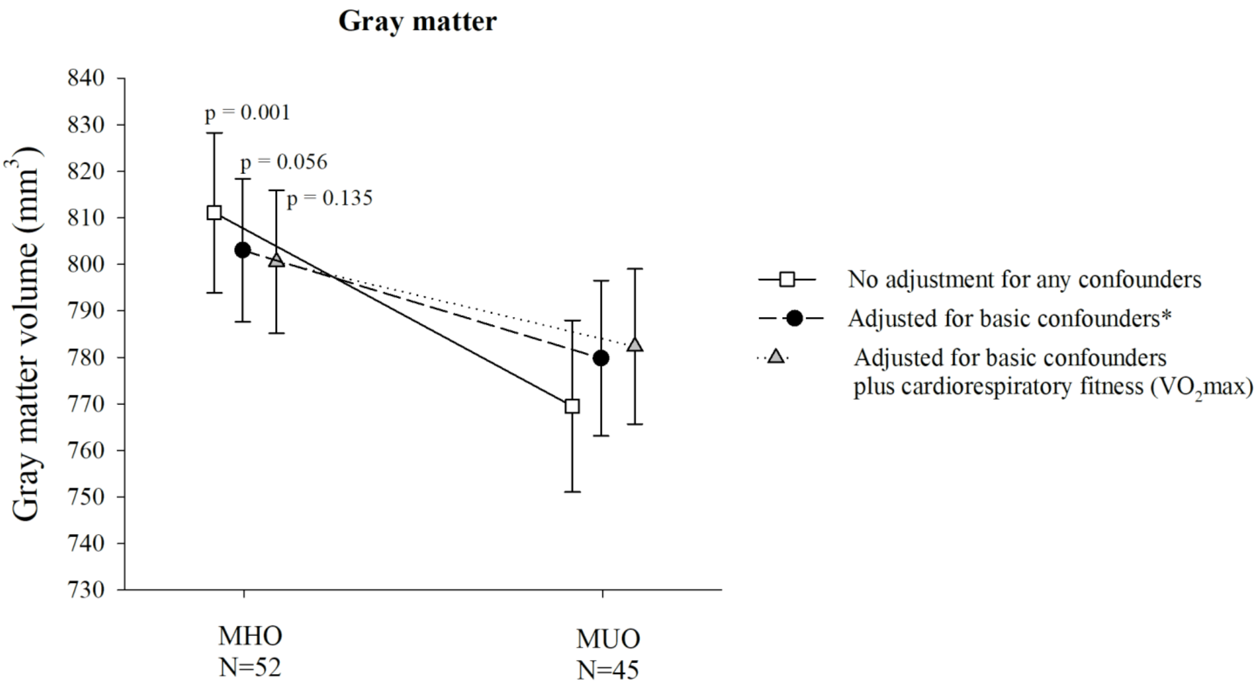


Figure 17. Differences in global gray matter between metabolically healthy and metabolically unhealthy overweight/obesity.

MHO: Metabolically healthy overweight/obesity. MUO: Metabolically unhealthy overweight/obesity.

*Basic confounders were sex, peak height velocity (years), parental education university level (none/one/ both of them), and body mass index (kg/m^2).

Regional gray matter volume

Table 30 shows the brain regions that depicted metabolic differences (i.e., MHO > MUO) in gray matter volume adjusted for confounders (model 1: sex, peak height velocity, parental education university level, BMI and total brain volume) and cardiorespiratory fitness (model 2). MHO children showed higher gray matter volume in six cortical regions after adjusting for confounders (all p's<0.001; k>60). When cardiorespiratory fitness was added to the model (Model 2), associations were generally attenuated (a reduction in the effect size Cohen-d ≥ 0.1) in each cortical region. **Figure 18** depicts the results shown for both models.

Table 30. Brain regions showing gray matter volume increases in metabolically healthy overweight/obesity compared to metabolically unhealthy overweight/obesity (n=97).

MHO (n=52) > MUO (n=45)								
Brain regions	***x***	***y***	***z***	***t***	***Cluster size***	***Hemisphere***	***Effect size***	
							Cohen's d	*95% CI*
Model 1:								
Fusiform gyrus	44	-33	-20	4.33	2008	Right	0.82	0.23,1.05
	-41	-30	-27	4.35	1581	Left	0.57	0.16,0.98
Calcarine	-12	-83	2	4.33	948	Left	0.72	0.31,1.13
	18	-66	14	3.69	662	Right	0.61	0.20,1.01
Lingual gyrus	-20	-68	-5	4.42	893	Left	0.67	0.26,1.08
	20	-77	-6	4.85	386	Right	0.80	0.38,1.21
Middle occipital gyrus	41	-80	14	3.72	120	Right	0.57	0.17,0.98
Superior temporal gyrus	36	20	-33	3.99	93	Right	0.64	0.23,1.05
Inferior temporal gyrus	-38	-6	-35	3.74	76	Left	0.53	0.12,0.93
Model 2:								
Fusiform gyrus	-41	-30	-27	4.08	700	Left	0.39	0.06,0.79
	44	-33	-20	4.64	427	Right	0.68	0.27,1.09
Calcarine	-12	-83	2	4.19	487	Left	0.58	0.17,0.98
Lingual gyrus	-20	-68	-5	4.15	388	Left	0.50	0.09,0.90
	20	-75	-6	4.62	256	Right	0.62	0.20,1.02
Middle occipital gyrus	41	-78	12	3.67	94	Right	0.47	0.06,0.87
Superior temporal gyrus	ns	ns	ns	ns	ns	ns	ns	ns
Inferior temporal gyrus	ns	ns	ns	ns	ns	ns	ns	ns

MHO: Metabolically healthy overweight/obesity. MUO: Metabolically unhealthy overweight/obesity; ns: No significant difference. 95% CI: 95% of confidence interval.

Model 1 was adjusted by sex, peak height velocity (years), parental education university level (none/one/ both of them), body mass index (kg/m^2), and total brain volume (mm^3).

Model 2 was adjusted for model 1 plus cardiorespiratory fitness (VO_2max).

All contrasts were thresholded using AlphaSim at P<0.001 with *k*= 60 voxels (model 1) and *k*=54 voxels (model 2) and surpassed Hayasaka correction. Anatomical coordinates (x, y, z) are given in Montreal Neurological Institute (MNI) Atlas space.

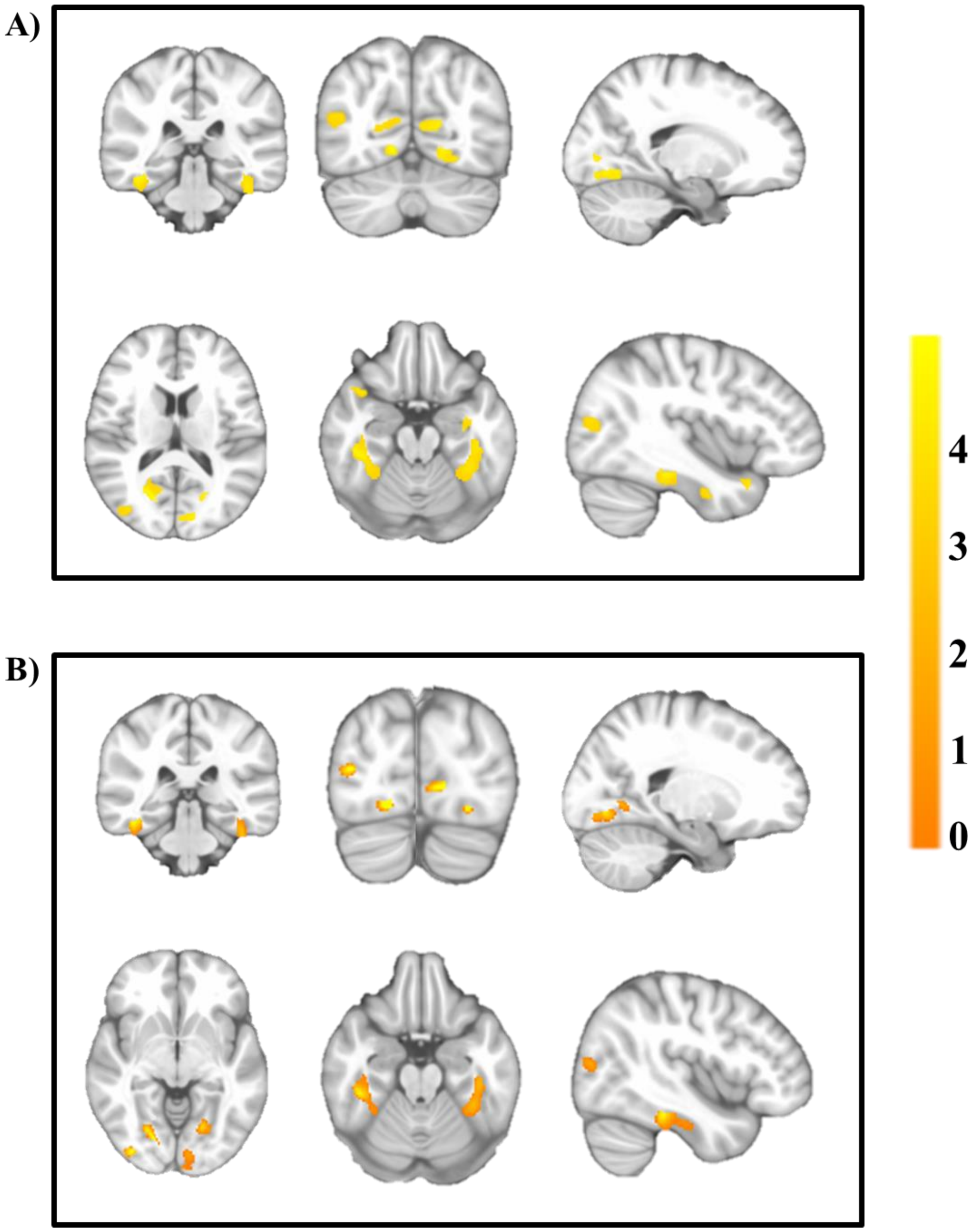


Figure 18. Brain regions showing higher gray matter volume in metabolically healthy overweight/obesity children (n=52) compared to those metabolically unhealthy overweight/obesity (n=45).

Analyses were adjusted for sex, peak height velocity (years), parental education university level (none/one/ both of them), body mass index (kg/m^2) and total brain volume (mm^3) (Panel A) and additionally for cardiorespiratory fitness (VO_2max) (Panel B). Results were thresholded using AlphaSim at P < 0.001 with k=60 (model 1) and k=54 (model 2) voxels, and surprassed Hayasaka correction (see Table 30). The color bar represents the associations (the lighter the yellow colour, the higher the association). Images are displayed in neurological convention, and thus, the right hemisphere corresponds to the right side in coronal displays. Sagittal planes show the left hemisphere (top of the figure) and right hemisphere (bottom of the figure).

Gray matter and academic achievement

Figure 19 shows the results from a linear regression analysis between gray matter volume and academic achievement. Positive associations were found between gray matter volume and academic achievement (ß=0.237, p=0.036) after adjusting for basic confounders (i.e., sex, age, parental education university level, and BMI). When cardiorespiratory fitness was added to the model, the relationship between gray matter and academic achievement was no longer significant (ß=0.210, p=0.064).

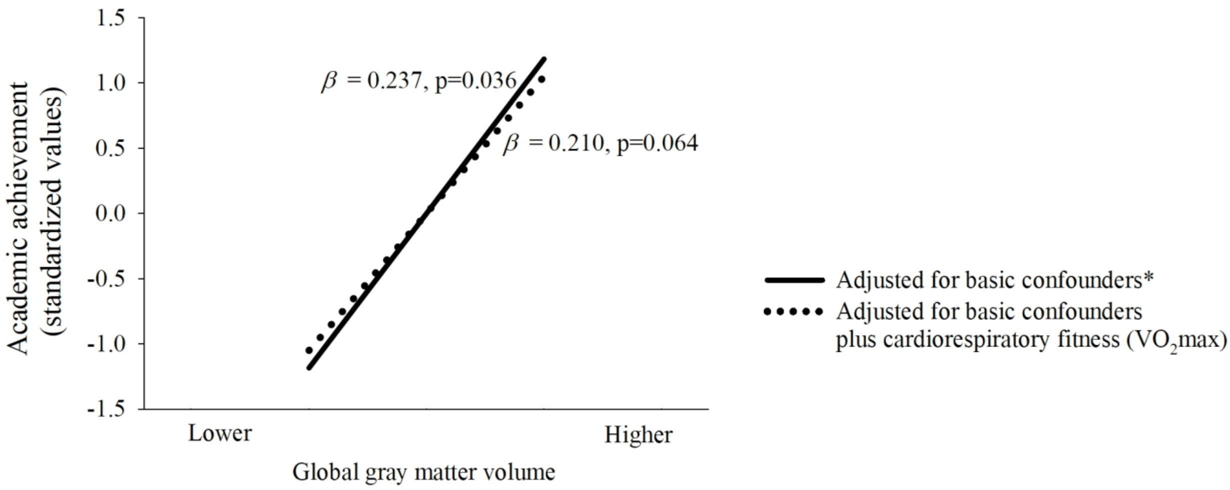


Figure 19. Linear regression analysis between global gray matter volume and academic achievement.

ß= beta standardized coefficients.

*Basic confounders were sex, age, parental education university level (none/one/ both of them), and body mass index (kg/m^2).

DISCUSSION

DISCUSSION

Summary of the main findings

Section 1: Physical fitness and fatness in preschool children

The main findings of the present International Doctoral Thesis suggest that: **I)** Most of the studies found in our systematic review were focused on reliability of the fitness tests while very few were focused on their validity or relationship with health outcomes in preschool children. Motor fitness, particularly balance, was the most studied fitness component, while cardiorespiratory fitness was the least studied in this population. We propose the PREFIT battery, field-based FITness testing in PREschool children; **II)** The intra-instrument test–retest reliability was excellent for both analog and digital dynamometers. The combined systematic error of ≤0.3 kg is relatively low for all devices, especially for the analog dynamometer, which has a precision of measure of 0.5 kg. Also, the systematic error between different instruments of the same model and between different models (digital vs. analog) ranged from |0.3 kg| to |0.6 kg|, which we consider acceptable for the measurement of muscular fitness. However, the systematic error between new and old dynamometers ranged from |0.8 kg| to |1.1 kg|, which in our opinion is a relatively large error and could substantially affect the estimations of strength; **III)** The adaptation of the original 20m shuttle run test, i.e., PREFIT 20m shuttle run test, consisted mainly in reducing the initial speed of 8.5 km/h to 6.5 km/h. This test was feasible and well accepted in both boys and girls from 3 to 5 years-old. Likewise, the MHR achieved for the entire sample was equivalent to 97% of the estimated theoretical MHR, and no significant differences by sex or age have been found. Moreover, the test showed a high reliability (mean difference of 2 laps) with no significant difference between sex or age; **IV)** The PREFIT battery is feasible and well accepted by preschoolers. Likewise, weight, height, BMI, waist circumference, PREFIT 20m shuttle run, handgrip strength, and 4x10m shuttle run tests are reliable in preschoolers, yet the one-leg stance test showed a poorer reliability and the variability error increased as the performance in the test was higher. Moreover, mixed findings were observed for the standing long jump, a systematic negative error was observed, which means that during the second assessment, they performed significantly worse than during the first assessment (2 weeks earlier). Learning effects have been previously reported in the literature, but a worsening effect in a second assessment when children are more familiar with the tests was an odd result. Practical recommendations to help the researchers to implement the PREFIT battery successfully in preschoolers were provided; **V)** For the first time, we provide reference standards for the main fitness components (i.e., cardiorespiratory fitness, muscular strength, speed-agility, and balance) by sex and age (for each 0.025 year of age increment, i.e. 9 days). Boys performed better than girls in cardiorespiratory fitness, muscular strength, and speed-agility tests, whereas girls performed slightly better in balance tests. Older preschool children performed better in all fitness tests than younger preschoolers. Furthermore, we observed sex-related differences in fitness across all preschool ages and percentiles, being greater in older children; **VI)** The prevalence of overweight/obesity in Spanish preschool children is 21.4% and 34.8% according to the WOF and WHO criteria, respectively. Among the whole study population (i.e., 3.178 Spanish preschool children), 6% of the children had obesity according to the previously mentioned criteria. Further, severe or morbid obesity was found in 2.5% of children. The prevalence of underweight was found in 3-5% of children. Moreover, sex-and age-specific reference standards for weight, height, BMI, waist circumference, and waist-to-height ratio from a relatively large sample of preschool children geographically distributed across Spain are provided; and **VII)** Preschoolers from Sweden had a lower prevalence of overweight/obesity than those from Spain (i.e., 21 vs. 31%). In regards to fitness level, preschoolers from Spain are fitter in terms of cardiorespiratory fitness, upper-muscular strength, speed-agility, and balance; and had worse performance in lower-muscular strength compared with those from Sweden.

Section 2: Physical fitness, academic achievement and brain in preadolescent children

In regards to the *Section 2* of the present International Doctoral Thesis, the main findings were: **I)** Various components of fitness were positively associated with academic achievement, measured using both standardized achievement tests and school grades, in overweight/obese children. Specifically, higher levels of cardiorespiratory fitness, muscular strength, and speed-agility were related to higher academic achievement in overweight/obese children. However, all significant associations found were attenuated or disappeared after additional adjustments for BMI and other fitness components (i.e., cardiorespiratory fitness), suggesting that they are inter-dependent in their association with academic achievement. On the other hand, physical activity did not demonstrate to be associated with any of the academic outcomes studied, suggesting a different relationship between these health factors (i.e., physical fitness and physical activity) with academic achievement; and **II)** MHO children showed marginally higher global gray matter volume and significantly higher total brain volume compared to MUO children. Likewise, MHO had higher gray matter in six cortical brain structures (i.e., fusiform gyrus, calcarine, lingual gyrus, and middle occipital, superior temporal and inferior temporal gyrus) compared MUO children. Global gray matter volume and total brain volume were positively associated with academic achievement. All observed differences were attenuated or disappeared after an additional adjustment for cardiorespiratory fitness, suggesting a role of fitness in the association of the MHO phenotype with brain structure and academic achievement.

Discussion of main findings with previous literature

Section 1: Physical fitness and fatness in preschool children

Study I. Systematic review of physical fitness in preschoolers

Reliability, validity and relation with health outcomes

Cardiorespiratory fitness

Although a small number of reliability studies were found, they suggest that either the 20m shuttle run, the 1/2-mile walk/run and the distance run in 3min tests provide reliable measures in preschool children aged 4 to 6 years. No information is available about the reliability of these tests in younger children, and little or no information either about their validity or relationship with health outcomes in preschool children.

Musculoskeletal fitness: muscular strength and flexibility

We found a few reliability studies, which suggest that both the handgrip strength and the standing long jump tests might be reliable measures in preschool children. However, most of this studies were conducted on 4 to 5 year-old[138-141], except for one study that included 3 year-olds[120]; so very little information is available about how reliable these tests are in 3 year-old children. Future studies should test the reliability of these tests separately at 3, 4 and 5 years of age. It is important to highlight that the Lode dynamometer was more reliable than the Martin vigorimeter and no information was found about other well-known dynamometers, such as Jamar, DyNex or TKK dynamometer. The sit-and-reach test seems also to be a reliable measure of flexibility in preschool children[123,140]. A major finding was however that no criterion-related validity or health-related study of single field-based tests in preschool children was found, which hampers further conclusion about the validity and health-relationship of the tests.

Motor fitness: speed-agility and balance

The studies reviewed support that the 4x10m shuttle run test for assessing speed-agility and different versions of the single-leg stance test for assessing balance are reliable tests to be used in preschool children. Although several studies examined the construct (convergent) validity of these speed-agility and balance tests, most of them did it for the overall scores of the subset, rather than for the single test so that little information is available about which test is the most valid in this age group. In addition, no criterion-related validity or health-related study (either longitudinal or cross-sectional) was found.

Feasibility of fitness testing in preschool children

During our search we identified a number of studies that did not meet the inclusion criteria or did not study reliability, validity or relationship with health of fitness tests, but they used the handgrip strength[150-155] and standing long jump[135,155-164] tests for measuring musculoskeletal fitness, different versions of the shuttle run test for measuring speed-agility[124,135,156,158,162-164], or the single-leg stance test (different versions) for measuring balance[135,160,162,165,166] in preschool children, and no problem was reported during testing sessions. Some physical activity-based intervention studies have also examined the effect of the intervention on physical fitness components in preschoolers, such as cardiorespiratory fitness, using different versions of the 20m-shuttle run test [167-169]. These studies do not provide information about reliability, validity or relationship with health of the tests, but provide further support that these tests are feasible to be used at these young ages.

The National Institutes of Health (NIH) Blueprint for Neuroscience Research has recently funded an initiative to create the NIH Toolbox Assessment of Neurological and Behavioral Function[170]. This initiative has tested the feasibility, reliability and validity of single tests to be used in people from 3 to 85 years. The authors proposed handgrip strength test for assessing muscular strength, standing balance test for assessing balance and the 2min walk test to assess cardiorespiratory fitness, in addition to other measures of dexterity and locomotion. This study is not included in our review because it does not provide statistics specifically for preschool children, but for the whole sample aged 3 to 85 or a subgroup aged 3 to 13. Nevertheless, this study further strengthens the feasibility of these tests to be used in preschool children.

Proposal of a field-based fitness-test battery for preschool children: the PREFIT battery

Although the information found about fitness testing in preschoolers is limited, a large number of studies (both observational and randomized controlled trials) are being conducted at the moment and researchers and practitioners need to know, based on the information currently available, which physical fitness tests are “better” than others. We used the information retrieved in the present systematic review and put it into context using existing evidence (i.e., recent systematic reviews[171-173] and methodological studies[174]) about reliability, validity and health-related fitness tests in older children and adolescents in order to make a proposal for fitness testing in preschoolers. Our proposal, named the PREFIT battery, field-based FITness testing in PREschool children is composed of the following tests: the 20m shuttle run test for assessing cardiorespiratory fitness, the handgrip strength and the standing long jump tests for assessing muscular musculoskeletal fitness, and 4x10m shuttle run and the one-leg stance tests for assessing motor fitness, i.e., speed-agility and balance respectively.

Study II. Reliability and validity of handgrip dynamometers

The main message drawn from these findings is that whenever possible, researchers, clinicians and practitioners should use the same dynamometer in repeated measures to minimize the systematic error among different apparatus. If repeated measures are taken using different instruments (e.g., new digital 2 minus new digital 1) or different models (e.g., new digital minus new analog), the systematic error is expected to range from 0.3 to 0.6 kg, with higher errors of measure observed between old and new dynamometers (both digital and analog). This finding is in accordance with the finding by Härkönen et al.[175], that older Jamar dynamometers were less accurate than newer ones. This fact raises a point of caution when interpreting results of handgrip strength levels in long-term follow-up studies because slight improvements or impairments (roughly 1 kg) in hand therapy or exercise programs could be attributable to a systematic error between new and old dynamometers. Thus, these findings should be interpreted and generalized cautiously because the systematic error might differ depending on the frequency of use, the way the dynamometer is used, and how old the dynamometer is.

The criterion-related validity analyses suggest that all dynamometers studied provide lower values than the SECA scale (−0.94 to −2.64 kg), which could be interpreted as a slight underestimation. This result is in contrast with what we observed in our previous study[176], in which the TKK (only the digital model) overestimated handgrip strength levels (0.49 kg). The difference between studies could be partially attributed to the model of the weight scale used (SECA 769 vs. 861) or to inter-instrument error. Nevertheless, we believe this last finding is not very relevant because handgrip strength will never be measured using a body weight scale. In addition, and in line with previous research[177], we assumed that the SECA scale was the gold standard, and therefore we calibrated the weights using this scale; however, we could not know how valid the SECA scale is because no better gold standard was available to test it. Consequently, we can conclude only that the TKK dynamometer provides values 1–2 kg lower than the SECA scale. Overall, heteroscedasticity analysis showed significant differences for the digital and analog dynamometers indicating that the error of the measures increased as the magnitude of the measures increased, as previously described[176].

Our results strongly support the use of the same instrument in repeated measures to minimize the systematic error to 0.3 kg or less. If different instruments, models, or dynamometers of different ages are used in repeated measures, the systematic error might increase up to 1 kg. This information is useful for studies in which the aim is to compare handgrip strength levels over different time periods (e.g., before and after a hand therapy programme, in intervention or follow-up studies). For example, the AVENA and HELENA studies Moliner-Urdiales et al.[178], analyzed secular trends in adolescents' physical fitness between 2001 and 2007. Adolescents showed a change of 4 kg in strength values measured by the handgrip strength test. If the researchers used the same dynamometer 6 years later, the expected error would be 1 kg. Some additional variance would be attributable to the biological variability of each participant, but the 4 kg of secular change reported is likely to reflect a real change in strength.

Study III. Adaptation of the original 20m shuttle run test in preschool children

Feasibility of the test

We observed that the test was performed correctly and was accepted for all age groups. Participants completed a minimum of 3 laps and maximum of 71, that is, equivalent to 8-10 min of the duration of the test as maximum. This duration is similar to the one observed in other studies in adolescents using the original test[25]. We have only found two studies[114,115] that have used the 20m shuttle run test in 4 and 5 years old children, and no studies were found in children aged 3. Therefore, to the best of our knowledge, the adaptation of the PREFIT 20m shuttle run test would be the first alternative for assessing field cardiorespiratory fitness in preschoolers.

Niederer et al.[114], observed that at least one evaluator should run with the children during the test to keep the correct pace of the test. After our study, we agree with the idea of the evaluator running with the children and we add that ideally, two evaluators should run with them. If the test is performed with one evaluator, we recommend reduce the group of the children assessed to 1-2 participants (running together with the evaluator, one in each side).

Maximality of the test

The motivation is a relevant factor for the performance of the test[179]. The results obtained showed that the PREFIT 20m shuttle run test could be considered a maximum test for assessing cardiorespiratory fitness in preschoolers, due to the children did a maximal effort, finishing the test in 200 beats/min that is equivalent to their 97% of the theoretical MHR. We have not observed significant differences between sexes or age group.

Reliability of the test

Mean difference (retest minus test, systematic error) in the PREFIT 20m shuttle run test was +2 laps. This difference was statistically significant suggesting a systematic error of +2 laps when the test is performed the second time, known as learning effect. However, the decision of whether the test is reliable or not must base on a scientific judge, due to the statistical alone cannot reply this question. As an example, when we test the reliability of a glucose measure performed two times for the same machine with 1 minute of difference, the variability expected cannot be the same as the one measured with a maximum test with 2 weeks apart where the own biological variability could have an important effect.

To our opinion, a variation of 2 laps having to two weeks apart between test and retest and the age of the participants, the test show a good reliability. In consonance with our results, others authors reported a good reliability when they used the original test in preschoolers[114,115]. However, is not possible to compare the findings because, on one hand, they applied the original test in children from 4 to 6 years old and on the other hand, they used different statistical methods (i.e. correlations). To compared the results with Niederer et al.[114], we performed a Pearson correlation (test-retest), obtaining the same values (r=0.84 for both studies, p<0.001). Yet, the correlation is a measurement of the strength of the association between two variables but not necessarily a measurement of agreement (reliability). It use is inappropriate because, first, it does not allow us to assess the systematic error, and second, it depends on the range of the values of the sample[101,180]. For instance, if one evaluator overestimates (positive systematic error) the punctuation of one test 20% compared with another evaluator, the correlation between measurements would be perfect, but they will never agree. Moreover, as more heterogeneous is the sample, the more the correlation. For this reason, we calculated Lin's concordance correlation.

If we compare the reliability of this test in adolescent population, we observed proportionally similar findings[181]. Due to in the adolescent study the test was measured in one minute stage, we transformed our data in stage (Table S8). In this sense, in accordance to Ortega et al.[181], we did not find significant difference between sex and age, although the systematic error of our study by sex is slightly higher to the one obtained in adolescents (HELENA study: -0.1 and 0.0; PREFIT: 0.1 and 0.5 stages for boys and girls respectively. However, the random error (SD of the mean difference) was similar and even slightly lower (HELENA study: 1.5 and 1.1; PREFIT: 1.0 and 1.1 stages for boys and girls respectively).

Study IV. Feasibility, reliability and practical recommendations for the PREFIT battery

Feasibility of the PREFIT battery

All the physical fitness tests were considered easy to measure, except standing long jump test for the youngest children, because it requires plenty of coordination and consequently more trials to perform correctly. We did not find any studies that have used these fitness tests in preschoolers as young as 3 years. Regarding the number of evaluators, the PREFIT 20m and 4×10m shuttle run tests should be performed with a minimum of two evaluators. Niederer et al.[114], performed the 20m shuttle run test with one evaluator running with children at 4-6 years. However, after our previous experiences, we recommend that, ideally two evaluators should accompany preschoolers during the test. In the 4×10m shuttle run test, we decided that in order to simplify the instructions and increase motivation.

Reliability of the PREFIT battery

Anthropometric measures showed high reliability results in all sample, boys and girls and 3, 4 and 5 years group. We observed that mean difference of +2 laps in the PREFIT 20m shuttle run test, considering the different groups of age of the participants and that there were 2 weeks apart, might not be meaningful. It is important to note that there were not differences between sex and age group, and no pattern of heteroscedasticity was observed. These findings concur with those reported by Niederer et al.[114].

The handgrip strength test showed a mean difference of |0.24kg|, that is, lower than the precision of the analog dynamometer (i.e. 0.5kg). Benefice et al.[117], observed a high correlation in handgrip strength test with only one day apart both in girls and boys 4-6.5 years (r=0.84 and 0.71 respectively), whereas in the present study we observed a greater correlation (r=0.86 for girls and boys). Differences among instruments (Martin vigorimeter vs. TKK) and the duration between measures (1 vs. 14 days) may explain these slight differences.

Likewise, for the standing long jump test, we observed a systematic error of |7.31cm| and |2.33cm| for S1 and S2 respectively. Oja and Jürimäe[118] observed a high correlation (test-retest-1 week apart-) in 4 years girls and boys (ICC=0.73-0.90). However, the correlation was lower when they assessed 5 years-old girls (ICC=0.57-0.69). In consonance with these authors and considering the standing long jump test S2, we observed the highest correlation in 4 years group (boys and girls r=0.77). The lowest correlation was observed in 3 years group (boys and girls r=0.60). However, due to the higher coordination patterns needed for standing long jump test and the difficulty observed in the preschool stage to perform it correctly, the reliability of this test is questionable. The results were however slightly better when the protocol was changed (replication study). Therefore, we recommend to perform the standing long jump test S2 (i.e. without placing a stick in the starting line).

The mean difference between measures in the 4×10m shuttle run test was only |0.12s|. Other authors[117] investigated the reliability of this test in preschoolers aged 4-5 years and concluded that the test showed an acceptable reliability (r=0.50 for girls, r=0.58 for boys). We observed that the 4×10m shuttle run test adapted for preschoolers has a good correlation in boys and girls (r=0.86).

The one-leg stance test presented a low reliability. We observed that the difference of the measures was 8.01s, which might be due to learning effect. Moreover, differences observed could be due to the child's personality and willingness to do the test. The range of the systematic error is approximately similar in boys and girls whilst the error increases linearly with the age (from 2.43 to 12.76s). Also, we confirmed that there were patterns of heteroscedasticity when we analysed the entire sample. Hence, due to the low reliability and high variability of the error, we decided to remove this test from the PREFIT battery. Although several studies[126,129,130] observed a high correlation in preferred leg and non-preferred leg (ICC from 0.73 to 0.99), our results showed a weakness in this test (r=from 0.55 to 0.68; mean differences from 2.43 to 12.76s). In addition, low levels of concordance correlation were observed in this test.

Likewise, researchers should take into account that, after an intervention, changes in handgrip strength has shown not to be sensitive to detect changes in strength unless the hand flexors are specifically trained[182,183], making it a good test for monitoring but less interesting for intervention studies.

When we compare our results in preschoolers with the reliability results of the HELENA study in adolescents (Table S8)[181], we observed similar systematic and random errors for cardiorespiratory fitness (transformed laps into stages), handgrip strength test, standing long jump S2 and 4×10m shuttle run test. Higher differences were found in the standing long jump test S1.

Practical recommendations

Practical recommendations (See Table S3) are provided to help the researchers to implement the PREFIT battery successfully in preschoolers.

Study V. Physical fitness reference standards for Spanish preschool children

To the best of our knowledge, our study is the first in providing sex- and age-specific reference standards of a complete set of physical fitness components in preschool children. A strength is that the applied tests were selected based on a systematic review in preschool children (see *Study I*), together with the existing evidence in older children and adolescents[2]. Previous studies[24,25,184] showed reference data for cardiorespiratory fitness, muscular strength, speed-agility, and flexibility in children and adolescents (>6 years old) using evidence-based fitness tests batteries (e.g. ALPHA). Nevertheless, our data are not fully comparable due to the differences in age groups and tests applied.

In regards to cardiorespiratory fitness, we observed that the differences of the 50th percentiles (P50) between both sexes were higher as age increased. Although we conducted different tests (original 20m shuttle run vs. PREFIT 20m shuttle run) and analyzed different age groups, other studies in children and adolescents showed a similar trend in P50 (e.g. sex-differences in 18-year-old adolescents: +38 laps for boys)[24,185]. Higher age-related differences by sex observed in adolescents compared to children or even preschool children might be explained by more pronounced physiological changes (due to the pubertal development) that occur as age increases.

Upper-limb muscular strength, assessed by the handgrip strength test, showed sex-differences in P50 of 1kg approximately as age increased for 0.5 years. Despite differences between studies and devices (analog vs. digital dynamometer), our results are in accordance with the results of De Miguel-Etayo et al.[24], who observed the same sex-differences in children aged 6-9 years old. Given that the methodology of the test is the same and inter-instrument reliability, as we have previously seen in the *Study II,* is high (mean difference, digital minus analog dynamometer = -0.35kg), our results are comparable in groups of 6-year-olds. Preschool children from the PREFIT project were stronger than the IDEFICS children (mean differences: 1.28 and 1.29kg for boys and girls, respectively). Roriz et al.[186] also provided reference values for Portuguese children aged 6 to 10 (age range 1-year), showing similar sex-differences (nearly 1kg). They also reported upper-limb muscular strength reference standard in P50, which is slightly lower compared to our results. The differences between studies might be due to limitations of the digital dynamometer, since the range of measurement is from 5kg to 100kg and the fact that 6-year-olds had several attempts below this range (i.e., 0kg). Another remarkable difference between the IDEFICS and the PREFIT studies is that overweight and obese children were not included in the calculation of reference standards in IDEFICS. Indeed, children and adolescents with higher BMI performed better in tests assessing absolute strength (i.e. also called non-weight bearing test)[187] compared to those with lower BMI.

According to lower-limb muscular strength, we observed that the differences in the P50 between both sexes ranged from 6 to 8 cm as the age increased every 0.5 years. La Torre et al.[27] provided reference data in standing long jump in a group of 3 to 6-year-olds from Jaén (a region from the south of Spain). It can be observed that the P50 of the PREFIT reference data was slightly higher in boys and girls and for all age groups than in the aforementioned study (differences ranging from 3 to 6 cm). Likewise, P50 depicts higher differences in the group of 3 to 4-year-olds (from +11 to +18 cm for boys and girls) and lower age differences in older groups (i.e. in 5 to 6-year-olds: from 4 to 8 cm). Other studies[24,186] provided reference values for children from 6 to 10 years old, showing lower variances between 0.5- and 1-year of difference. Moreover, the P50 of 6-year-olds from the European IDEFICS study[24] was higher than in both Spanish preschoolers participating in the PREFIT and in Portuguese children[186]. Once again, this result could be due to the exclusion of overweight and obese children from the data analyses in the IDEFICS study[24]. In contrast, the Portuguese study and the PREFIT study provide reference values for the whole sample, including all weight status categories[186]. Standing long jump test is a weight-bearing test where children have to move and lift their body mass. As a result, heavier children usually perform worse than their counterparts with lower body mass. In line with this assumption, Henriksson et al.[188] confirmed that a better performance on weight-bearing tests in preschool children was associated with a lower fat mass index. The researchers concluded that the more favorable body composition you have, the fitter you are.

Given the differences in the applied methodology, it is not possible to compare our results on motor fitness with any previous study neither in preschoolers nor in children and adolescents. In 4x10m shuttle run test, P50 showed the same trend for boys and girls, increasing their performance with age, and the range of difference between ages being practically systematic (from 0.5 to nearly 2 seconds). The reason for the greater performance in older children could be related to the development produced in motor coordination during the preschool period and childhood. As an example, this improvement was also demonstrated in European children from 6 to 9 years old (P50) who reduced the performance time by one second in 40m sprint[24]. Concerning balance, little is known about the reference standards of this fitness component in preschoolers. Between sexes, the P50 of boys and girls followed a similar pattern. Differences between younger and older preschoolers are greater in P50 both for boys and girls (approximately 18 seconds of difference). It is important to note that, although this test showed low reliability according to the data reported in the *Study IV*, we decided to provide its reference standard in order to help professionals to detect low levels of this fitness component. Nevertheless, researchers should be cautious when comparing pre-post values on two different occasions or after an intervention programme, due to the low reliability found.

The sex differences observed showed that already from preschool ages differences between boys and girls increased with age. Similar findings were reported by Castro-Piñero et al.[189], who observed sex differences during the stage from childhood to adolescence. Growth and, particularly, the early maturational status of girls play an important role. Our results are novel because they add to the literature the existence of sexual dimorphism in preschool children, and characterize the pattern of the different development course in both sexes with the age. Nevertheless, further studies are needed in order to corroborate or contrast these findings.

Overall, the present physical fitness reference standards allow other researchers or professionals to classify preschool children in sex- and age-percentiles. Preschool children can also be classified in fitness categories such as very low ($X < P10$), low ($P10 \leq X < P25$), medium ($P25 \leq X < P75$), high ($P75 \leq X < P95$), and very high ($X \geq P95$) and also scaling them from 0 to 10. In line with this assumption, we will upload an excel-based calculator to the website. With this tool, the researcher or practitioner can copy and paste age, sex, and the result of the fitness test, and the calculator will inform at which percentile that fitness value belongs to. The calculator functions entering either the data of one child, or copying and pasting columns from a data set, for instance of 3000 participants (see the Thesis CD; -will be freely available at http://profith.ugr.es/recursos/prefit). Thereby, professionals (sports practitioners, teachers, health care, trainers, etc.) can identify and help young children classified into the lowest categories, implementing strategies to promote physical fitness and physical activity to prevent or reduce future health-related problems. This paper provides national specific valid reference standards for preschool children and thus, our results are valid to compare with Spanish preschool children. However, since no data are available from other countries concerning this population, these reference standards could help and guide professionals in other countries in fitness classification until their own and/or international reference standards are available.

Study VI. Weight status and fatness reference standards for Spanish preschool children

Prevalence of weight status categories

About 21 and 35% (according to the WOF and WHO criteria, respectively) of the examined preschool children had overweight/obesity. Our results are in line with those found in European children (2.0-9.9 years old), where the total overweight/obesity prevalence was 20% and 28.4% based on WOF and WHO criteria, respectively[34]. Particularly, WOF data showed that Spain (including only one city, i.e. Zaragoza) presented the same overweight/obesity rates as we found in the entire sample of Spanish preschool children (i.e. overweight/obese: 21%, 15% overweight and 6% obese children)[34]. Based on data of 144 countries, de Onis et al.[190] calculated a global prevalence of overweight/obesity in preschool children of 12.9% according to the WHO criteria. Such differences between prevalence could be explained by (1) the different z-score criteria used for overweight/obesity classification, which did not include those at risk of overweight, i.e. z-score>1 and ≤2 in overweight category, (2) the timeframe of 15-25 years of difference between measurements (de Onis et al. study: 1990-2000 and PREFIT study: 2014-2015), and (3) the sample characteristics measured (de Onis et al. study: developing and developed countries *vs*. PREFIT study: developed country).

A further exploratory analysis using the same categorization as de Onis et al.[190] was conducted. This analysis showed a difference between PREFIT and the global situation in developed countries, i.e. higher overweight/obesity prevalence in PREFIT (18.2% *vs.* 12.9%). The higher proportion of overweight/obesity in Spanish preschoolers than those of the global average is not surprising given the fact that obesity is unequally distributed in Europe, being preschoolers and adolescents from the south heavier compared to their peers from the north[45]. Health-related factors such as higher levels of sedentary behaviour and lower levels of physical activity might explain these differences[45,191].

Overall, based on the WOF and WHO categories, our data showed a 5-6 percentage point higher prevalence of overweight/obesity in girls compared to boys (WOF: 23.9% *vs*. 19.3% and WHO: 37.9% *vs*. 32.0%, respectively). This tendency is partially in agreement with previous studies in European children[34]. In contrast to our results, several authors observed that using WHO categorization the overweight and obese rates were higher in boys compared to girls[34]. Differences between the most known international criteria (WOF and WHO) might be due to the development (characteristics of the study sample measured, countries evaluated, and sample sizes) and definition of overweight and obesity categories. In this sense, although we cannot express preference for one of the criteria, we see some advantages on using the BMI cut-offs provided by the WOF: 1) these cut-points are linked to the well-accepted BMI cut-points in adults for defining overweight and obesity (i.e. 25 and 30 kg/m^2) and 2) it provides the opportunity to study different degrees of obesity (mild, severe, and the recently published morbid obesity cut-offs)[40]. In contrast, WHO cut-off defines obesity as one group, without degrees of severity. We consider that analysing different degrees of obesity is a novel result from our study, in which we observed that about 1.3% of the study sample presented morbid obesity at preschool ages, and 2.5% of the children were either severe or morbid obese. Indeed, considering the whole population in Spain from 3 to 5 years old, 1.3% with morbid obesity would represent more than 14.000 preschool children. Thus, considering severe and morbid obesity together, this figure would increase up to ~30.000[192]. This result is crucial since it is well-known that severe and morbid obesity are among the most pathogenic conditions.

In addition, obesity at this age tracks into adulthood, and obesity-related non-communicable diseases such as cardiovascular diseases, diabetes, physical, and mental problems might appear in adulthood[193]. For example, longitudinal studies showed that overweight/obesity status in adolescents was strongly associated with increased mortality in adulthood due to cardiovascular diseases or other obesity-related risk factors[194]. In addition, it has been reported that the risk of cardiovascular disease mortality increased by 7% for every 2 additional years lived with obesity[195]. For this reason, early prevention of obesity in preschool ages is a major public health aim worldwide.

In regards to underweight status, only a few studies examined the prevalence in preschool children[42,44]. In Spain, a trend analysis (1983-2011) showed that underweight prevalence increased from 13.7 to 22.6% in preschoolers aged 2-5 years old[44]. Our results showed much lower rates of underweight (3-5%), in line with previous studies in preschoolers of developed countries (1%)[42] and also in Spanish and Portuguese children aged 6-8 years old (1-6%)[43,196]. Such differences could be explained by the different characteristics of the study samples, the criteria selected for grouping weight status, as well as by the dates of the measurements. Nevertheless, regardless of the variation observed among studies, it is interesting to highlight that our study shows that underweight prevalence was slightly lower than those obtained in obesity (mild, severe, and morbid), and that most of the children classified as underweight belonged to the milder category of underweight (i.e. type I, BMI 17-18.4 kg/m^2 at 18 years of age).

Little is known about the increased risk for diseases related to underweight compared to obesity in children, and more studies are needed in this direction. In this context, one study observed greater morbidity in underweight compared to obesity in preschoolers[41]. Therefore, at these early ages, low body weight is also an important factor to be considered from a public health point of view.

Sex-and age-specific reference standards

Available information providing sex-and age-specific anthropometric reference standards in preschool children are limited to few studies[29,31]. Moreover, due to the differences in the applied methodologies (different inclusion or exclusion criteria of the study sample, anthropometric tests used, etc.), our results are not fully comparable with other studies.

Overall, weight and height values were slightly higher in boys compared to girls, whereas no differences were observed in BMI. The trend of the weight and height growth curves was practically similar for boys and girls. In comparison with the data provided in previous studies in 5 year-old children, preschoolers from Spain were approximately 1kg heavier (50th percentile:19.8 and 19.6 for boys and girls, respectively) compared to those peers measured by the WHO[87], and about 0.5kg in comparison with Colombian counterparts[28]. However, taking together the WHO and Colombian reference data in 5 year-olds, we observed that Spanish preschool children were slightly taller than those who belonged to the WHO data (differences range from 0.5-1 cm)[87] and Colombian preschoolers (differences range from 1.5-1.9 cm)[28].

Our study provides an estimate of the rate of growths. For the 50th percentile curve of 3 to 6 year-olds, the 1-year increase is between 2-3 kg for weight and approximately 7 cm for height. In this period of life, preschool children experience many physiological and developmental changes and thus, the reference standards were provided by 0.025 year-steps (every 9 days/1.5 week of age approximately). BMI showed a slightly different pattern for boys and girls. In girls, the 50th percentile curve showed a plateau effect across all age groups. In boys, although changes are not remarkably different, BMI starts declining at the age of 3 and rises up to the starting point at the age of 6. Boys also show a plateau effect between the ages of 4 and 5.5 years old. It is also important to note the J-curve trend depicted in the highest percentiles (95th and 99th percentiles) of the boys' BMI. Particularly, the peak of adiposity rebound was found at the age of 4 in 99th percentile for boys. Similar ages for adiposity rebound were observed in the WHO (4 years old)[87] and Colombian reference standard (5 years old)[28]. In girls, neither the curves nor the reference data depicted a clear pattern in the highest percentiles to detect the age for adiposity rebound. It is of interest to highlight the relevance of the identification of critical periods for the development of childhood obesity for targeting prevention measures, given the negative consequence of an early adiposity rebound for obesity in adults[197,198].

Waist circumference is a well-known estimate of abdominal fat in adults, while less is known about it in children[31]. In this sense, Brambrilla et al.[199] assessed the association between waist circumference and visceral and subcutaneous abdominal adipose tissue measured by magnetic resonance imaging in children from 7 to 16 years old. The researchers concluded that waist circumference can be a good predictor of abdominal adiposity in children and adolescents. Our results showed a similar curve trend between sexes with slightly higher values in girls than in boys. Likewise, waist circumference increased with age. Based on 50th percentile curve we can observe that differences between years were about 0.3 and 0.6 cm per year for boys and girls, respectively. Nagy et al.[31] established percentile reference curves for European children from 2.0 to 10.9 years old based on the IDEFICS study providing approximately the same trend between years (differences around 0.7 cm) and showing higher waist circumference in girls than in boys. However, these results cannot be compared with our results since overweight and obese children were excluded[31], hence not representing the whole population.

The same patterns in boys and girls were observed for both waist-to-height ratio and waist circumference. In contradiction to the waist circumference percentile curves, the waist-to-height ratio decreased as age increased. Although we cannot directly compare our percentile curves with the curves from the IDEFICS study since the eligibility criteria were different, it is interesting to mention that the results depicted the same trend. Moreover, our study could contribute to create international standards, for instance, for waist circumference and waist-to-height ratio in order to define abdominal obesity by connecting the percentile curves to adults' cut-points as it was done for BMI by Cole and colleagues[85].

Overall, the present anthropometric reference standards allow other researchers or professionals to classify preschool children in sex-and age-specific percentiles. In line with this assumption, as occurs with physical fitness, an excel-based calculator will be available at our group's website. With this tool, researchers or practitioners can simply copy and paste sex, age, and the result of the anthropometric tests, and the calculator will inform which percentile belongs to the entered anthropometric value compared with the reference data presented in this article. The calculator will work for entering either the data of one child, or to copy and paste whole columns from a data set, for instance of 2500 subjects (see theThesis CD or temporally available for thesis purposes at Dropbox: https://www.dropbox.com/sh/les8w15j03md1fj/AADID9TTKRql9EoND_tUQzdra?dl=0).

Study VII. Comparison of fatness and physical fitness between Swedish and Spanish preschool children

Currently, available information comparing the anthropometric and fitness measures from different countries is limited to adolescents[45]. To the best of our knowledge, this is the first study that compares anthropometric and fitness scores between preschoolers from two clearly separate regions in Europe (Sweden and Spain). Our findings concur with those observed in adolescents, in which those from the center-north (including Sweden) were taller and had lower BMI than those from the south (including Spain)[45]. Ortega et al.[33] showed that the prevalence of overweight and obesity in adolescents was approximately double in southern compared to central-northern Europe, exactly the same as we observed in preschoolers. Furthermore, van Stralen et al.[49] showed in preschoolers (4-7 years) the highest rates of overweight/obesity in the southern European countries (i.e. Spain and Greece). Other authors provided prevalence data from different countries separately, and have shown higher rates of prevalence of overweight/obesity in southern Europe (WOF, Spain=32% and Portugal=37%)[200,201] compared with countries from northern Europe (BMI $\geq P_{85}$, Great Britain=19%)[202]. Since it is well-known that obesity at early ages tracks into adulthood and obesity at any age predicts higher premature mortality due to cardiovascular disease and cancer, early prevention of obesity is a major public health goal worldwide. Spain is one of the leading countries in the European ranking of childhood obesity at the age of 5-10 years. Our data contributes to the existing information by directly comparing the prevalence of overweight/obesity in preschoolers aged 4 years using the WOF cut-points[85] and showing that the high prevalence of childhood obesity might start already at preschool age.

It is also relevant to assess fitness at young ages due to the relation between the fitness components on health outcomes[3,19,203,204]. Our study showed that preschoolers from Spain had a better performance in cardiorespiratory fitness, upper-muscular strength, speed-agility, and static balance and worse results in lower-muscular strength compared with those from Sweden. The results regarding cardiorespiratory fitness showed that in these ages there were significant differences between Swedish and Spanish preschoolers (both in boys and girls). Within this context, several studies have reviewed cardiorespiratory fitness and have compared the data between countries[45,47,179]. In contrast with our results, they concluded that Spanish or adolescents from the south of Europe had worse performance in this component of fitness than adolescents from the north (both boys and girls)[45,47,179]. These results must be interpreted with caution because we used different methods to assess cardiorespiratory fitness and although we applied correction factors to account for the differences in the methods used, exact comparisons between both studies are not possible for this fitness test.

It might seem contradictory that the two muscular strength tests included handgrip strength and standing long jump, showed opposite results. However, there is a plausible explanation for these results. It has been reported that children and adolescents with higher BMI perform better in tests assessing absolute strength (i.e. not influenced by body weight) such as the handgrip strength test[187,205]. This notion is supported by our findings that preschoolers from Spain performed better in the handgrip strength test compared with those from Sweden, which was due to their higher BMI, since such differences were no longer significant after adjustment for BMI.

We observed that preschoolers from Sweden compared to those from Spain had more strength in their lower limbs, as assessed by the standing long jump test. We hypothesized that this could be due to the fact that they are taller, since we observed in exploratory analyses (data not shown) that the tallest preschoolers had better performance in this test. However, after further adjustment for height in the ANCOVA model no differences were found. Interestingly, when comparing the fitness levels of adolescents from the center-north vs. the south of Europe, we observed the largest difference (in favour to the north) for the standing long jump (i.e. 0.4 SDs compared with 0.1 SDs observed for cardiorespiratory fitness). The fact that young people from Sweden have a better performance in lower-limb muscular strength could be partly, explained by the fact that, on average, they are breastfed for a longer period than young people from Spain. According to the data of *the State of the World's Mother report*[206], the prevalence of breastfeeding in Sweden is 32% higher than in Spain (any at six months=72% and 40%, respectively), and it has been shown that adolescents with longer breastfeeding have higher muscular strength in their lower limbs as assessed by the standing long jump test[207], which could be seen already at preschool age.

With regards to motor fitness, preschoolers from Spain are fitter in terms of speed-agility and balance compared to their peers from Sweden. In contrast, Ortega et al.[33] observed that adolescents from the north of Europe presented better performance in speed-agility compared to those from the south. Nevertheless, it is important to mention that speed-agility and cardiorespiratory fitness highly and positively correlate with each other[47] in young people and the south-north differences observed for these two components are in line, supporting that these differences actually exist. For the balance test, when we adjusted for BMI the differences persisted, it is interesting to note that Niederer et al.[114] in their study with overweight and normal weight preschoolers observed that the overweight group had higher static balance compared to the normal weight group. Comparing with Niederer et al.[114], the reason that we did not find a difference when we adjusted for BMI might be due to the differences in BMI between groups was not significant enough (i.e. preschoolers were in normal weight category).

Section 2: Physical fitness, academic achievement and brain in preadolescent children

Study VIII. Physical fitness, physical activity and academic achievement preadolescent children

Physical fitness and academic achievement

Our findings are in accordance with previous reports that showed positive associations between fitness and academic achievement in similarly aged children[52,54]. In our study, a field measure of cardiorespiratory fitness was associated with language-related skills (i.e. writing and written expression). This result is in line with that of Telford et al.[208], who observed a positive relationship between the 20m shuttle-run test and writing in children aged 8.5-10.5. In accordance with this result, growing evidence depicts the role of cardiorespiratory fitness on brain and cognition in children during preadolescent development[12,209-211]. However, the mechanisms for this association remain unclear. It has been suggested that aerobic activity influences growth factors, including brain derived neurotrophic factor (BDNF), insulin-like growth factor, and vascular endothelial growth factor, which are involved in neurogenesis, angiogenesis, cellular proliferation, and neural plasticity processes that result in healthier brains[212-214]. Likewise, a greater level of fitness has positive effects on molecular and cellular aspects of brain structure and function, particularly in specific regions and networks (e.g., prefrontal cortex and associated executive control network [basal ganglia], hippocampus, etc.) that underlie specific cognitive functions (e.g., mathematics, language, etc.)[13,215]. These changes might be related to improvements on brain, cognition and academic achievement.

In contrast with our results, other studies have observed that cardiorespiratory fitness (measured via both field and laboratory tests) was related to higher reading skills (as a measure of language)[66,216,217] and mathematics[66]. In particular, the only study which analyzed fitness and academic achievement in overweight children is not in agreement with our findings, since they found a positive association for reading and mathematics, whilst we did not[66]. Such discrepancies among results could be explained by the characteristics of the study sample (Davis et al.: overweight vs. this study: overweight and obese) and the different confounders used (Davis et al.: race, sex, primary caregiver's education level vs. this study: sex, age, maternal education). Moreover, we found differences in the results of our study between measures of cardiorespiratory fitness (i.e. field-based and laboratory tests), which could be explained by: 1) the nature of the measurements, being one more focused on performance measurement (greater number of laps, higher performance) and the other on a physiological marker (VO_2max). As an example of the different nature of these two variables, laps increase with age whilst VO_2max decreases with age from childhood to adolescence[23]; 2) the natural condition of running in a field-based test (running in a playground, with general fatigue as a reason to terminate the maximum test, etc.) *versus* the non-natural conditions of the laboratory test (e.g. wearing a mask, walking on a treadmill at a steep slope, with higher local muscular fatigue as a reason to terminate the maximum test, etc.); and 3) the higher MHR observed during the field test compared to the laboratory test (mean differences: +4.2 beats per minute, p<0.01).

In regard to muscular strength, our findings in the field and laboratory tests showed significant associations with mathematics skills in Woodcock-Muñoz and school grades (grade point average, natural and social sciences, and foreign language). The relationship observed for the muscular strength and mathematics scores is consonant with evidence from previous cross-sectional findings in children[218]. However, other studies have also reported a lack of significant findings between muscular strength and these academic components[54,55,219,220]. The discrepancies between findings could be due to the different characteristics of the study sample, fitness, and academic measurements used as well as the covariates applied in the model. Although continued work is necessary to expose the exact mechanisms relating muscular strength and cognitive and academic performance in children, previous studies in the elderly have shown the role of resistance training on these outcomes[221-223]. For instance, after a 6-month trial, those participants involved in the resistance training programme demonstrated benefits in memory performance and verbal concept formation[223]. These findings raise the possibility that a broader spectrum of cognitive functions could also show improvement with resistance training. To the best of our knowledge, no study has examined science components individually and, therefore, it is difficult to compare our results with previous findings given the differences in content. Likewise, another explanation for differences between academic tests is that school grades are influenced by subjective teacher evaluations, making them a less objective outcome than standardized tests. Overall, although the relationship between muscular strength and academic achievement remains uncertain[224], our study supports the relevance of muscular strength during preadolescence.

Speed-agility fitness has been the least studied component in the literature. In fact, only one study has examined this relationship using the 4x10m shuttle-run test and academic outcomes (i.e. school grades) in normal weight children[54]. Similar to our results for writing and written expression (language-related skills), Esteban-Cornejo et al.[54] found a significant association between speed-agility fitness and language. The higher demands for coordination, agility, and memorization in the 4x10m shuttle-run test, which involves the temporal lobe (also crucial for memory and language)[12], may be one potential explanation of these findings. Likewise, a recent study[12] conducted by our group showed that those with higher speed-agility fitness presented higher gray matter volume in the inferior frontal gyrus and the superior temporal gyrus, which, in turn, were related to better academic performance (total academic achievement, reading, and academic fluency).

Moreover, to the best of our knowledge, this is the first study to analyze the unique relationship of cardiorespiratory, muscular strength, and speed-agility fitness. Our results shed light on the role of cardiorespiratory fitness as a confounder in the relationship between muscular strength, while speed-agility fitness also confounds cardiorespiratory fitness (i.e., confounding one another) on academic outcomes. Overall, our findings suggest that, after the additional adjustment of fitness, the unique relationship of cardiorespiratory fitness, muscular strength, and speed-agility with academic achievement was null or attenuated, indicating the role of fitness in driving the significant findings.

Physical activity and academic achievement

Alternatively, we did not observe any association of objectively-measured physical activity with academic achievement, which is in agreement with most previous reports[60,63-65]. However, it should be noted that examples in the literature do exist for both positive[60,61] and negative[62] relationships between physical activity and academic outcomes. The explanation for the conflicting results may be the differences in the study sample, the method of measurement in collecting physical activity data (self-reported or accelerometry), the accelerometer data collection and processing criteria (cut-points, placement of the accelerometer, metric used, etc.), the control of confounding variables, and the academic achievement tests and outcomes studied (standardized *versus* school grades).

Considering the physical fitness and physical activity findings together, the differences observed in their association with academic outcomes in our study could be explained because fitness is considered a physiological condition or state, and thus, it shows less day-to-day variation than physical activity, which is a behavior with higher day-to-day variability. As stated previously, physical activity is strongly related to physical fitness, being the main focus of the interventions intended to improve physical fitness.

Study IX. Differences in brain in metabolic healthy vs. unhealthy: associations with academic achievement and role of cardiorespiratory fitness

In the last few years, increased information has emerged on obesity-related comorbidities and metabolic alterations and their relationship with structural brain abnormalities in adolescents[225,226]. Interestingly, we observed that those who presented a MHO, global gray matter volume was marginally greater compared to those with a metabolically unhealthy profile. Moreover, we observed a similar trend for global white matter and we observed a significantly higher total brain volume in MHO compared to MUO, after adjustment for relevant confounders such as sex, age, parental education university level and BMI. However, all these differences disappear after accounting for cardiorespiratory fitness, suggesting that cardiorespiratory fitness play an important role in the relationship between the MHO and the brain. We cannot directly compare our results with previous studies due to the lack of the literature on MHO and whole-brain analyses, but our results are in line with the study by Yau et al. (2012)[83], who observed that that adolescents with metabolic syndrome have reduced hippocampal volume and academic achievement. Endothelial dysfunction, a failure of glucose homeostasis and/or an alteration of vascular integrity have been linked with obesity and cognitive impairment[227-230]. Therefore, the mechanisms underlying obesity (metabolic risk factors, insulin resistance, diabetes, inflammation, etc.) may influence brain health and explain such differences. In **Figure 20** we illustrated the main findings of this study and outlined potential mechanisms that could explain these novel findings. In support of this assumption, VanWagner et al.[231] examined global gray matter volume in obese middle-aged adults with and without non-alcoholic fatty liver disease (NAFLD), suggesting that NAFLD was negatively associated with brain volume. This study provided insight into the potential metabolic role of liver fat in brain health. Thus, based on the previous findings indicated above regarding the association between MUO and excess liver fat, we could hypothesize that these brain alterations could also have been present in the MUO children assessed in our study.

With regard to regional brain analyses, our results indicated that MHO presented higher gray matter in six cortical brain regions (i.e., superior temporal gyrus, inferior temporal gyrus, fusiform gyrus, calcarine, lingual gyrus, and middle occipital gyrus) compared to MUO. Differences in visceral fat accumulation, birth weight, adipose cell size, gene expression-encoding marker of adipose[232] health differentiation may favour the development of a healthy metabolic phenotype, which in turn might help maintain lower levels of inflammation and insulin resistance[81,233], and thus, lower the levels of growth factors, such as BDNF, vascular endothelial growth factor (VEGF) and insulin-like growth factor (IGF), which contribute to gray matter development[212]. However, such speculation should be taken with caution since this study is the first to analyse both global and regional gray matter volume based on metabolic phenotypes (i.e., MHO and MUO).

Another important aspect of the present study was the analysis examining the relationship between gray matter volume and academic achievement among overweight and obese youth, observing a positive association between then. Likewise, as exploratory analysis (Figure S4), we also observed the positive association between total brain volume and academic achievement. We did not find previous studies examining these associations; therefore, future studies will confirm or contrast these findings. In a previous study, also from the ActiveBrains project, we observed that gray matter volume in three brain regions, previously associated with cardiorespiratory fitness, were also positively related to academic achievement in overweight/obese children[12]. In the present study, we contributed to the previous one[12] by showing a positive association between global gray matter, total brain volume and academic achievement, in addition to the regional associations previously shown. Although this gap in the literature is not well documented, with the current data we demonstrate that a relationship exists between global gray matter and total brain volumes and academic achievement.

The role of cardiorespiratory fitness over metabolic phenotypes (i.e., MHO and MUO) has been previously examined in relation to several health outcomes[67,79,84]. Our group recently showed that the higher risk in cardiovascular prognosis observed in MHO adults compared to metabolically healthy normal-weight seem to be explained by cardiorespiratory fitness[80,84]. In the present study, our results showed no significant differences between metabolic phenotypes in global gray matter volume and an attenuation from six to four brain regions after controlling for cardiorespiratory fitness. It is interesting to highlight that those regions that disappeared were observed in a previous study strongly associated with cardiorespiratory fitness in overweight/obese children[12] which could explain our results, showing the role of cardiorespiratory fitness on brain structure. However, they also shown associations in fusiform gyrus and calcarine[12], while we did not. Of note, the non-significant differences found in global gray matter was not exclusively related to the inclusion of cardiorespiratory fitness in the model, but also due to the inclusion of confounders, which resulted in only marginally significant effects. Yet, in an exploratory analysis, when global structure volumes were combined together (i.e., total brain volume), our results clearly showed the role of fitness between metabolic phenotypes. Interestingly, in regional brain volume, the four brain regions showed a lower cluster size and effect size in MUO children in comparison to MHO after controlling for cardiorespiratory fitness. In accordance with these findings, the association between gray matter and total brain volume and academic achievement also disappeared when we considered cardiorespiratory fitness in the model. Therefore, based on previous literature, the current findings support cardiorespiratory fitness as a potential association-modifier in global and regional brain volume as well as academic achievement in overweight/obese children.

Further, it is also known that obesity affects these physiological processes, such that increases in obesity are related to decreases in brain health (lower brain volume and academic achievement)[234]. However, previous studies have tested the 'fat but fit' paradox, showing that higher fitness attenuates the negatives consequences of excess body mass on cardiovascular disease risk[67,235], however it was unknown whether fitness also attenuate the adverse consequences of obesity on other outcomes such brain and brain-related outcomes. Our findings add to this body of knowledge by supporting that there are differences in brain between MHO and MUO children, and that these differences are at least partially explained by cardiorespiratory fitness. Our data also suggest that the association of global gray matter and total brain volume with academic achievement was attenuated after accounting for cardiorespiratory fitness.

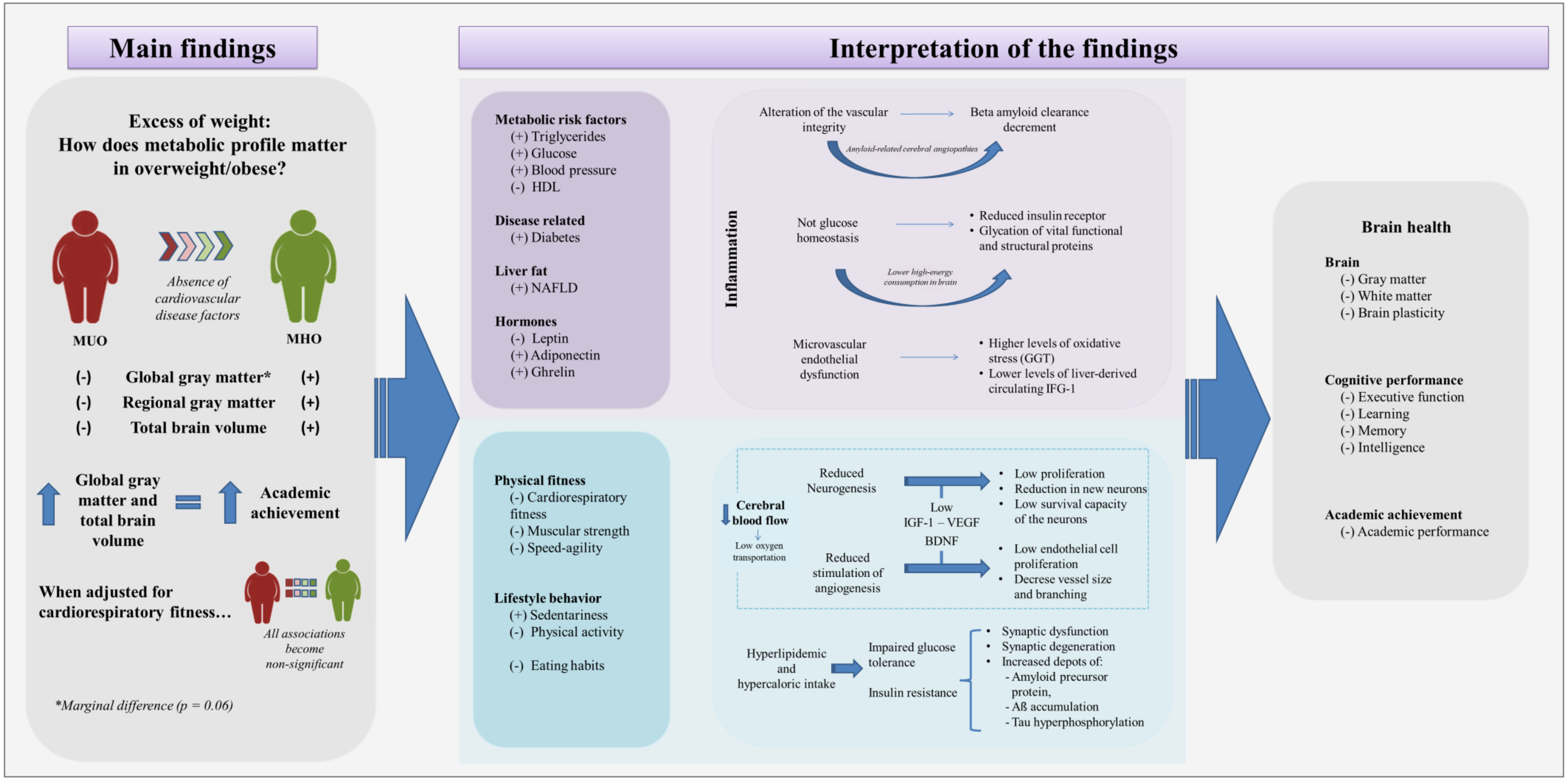


Figure 20. Graphical illustration of the main findings of the present study and plausible physiological interpretation: Factors involved with metabolic profile and its association with brain and cognitive/academic outcomes.

Higher levels of the factors are indicated as (+). Lower levels of the factors are indicated as (-). MHO: Metabolically healthy overweight/obesity. MUO: Metabolically unhealthy overweight/obesity. HDL: High density lipoprotein. NAFLD: Non-Alcoholic Fatty Liver Disease. IGF-1: Insulin-like Growth Factor-1. GGT: Gamma-Glutamil Transferasa. VEGF: Vascular Endothelial Growth Factor. BDNF: Brain Derived Neurotrophic Factor.

LIMITATIONS AND STRENGTHS

Limitations

The studies comprising this International Doctoral Thesis have several limitations that that must be underlined:

In regards to the *Study I*, quality assessment of each study was not carried out in the systematic review due to the small number of studies identified in preschool children for each test. Consequently, levels of evidence based on quality assessment could not be constructed. Due to the limited information found in preschool children, the test battery hereby proposed is a combination of the information found in preschool children and existing information in older children and adolescents. Once more evidence is accumulated in preschool children, this proposal should be revisited.

A limitation of the *Study II* is the absence of a group of people with different characteristics to examine the reliability of different models of dynamometers.

For the *Study III*, the limited sample size particularly in 3 and 4 years old is a limitation to acknowledge. Likewise, the performance in the running test could be influenced by the motor and physiological development of each children, yet we could not assess objectively those movement patterns.

In the *Study IV*, the relatively small sample size of 3 years group (n=29) is the main limitation. Although we invited to participate in the study the same number of classes, parents of the preschoolers of 3 years group were less agreeable to participate (participation rate=58% vs. 77% for 4-5 years group). This limit the generalisation of the results found especially for the youngest group.

The cross-sectional design of the *Study V* and *Study VI* do not allow to examine inter- and intra-individual differences, resulting in the need for longitudinal studies. We also consider a limitation the lack of sensitivity analyses for obese children, which does not allow to provide specific reference standards for them. In relation to the sample analyzed, it is not representative of the Spanish population, although it is a relatively large sample size that covers cities from north to south and from east to west in Spain.
Specifically, in regards to the *Study V*, the lack of validation studies of the physical fitness applied in preschool children due to the logistic problems inherent to the age of the children is another limitation. In addition, the difficulty to differentiate between motivation and performance limitations is another study limitation to acknowledge. In relation with the *Study VI* the lack of other adiposity markers (e.g. fat mass, fat free mass) accurately measured with gold standard methods is another limitation.

One of the limitations of the *Study VII* was the relatively small sample size (especially from Spain) which may limit generalizability. The possibility that the areas selected for this study might not be nationally representative is another limitation. Also, this study did not adjust for others variables such as socioeconomic factors, objectively measured physical activity or diet. Furthermore, using BMI as a measure of total body fat in model 2 may not have fully adjusted for differences in body composition since BMI does not separate fat and lean mass. Finally, the different initial speed in the PREFIT 20m shuttle run test and the application of two correction factors for the tests need also to be acknowledged as limitations.

The main limitation of the *Study VIII* is its cross-sectional design, which cannot draw causality. However, these are the baseline data from an ongoing longitudinal study, and it is important to understand between-group differences that are observable prior to the intervention. The limited sample size, especially in the 1-RM or school grade outcomes, reduces the power of our analyses. Likewise, children from different schools may have been graded differently, being another limitation to consider.

Lastly, for the *Study IX*, the cross-sectional design does not allow for causal inference and the lack of a normal weight group preclude us from do comparisons between overweight/obese and normal-weight children.

Strengths

For the *Study I*, two major databases were used for the search (PubMed and Web of Science) and there is the possibility that some additional studies could have been identified if using other electronic databases. Nevertheless, PubMed is currently the largest electronic database for the life sciences, and now most journals in sports sciences field are indexed in the Journal Citations Report. In addition, Web of Science includes all journals in all fields, with an impact factor, indexed in the Journal Citation Report. We believe these two databases, together with the review of the reference lists of the papers reviewed, gave us a good overall picture of the studies conducted in preschool children using fitness tests.

The main strength of the *Study II* is that we used the TKK dynamometer (digital and analog version) which is shown that it is reliable and valid and also, the characteristics of the device such as precision could help to detect minimum changes in comparison with other dynamometers (e.g. Jamar).

In regards of the *Study III and Study IV*, to the best of our knowledge this is the first study that analyse in depth the reliability of fitness tests in preschool children. Also, we included statistical methods not used before such as Bland and Altman method, the heteroscedasticity analyses and the Lin's concordance correlation coefficient.

Notable strengths of the *Study V* and *Study VI* are the relatively large sample size in preschool children geographically distributed in Spain, the sex-and age-specific physical fitness and anthropometric standards, the derivation of the percentile curves using GAMLSS, and the standardization of measurements across all the involved centres.

Additionally, for the *Study VI*, the prevalence reported in all obesity categories should be acknowledged.

The main strength of the *Study VII* is that we standardized the methodological procedures between the MINISTOP and PREFIT projects which allowed us to make accurate comparisons (except for the test).

The strengths of the *Study VIII* are the inclusion of valid[18], reliable[17], and health-related field and laboratory fitness tests[2]; the novel analyses of physical activity based on different locations and metrics; and a comprehensive academic achievement battery measured via standardized tests and school grades.

Lastly, the novelty of the investigation, the use of magnetic resonance imaging to assess brain structure in nearly 100 children and the use of a thorough (roughly 2h testing per child) standardized academic achievement outcomes are the strengths of the *Study IX*.

FUTURE DIRECTIONS

FUTURE DIRECTIONS

Study I

The systematic review carried out as the first study of the present International Doctoral Thesis, has identified a number of future research directions, as follows: A major finding of this review is the lack of information regarding criterion-related validity, indicating the need for future research in this direction. Information regarding physical fitness tests in preschool children and health outcomes during that period (cross-sectional studies) and later in life (longitudinal studies) was almost non-existent, which could be due to the fact that researchers did not have enough information about how fitness should be tested in this age group. The present systematic review will contribute to better inform researchers about which physical fitness tests can be included in their studies, which in turn could result in more and improved future studies relating physical fitness with health outcomes in preschool children. Studies focused on fitness in preschool children should take into account that, at these early ages, a 1-year difference implies marked changes in fitness and development.

Study II

Further studies involving human samples and different types of dynamometers (e.g., Jamar, Dynex) will confirm or contrast our findings.

Study III and Study IV

Studies with larger samples will contrast our findings. Likewise, studies focused on validity of these fitness tests (i.e., PREFIT 20m shuttle run, handgrip strength, standing long jump and 4x10m shuttle runt tests) in preschool children are also needed.

Study V to Study VI

Further studies should examine and provide reference standards at international level. Likewise, studies implementing an exercise intervention will show the expected changes in physical fitness in this age group.

Study VII

Further studies involving larger sample sizes, different countries from the north and south of Europe and adding the use of other direct measures of body composition (e.g., bioelectrical impedance) will confirm or contrast these findings.

Study VIII

Further studies with larger samples are needed, and further examination of the effect of a physical exercise programme on academic achievement in overweight/obese children should be conducted to corroborate the observed results. Likewise, further randomized controlled trials should focus on the inter-relationship between physical activity and fitness and how it affects academic achievement.

Study IX

Future studies should examine differences between healthy and unhealthy metabolic phenotypes in healthy weight children to better understand the role of metabolic phenotypes in overweight/obese children.

CONCLUSIONES / CONCLUSIONS

CONCLUSIONES

Conclusiones generales

Los resultados de esta Tesis Doctoral Internacional aportan nuevos métodos para una mejor evaluación de la condición física en prescolares, ofreciendo además valores de referencia para poder interpretarla. Igualmente, esta tesis concluye que no sólo la capacidad cardiorrespiratoria se asocia con el rendimiento académico sino también la fuerza muscular y velocidad-agilidad. Los resultados sugieren también que los niños con sobrepeso/obesidad con un fenotipo metabólicamente sano tienen mayor materia gris cerebral y mayor volumen total cerebral que aquellos con un fenotipo metabólicamente no sano, lo cual se asoció a su vez con un mejor rendimiento académico. Estos resultados darán lugar a futuros estudios prospectivos y de intervención centrados en la salud física y cerebral en la infancia y en el futuro.

Conclusiones específicas

Las principales conclusiones de la presente Tesis Doctoral International son:

1. La revisión sistemática de la literatura propone los siguientes test para la batería PREFIT: el test de 20m de ida y vuelta PREFIT para la evaluación de la capacidad cardiorrespiratoria, los test de prensión manual y salto horizontal con pies juntos para la evaluación de la fuerza muscular, y el test de 4x10m ida y vuelta y el test de equilibro a una pierna para la evaluación de la capacidad motriz, es decir, velocidad-agilidad y equilibrio, respectivamente.

2. Nuestros resultados apoyan el uso del mismo dinamómetro, ya sea analógico o digital, para la medición de fuerza en medidas repetidas para minimizar el error sistemático a un máximo de 0,3kg. Si se utilizan diferentes instrumentos, modelos o dinamómetros de diferentes años en medidas repetidas, el error sistemático incrementará hasta 1kg. Este estudio permitirá distinguir entre qué es un cambio significativo clínicamente y qué cambio es probable que sea debido a la variabilidad de la medición del dinamómetro.

3. El test de 20m de ida y vuelta PREFIT es viable, máximo y fiable en niños de edad preescolar. Estudios longitudinales o de intervención que utilicen este test deberían tener en cuenta que cambios en el rendimiento de este test de más de 2 vueltas podría ser debido a la variabilidad de la medida, mientras cambios mayores podrían ser atribuidos a la intervención o a cambios asociados con el crecimiento.

4. En general, la batería PREFIT es viable y fiable para evaluar la condición física en preescolares. Sin embargo, la fiabilidad del test de salto horizontal con pies juntos mostró resultados inconsistentes y requiere de más estudios que examinen

su fiabilidad. El test de equilibrio a una pierna mostró una fiabilidad baja, no apoyando por tanto su uso en niños de 3 a 5 años.

5. La presente Tesis proporciona por primera vez valores de referencia de condición física específicos para sexo y edad en preescolares españoles. Este estudio en preescolares de 3 a 5 años amplia aquellos ya existentes en niños[22-24] y adolescentes[22,23,25]. Las diferencias entre sexo son detectables ya a los 3 años e incrementan con la edad. Estos resultados ayudarán a profesionales de la salud, del deporte y de la educación para identificar a preescolares con un alto/bajo nivel de condición física y examinar los cambios a lo largo del tiempo incluyendo aquellos obtenidos debidos a los efectos de la intervención.

6. Existe una alta prevalencia de sobrepeso/obesidad en niños preescolares españoles (21% según la Federación Mundial de Obesidad y 35% según la Organización Mundial de la Salud). Particularmente es que el 6% de todos los niños evaluados (n = 3.178) eran obesos y un 2,5% tenían obesidad severa/mórbida (1,3% obesidad mórbida). Así mismo, la prevalencia de preescolares con bajo peso fue ligeramente menor que la de obesos (i.e., 3-5% vs. 6%, respectivamente). Además, proporcionamos valores de referencia antropométricos específicos para sexo y edad en preescolares de 3 a 5 años que nos permitirá una mejor interpretación de la evaluación antropométrica en niños preescolares españoles.

7. Observamos una mayor prevalencia de sobrepeso/obesidad en niños españoles en comparación con niños suecos de 4 años de edad, sugiriendo que las diferencias en prevalencia de sobrepeso/obesidad entre el sur y norte de Europa podrían estar presentes ya desde edad preescolar, no observando diferencias consistentes en la condición física.

8. Además de la capacidad cardiorrespiratoria, otros componentes de la condición física tales como la fuerza muscular y la velocidad-agilidad podrían contribuir a un mejor rendimiento académico. No se observó ninguna relación entre la actividad física y el rendimiento académico.

9. Los niños con sobrepeso/obesidad y un fenotipo metabólicamente sano presentaron mayor volumen de materia gris y mayor volumen total cerebral en comparación con aquellos niños con sobrepeso/obesidad que fueron categorizados como metabólicamente no sanos, lo que a su vez se relacionó con un mayor rendimiento académico. Nuestros resultados también sugieren que la capacidad cardiorrespiratoria podría jugar un rol importante en las diferencias en el cerebro observadas entre fenotipos metabólicos.

CONCLUSIONS

General conclusions

The results of the present International Doctoral Thesis provide new methods for a better assessment of physical fitness in preschool children, providing reference standards to interpret it. Likewise, this Thesis concludes that not only cardiorespiratory fitness is associated with academic achievement but also muscular strength and speed-agility. Moreover, the results of the present Thesis suggest that overweight/obese children with a metabolically healthy phenotype present a higher gray matter volume and higher total brain volume compared to those with an unhealthy phenotype, which was associated with a better academic achievement. These results will lead to future prospective and intervention investigations on the physical and brain health at childhood and later in life.

Specific conclusions

The main conclusions of the present International Doctoral Thesis are:

1. The systematic review conduced propose the following fitness tests for the PREFIT battery:the 20m shuttle-run test to assess cardiorespiratory fitness, the handgrip-strength and the standing long-jump tests to assess musculoskeletal fitness, and the 4x10m shuttle run and the one-leg-stance tests to assess motor fitness, i.e., speed/agility and balance, respectively.

2. This Thesis supports the use of the same dynamometer (either analog or digital) for the measurement of muscular fitness in repeated measures to minimize the systematic error to 0.3kg or less. If different instruments, models or dynamometers of different ages are used in repeated measures, the systematic error might increase up to 1kg. These findings will allow distinguishing between what is a clinically meaningful change from what is just the variability error of the dynamometers.

3. The PREFIT 20m shuttle run test is feasible, maximum and reliable in preschool children. Future longitudinal or intervention studies using this test should take into account that changes in the test performance of 2 laps may be due to the variability of the measure, while wider changes would be attributable to the intervention or changes associated with growth.

4. Overall, the PREFIT battery is feasible and reliable to assess physical fitness in preschoolers yet standing long jump showed mixed findings and requires further studies to assess its reliability. The one-leg stance test showed poor reliability in our study and thus, its use in 3 to 5 years-old would not be recommended.

5. The present Thesis provides for first time sex- and age-specific physical fitness reference standards in Spanish preschool children. This study in children from 3 to 5 years old extends to preschoolers the already existing reference standards in older children[22-24] and adolescents[22,23,25]. Sexual dimorphism is detectable already at the age of 3 and increases with age. These findings will help health, sport, and school professionals to identify preschool children with a high/very low fitness level and to examine changes over time, including those obtained due to intervention effects.

6. A high prevalence of overweight/obesity in preschool Spanish children (WOF: 21% and WHO: 35%) was observed, being 6% of them obese and2.5% severe/morbid obese (1.3% morbid obesity). Likewise, the prevalence of underweight was slightly lower than obesity in preschoolers (i.e., 3-5% vs. 6% respectively). Moreover, we provide sex- and age-specific anthropometric reference standards in preschoolers between the ages of 3 and 5 years old which will allow a better interpretation of anthropometric assessment in Spanish preschool children.

7. We observed a significantly higher prevalence of overweight/obesity in Spanish children compared to Swedish children at 4 years of age, suggesting that differences regarding the prevalence of overweight/obesity in Southern and Northern Europe may be present already at preschool ages. However, differences in fitness were not consistent.

8. Besides of cardiorespiratory fitness, other physical fitness components such as muscular and speed-agility fitness may contribute to better academic achievement. No relationships were observed for physical activity and academic achievement.

9. MHO children have greater gray matter and total brain volume than their MUO peers, which in turn was related to better academic achievement. Our data also suggest that cardiorespiratory fitness might play an important role in the differences in brain observed between metabolic phenotypes.

REFERENCES

1. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Reports. 1985;100(2):126-131.
2. Ruiz JR, Castro-Pinero J, Artero EG, et al. Predictive validity of health-related fitness in youth: a systematic review. British Journal of Sports Medicine. 2009;43(12):909-923.
3. Ortega FB, Ruiz JR, Castillo MJ, Sjostrom M. Physical fitness in childhood and adolescence: a powerful marker of health. International Journal of Obesity (2005). 2008;32(1):1-11.
4. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. The International Journal of Behavioral Nutrition and Physical Activity. 2017;14(1):75.
5. World Health Organization. Physical activity strategy for the WHO European region 2016-2025. 2016. URL http://www.euro.who.int/en/publications/abstracts/physical-activity-strategy-for-the-who-european-region-20162025 (accessed 23 December 2017).
6. Truelove S, Bruijns BA, Vanderloo LM, O'Brien KT, Johnson AM, Tucker P. Physical activity and sedentary time during childcare outdoor play sessions: A systematic review and meta-analysis. Preventive Medicine. 2018;108:74-85.
7. Flodmark CE, Lissau I, Moreno LA, Pietrobelli A, Widhalm K. New insights into the field of children and adolescents' obesity: the European perspective. International Journal of Obesity and Related Metabolic Disorders. 2004;28(10):1189-1196.
8. Gutin B, Johnson MH, Humphries MC, et al. Relationship of visceral adiposity to cardiovascular disease risk factors in black and white teens. Obesity. 2007;15(4):1029-1035.
9. de Koning L, Denhoff E, Kellogg MD, de Ferranti SD. Associations of total and abdominal adiposity with risk marker patterns in children at high-risk for cardiovascular disease. BMC Obesity. 2015;2:15.
10. Donnelly JE, Hillman CH, Castelli D, et al. Physical Activity, Fitness, Cognitive Function, and Academic Achievement in Children: A Systematic Review. Medicine and Science in Sports and Exercise. 2016;48(6):1197-1222.
11. Alvarez-Bueno C, Pesce C, Cavero-Redondo I, Sanchez-Lopez M, Garrido-Miguel M, Martinez-Vizcaino V. Academic Achievement and Physical Activity: A Meta-analysis. Pediatrics. 2017;140(6).
12. Esteban-Cornejo I, Cadenas-Sanchez C, Contreras-Rodriguez O, et al. A whole brain volumetric approach in overweight/obese children: Examining the association with different physical fitness components and academic performance. The ActiveBrains project. NeuroImage. 2017;159:346-354.
13. Chaddock L, Erickson KI, Prakash RS, et al. A neuroimaging investigation of the association between aerobic fitness, hippocampal volume, and memory performance in preadolescent children. Brain Research. 2010;1358:172-183.
14. Schaeffer DJ, Krafft CE, Schwarz NF, et al. An 8-month exercise intervention alters frontotemporal white matter integrity in overweight children. Psychophysiology. 2014;51(8):728-733.
15. Chaddock L, Erickson KI, Prakash RS, et al. A functional MRI investigation of the association between childhood aerobic fitness and neurocognitive control. Biological Psychology. 2012;89(1):260-268.
16. Krafft CE, Schaeffer DJ, Schwarz NF, et al. Improved frontoparietal white matter integrity in overweight children is associated with attendance at an after-school exercise program. Developmental Neuroscience. 2014;36(1):1-9.
17. Artero EG, Espana-Romero V, Castro-Pinero J, et al. Reliability of field-based fitness tests in youth. International Journal of Sports Medicine. 2011;32(3):159-169.
18. Castro-Pinero J, Artero EG, Espana-Romero V, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. British Journal of Sports Medicine. 2010;44(13):934-943.

19. Ruiz JR, Castro-Pinero J, Espana-Romero V, 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):518-524.
20. IOM (Institute of Medicine). Fitness measures and health outcomes in youth. Washington, DC: The National Academies Press; 2012.
21. Ortega FB, Ruiz J. Fitness in Youth: Methodological Issues and Understanding of Its Clinical Value. American Journal of Lifestyle Medicine. 2015.
22. Ruiz JR, Cavero-Redondo I, Ortega FB, Welk GJ, Andersen LB, Martinez-Vizcaino V. Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents; what level of fitness should raise a red flag? A systematic review and meta-analysis. British Journal of Sports Medicine.2016.
23. Tomkinson GR, Lang JJ, Tremblay MS, et al. International normative 20 m shuttle run values from 1 142 026 children and youth representing 50 countries. British Journal of Sports Medicine. 2016.
24. De Miguel-Etayo P, Gracia-Marco L, Ortega FB, et al. Physical fitness reference standards in European children: the IDEFICS study. International Journal of Obesity (2005). 2014;38 Suppl 2:S57-66.
25. Ortega FB, Artero EG, Ruiz JR, et al. Physical fitness levels among European adolescents: the HELENA study. British Journal of Sports Medicine. 2011;45(1):20-29.
26. Tomkinson GR, Carver KD, Atkinson F, et al. European normative values for physical fitness in children and adolescents aged 9-17 years: results from 2 779 165 Eurofit performances representing 30 countries. British Journal of Sports Medicine. 2017.
27. Latorre-Roman PA, Garcia-Pinillos F, Mora-Lopez D. Reference Values of Standing Long Jump in Preschool Children: A Population-Based Study. Pediatric Exercise Science. 2016.
28. Duran P, Merker A, Briceno G, et al. Colombian reference growth curves for height, weight, body mass index and head circumference. Acta Paediatrica. 2016;105(3):e116-125.
29. Wells JC, Williams JE, Chomtho S, et al. Body-composition reference data for simple and reference techniques and a 4-component model: a new UK reference child. The American Journal of Clinical Nutrition. 2012;96(6):1316-1326.
30. Moreno LA, Mesana MI, Gonzalez-Gross M, et al. Body fat distribution reference standards in Spanish adolescents: the AVENA Study. International Journal of Obesity (2005). 2007;31(12):1798-1805.
31. Nagy P, Kovacs E, Moreno LA, et al. Percentile reference values for anthropometric body composition indices in European children from the IDEFICS study. International Journal of Obesity (2005). 2014;38 Suppl 2:S15-25.
32. Ahrens W, Moreno LA, Marild S, et al. Metabolic syndrome in young children: definitions and results of the IDEFICS study. International Journal of Obesity (2005). 2014;38 Suppl 2:S4-14.
33. World Health Organization. Global Strategy on Diet, Physical Activity and Health. 2014. URL http:// www.who.int/dietphysicalactivity/childhood/en/ (accessed 23 December 2017).
34. Ahrens W, Pigeot I, Pohlabeln H, et al. Prevalence of overweight and obesity in European children below the age of 10. International Journal of Obesity (2005). 2014;38 Suppl 2:S99-107.

35. Bluher S, Meigen C, Gausche R, et al. Age-specific stabilization in obesity prevalence in German children: a cross-sectional study from 1999 to 2008. International Journal of Pediatric Obesity. 2011;6(2-2):e199-206.
36. Olds T, Maher C, Zumin S, et al. Evidence that the prevalence of childhood overweight is plateauing: data from nine countries. International Journal of Pediatric Obesity. 2011;6(5-6):342-360.
37. Ortega FB, Ruiz JR, Labayen I. Trends in the prevalence of morbid obesity in Australian children and adolescents from 1985 to 2008: what do we know about? International Journal of Obesity (2005). 2011;35(10):1331; author reply 1332-1333.
38. 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. Lancet. 2017;390(10113):2627-2642.
39. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384(9945):766-781.
40. Bervoets L, Massa G. Defining morbid obesity in children based on BMI 40 at age 18 using the extended international (IOTF) cut-offs. Pediatric Obesity. 2014;9(5):e94-98.
41. Wake M, Clifford SA, Patton GC, et al. Morbidity patterns among the underweight, overweight and obese between 2 and 18 years: population-based cross-sectional analyses. International Journal of Obesity (2005). 2013;37(1):86-93.
42. de Onis M, Blossner M, Borghi E, Frongillo EA, Morris R. Estimates of global prevalence of childhood underweight in 1990 and 2015. JAMA. 2004;291(21):2600-2606.
43. Rito A, Wijnhoven TM, Rutter H, et al. Prevalence of obesity among Portuguese children (6-8 years old) using three definition criteria: COSI Portugal, 2008. Pediatric Obesity. 2012;7(6):413-422.
44. de Ruiter I, Olmedo-Requena R, Sanchez-Cruz JJ, Jimenez-Moleon JJ. Trends in Child Obesity and Underweight in Spain by Birth Year and Age, 1983 to 2011. Revista Espanola de Cardiologia. 2017.
45. Ortega FB, Ruiz JR, Labayen I, et al. Health inequalities in urban adolescents: role of physical activity, diet, and genetics. Pediatrics. 2014;133(4):e884-895.
46. Lang JJ, Tremblay MS, Leger L, Olds T, Tomkinson GR. International variability in 20 m shuttle run performance in children and youth: who are the fittest from a 50-country comparison? A systematic literature review with pooling of aggregate results. British Journal of Sports Medicine. 2016.
47. Ortega FB, Ruiz JR, Castillo MJ, et al. [Low level of physical fitness in Spanish adolescents. Relevance for future cardiovascular health (AVENA study)]. Revista Espanola de Cardiologia. 2005;58(8):898-909.
48. World Obesity Federation. Overweight and Obesity in Pre (aged 5–10 years) and Post (14–17 yrs) Adolescent children. 2014. URL http://www.worldobesity.org/site_media/library/resource_images/Pre_and_post_adolescent_chart_June_14.pdf (accessed 27 December 2017).
49. van Stralen MM, te Velde SJ, van Nassau F, et al. Weight status of European preschool children and associations with family demographics and energy balance-related behaviours: a pooled analysis of six European studies. Obesity Reviews. 2012;13 Suppl 1:29-41.
50. Diamond A. Executive functions. Annual Review of Psychology. 2013;64:135-168.

51. Chaddock-Heyman L, Erickson KI, Kienzler C, et al. The role of aerobic fitness in cortical thickness and mathematics achievement in preadolescent children. PloS one. 2015;10(8):e0134115.
52. Torrijos-Nino C, Martinez-Vizcaino V, Pardo-Guijarro MJ, Garcia-Prieto JC, Arias-Palencia NM, Sanchez-Lopez M. Physical fitness, obesity, and academic achievement in schoolchildren. The Journal of Pediatrics. 2014;165(1):104-109.
53. Aguilar MM, Vergara FA, Velasquez EJ, Marina R, Garcia-Hermoso A. Screen time impairs the relationship between physical fitness and academic attainment in children. Jornal de Pediatria. 2015;91(4):339-345.
54. Esteban-Cornejo I, Tejero-Gonzalez CM, Martinez-Gomez D, et al. Independent and combined influence of the components of physical fitness on academic performance in youth. The Journal of Pediatrics. 2014;165(2):306-312 e302.
55. Kao SC, Westfall DR, Parks AC, Pontifex MB, Hillman CH. Muscular and Aerobic Fitness, Working Memory, and Academic Achievement in Children. Medicine and Science in Sports and Exercise. 2017;49(3):500-508.
56. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. The International Journal of Behavioral Nutrition and Physical Activity. 2010;7:40.
57. Esteban-Cornejo I, Tejero-Gonzalez CM, Sallis JF, Veiga OL. Physical activity and cognition in adolescents: A systematic review. Journal of Science and Medicine in Sport. 2015;18(5):534-539.
58. Fedewa AL, Ahn S. The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: a meta-analysis. Research Quarterly for Exercise and Sport. 2011;82(3):521-535.
59. Alvarez-Bueno C, Pesce C, Cavero-Redondo I, Sanchez-Lopez M, Pardo-Guijarro MJ, Martinez-Vizcaino V. Association of physical activity with cognition, metacognition and academic performance in children and adolescents: a protocol for systematic review and meta-analysis. BMJ Open. 2016;6(6):e011065.
60. Syvaoja HJ, Kantomaa MT, Ahonen T, Hakonen H, Kankaanpaa A, Tammelin TH. Physical activity, sedentary behavior, and academic performance in Finnish children. Medicine and Science in Sports and Exercise. 2013;45(11):2098-2104.
61. Dwyer T, Sallis JF, Blizzard L, Lazarus R, Dean K. Relation of Academic Performance to Physical Activity and Fitness in Children. Pediatr. Exerc. Sci. 2001;13(3):225-237.
62. Esteban-Cornejo I, Tejero-Gonzalez CM, Martinez-Gomez D, et al. Objectively measured physical activity has a negative but weak association with academic performance in children and adolescents. Acta Paediatrica. 2014;103(11):e501-506.
63. LeBlanc MM, Martin CK, Han H, et al. Adiposity and physical activity are not related to academic achievement in school-aged children. Journal of Developmental and Behavioral Pediatrics. 2012;33(6):486-494.
64. Pindus DM, Drollette ES, Scudder MR, et al. Moderate-to-Vigorous Physical Activity, Indices of Cognitive Control, and Academic Achievement in Preadolescents. The Journal of Pediatrics. 2016;173:136-142.
65. Hansen DM, Herrmann SD, Lambourne K, Lee J, Donnelly JE. Linear/nonlinear relations of activity and fitness with children's academic achievement. Medicine and Science in Sports and Exercise. 2014;46(12):2279-2285.
66. Davis C, Cooper S. Fitness, fatness, cognition, behavior, and academic achievement among overweight children: do cross-sectional associations correspond to exercise trial outcomes? Preventive Medicine. 2011;52 Suppl 1:S65-69.

67. Ortega FB, Lavie CJ, Blair SN. Obesity and Cardiovascular Disease. Circulation Research. 2016;118(11):1752-1770.
68. Umer A, Kelley GA, Cottrell LE, Giacobbi P, Jr., Innes KE, Lilly CL. Childhood obesity and adult cardiovascular disease risk factors: a systematic review with meta-analysis. BMC Public Health. 2017;17(1):683.
69. Medic N, Ziauddeen H, Ersche KD, et al. Increased body mass index is associated with specific regional alterations in brain structure. International Journal of Obesity (2005). 2016;40(7):1177-1182.
70. Brooks SJ, Benedict C, Burgos J, et al. Late-life obesity is associated with smaller global and regional gray matter volumes: a voxel-based morphometric study. International Journal of Obesity (2005). 2013;37(2):230-236.
71. Kamijo K, Khan NA, Pontifex MB, et al. The relation of adiposity to cognitive control and scholastic achievement in preadolescent children. Obesity. 2012;20(12):2406-2411.
72. O'Brien PD, Hinder LM, Callaghan BC, Feldman EL. Neurological consequences of obesity. The Lancet Neurology. 2017;16(6):465-477.
73. Jolliffe CJ, Janssen I. Development of age-specific adolescent metabolic syndrome criteria that are linked to the Adult Treatment Panel III and International Diabetes Federation criteria. Journal of the American College of Cardiology. 2007;49(8):891-898.
74. Fan J, Song Y, Chen Y, Hui R, Zhang W. Combined effect of obesity and cardio-metabolic abnormality on the risk of cardiovascular disease: a meta-analysis of prospective cohort studies. International Journal of Cardiology. 2013;168(5):4761-4768.
75. Kramer CK, Zinman B, Retnakaran R. Are metabolically healthy overweight and obesity benign conditions?: A systematic review and meta-analysis. Annals of Internal Medicine. 2013;159(11):758-769.
76. Roberson LL, Aneni EC, Maziak W, et al. Beyond BMI: The "Metabolically healthy obese" phenotype & its association with clinical/subclinical cardiovascular disease and all-cause mortality -- a systematic review. BMC Public Health. 2014;14:14.
77. Zheng R, Zhou D, Zhu Y. The long-term prognosis of cardiovascular disease and all-cause mortality for metabolically healthy obesity: a systematic review and meta-analysis. Journal of Epidemiology and Community Health. 2016;70(10):1024-1031.
78. Eckel N, Meidtner K, Kalle-Uhlmann T, Stefan N, Schulze MB. Metabolically healthy obesity and cardiovascular events: A systematic review and meta-analysis. European Journal of Preventive Cardiology. 2016;23(9):956-966.
79. Ortega FB, Cadenas-Sanchez C, Sui X, Blair SN, Lavie CJ. Role of Fitness in the Metabolically Healthy But Obese Phenotype: A Review and Update. Progress in Cardiovascular Diseases. 2015.
80. Lavie CJ, Kokkinos PF, Ortega FB. Impact of Physical Activity and Fitness in Metabolically Health Obesity. Journal of the American College of Cardiology. 2018.
81. Bluher S, Schwarz P. Metabolically healthy obesity from childhood to adulthood - Does weight status alone matter? Metabolism. 2014;63(9):1084-1092.
82. Cadenas-Sanchez C, Ruiz JR, Labayen I, et al. Prevalence of Metabolically Healthy but Overweight/Obese Phenotype and Its Association With Sedentary Time, Physical Activity, and Fitness. Journal of Adolescent Health. 2017.
83. Yau PL, Castro MG, Tagani A, Tsui WH, Convit A. Obesity and metabolic syndrome and functional and structural brain impairments in adolescence. Pediatrics. 2012;130(4):e856-864.

84. Ortega FB, Lee DC, Katzmarzyk PT, et al. The intriguing metabolically healthy but obese phenotype: cardiovascular prognosis and role of fitness. European Heart Journal. 2013;34(5):389-397.
85. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatric Obesity. 2012;7(4):284-294.
86. Organization WH. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization; 2006. (Accessed 10 october 2017).
87. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatrica. 2006;450:76-85.
88. 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.
89. Sanchez-Delgado G, Cadenas-Sanchez C, Mora-Gonzalez J, et al. Assessment of handgrip strength in preschool children aged 3 to 5 years. The Journal of Hand Surgery, European Volume. 2015.
90. Vicente-Rodriguez G, Rey-Lopez JP, Ruiz JR, et al. Interrater reliability and time measurement validity of speed-agility field tests in adolescents. Journal of Strength and Conditioning Research. 2011;25(7):2059-2063.
91. Mirwald RL, Baxter-Jones AD, Bailey DA, Beunen GP. An assessment of maturity from anthropometric measurements. Medicine & Science in Sports & Exercise. 2002;34(4):689-694.
92. Davis CL, Pollock NK, Waller JL, et al. Exercise dose and diabetes risk in overweight and obese children: a randomized controlled trial. JAMA. 2012;308(11):1103-1112.
93. Faigenbaum AD, Milliken LA, Westcott WL. Maximal strength testing in healthy children. Journal of Strength and Conditioning Research. 2003;17(1):162-166.
94. van Hees VT, Gorzelniak L, Dean Leon EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PloS one. 2013;8(4):e61691.
95. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Medicine and Science in Sports and Exercise. 2014;46(9):1816-1824.
96. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. Scandinavian Journal of Medicine & Science in Sports. 2016.
97. McGrew K, Woodcock R. Woodcock-Johnson III: Technical Manual. Itasca, IL: Riverside Publishing Company; 2001.
98. Ashburner J, Friston KJ. Unified segmentation. NeuroImage. 2005;26(3):839-851.
99. Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage. 2007;38(1):95-113.
100. Ashburner J, Friston KJ. Voxel-based morphometry--the methods. NeuroImage. 2000;11(6 Pt 1):805-821.
101. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-310.
102. Mahon AD, Marjerrison AD, Lee JD, Woodruff ME, Hanna LE. Evaluating the prediction of maximal heart rate in children and adolescents. Research Quarterly for Exercise and Sport. 2010;81(4):466-471.
103. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45(1):255-268.

104. Cohen J. A power primer. Psychological Bulletin. 1992;112(1):155-159.
105. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2005;54:507–554. .
106. van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. Statistics in Medicine. 2001;20(8):1259-1277.
107. Intemann T PH, Herrmann D, Ahrens W, Pigeot I. . Estimating age-and height-specific percentile curves for children using GAMLSS in the IDEFICS study. In: Wilhelm AF, Kestler HA, Herausgeber. Analysis of large and complex data. Studies in classification, data analysis, and knowledge organization. Baier, D.; Critchley, F.; Decker, R.; Diday, E.; Greenacre, M.; Lauro, C. N.; Meulman, J.; Monari, P.; Nishisato, S.; Ohsumi, N.; Opitz, O.; Ritter, G.; Schader, M. Heidelberg: Springer International Publishing; 2016:385-394.
108. McCarthy HD, Ashwell M. A study of central fatness using waist-to-height ratios in UK children and adolescents over two decades supports the simple message--'keep your waist circumference to less than half your height'. International Journal of Obesity (2005). 2006;30(6):988-992.
109. Kerr J, Marinac CR, Ellis K, et al. Comparison of Accelerometry Methods for Estimating Physical Activity. Medicine and Science in Sports and Exercise. 2017;49(3):617-624.
110. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. Sports Medicine. 2017.
111. Song X-W, Dong Z-Y, Long X-Y, et al. REST: A Toolkit for Resting-State Functional Magnetic Resonance Imaging Data Processing. PloS one. 2011;6(9):e25031.
112. Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage. 2004;22(2):676-687.
113. Byrne BM. Structural equation modeling with AMOS, basic concepts, applications, and programming. . Hillsdale, New Jersey: Lawrence Eribaum Associates; 2001.
114. Niederer I, Kriemler S, Zahner L, et al. BMI group-related differences in physical fitness and physical activity in preschool-age children: a cross-sectional analysis. Research Quarterly for Exercise and Sport. Mar 2012;83(1):12-19.
115. Reeves L, Broeder CE, Kennedy-Honeycutt L, East C, Matney L. Relationship of fitness and gross motor skills for five- to six-yr.-old children. Perceptual and Motor Skills. 1999;89(3 Pt 1):739-747.
116. Rikli RE, Petray C, Baumgartner TA. The reliability of distance run tests for children in grades K-4. Research Quarterly for Exercise and Sport. 1992;63(3):270-276.
117. Benefice E, Fouere T, Malina RM. Early nutritional history and motor performance of Senegalese children, 4-6 years of age. Annals of Human Biology. 1999;26(5):443-455.
118. Oja L, Jürimäe T. Assessment of motor ability of 4- and 5-year-old children. American Journal of Human Biology. 1997;9(5):659-664.
119. Molenaar HM, Zuidam JM, Selles RW, Stam HJ, Hovius SE. Age-specific reliability of two grip-strength dynamometers when used by children. The Journal of Bone and Joint Surgery. American volume. 2008;90(5):1053-1059.
120. Krombholz H. Testbatterie zur Erfassung motorischer Leistungen im Vorschulalter MoTB 3-7. Beschreibung, Gütekriterien, Normwerte und ausgewählte Ergebnisse [A test battery for measuring motor performance of preschool children MoTB 3-7.

Description, quality criteria, norms, and selected results]. 2011. http://psydok.sulb.uni-saarland.de/volltexte/2011/2711/. (Accessed 3rd May 2017).
121. Molenaar HM, Selles RW, Schreuders TA, Hovius SE, Stam HJ. Reliability of hand strength measurements using the Rotterdam Intrinsic Hand Myometer in children. The Journal of Hand Surgery. 2008;33(10):1796-1801.
122. Rose KJ, Burns J, Ryan MM, Ouvrier RA, North KN. Reliability of quantifying foot and ankle muscle strength in very young children. Muscle Nerve. 2008;37(5):626-631.
123. Koslow RE. Sit and Reach Flexibility Measures for Boys and Girls Aged 3 through 8 Years. Perceptual and Motor Skills. 1987;64(3):1103-1106.
124. Deitz JC, Kartin D, Kopp K. Review of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2). Physical Occupational Therapy in Pediatrics. 2007;27(4):87-102.
125. Nguyen T, Obeid J, Timmons BW. Reliability of fitness measures in 3- to 5-year-old children. Pediatric Exercise Science. 2011;23(2):250-260.
126. Hua J, Gu G, Meng W, Wu Z. Age band 1 of the Movement Assessment Battery for Children-Second Edition: exploring its usefulness in mainland China. Research in Developmental Disabilities. 2013;34(2):801-808.
127. Kasuga K, Demura S, Aoki H, Shin S, Sugiura H, Uchida Y. Sex and age-level differences of walking time in preschool children on an obstacle frame. Journal of Physiological Anthropology. 2012;31:8.
128. De Kegel A, Baetens T, Peersman W, Maes L, Dhooge I, Van Waelvelde H. Ghent developmental balance test: a new tool to evaluate balance performance in toddlers and preschool children. Physical Therapy. 2012;92(6):841-852.
129. Smits-Engelsman BC, Niemeijer AS, van Waelvelde H. Is the Movement Assessment Battery for Children-2nd edition a reliable instrument to measure motor performance in 3 year old children? Research in Developmental Disabilities. 2011;32(4):1370-1377.
130. Ellinoudis T, Evaggelinou C, Kourtessis T, Konstantinidou Z, Venetsanou F, Kambas A. Reliability and validity of age band 1 of the Movement Assessment Battery for Children--second edition. Research in Developmental Disabilities. 2011;32(3):1046-1051.
131. Sun SH, Zhu YC, Shih CL, Lin CH, Wu SK. Development and initial validation of the Preschooler Gross Motor Quality Scale. Research in Developmental Disabilities. 2010;31(6):1187-1196.
132. Kroes M, Vissers YL, Sleijpen FA, et al. Reliability and validity of a qualitative and quantitative motor test for 5- to 6-year-old children. European Journal of Paediatric Neurology. 2004;8(3):135-143.
133. Broadstone BJ, Westcott SL, Deitz JC. Test-retest reliability of two tiltboard tests in children. Physical Therapy. 1993;73(9):618-625.
134. Greenspan KA. Relationships among age, sex, and lateral dominance for 3- to 6-year-old children performing unilateral stance. Perceptual and Motor Skills. 1990;71(2):615-623.
135. Kim CI, Han DW, Park IH. Reliability and validity of the Test of Gross Motor Development-II in Korean preschool children: applying AHP. Research in Developmental Disabilities. 2014;35(4):800-807.
136. Cools W, Martelaer KD, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. Journal of Sports Science & Medicine. 2009;8(2):154-168.

137. Niederer I, Kriemler S, Zahner L, et al. BMI group-related differences in physical fitness and physical activity in preschool-age children: a cross-sectional analysis. Research Quaterly for Exercise and Sport. 2012;83(1):12-19.

138. Reeves L, Broeder CE, Kennedy-Honeycutt L, East C, Matney L. Relationship of fitness and gross motor skills for five- to six-yr.-old children. Perceptual and Motor Skills. 1999;89(3 Pt 1):739-747.

139. Benefice E, Fouere T, Malina RM. Early nutritional history and motor performance of Senegalese children, 4-6 years of age. Ann. Hum. Biol. Sep-Oct 1999;26(5):443-455.

140. Oja L, Jürimäe T. Assessment of motor ability of 4- and 5-year-old children. Annals of Human Biology. 1999;26(5):443-455.

141. Molenaar HM, Zuidam JM, Selles RW, Stam HJ, Hovius SE. Age-specific reliability of two grip-strength dynamometers when used by children. Journal of Bone Joint Surgery. American. 2008;90(5):1053-1059.

142. Hua J, Gu G, Meng W, Wu Z. Age band 1 of the Movement Assessment Battery for Children-Second Edition: exploring its usefulness in mainland China. Research in Developmental Disabilities. 2013;34(2):801-808.

143. Kakebeeke TH, Caflisch J, Chaouch A, Rousson V, Largo RH, Jenni OG. Neuromotor development in children. Part 3: motor performance in 3- to 5-year-olds. Developmental Medicine and Child Neurology. 2013;55(3):248-256.

144. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-174.

145. De Kegel A, Baetens T, Peersman W, Maes L, Dhooge I, Van Waelvelde H. Ghent developmental balance test: a new tool to evaluate balance performance in toddlers and preschool children. Physical Therapy. 2012;92(6):841-852.

146. Ellinoudis T, Evaggelinou C, Kourtessis T, Konstantinidou Z, Venetsanou F, Kambas A. Reliability and validity of age band 1 of the Movement Assessment Battery for Children--second edition. Research in Developmental Disabilities. 2011;32(3):1046-1051.

147. Smits-Engelsman BC, Niemeijer AS, van Waelvelde H. Is the Movement Assessment Battery for Children-2nd edition a reliable instrument to measure motor performance in 3 year old children? Research in Developmental Disabilities. 2011;32(4):1370-1377.

148. Nguyen T, Obeid J, Timmons BW. Reliability of fitness measures in 3- to 5-year-old children. Pediatric Exercise Science. 2011;23(2):250-260.

149. Sun SH, Zhu YC, Shih CL, Lin CH, Wu SK. Development and initial validation of the Preschooler Gross Motor Quality Scale. Research in Developmental Disabilities. 2010;31(6):1187-1196.

150. Lee-Valkov PM, Aaron DH, Eladoumikdachi F, Thornby J, Netscher DT. Measuring normal hand dexterity values in normal 3-, 4-, and 5-year-old children and their relationship with grip and pinch strength. Journal of Hand Therapy. 2003;16(1):22-28.

151. Butterfield SA, Lehnhard RA, Coladarci T. Age, sex, and body mass index in performance of selected locomotor and fitness tasks by children in grades K-2. Perceptual and Motor Skills. 2002;94(1):80-86.

152. Butterfield SA, Lehnhard RA, Loovis EM, Coladarci T, Saucier D. Grip strength performances by 5- to 19-year-olds. Perceptual and Motor Skills. 2009;109(2):362-370.

153. Haishi K, Okuzumi H, Kokubun M, Komatsu A, Kitajima Y, Hosobuchi T. Verbal regulation of grip force in preschoolers. Perceptual and Motor Skills. 2009;108(2):540-548.

154. Molenaar HM, Selles RW, Zuidam JM, Willemsen SP, Stam HJ, Hovius SE. Growth diagrams for grip strength in children. Clinical Orthopaedics and Related Research. 2010;468(1):217-223.
155. Yoshizawa S, Ishizaki T, Honda H. Physical fitness of children aged 5 and 6 years. Journal of Human Ergology. 1977;6(1):41-51.
156. Fjortoft I, Pedersen AV, Sigmundsson H, Vereijken B. Measuring physical fitness in children who are 5 to 12 years old with a test battery that is functional and easy to administer. Physical Therapy. 2011;91(7):1087-1095.
157. Duger T, Bumin G, Uyanik M, Aki E, Kayihan H. The assessment of Bruininks-Oseretsky test of motor proficiency in children. Pediatric Rehabilitation. 1999;3(3):125-131.
158. Venetsanou F, Kambas A, Aggeloussis N, Fatouros I, Taxildaris K. Motor assessment of preschool aged children: A preliminary investigation of the validity of the Bruininks-Oseretsky test of motor proficiency - short form. Human Movement Science. 2009;28(4):543-550.
159. Woll A, Worth A, Mundermann A, Holling H, Jekauc D, Bos K. Age- and sex-dependent disparity in physical fitness between obese and normal weight children and adolescents. Journal of Sports Medicine and Physycal Fitness. 2013;53(1):48-55.
160. Woll A, Kurth BM, Opper E, Worth A, Bos K. The 'Motorik-Modul' (MoMo): physical fitness and physical activity in German children and adolescents. European Journal of Pediatrics. 2011;170(9):1129-1142.
161. Krombholz H. Physical performance in relation to age, sex, birth order, social class, and sports activities of preschool children. Perceptual amd Motor Skills. 2006;102(2):477-484.
162. Kambas A, Aggeloussis N. Construct validity of the Bruininks-Oseretsky Test of Motor Proficiency-short form for a sample of Greek preschool and primary school children. Perceptual amd Motor Skills. 2006;102(1):65-72.
163. Welon Z, Sekita B. Physical fitness, body size, and body build in pre-school children. Studies in Physical and Anthropometry. 1975(2):25-32.
164. Kondric M, Trajkovski B, Strbad M, Foretic N, Zenic N. Anthropometric influence on physical fitness among preschool children: gender-specific linear and curvilinear regression models. Collegium Antropologicum. 2013;37(4):1245-1252.
165. Atwater SW, Crowe TK, Deitz JC, Richardson PK. Interrater and test-retest reliability of two pediatric balance tests. Physical Therapy. 1990;70(2):79-87.
166. An MH, Yi CH, Jeon HS, Park SY. Age-related changes of single-limb standing balance in children with and without deafness. International Journal of Pediatrics Otorhinolaryngology. 2009;73(11):1539-1544.
167. Nemet D, Geva D, Eliakim A. Health promotion intervention in low socioeconomic kindergarten children. Journal of Pediatrics. 2011;158(5):796-801 e791.
168. Puder JJ, Marques-Vidal P, Schindler C, et al. Effect of multidimensional lifestyle intervention on fitness and adiposity in predominantly migrant preschool children (Ballabeina): cluster randomised controlled trial. BMJ. 2011;343:d6195.
169. Booth ML, Macaskill P, Phongsavan P, McLellan L, Okely T. Methods of the NSW Schools Fitness and Physical Activity Survey, 1997. Journal of Science and Medicine in Sport. 1998;1(2):111-124.
170. Reuben DB, Magasi S, McCreath HE, et al. Motor assessment using the NIH Toolbox. Neurology. 2013;80(11 Suppl 3):S65-75.

171. Artero EG, Espana-Romero V, Castro-Pinero J, et al. Reliability of field-based fitness tests in youth. International Journal of Sports Medicine. 2011;32(3):159-169.
172. Castro-Pinero J, Artero EG, Espana-Romero V, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. British Journal of Sports Medicine. 2010;44(13):934-943.
173. Ruiz JR, Castro-Pinero J, Artero EG, et al. Predictive validity of health-related fitness in youth: a systematic review. British Journal of Sports Medicine. 2009;43(12):909-923.
174. Ruiz JR, Castro-Pinero J, Espana-Romero V, 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):518-524.
175. Harkonen R, Harju R, Alaranta H. Accuracy of the Jamar dynamometer. Journal of Hand Therapy. 1993;6(4):259-262.
176. Espana-Romero V, Ortega FB, Vicente-Rodriguez G, Artero EG, Rey JP, Ruiz JR. Elbow position affects handgrip strength in adolescents: validity and reliability of Jamar, DynEx, and TKK dynamometers. Journal of Strength and Conditioning Research. 2010;24(1):272-277.
177. Mullany LC, Darmstadt GL, Khatry SK, Leclerq SC, Tielsch JM. Relationship between the surrogate anthropometric measures, foot length and chest circumference and birth weight among newborns of Sarlahi, Nepal. European Journal of Clinical Nutrition. 2007;61(1):40-46.
178. Moliner-Urdiales D, Ruiz JR, Ortega FB, et al. Secular trends in health-related physical fitness in Spanish adolescents: the AVENA and HELENA studies. Journal of Science and Medicine in Sport. 2010;13(6):584-588.
179. Olds T, Tomkinson G, Leger L, Cazorla G. Worldwide variation in the performance of children and adolescents: an analysis of 109 studies of the 20-m shuttle run test in 37 countries. Journal of Sports Sciences. 2006;24(10):1025-1038.
180. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine. 1998;26(4):217-238.
181. Ortega FB, Artero EG, Ruiz JR, et al. Reliability of health-related physical fitness tests in European adolescents. The HELENA Study. International Journal of Obesity (2005). 2008;32 Suppl 5:S49-57.
182. Rhodes EC, Martin AD, Taunton JE, Donnelly M, Warren J, Elliot J. Effects of one year of resistance training on the relation between muscular strength and bone density in elderly women. British Journal of Sports Medicine. 2000;34(1):18-22.
183. Tieland M, Verdijk LB, de Groot LC, van Loon LJ. Handgrip strength does not represent an appropriate measure to evaluate changes in muscle strength during an exercise intervention program in frail older people. International Journal of Sport Nutrition and Exercise Metabolism. 2015;25(1):27-36.
184. Gulias-Gonzalez R, Sanchez-Lopez M, Olivas-Bravo A, Solera-Martinez M, Martinez-Vizcaino V. Physical fitness in Spanish schoolchildren aged 6-12 years: reference values of the battery EUROFIT and associated cardiovascular risk. The Journal of School Health. 2014;84(10):625-635.
185. Santos R, Mota J, Santos DA, Silva AM, Baptista F, Sardinha LB. Physical fitness percentiles for Portuguese children and adolescents aged 10-18 years. Journal of Sports Sciences. 2014;32(16):1510-1518.

186. Roriz De Oliveira MS, Seabra A, Freitas D, Eisenmann JC, Maia J. Physical fitness percentile charts for children aged 6-10 from Portugal. The Journal of Sports Medicine and Physical Fitness. 2014;54(6):780-792.
187. Artero EG, Espana-Romero V, Ortega FB, et al. Health-related fitness in adolescents: underweight, and not only overweight, as an influencing factor. The AVENA study. Scandinavian Journal of Medicine & Science in Sports. 2010;20(3):418-427.
188. Henriksson P, Cadenas-Sanchez C, Leppanen MH, et al. Associations of Fat Mass and Fat-Free Mass with Physical Fitness in 4-Year-Old Children: Results from the MINISTOP Trial. Nutrients. 2016;8(8).
189. Castro-Pinero J, Gonzalez-Montesinos JL, Mora J, et al. Percentile values for muscular strength field tests in children aged 6 to 17 years: influence of weight status. Journal of Strength and Conditioning Research. 2009;23(8):2295-2310.
190. de Onis M, Blossner M, Borghi E. Global prevalence and trends of overweight and obesity among preschool children. The American Journal of Clinical Nutrition. 2010;92(5):1257-1264.
191. Ruiz JR, Ortega FB, Martinez-Gomez D, et al. Objectively measured physical activity and sedentary time in European adolescents: the HELENA study. American Journal of Epidemiology. 2011;174(2):173-184.
192. Statistic NIo. Population statistics by age and sex. 2016 (accessed 16th September 2017).
193. Lakshman R, Elks CE, Ong KK. Childhood obesity. Circulation. 2012;126(14):1770-1779.
194. Twig G, Afek A, Derazne E, et al. Diabetes risk among overweight and obese metabolically healthy young adults. Diabetes Care. 2014;37(11):2989-2995.
195. Abdullah A, Wolfe R, Stoelwinder JU, et al. The number of years lived with obesity and the risk of all-cause and cause-specific mortality. International Journal of Epidemiology. 2011;40(4):985-996.
196. Gulias-Gonzalez R, Martinez-Vizcaino V, Garcia-Prieto JC, Diez-Fernandez A, Olivas-Bravo A, Sanchez-Lopez M. Excess of weight, but not underweight, is associated with poor physical fitness in children and adolescents from Castilla-La Mancha, Spain. European Journal of Pediatrics. 2014;173(6):727-735.
197. Rolland-Cachera MF, Deheeger M, Maillot M, Bellisle F. Early adiposity rebound: causes and consequences for obesity in children and adults. International Journal of Obesity (2005). 2006;30 Suppl 4:S11-17.
198. Bornhorst C, Tilling K, Russo P, et al. Associations between early body mass index trajectories and later metabolic risk factors in European children: the IDEFICS study. European Journal of Epidemiology. 2016;31(5):513-525.
199. Brambilla P, Bedogni G, Moreno LA, et al. Crossvalidation of anthropometry against magnetic resonance imaging for the assessment of visceral and subcutaneous adipose tissue in children. International Journal of Obesity (2005). 2006;30(1):23-30.
200. Gomes S, Espanca R, Gato A, Miranda C. Obesity in preschool age - too early to be too heavy!. Acta Medica Portuguesa. 2010;23(3):371-378.
201. Serra-Majem L, Aranceta Bartrina J, Perez-Rodrigo C, Ribas-Barba L, Delgado-Rubio A. Prevalence and deteminants of obesity in Spanish children and young people. The British Journal of Nutrition. 2006;96 Suppl 1:S67-72.
202. Jebb SA, Rennie KL, Cole TJ. Prevalence of overweight and obesity among young people in Great Britain. Public Health Nutrition. 2004;7(3):461-465.

203. Ortega FB, Labayen I, Ruiz JR, et al. Improvements in fitness reduce the risk of becoming overweight across puberty. Medicine and Science in Sports and Exercise. 2011;43(10):1891-1897.
204. Ortega FB, Silventoinen K, Tynelius P, Rasmussen F. Muscular strength in male adolescents and premature death: cohort study of one million participants. BMJ. 2012;345:e7279.
205. Voss C, Sandercock G, Wharf Higgins J, et al. A cross-cultural comparison of body composition, physical fitness and physical activity between regional samples of Canadian and English children and adolescents. Canadian Journal of Public Health. 2014;105(4):e245-250.
206. Save the Children. State of the World's Mother report. 2012. URL http://www.savethechildren.org (accessed 12 December 2017).
207. Artero EG, Ortega FB, Espana-Romero V, et al. Longer breastfeeding is associated with increased lower body explosive strength during adolescence. The Journal of Nutrition. 2010;140(11):1989-1995.
208. Telford RD, Cunningham RB, Telford RM, Abhayaratna WP. Schools with fitter children achieve better literacy and numeracy results: evidence of a school cultural effect. Pediatric Exercise Science. 2012;24(1):45-57.
209. Hillman CH, Pontifex MB, Castelli DM, et al. Effects of the FITKids randomized controlled trial on executive control and brain function. Pediatrics. 2014;134(4):e1063-1071.
210. Chaddock L, Pontifex MB, Hillman CH, Kramer AF. A review of the relation of aerobic fitness and physical activity to brain structure and function in children. Journal of the International Neuropsychological Society. 2011;17(6):975-985.
211. Ortega FB, Campos D, Cadenas-Sanchez C, et al. Physical fitness and shapes of subcortical brain structures in children. The British Journal of Nutrition. 2017:1-10.
212. Cotman CW, Berchtold NC, Christie LA. Exercise builds brain health: key roles of growth factor cascades and inflammation. Trends in Neurosciences. 2007;30(9):464-472.
213. Vaynman SS, Ying Z, Yin D, Gomez-Pinilla F. Exercise differentially regulates synaptic proteins associated to the function of BDNF. Brain Research. 2006;1070(1):124-130.
214. van Praag H, Kempermann G, Gage FH. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. Nature Neuroscience. 1999;2(3):266-270.
215. Chaddock L, Erickson KI, Prakash RS, et al. Basal ganglia volume is associated with aerobic fitness in preadolescent children. Developmental Neuroscience. 2010;32(3):249-256.
216. Rauner RR, Walters RW, Avery M, Wanser TJ. Evidence that aerobic fitness is more salient than weight status in predicting standardized math and reading outcomes in fourth- through eighth-grade students. The Journal of Pediatrics. 2013;163(2):344-348.
217. Van Dusen DP, Kelder SH, Kohl HW, 3rd, Ranjit N, Perry CL. Associations of physical fitness and academic performance among schoolchildren. The Journal of School Health. 2011;81(12):733-740.
218. Eveland-Sayers BM, Farley RS, Fuller DK, Morgan DW, Caputo JL. Physical fitness and academic achievement in elementary school children. Journal of Physical Activity & Health. 2009;6(1):99-104.

219. de Greeff JW, Hartman E, Mullender-Wijnsma MJ, Bosker RJ, Doolaard S, Visscher C. Physical fitness and academic performance in primary school children with and without a social disadvantage. Health Education Research. 2014;29(5):853-860.
220. Edwards JU, Mauch L, Winkelman MR. Relationship of nutrition and physical activity behaviors and fitness measures to academic performance for sixth graders in a midwest city school district. The Journal of School Health. 2011;81(2):65-73.
221. Liu-Ambrose T, Donaldson MG. Exercise and cognition in older adults: is there a role for resistance training programmes? British Journal of Sports Medicine. 2009;43(1):25-27.
222. Liu-Ambrose T, Nagamatsu LS, Graf P, Beattie BL, Ashe MC, Handy TC. Resistance training and executive functions: a 12-month randomized controlled trial. Archives of Internal Medicine. 2010;170(2):170-178.
223. Cassilhas RC, Viana VA, Grassmann V, et al. The impact of resistance exercise on the cognitive function of the elderly. Medicine and Science in Sports and Exercise. 2007;39(8):1401-1407.
224. Santana CCA, Azevedo LB, Cattuzzo MT, Hill JO, Andrade LP, Prado WL. Physical fitness and academic performance in youth: A systematic review. Scandinavian Journal of Medicine & Science in Sports. 2017;27(6):579-603.
225. Bruehl H, Sweat V, Tirsi A, Shah B, Convit A. Obese Adolescents with Type 2 Diabetes Mellitus Have Hippocampal and Frontal Lobe Volume Reductions. Neuroscience and Medicine. 2011;2(1):34-42.
226. Nouwen A, Chambers A, Chechlacz M, et al. Microstructural abnormalities in white and gray matter in obese adolescents with and without type 2 diabetes. NeuroImage. 2017;16:43-51.
227. Stabouli S, Kotsis V, Karagianni C, Zakopoulos N, Konstantopoulos A. Blood pressure and carotid artery intima-media thickness in children and adolescents: the role of obesity. Hellenic Journal of Cardiology. 2012;53(1):41-47.
228. Zebekakis PE, Nawrot T, Thijs L, et al. Obesity is associated with increased arterial stiffness from adolescence until old age. Journal of Hypertension. 2005;23(10):1839-1846.
229. Zhu W, Huang X, He J, Li M, Neubauer H. Arterial intima-media thickening and endothelial dysfunction in obese Chinese children. European Journal of Pediatrics. 2005;164(6):337-344.
230. Pandit D, Kinare A, Chiplonkar S, Khadilkar A, Khadilkar V. Carotid arterial stiffness in overweight and obese Indian children. Journal of Pediatric Endocrinology & Metabolism. 2011;24(1-2):97-102.
231. VanWagner LB, Terry JG, Chow LS, et al. Nonalcoholic fatty liver disease and measures of early brain health in middle-aged adults: The CARDIA study. Obesity. 2017;25(3):642-651.
232. Primeau V, Coderre L, Karelis AD, et al. Characterizing the profile of obese patients who are metabolically healthy. International Journal of Obesity (2005). 2011;35(7):971-981.
233. Bluher M. Are metabolically healthy obese individuals really healthy? European Journal of Endocrinology. 2014;171(6):R209-219.
234. Phillips CM. Metabolically healthy obesity across the life course: epidemiology, determinants, and implications. Annals of the New York Academy of Sciences. 2017;1391(1):85-100.
235. Ortega FB, Ruiz JR, Labayen I, Lavie CJ, Blair SN. The Fat but Fit paradox: what we know and don't know about it. British Journal of Sports Medicine. 2017.

SHORT CURRICULUM VITAE

CRISTINA CADENAS SÁNCHEZ

Date of birth: 30/05/1989

Email: cadenas@ugr.es

Academic training

- **2007-2010.** GRADUATE DEGREE IN TEACHING IN PHYSICAL EDUCATION. Teaching School La Inmaculada, University of Granada, Granada, Spain.
- **2010-2012.** GRADUATE DEGREE IN SPORT SCIENCES. Faculty of Sport Sciences, University of Granada, Granada, Spain.
- **2012-2013**. MASTER DEGREE IN RESEARCH ON PHYSICAL ACTIVITY AND SPORT. Faculty of Sport Sciences, University of Granada, Granada, Spain.

Grants

Cristina Cadenas-Sanchez has applied and competed for nearly all the scholarships and grants provided by the Spanish Research and Education system as well as by the University of Granada, succeeding in most of them. Out of 8 scholarships granted, we highlighted below the most relevant and research-related:

- **2015-2019.** Research Fellowship: **Formación del Personal Investigador (FPI).** Spanish Ministry of Economy and Competitiveness.
- **2015-(Feb-Mar).** Research Fellowship: **Formación del Profesorado Universitario (FPU)**. University of Granada.
- **2012-2013. Starting-up research grant.** University of Granada.

Awards

- **2016**. **Award to the best GPA in the Master Degree in Research on Physical Activity and Sport.** University of Granada.

- **2016**. **Award to the best Master Thesis Work in the Master Degree in Research on Physical Activity and Sport.** University of Granada.

- **2013**. **Book award to the 150 best academic performances from the University of Granada (2011-2012).** University of Granada.

Research stays

International research stays

- **To happen in 2018 (already funded).** Center for the Neural Basis of Cognition. Department of Psychology. **University of Pittsburgh.** Pittsburgh, United States. Duration: 4 months (August-December 2018)
- **2017.** Center for Cognitive and Brain Health, Department of Psychology. **Northeastern University**. Boston, United States.
  Duration: 4 months.
- **2016**. Department of Biosciences and Nutrition, Unit for Preventive Nutrition. **Karolinska Institutet**. Stockholm, Sweden.
  Duration: 3 months.
- **2015**. Centre for Active Lifestyles, Institute of Sport, Physical Activity and Leisure. **Leeds Beckett University**. Leeds, United Kingdom.
  Duration: 1.5 month.
- **2014**. Department of Biosciences and Nutrition, Unit for Preventive Nutrition. **Karolinska Institutet**. Stockholm, Sweden.
  Duration: 1 month.
- **2013.** School of Sport and Exercise Sciences. **Liverpool John Moores University**. Liverpool, United Kingdom.
  Duration: 1 week.

National research stays

- **2016.** Faculty of Pharmacy. **University of Basque Country.** Vitoria, España.
  Duration: 1 week.
- **2016.** Centro de Estudios Socio Sanitarios. **University of Castilla La-Mancha**. Cuenca, España.
  Duration: 1 week.

Research Projects

Cristina has participated in 12 research projects. To note, she is the project manager and work-field coordinator of two projects: PREFIT (assessing fitness in preschool children) and ActiveBrains (detailed below).

1. Project: Estudio **PREFIT**: evaluación del FITness en PREscolares
   Funding Source: Ayudas a investigadores Ramón y Cajal a la iniciación de líneas de investigación MINECO.
   Duration: from 2013 to now
   Principal Investigator: Francisco B. Ortega Porcel
   Role of PhD candidate: **Project manager**, design of the study, data collection, data analysis, interpretation and scientific dissemination of results.

2. Project: **ActiveBrains**: Efecto de un ensayo aleatorizado basado en ejercicio físico sobre el rendimiento cognitivo y cerebro (funcional y estructuralmente) en preadolescentes con sobrepeso/obesidad.
   Funding Source: Ministerio de Economía y Competitividad
   Duration: from 2014 to now

Principal Investigator: Francisco B. Ortega Porcel
Role of PhD candidate: **Project manager**, design of the study, data collection, data analysis, interpretation and scientific dissemination of results.

Publications

Cristina Cadenas-Sanchez has already published 33 scientific articles (58% as first author) in Journals indexed in the JCR/PubMed. To highlight, Cristina is first or second author in 76% of these papers. The total number of citations is 277 and the h-index is 9 (based on Google Scholar). Likewise, she has published 1 book and 2 book chapters. Only the scientific publications as a first author are shown below.

Scientific publications included in the present International Doctoral Thesis

1. Ortega FB, **Cadenas-Sánchez C**, Sánchez-Delgado G, Mora-González J, Martínez-Téllez B, García-Artero E, Castro-Piñero J, Labayen I, Chillón P, Löf M, Ruiz JR. 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. Impact Factor: 6.8 (Q1, Sport Sciences). (Referred in the Thesis as *Study I*).

2. **Cadenas-Sánchez C**, Sánchez-Delgado G, Martínez-Téllez B, Mora-González J, Löf M, España-Romero V, Ruiz JR, Ortega FB. Reliability and validity of different models of TKK hand dynamometers. *American Journal of Occuppational Therapy*. 2016. July-August 70 (4). Impact Factor: 2.1 (Q1, Rehabilitation). (Referred in the Thesis as *Study II*).

3. **Cadenas-Sánchez C,** Alcántara-Moral F, Sánchez-Delgado G, Mora-González J, Martínez-Téllez B, Herrador-Colmenero M, Jiménez-Pavón D, Femia P, Ruiz JR, Ortega FB. Assessment of in preschool children: adaptation of the 20 metres shuttle run test. *Nutrición Hospitalaria*. 2014; 30(6). Impact Factor: 1.1 (Q4, Nutrition and Dietetics). (Referred in the Thesis as *Study III*).

4. **Cadenas-Sanchez C**, Martinez-Tellez B, Sanchez-Delgado G, Mora-Gonzalez J, Castro-Piñero J, Löf M, Ruiz JR, Ortega FB. 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):910-915. Impact Factor: 4.8 (Q1, Sport Sciences). (Referred in the Thesis as *Study IV*).

5. **Cadenas-Sanchez C**, Intemann T, Labayen I, Peinado AB, Vidal-Conti J, Sanchís-Moysi J, Moliner-Urdiales D, Rodriguez Perez MA, Cañete Garcia-Prieto J, Fernández-Santos JR, Martinez-Tellez B, Vicente-Rodríguez G, Löf M, Ruiz JR, Ortega FB. Physical fitness reference standards for preschool children: The PREFIT Project. *Submitted.* (Referred in the Thesis as *Study V*).

6. **Cadenas-Sanchez C**, Intemann T, Labayen I, Artero EG, Alvarez-Bueno C, Sanchís-Moysi J, Benito PJ, Beltran-Valls MR, Perez-Perez A, Sanchez-Delgado G, Palou P, Vicente-Rodríguez G, Moreno LA, Ortega FB. Prevalence

of severe/morbid obesity and anthropometric reference standards in preschool children. *Submitted.* (Referred in the Thesis as *Study VI*).

7. **Cadenas-Sánchez C**, Nyström Delisle C, Sanchez-Delgado G, Martinez-Tellez B, Mora-Gonzalez J, Risinger AS, Ruiz JR, Ortega FB, Löf M. Prevalence of overweight/obesity and fitness level in preschool children from the north compared with the south of Europe: an exploration with two countries. *Pediatric Obesity*. 2015; 11(5):403.10. Impact Factor: 3.7 (Q1, Pediatrics). (Referred in the Thesis as *Study VII*).

8. **Cadenas-Sánchez C,** Mora-Gonzalez J, Migueles JH, Martin-Matillas M, Gómez-Vida J, Escolano-Margarit MV, Maldonado J, Enriquez GM, Pastor-Villaescusa B, de Teresa C, Navarrete S, Lozano RM, de Dios Beas-Jiménez J, Estévez-López F, Mena-Molina A, Heras MJ, Chillón P, Campoy C, Muñoz-Hernández V, Martínez-Ávila WD, Merchan ME, Perales JC, Gil Á, Verdejo-García A, Aguilera CM, Ruiz JR, Labayen I, Catena A, Ortega FB. An exercise-based randomized controlled trial on brain, cognition, physical health and mental health in overweight/obese children (ActiveBrains project): Rationale, design and methods. *Contemporary Clinical Trials.* 2016; 47:315-324. Impact Factor: 2.6 (Q3, Medicine, Research and Experimental). (Referred in the Thesis as *Study 0,* presented in the method of *Section 2*).

9. **Cadenas-Sanchez C**, Esteban-Cornejo I, Migueles JH, Mora-Gonzalez J, Henriksson P, Rodríguez-Ayllon M, Molina-García P, Löf M, Labayen I, Hillman CH, Catena A, Ortega FB. Fitness, physical activity, and academic achievement in overweight/obese children. *Submitted.* (Referred in the Thesis as *Study VIII*)

10. **Cadenas-Sanchez C**, Esteban-Cornejo I, Verdejo-Roman J, Migueles JH, Mora-Gonzalez J, Henriksson P, Gómez-Vida J, Maldonado J, Labayen I, Kramer A, Hillman CH, Catena A, Ortega FB. Differences between metabolically healthy and unhealthy overweight/obesity in brain volumes of children: role of fitness. *Submitted.* (Referred in the Thesis as *Study IX*)

Scientific publications directly related but not included in the present International Doctoral Thesis

11. **Cadenas-Sanchez C,** Henriksson P, Henriksson H, Delisle Nyström C, Pomeroy J, Ruiz JR, Ortega FB, Löf M. Parental body mass index and its association with body composition, physical fitness and lifestyle factors in their 4-year-old children: Results from MINISTOP trial. *European Journal Clinical Nutrition.* 2017; 71(10): 1200-1205. Impact Factor: 3.1 (Q2, Nutrition and Dietetics).

12. **Cadenas-Sánchez C**, Artero EG, Concha F, Leyton B, Kain J. Anthropometric characteristics and physical fitness level in relation to body weight status in

Chilean preschool children. *Nutrición Hospitalaria*. 2015; 32(1):346-353. Impact Factor: 1.5 (Q3, Nutrition and Dietetics).

13. **Cadenas-Sánchez C,** Ruiz JR, Labayen I, Huybrechts I, Manios Y, González-Gross M, Breidenassel C, Kafatos A, De Henauw S, Vanhelst J, Widhalm K, Molnar D, Bueno G, Censi L, Plada M, Sjöström M, Moreno LA, Castillo MJ, Ortega FB. Prevalence of metabolically healthy but overweight/obese phenotype and its association with sedentary time, physical activity and fitness. *Journal of Adolescent Health.* 2017; 61(1):107-114. Impact Factor: 3.9 (Q1, Pediatrics).

14. **Cadenas-Sanchez C**, Ruiz JR, Labayen I, Ortega FB. Response to "The obesity phenotypes in adolescents: some lessons from the HELENA study" by Dr. Rey-Lopez and Dr. de Rezende. *Journal of Adolescent Health*. 2017; 61:266-7. Impact Factor: 3.9 (Q1, Pediatrics).

15. **Cadenas-Sánchez C,** Vanhelst J, Ruiz JR, Castillo-Gualda R, Libuda L, Labayen I, De Miguel-Etayo P, Marcos A, Molnár E, Moreno LA, Sjöström M, Gottrand F, Widhalm K, Ortega FB. Fitness and fatness in relation with attention capacity in European adolescents: the HELENA study. *Journal of Science and Medicine in Sport*. 2017; 20(4): 373-379. Impact Factor: 4.8 (Q1, Sport Sciences).

Other scientific publications not related to the present International Thesis

16. **Cadenas Sánchez C**, Arellano R, Vanrenterghem J, López-Contreras G. Kinematic adaptations of forward and backward walking on land and in water. *Journal of Human Kinetics*. 2015; 30; 49:15-24. Impact Factor: 0.8 (Q4, Sport Sciences).

17. **Cadenas-Sánchez C**, Arellano R, Taladriz S, López-Contreras G. Biomechanical characteristics of adults walking forward and backward in water at different stride frequencies. *Journal of Sports Sciences*. 2016; 34(3): 224-231. Impact Factor: 2.5 (Q1, Sport Sciences).
In

18. **Cadenas-Sánchez C**, Ruiz JR. Effect of a physical activity programme in patients with fibromyalgia: A sistematic review. *Medicina Clínica.* 2014; 143(12):548-553. Impact Factor: 1.4 (Q2 Medicine, General and Internal).

Other scientific publications as a co-author

19. Gonzalez-Gil E, **Cadenas-Sanchez C** … Moreno LA. Inflammation in metabolically healthy and metabolically abnormal adolescents: the HELENA study. *Nutrition, Metabolism and Cardiovascular Diseases;* 2018; 28(1):77-83. Impact Factor: 3.7 (Q2, Nutrition and Dietetics).

20. Labayen Goñi I, Arenaza L, Medrano M, García N, **Cadenas-Sanchez C,** Ortega FB. Associations between the adherence to the Mediterranean diet and cardiorespiratory fitness with total and central obesity in preschool children: the

PREFIT Project. *European Journal of Nutrition*. 2017. Impact Factor: 4.5 (Q1, Nutrition and Dietetics).

21. Rodriguez-Ayllon M, **Cadenas-Sanchez C,** Esteban-Cornejo I, Migueles JH, Mora-Gonzalez J, Henriksson P, Martín-Matillas M, Mena-Molina A, Molina-García P, Estévez-López F, Enriquez GM, Perales JC, Ruiz JR, Catena A, Ortega FB. Physical fitness and psychological health in overweight/obese children: A cross-sectional study from the ActiveBrains project. *Journal of Science and Medicine in Sport*. 2018; 21(2):179-184. Impact Factor: 4.8 (Q1, Sport Sciences).

22. Mora-Gonzalez J, **Cadenas-Sanchez C** … Ortega FB. Assessment of cardiorespiratory fitness in 5-6 year-old: Convergent validity of the preschool-adapted (PREFIT) 20 meters shuttle-run test against the original test. *European Journal of Applied Physiology.* Impact Factor: 2.1 (Q2, Sport Sciences).

23. Delisle Nyström C, Henriksson P, Martínez-Vizcaíno V, Medrano M, **Cadenas-Sanchez C**, … Ortega FB. Does cardiorespiratory fitness attenuate the adverse effects of severe/morbid obesity on cardiometabolic risk and insulin resistance in children? A pooled analysis. *Diabetes Care,* 2017; 40:1-8. Impact Factor: 11.9 (Q1, Endocrinology and metabolism).

24. Esteban-Cornejo…**Cadenas-Sanchez C**. A whole brain volumentri approach in overweight/obese children: examining the association with different physical fitness components and academic performance. The ActiveBrains project. *Neuroimage*, 2017, 159: 346-54. Impact Factor: 5.8 (Q1, Neurosciences).

25. Esteban-Cornejo I, **Cadenas-Sanchez C** … Ortega FB. Attention in European adolescents: role of different health-related factors. The HELENA study. *European Journal of Pediatrics*, 2017. Impact Factor: 1.9 (Q2, Pediatrics).

26. Leppänen M … **Cadenas-Sanchez C**, Löf M. Longitudinal physical activity, body composition and physical fitness in preschoolers. *Medicine and Science in Sports and Exercise*, 2017, May 20. Impact Factor: 4.1 (Q1, Sport Sciences).

27. Esteban-Cornejo I, Henriksson P, **Cadenas-Sanchez C** … Ortega FB. Early life programming of attention capacity in adolescents: The HELENA Study. *Maternal and Child Nutrition.* 2017. Impact Factor: 2.5 (Q1, Pediatrics).

28. Hidalgo JH, **Cadenas-Sanchez C** … Ortega FB. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Medicine*. 2017. Impact Factor: 6.8 (Q1, Sport Sciences).

29. Hidalgo JH, **Cadenas-Sanchez C**, Ortega FB. Critique of "Physical activity assessment between consumer-and research-grade accelerometers: a comparative study in free-living conditions". *JMIR Mhealth Uhealth.* 2017. 27, 5(2):e15. Impact Factor: 4.6 (Q1, Health Care Sciences and Services).

30. Vanlancker T, Schaubroeck E, Vyncke K, **Cadenas-Sanchez C** …Huybrechts I. Comparison of definitions for the metabolic syndrome in adolescents. The HELENA-study. *European Journal of Pediatrics*. 2017. Feb, 176(2): 241-252. Impact Factor: 1.9 (Q2, Pediatrics).

31. Mora-Gonzalez J, Rodriguez-Lopez **Cadenas-Sanchez C** … Chillón P. Active Commuting to school and academic achievement in children and adolescents. *Acta Paediatrica.* 2017. Feb, 106(2): 334-340. Impact Factor: 2.0 (Q2, Pediatrics).

32. Ortega FB, Campos D, **Cadenas-Sánchez C** et al. Physical fitness and shape of subcortical brain structures in children. *British Journal of Nutrition*. 2017. Mar, 27:1-10. Impact Factor: 3.7 (Q1, Nutrition and Dietetics).

33. Henriksson P, **Cadenas-Sanchez C**, Leppänen MH, Delisle Nyström C, Ortega FB, Pomeroy J, Ruiz JR, Löf M. Associations of fat mass and fat-free mass with physical fitness in 4-year-old children: results from the MINISTOP Trial. *Nutrients.* 2016. July 30; 8(8). Impact Factor: 3.6 (Q1, Nutrition and Dietetics).

34. Leppänen MH, Nyström CD, Henriksson P, Pomeroy J, Ruiz JR, Ortega FB, **Cadenas-Sanchez C**, Löf M. Physical activity intensity, sedentary behavior, body composition and physical fitness in 4-year-old children: results from the MINISTOP trial. *International Journal of Obesity*. 2016. July 40(7): 1126-33. Impact Factor: 5.5 (Q1, Nutrition and Dietetics).

35. Martinez-Tellez B, Sanchez-Delgado G, **Cadenas-Sanchez C,** Mora-Gonzalez J, Martín-Matillas M, Löf M, Ortega FB, Ruiz JR. Health-related physical fitness is associated with total and central body fat in preschool children aged 3 to 5 years. *Pediatric Obesity.* 2015. Impact Factor: 3.7 (Q1, Pediatrics).

36. Sánchez-Delgado G, **Cadenas-Sánchez C**, Mora-González J, Martínez-Téllez B, Chillón P, Löf M, Ortega FB, Ruiz JR. Assessment of handgrip sterngth in preschool children aged 3-5 years. *Journal of Hand Surgery (European Volume).* 2015. Impact Factor: 1.9 (Q2, Surgery).

37. Ortega FB, **Cadenas-Sánchez C**, Sui X, Blair SN, Lavie CJ. Role of Fitness in Metabolically Healthy but Obese Phenotype: A review and Update. *Progress in Cardiovascular Disease*. 2015; 58(1):76-86. Impact Factor: 4.6 (Q1, Cardiac and cardiovascular systems).

Other merits

- 14 **international** and **national invited talks.**
- **Reviewer** for 12 JCR-journals, including some top-journals in Sport field such as Sports Medicine.
- More than 50 **oral communications** or **posters** both in **national** and **international conferences.**

AGRADECIMIENTOS / ACKNOWLEDGEMENTS

AGRADECIMIENTOS / ACKNOWLEDGEMENTS

La presente Tesis Doctoral Internacional está apoyada gracias a la financiación concedida por parte del Ministerio de Ciencia e Innovación (RYC-2011-09011) y por el Ministerio de Economía y Competitividad (DEP2013-47540 y BES-2014-068829). También queremos agradecer la colaboración de todos los participantes y a aquellas personas involucradas en la toma de datos de los proyectos.

The present International Doctoral Thesis was supported by grant from the Spanish Ministry of Science and Innovation (RYC-2011-09011) and by the Spanish Ministry of Economy and Competitiveness (DEP2013-47540 and BES-2014-068829). We also acknowledge the collaboration of all participants and people involved in data measurements.

ANNEXES

ANNEXES

Table S1. Search strategy used and number of articles found in **PubMed**.

Search criteria 1	**MeSH Entry Terms for Criteria 1**	**Search criteria 2**	**MeSH Entry Terms for Criteria 2**	**Items found (Combining criteria 1 "AND" 2)**
Child, preschool (MeSH)	Preschool Child Children, Preschool Preschool Children	Physical fitness (MeSH)	Fitness, Physical Physical Conditioning, Human Conditioning, Human Physical Conditionings, Human Physical Human Physical Conditioning Human Physical Conditionings Physical Conditionings, Human	422
		Muscle Strength (MeSH)	Strength, Muscle	438
		Muscle strength dynamometer (MeSH)	Dynamometer, Muscle Strength Dynamometers, Muscle Strength Muscle Strength Dynamometers	39
		Range of motion, articular (MeSH Major Topic*)	Joint Range of Motion Joint Flexibility Flexibility, Joint Range of Motion Passive Range of Motion	212
		Postural Balance (MeSH)	Musculoskeletal Equilibrium Equilibrium, Musculoskeletal Postural Equilibrium Equilibrium, Postural Balance, Postural	369
			-	25
		Cardiovascular fitness	-	18
		Aerobic fitness	-	32

	Aerobic capacity	-	35
	Maximal oxygen consumption	-	19
	VO_2max	-	19
	Running Speed	-	11
	Agility	-	41
	*Fitness-test Batteries: Acronym. Society/organization***		
	EUROFIT. Council of Europe Committee for the Development of Sport		0
	FITNESSGRAM. Council of Europe Committee for the Development of Sport		4
	PHYSICAL BEST, American Association for Health, Physical Education, and Recreation (AAHPER)		0
	PCHF. President's Challenge: Health Fitness. The President's Council on Physical Fitness and Sports/American Association for Health, Physical Education, and Recreation (AAHPER)		0
	PCPF. President's Challenge: Physical Fitness. The President's Council on Physical Fitness and Sports/American Association for Health, Physical Education, and Recreation (AAHPER)		1
	AAUTB. Amateur Athletic Union Test Battery. Chrysler Foundation/Amateur Athletic Union		0
	NZFT. New Zealand Fitness Test. Rusell/Department of Education		0
	HRFT. Health-Related Fitness Test, American Association for Health, Physical Education, and Recreation (AAHPER)		0
	YMCAYFT. YMCA Youth Fitness Test		0
	NYPFP. National Youth Physical Program. The United States Marines Youth Foundation		0
	IPFT. International Physical Fitness Test (United States Sports Academic/ General Organization of Youth and Sport of Bahrain)		4
	CAHPER. Fitness Performance Test II. Canadian Association for Health, Physical Education and Recreation (CAHPER)		0
	CPAFLA. The Canadian Physical Activity, Fitness & Lifestyle Approach (Canadian Society for Exercise Physiology)		0
	NFTP-PRC. National Fitness Test Program in the Popular Republic China		0

	(China's National Sport and Physical Education Committee)	
	AFEA. Australian Fitness Education Award. The Australian Council for Health, Education and Recreation, ACHER	1
Total items found	All the searches were combined using the operator "OR" so that duplicated articles were excluded from the final number of articles found. Example: ("Child, Preschool"[MeSH] AND "Physical Fitness"[MeSH]) OR ("Child, Preschool"[MeSH] AND "Muscle Strength"[MeSH])	**1498**

The search recruited articles published until April 1st 2014; no starting date limit was set for the search.

* For this search term, we noted that the search retrieved too many articles (n=1571) that had nothing to do with the topic of this review. In this case, the command "Major Topic" was attached to the search, restricting the search to those studies in which the search term is the major focus of the study (n=212).

** Searches conducted using the acronyms, e.g. EUROFIT was entered in the search browser.

MeSH (Medical Subject Headings) is the National Library of Medicine controlled vocabulary thesaurus used for indexing articles for PubMed.

Table S2. Search strategy used and number of articles found in Web of Science.

Search criteria 1	Search criteria 2	Search criteria 3	Items found (Combining criteria 1, 2 "AND" 3)
"Preschool child" OR "Preschool Children" OR Preschooler* OR Preschool	Physical Conditioning Physical fitness	-	500
	Muscle Strength Muscular Strength	-	501
	Joint Range of motion Joint flexibility Range of motion	Test*	355
	Musculoskeletal Equilibrium Postural Balance Postural Equilibrium	-	375
	Cardiovascular fitness Aerobic fitness Aerobic capacity Maximal oxygen consumption VO_2max	-	155
	Running Speed Agility	-	59
	*Fitness-test Batteries: Acronym. Society/organization***		
	EUROFIT. Council of Europe Committee for the Development of Sport		3
	FITNESSGRAM. Council of Europe Committee for the Development of Sport		4
	PHYSICAL BEST, American Association for Health, Physical Education, and Recreation (AAHPER)		1
	PCHF. President's Challenge: Health Fitness. The President's Council on Physical Fitness and Sports/American Association for Health, Physical Education, and Recreation (AAHPER)		0
	PCPF. President's Challenge: Physical Fitness. The President's Council on		1

	Physical Fitness and Sports/American Association for Health, Physical Education, and Recreation (AAHPER)	
	AAUTB. Amateur Athletic Union Test Battery. Chrysler Foundation/Amateur Athletic Union	0
	NZFT. New Zealand Fitness Test. Rusell/Department of Education	0
	HRFT. Health-Related Fitness Test, American Association for Health, Physical Education, and Recreation (AAHPER)	0
	YMCAYFT. YMCA Youth Fitness Test	0
	NYPFP. National Youth Physical Program. The United States Marines Youth Foundation	0
	IPTF. International Physical Fitness Test (United States Sports Academic/ General Organization of Youth and Sport of Bahrain)	0
	CAHPER. Fitness Performance Test II. Canadian Association for Health, Physical Education and Recreation (CAHPER)	0
	CPAFLA. The Canadian Physical Activity, Fitness & Lifestyle Approach (Canadian Society for Exercise Physiology)	0
	NFTP-PRC. National Fitness Test Program in the Popular Republic China (China's National Sport and Physical Education Committee)	0
	AFEA. Australian Fitness Education Award. The Australian Council for Health, Education and Recreation, ACHER	1
Total items found	All the searches were combined using the operator "OR" so that duplicated articles were excluded from the final number of articles found. Example: ("Preschool child" OR "Preschool Children" OR Preschooler* OR Preschool) AND ("Physical fitness" OR "Physical Conditioning") OR ("Preschool child" OR "Preschool Children" OR Preschooler* OR Preschool) AND (Muscle strength OR Muscular strength)	**1732**

The search recruited articles published until April 1st 2014; no starting date limit was set for the search.

* For this search term, we noted that the search retrieved too many articles (n=1961) that had nothing to do with the topic of this review. In this case, the word "Test" (search criteria 3) was included in the search, restricting the search to those studies in fitness testing (n=355).

Table S3. Practical considerations when using the tests proposed in the PREFIT battery, field-based **FIT**ness testing in **PRE**school children aged 3 to 5 years group.

Fitness test	**Practical considerations/recommendations**
Overall recommendation	- Evaluators must be constantly encouraging and motivating the children. We recommend doing the evaluation like a game/fairy tale, presenting each test like an adventure or challenge. Telling them a story (preschoolers can be in a circle and the evaluator in the middle of it) plenty of adventures, can be a good strategy. See Figure S2 in supplementary material as an example. - The person who tell the fairy tale must to interact with the children when he/she is telling the adventures (e.g. for anthropometry test, “How can we dressed up Cofito? We need to do some measurement to make a suit that perfectly fits him”) and they have to imitate the gestures of dressing, etc. Later when testing, each evaluator remind to the kid in which adventure station he/she is and what is the task there. - A minimum of 2 evaluators are needed, although, ideally 5 evaluators are recommended. - Children were in their normal class room and researcher took them to a separate testing room in groups of 8 children. The 8 children are organized by couples, each couple going to one of the four stations (1= anthropometry, 2= handgrip 3= standing long jump, 4= one-leg stance test). PREFIT 20m shuttle run test was assessed at the end of the evaluation session, since this is a maximal test and children are exhausted after that. - The tests could be located in the same area, although whenever possible, we recommend performing the anthropometric measurements and one-leg stance test separately and in a quiet room. - We recommend performing an example before the test. - The scores obtained in all attempts are recorded, yet the best performance will be used for analyses, except for anthropometric measurements in which the average score is computed and used in analyses.
Weight, height and waist circumference	- For waist circumference, we recommend the measurement at the umbilical level, since typical body shape at these ages does not allow often to identify minimal/narrowest waist.
Handgrip strength test	- We recommend using the analog version (TKK model 5001), since the range of measurement is from 0 to 100 kg, while the digital version (TKK 5401) is from 5 to 100 kg, and some small kids perform less than 5 kg. (See methodological paper about these dynamometers[10]) - In a previous methodological paper, we observed that the optimal grip span for preschoolers was 4.0 cm[11]. - Children tend to take the dynamometer with both hands and squeeze simultaneously. It is something that needs to be corrected most of times.
Standing long jump test	- Young preschoolers (specially the 3 years-old group) have problems to jump with feet together and to land without falling. Therefore, one or two familiarization attempts before trial are recommended. - We also recommend performing 3 trials instead of 2 in comparison with the rest of the tests.

	- We recommend drawing footprints on the floor in order to guide the child to know the take-off line.
4×10 m shuttle run test	- To make this test simpler compared with older children and adolescents, do not use sponges to be exchanged when crossing the lines 10 m apart (4 times × 10 m). Instead of sponges, two evaluators are located at behind each of the line (10 m apart) and preschoolers have to touch the evaluator hand (place behind the line) and go back at maximum speed. - The evaluator could provide the instructions during the development of the test. For example: "Now, you have to clap my hand and turn quickly".
One-leg stance test	- There are many versions of this test (e.g. eyes open vs. close, standing on the floor vs. on a beam forefoot or with shoes). We recommend using the simplest one when implemented in preschoolers. The child stand on a non-slippery floor with eyes opened. The time (seconds with two decimal) that the child is able to maintain the requested position is scored. Each leg is tested once and the average of both legs is used for analyses.
PREFIT 20m shuttle run test	- Very young children have problems to keep the appropriate pace while doing this test, so it is recommended that one or two evaluators run with them. When two evaluators are available, then, one of them can run in front of the preschoolers and the other one behind, keeping children between them. - Groups should be up to 4-8 children. If only one evaluator is available, no more than 3-4 children are recommended. - While scoring this test, you may record the number of laps instead of completed stages as sometimes done in older children (conversions from laps to stages or final speed can be done *a posteriori* if needed. This will make the test more precise and discriminant.

Table S4. Definition of the academic achievement components measured by Woodcock-Muñoz standardized test.

Components	**Definitions**
Total achievement	Overall measure of the academic performance based on reading, mathematics and writing.
Reading	Broad measure of reading performance that includes word identification, reading speed and comprehension.
Oral language	Measure of linguistic competency, listening ability and oral comprehension.
Writing	Broad measure of written language performance that includes spelling, quality of written sentences and speed of writing.
Written expression	Combined measure between writing speed and the quality of the sentence (meaningful, coherence, etc.).
Mathematics	Broad measure of mathematics performance which includes calculation skills, problem solving and the ability to subtract, sum, multiply or divide quickly.
Mathematics calculation skills	Combined measured of mathematics based on do simple calculations quickly and the ability to perform mathematical computations.
Science	Measure the knowledge in the sciences, history, geography, government, economics, art, music and literature.

Table S5. Jolliffe and Janssen[73] criteria for classification of metabolically healthy and unhealthy phenotypes.

Years	Waist circumference (cm)	Triglycerides (mg/dL)	Glucose (mg/dL)	HDL cholesterol (mg/dL)	Systolic (mmHg)	Diastolic (mmHg)
Boys:						
12	≥ 85.1	≥ 127.40	≥ 100.97	≥ 43.70	≥ 121	≥ 76
Girls:						
12	≥ 72.5	≥ 141.60	≥ 100.97	≥ 48.34	≥ 121	≥ 80

HDL: High density lipoprotein.

Table S6. Generalized Additive Model for Location, Scale and Shape (GAMLSS) models used to calculate physical fitness and anthropometric reference standards.

	Sex	N (all)	Model Distribution	Parameters μ	$log(\sigma)$	υ	$log(\tau)$
PREFIT 20m shuttle run (laps)	Boys	1634	BCPE	cs(age)	cs(age)	cs(age)	const.
	Girls	1466	BCCG	cs(age)	const.	const.	-
Handgrip strength (kg)	Boys	1675	BCT	age	age	const.	Age
	Girls	1498	BCT	age	cs(age)	const.	cs(age)
Standing long jump (cm)	Boys	1669	BCPE	cs(age)	const.	const.	Age
	Girls	1494	BCPE	cs(age)	const.	const.	Age
4x10m shuttle run (s)	Boys	1669	BCT	cs(age)	cs(age)	const.	const.
	Girls	1482	BCPE	cs(age)	cs(age)	const.	cs(age)
One-leg stance (s)*	Boys	1667	BCT	cs(age)	const.	const.	const.
	Girls	1494	BCT	cs(age)	const.	const.	const.
Weight (kg)	Boys	1675	BCT	cs(age)	cs(age)	const.	const.
	Girls	1501	BCT	cs(age)	cs(age)	const.	cs(age)
Height (cm)	Boys	1675	BCCG	cs(age)	age	age	-
	Girls	1501	BCCG	cs(age)	const.	const.	-
Body mass index (kg/m²)	Boys	1675	BCT	cs(age)	cs(age)	const.	const.
	Girls	1501	BCPE	const.	cs(age)	const.	const.
Waist circumference (cm)	Boys	1676	BCPE	cs(age)	cs(age)	const.	cs(age)
	Girls	1500	BCPE	cs(age)	cs(age)	const.	cs(age)
Waist-to-height ratio	Boys	1674	BCCG	cs(age)	cs(age)	const.	-
	Girls	1500	BCT	cs(age)	cs(age)	const.	const.

BCT: Box-Cox t distribution; BCPE: Box-Cox power exponential; BCCG: Box-Cox Cole and Green (i.e the LMS method); cs: cubic splines; age: linear function of age; const.: constant.

* In one-leg stance test, outlier values (i.e. >150 seconds) were excluded.

Table S7. Stages, speed (km/h) and laps of the original 20m shuttle run test.

Stages	Speed (km/h)	Laps	Total laps
0.5	8.5	4	4
1	8.5	3	7
1.5	9.0	4	11
2	9.0	4	15
2.5	9.5	4	19
3	9.5	4	23
3.5	10.0	4	27
4	10.0	4	31
4.5	10.5	5	36
5	10.5	4	40
5.5	11.0	5	45
6	11.0	4	49
6.5	11.5	5	54
7	11.5	5	59
7.5	12.0	5	64
8	12.0	5	69
8.5	12.5	5	74
9	12.5	5	79
9.5	13.0	5	84
10	13.0	6	90
10.5	13.5	6	96
11	13.5	5	101
11.5	14.0	6	107
12	14.0	6	113
12.5	14.5	6	119
13	14.5	6	125
13.5	15.0	6	131
14	15.0	7	138
14.5	15.5	7	145
15	15.5	6	151
15.5	16.0	7	158
16	16.0	6	164
16.5	16.5	7	171
17	16.5	7	178
17.5	17.0	7	185
18	17.0	7	192
18.5	17.5	7	199
19	17.5	8	207
19.5	18.0	8	215
20	18.0	7	222

Table S8. Comparison of the reliability of fitness tests between preschoolers (PREFIT study) and adolescents (HELENA study)[181].

Tests	Mean differences			
	Boys		Girls	
	PREFIT	HELENA	PREFIT	HELENA
PREFIT 20m shuttle run (stages)	0.1 ± 1.0	-0.1 ± 1.5	0.5 ± 1.1*	0.0 ± 1.1
Handgrip strength (kg)	-0.4 ± 1.3	0.3 ± 2.5	-0.0 ± 1.2	0.0 ± 1.8
Standing long jump-S1 (cm)	-7.5 ± 14.3*	-0.3 ± 12.9	-7.0 ± 14.6*	0.3 ± 9.0
Standing long jump-S2 (cm)	-4.10 ±12.9*	-0.3 ± 12.9	-0.02 ±13.0	0.3 ± 9.0
4×10 m shuttle run (s)	0.1 ± 0.9	0.1 ± 0.7	0.2 ± 1.1	0.1 ± 0.8
One-leg stance test (s)	8.9 ± 20.4	-	6.7 ± 21.6	-

S1= Study 1, S2= Study 2

*In these cases we consider that the reliability of fitness tests markedly differs.

Standing long jump-S2 results corresponds to the replication study performed for this test in different preschoolers.

Supplemental Information 1. Detailed information on the data collection and processing criteria of the physical activity measurement.

The raw data collected at a sampling frequency of 100 Hz were processed in R (v. 3.1.2, https:// www.cran.r-project.org/) using the GGIR package (v. 1.5-12, https://cran.r-project.org/web/packages/GGIR/)[1] and in the ActiLife software (v.6.13.3, ActiGraph TM, Pensacola, FL). To process the data in the GGIR package 1) we auto-calibrated of the signal according to the local gravity[2]; 2) we derived the mean of the Euclidian Norm Minus One G (ENMO, 1G = 1 gravitational acceleration ~9.8 m/s^2) with negative values rounded to zero; 3) we imported the ActiGraph's activity counts over 5s epochs derived from the ActiLife software using the default filter; 4) we calculated the non-wear time using the approach described by Van Hees et al.[3] Briefly, 15-min blocks were classified as non-wear time if the standard deviation of 2 out of the 3 axes was lower than 13 mg during the surrounding 60-min moving window; 5) we identified clipping scores, i.e., atypical high accelerations related to malfunctioning of the accelerometers; 6) we detected imputations of the non-wear and clipped time by means of the rest of the days during the same time interval as the detected window[4]. If no data were collected for the certain window for the rest of the days, then, non-wear time was replaced by 0; and 7) we identified of the sleeping hours based on an automatized algorithm based on the anteroposterior angle estimated from the accelerometer[5] guided by the sleep onset and waking-up times reported by the participants.

The inclusion criterion was ≥600 min/day of waking hours and ≥240 min/day of sleeping hours for a valid day, and a minimum of 4 days (3 weekdays and 1 weekend day) to be included in the analyses. The compliance with wearing the accelerometer was high with 98% of sample wearing it ≥6 days.

References

1. van Hees VT, Gorzelniak L, Dean Leon EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PloS one 2013; 8(4): e61691.
2. van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. Journal of Applied Physiology (Bethesda, Md : 1985) 2014; 117(7): 738-44.
3. van Hees VT, Renstrom F, Wright A, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PloS one 2011; 6(7): e22922.
4. Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. Medicine and Science in Sports and Exercise 2012; 44(1 Suppl 1): S61-7.
5. Van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, Abell JG, Kivimäki M, Trenell MI, Singh-Manoux A. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PloS One 2015; 10(11):e0142533.

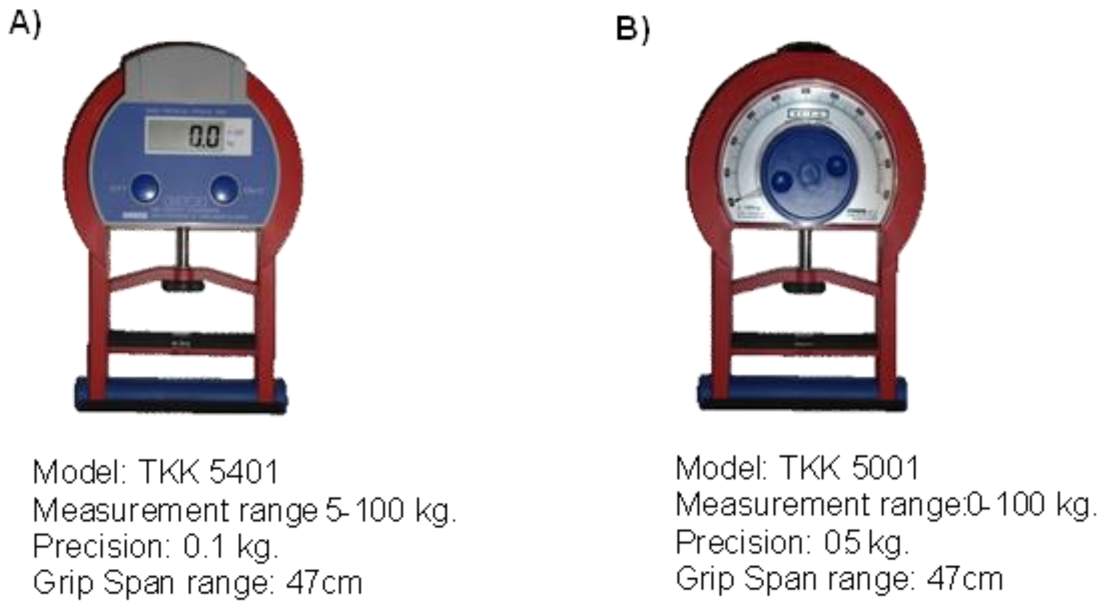
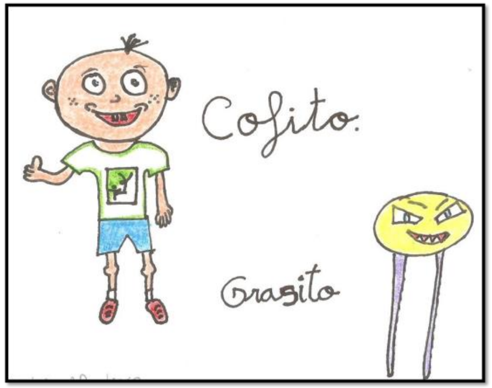
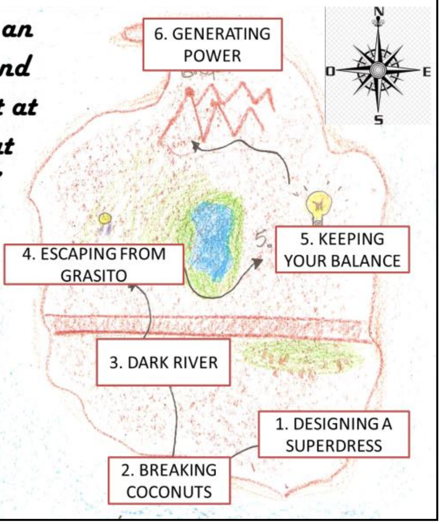


Figure S1. Digital (Figure S1a) and (FigureS1b) analog TKK hand dynamometers.

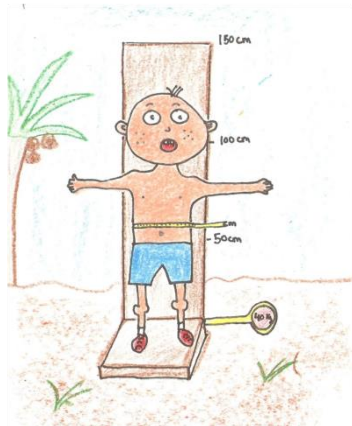
Help Cofito escape from Grasito



Cofito has an accident and gets caught at "The Fat Island"



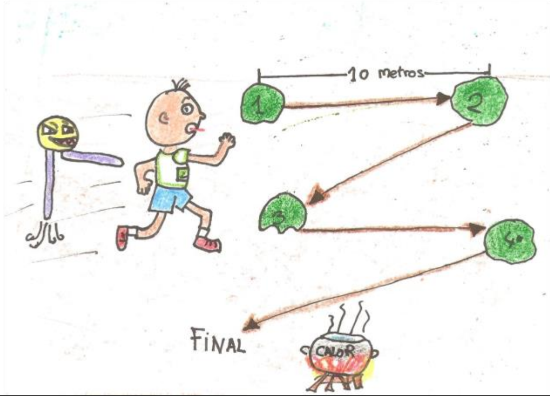
Dressmaking as Cofito…



Help Cofito breaking coconuts just with one hand

Cofito needs to jump the dark river, ¿Are you joining? ☺

Oh No!! Cofito is being following by Grasito. ¡Let's make him fall into the heat!



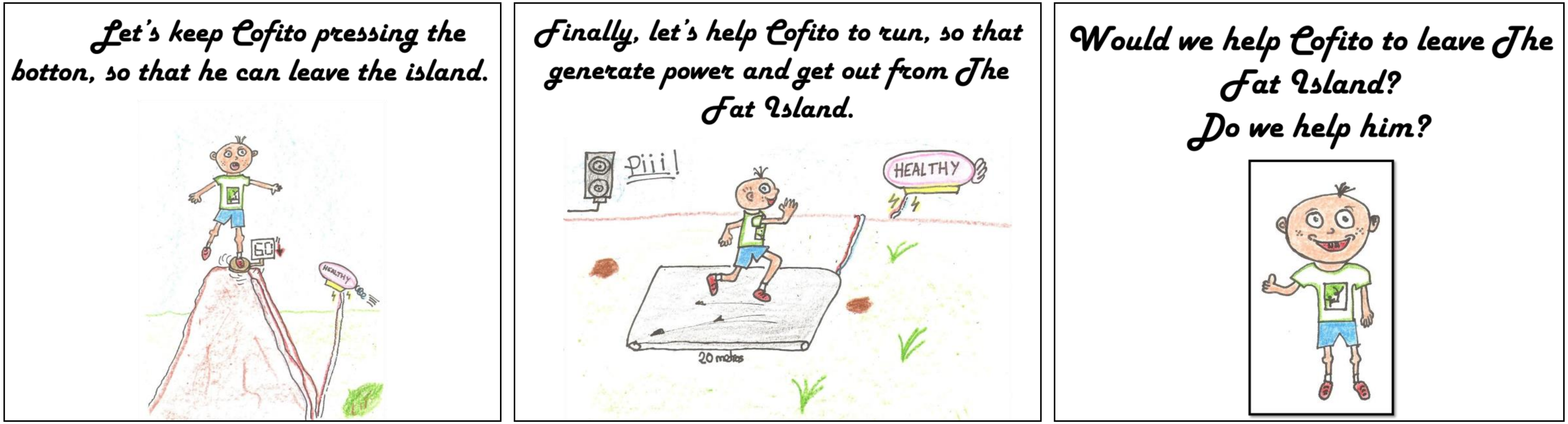


Figure S2. Cofito's fairy tale showing all the tests of PREFIT battery.

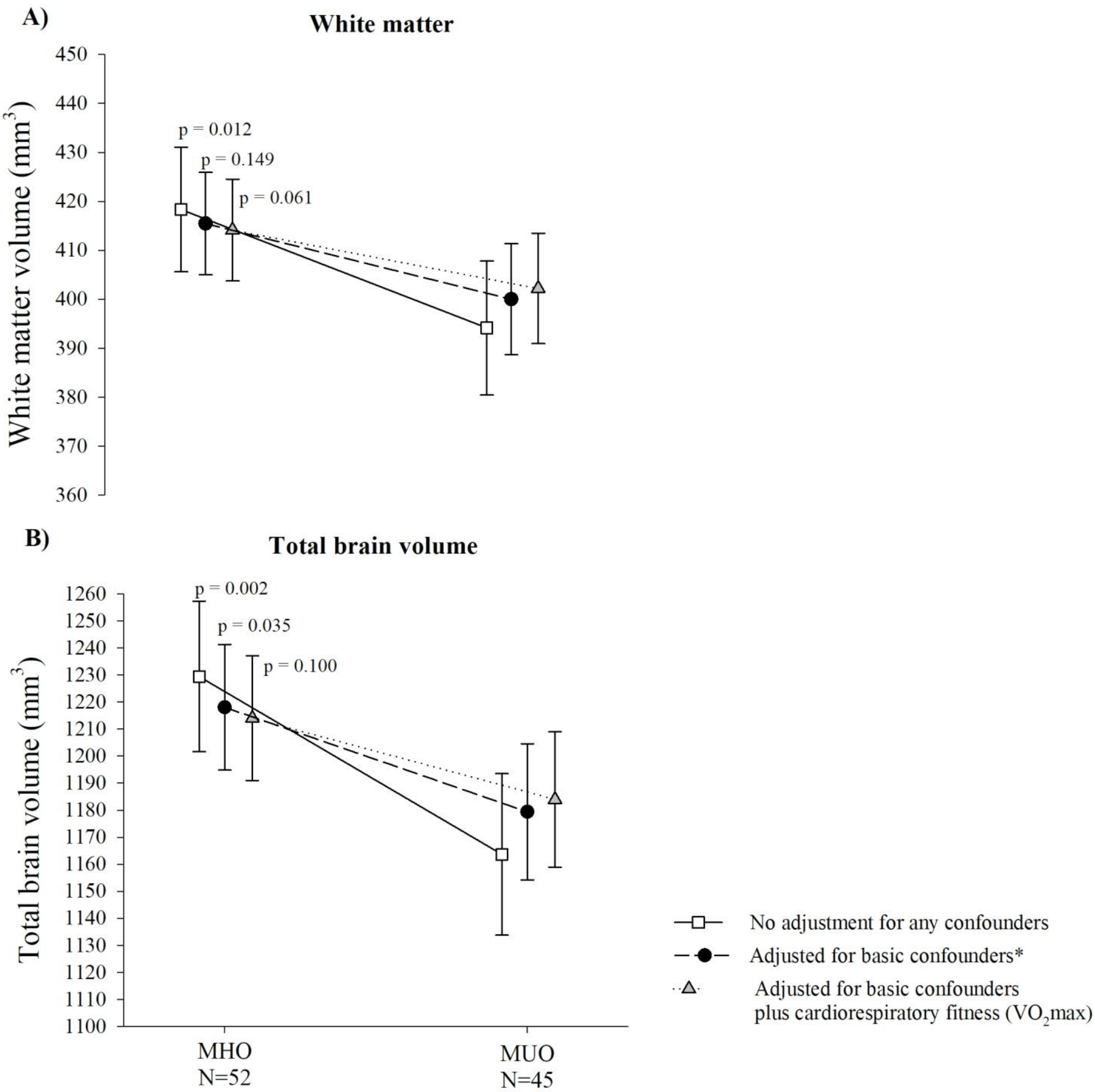


Figure S3. Differences in global white matter (Figure S3a) and total brain volume (Figure S3b) between metabolically healthy and metabolically unhealthy overweight/obesity.

MHO: Metabolically healthy overweight/obesity. MUO: Metabolically unhealthy overweight/obesity.

*Basic confounders were sex, peak height velocity (years), parental education university level (none/one/ both of them) and body mass index (kg/m^2).

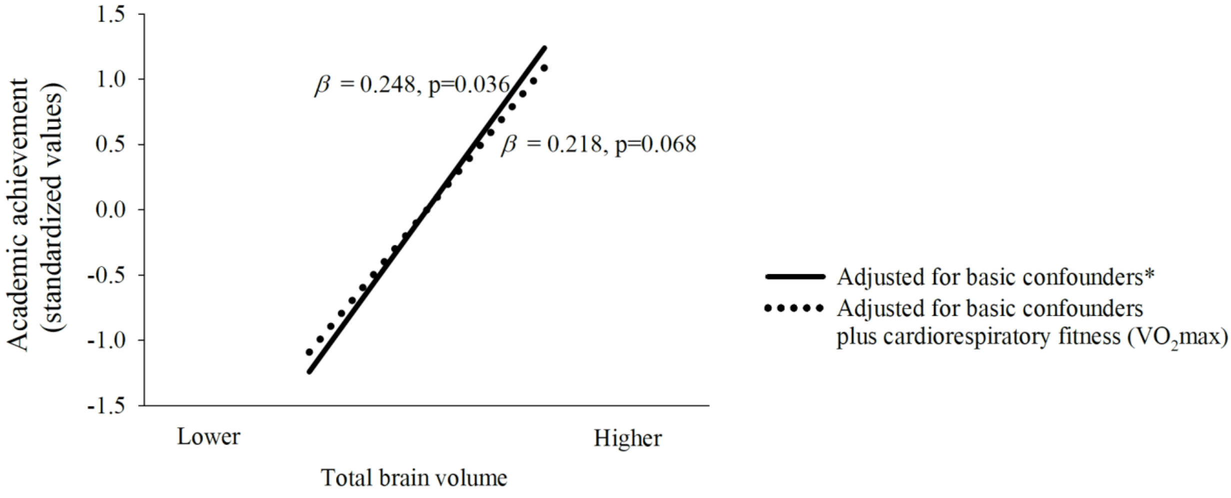


Figure S4. Linear regression analysis between total brain volume and academic achievement.

ß= beta standardized coefficients.

*Basic confounders were sex, age, parental education university level (none/one/ both of them), and body mass index (kg/m^2).