






## Ageing

# Development of a common scale for measuring healthy ageing across the world: results from the ATHLOS consortium

Albert Sanchez-Niubo <sup>1,2\*</sup>, Carlos G Forero,<sup>3</sup> Yu-Tzu Wu,<sup>4</sup> Iago Giné-Vázquez,<sup>1,2</sup> Matthew Prina,<sup>4,5</sup> Javier De La Fuente,<sup>6,7</sup> Christina Daskalopoulou,<sup>4</sup> Elena Critselis,<sup>8</sup> Alejandro De La Torre-Luque,<sup>2,6,7</sup> Demosthenes Panagiotakos,<sup>8</sup> Holger Arndt,<sup>9</sup> José Luis Ayuso-Mateos,<sup>2,6,7</sup> Ivet Bayes-Marin,<sup>1,2</sup> Jerome Bickenbach,<sup>10,11</sup> Martin Bobak,<sup>12</sup> Francisco Félix Caballero,<sup>13,14</sup> Somnath Chatterji,<sup>15</sup> Laia Egea-Cortés,<sup>1</sup> Esther García-Esquinas <sup>13,14</sup>, Matilde Leonardi,<sup>16</sup> Seppo Koskinen,<sup>17</sup> Ilona Koupil,<sup>18,19</sup> Blanca Mellor-Marsá,<sup>1,20</sup> Beatriz Olaya,<sup>1,2</sup> Andrzej Pająk <sup>21</sup>, Martin Prince,<sup>5,22</sup> Alberto Raggi,<sup>16</sup> Fernando Rodríguez-Artalejo,<sup>13,14</sup> Warren Sanderson,<sup>23,24</sup> Sergei Scherbov,<sup>23,25,26</sup> Abdonas Tamosiunas,<sup>27</sup> Beata Tobias-Adamczyk,<sup>28,29</sup> Stefanos Tyrovolas<sup>1,2</sup> and Josep Maria Haro<sup>1,2</sup>; the ATHLOS Consortium

<sup>1</sup>Research, Innovation and Teaching Unit, Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain, <sup>2</sup>Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, Madrid, Spain, <sup>3</sup>Department of Medicine, International University of Catalunya, Barcelona, Spain, <sup>4</sup>Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK, <sup>5</sup>Global Health Institute, King's College London, London, UK, <sup>6</sup>Department of Psychiatry, Universidad Autónoma de Madrid, Madrid, Spain, <sup>7</sup>Instituto de Investigación Sanitaria Princesa (IIS Princesa), Hospital Universitario de La Princesa, Madrid, Spain, <sup>8</sup>Department of Nutrition and Dietetics, School of Health Sciences and Education, Harokopio University, Athens, Greece, <sup>9</sup>Spring Techno GMBH & Co. KG, Bremen, Germany, <sup>10</sup>Swiss Paraplegic Research, Guido A. Zäch Institute (GZI), Nottwil, Switzerland, <sup>11</sup>Department of Health Sciences & Health Policy, University of Lucerne, Lucerne, Switzerland, <sup>12</sup>Department of Epidemiology and Public Health, University College London, London, UK, <sup>13</sup>Department of Preventive Medicine and Public Health, Universidad Autónoma de Madrid/Idipaz, Madrid, Spain, <sup>14</sup>Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, CIBERESP, Madrid, Spain, <sup>15</sup>Information, Evidence and Research, World Health Organization, Geneva, Switzerland, <sup>16</sup>Neurology, Public Health, Disability Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>17</sup>Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland, <sup>18</sup>Department of Public Health Sciences, Stockholm University, Stockholm, Sweden, <sup>19</sup>Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden, <sup>20</sup>Sanitary Research Institute, Hospital Clínico San Carlos, Madrid, Spain, <sup>21</sup>Department of Epidemiology and Population Studies, Jagiellonian University Medical College, Krakow, Poland, <sup>22</sup>Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College

London, London, UK, <sup>23</sup>Wittgenstein Centre for Demography and Global Human Capital, International Institute for Applied Systems Analysis, Laxenburg, Austria, <sup>24</sup>Department of Economics, Stony Brook University, Stony Brook, NY, USA, <sup>25</sup>Austrian Academy of Science, Vienna Institute of Demography, Vienna, Austria, <sup>26</sup>International Laboratory of Demography and Human Capital, Russian Presidential Academy of National Economy and Public Administration, Moscow, Russian Federation, <sup>27</sup>Department of Population Studies Institute of Cardiology, Lithuanian University of Health Sciences, Kaunas, Lithuania, <sup>28</sup>Department of Medical Sociology, Jagiellonian University Medical College, Krakow, Poland and <sup>29</sup>Department of Epidemiology, Jagiellonian University Medical College, Krakow, Poland

\*Corresponding author. Parc Sanitari Sant Joan de Déu, Dr Antoni Pujadas 42, Sant Boi de Llobregat, Barcelona 08830. Spain. E-mail: albert.sanchez@pssjd.org

Editorial decision 11 October 2020; Accepted 23 October 2020

## Abstract

**Background:** Research efforts to measure the concept of healthy ageing have been diverse and limited to specific populations. This diversity limits the potential to compare healthy ageing across countries and/or populations. In this study, we developed a novel measurement scale of healthy ageing using worldwide cohorts.

**Methods:** In the Ageing Trajectories of Health-Longitudinal Opportunities and Synergies (ATHLOS) project, data from 16 international cohorts were harmonized. Using ATHLOS data, an item response theory (IRT) model was used to develop a scale with 41 items related to health and functioning. Measurement heterogeneity due to intra-dataset specificities was detected, applying differential item functioning via a logistic regression framework. The model accounted for specificities in model parameters by introducing cohort-specific parameters that rescaled scores to the main scale, using an equating procedure. Final scores were estimated for all individuals and converted to T-scores with a mean of 50 and a standard deviation of 10.

**Results:** A common scale was created for 343 915 individuals above 18 years of age from 16 studies. The scale showed solid evidence of concurrent validity regarding various sociodemographic, life and health factors, and convergent validity with healthy life expectancy ( $r=0.81$ ) and gross domestic product ( $r=0.58$ ). Survival curves showed that the scale could also be predictive of mortality.

**Conclusions:** The ATHLOS scale, due to its reliability and global representativeness, has the potential to contribute to worldwide research on healthy ageing.

**Key words:** Healthy ageing, scale, functional ability, intrinsic capacity, item response theory, data integration

### Key Messages

- This study used an item response theory approach to develop a common scale for measuring healthy ageing, based on data harmonized and integrated from 16 international cohorts.
- The scale can measure the biopsychosocial aspects of health and functioning, since it covers domains of vitality, sensory functions, locomotion, cognition and activities of daily living that imply interaction with the individual's environment.
- The scale is intended to be universally applicable for evaluating healthy ageing, as it arises from the use of international cohorts covering 38 countries from all populated continents.
- Notwithstanding efforts at integration, as far as we know, no other study has yet produced a common measurement approach, based concomitantly on the combination of intrinsic capacity and functional ability, for assessing internationally healthy ageing.
- The development of this scale may help researchers and policy makers to have a better understanding of healthy ageing and will move forward in epidemiological research of healthy ageing.

## Introduction

Population ageing poses enormous challenges to the social welfare state as a result of greater needs for the elderly's health and social care.<sup>1,2</sup> To address this issue, research initiatives have analysed the concept of healthy, active or successful ageing internationally at several levels.<sup>3,4</sup> However, few of these research efforts have adequately tackled the complex challenge of addressing the concept of healthy ageing, since international consensus regarding how healthy ageing should be measured, to account for the diversity of populations globally, has not been achieved to date.<sup>5-7</sup> Existing indices or scores of healthy ageing address different concepts and are limited to specific populations.<sup>4-9</sup> Hence, efforts are needed to obtain a validated universally applicable measurement tool.<sup>7</sup>

The World Health Organization (WHO) proposed in 2015 to define healthy ageing as an 'ongoing process of developing and maintaining the functional ability that enables well-being in older age'.<sup>2,6</sup> This framework has moved away from focusing on the presence of disease experienced at a single time point to considering healthy ageing as a function of an individual's functional ability over time. Functional ability is determined by the interaction of individuals' intrinsic capacity and their environment. In turn, an individual's intrinsic capacity is comprehensively considered by addressing all physical and mental capacities; and the environment should at least include access to medications, personal support, assistive devices and physical barriers that may be either facilitate or hinder functional ability.<sup>10</sup> To this effect, a measure which combines an individual's intrinsic capacity and functional ability may be able to more broadly capture a person's healthy ageing level. In addition, such a measure may have the ability to stimulate more effective prevention strategies by fostering either intrinsic capacity or the resulting functional ability through environmental interventions.

Epidemiological studies on ageing tend to collect heterogeneous information on biopsychosocial aspects of health and functioning. Integrating data from multiple cohort studies can be a viable way to combine knowledge gained with a sustainable methodology and provide a nuanced understanding of ageing in different populations. It increases sample size and improves statistical power to accurately estimate health outcomes and their determinants. Additionally, it facilitates comparisons within and across study populations due to variety in geography, composition, socioeconomic status and other factors of interest. This provides significant opportunities for researchers to pool data from multiple studies and conduct data analyses

simultaneously.<sup>11</sup> Some key harmonization and integration activities have been conducted as the Health and Retirement Studies family, the Integrative Analysis of Longitudinal Studies on Aging, or the WHO Study on Global Ageing and Adult Health.<sup>12-14</sup> Yet there is a need to develop a common approach that will facilitate temporal and regional assessments of healthy ageing across diverse populations.

This study is based on the Ageing Trajectories of Health-Longitudinal Opportunities and Synergies (ATHLOS) project, which has produced a large harmonized dataset from 38 countries from all populated continents.<sup>15</sup> This study aims to develop a novel scale measuring healthy ageing using items about intrinsic capacity and functional ability and to provide evidence of validity. The scale is intended to be universal, since individual data from any one study can be used to estimate healthy ageing scores comparable to all of them.

## Methods

The study protocol was approved by the Committee on the Ethics of Clinical Research, CEIC Fundació Sant Joan de Déu (protocol number: PIC-22-15). All data were anonymized and electronic health record confidentiality was respected in accordance with national and international law.

## Data sources

The ATHLOS cohort is composed of harmonized datasets of international cohorts related to health and ageing. To this effect, data from the following 16 studies were considered: the 10/66 Dementia Research Group Population-Based Cohort Study (10/66)<sup>16</sup> with waves 1 and 2; the Australian Longitudinal Study of Aging (ALSA)<sup>17</sup> from wave 1 to 13; the China Health and Retirement Longitudinal Study (CHARLS)<sup>18</sup> with waves 1 and 2; Collaborative Research on Ageing in Europe (COURAGE)<sup>19</sup> with waves 1 and 2; the English Longitudinal Study of Ageing (ELSA)<sup>20</sup> from wave 1 to 7; Study on Cardiovascular Health, Nutrition and Frailty in Older Adults in Spain (ENRICA)<sup>21</sup> from wave 1 to 3; the Health, Alcohol and Psychosocial factors in Eastern Europe Study (HAPIEE)<sup>22</sup> with waves 1 and 2; the Health 2000/2011 Survey (H2000/11)<sup>23</sup> with waves 1 and 2; the Health and Retirement Study (HRS)<sup>24</sup> from wave 1 to 11; the Japanese Study of Aging and Retirement (JSTAR)<sup>25</sup> from wave 1 to 3, the Korean Longitudinal Study of

Ageing (KLOSA)<sup>26</sup> from wave 1 to 4; the pilot-study Longitudinal Aging Study in India (LASI);<sup>27</sup> the Mexican Health and Aging Study (MHAS)<sup>28</sup> from wave 1 to 3; the Study on Global Ageing and Adult Health (SAGE)<sup>29</sup> with only wave 1; the Survey of Health, Ageing and Retirement in Europe (SHARE)<sup>30</sup> from wave 1 to 5; and the Irish Longitudinal Study of Ageing (TILDA)<sup>31</sup> with waves 1 and 2. The above studies include populations from 38 countries across five continents.

The data harmonization aimed to transform study-specific variables in a homogeneous definition and format across studies.<sup>32</sup> More detailed information about the harmonization process in the ATHLOS project can be found elsewhere.<sup>15</sup>

### Intrinsic capacity and functional ability items

To develop the healthy ageing scale, the ATHLOS consortium agreed on a comprehensive list of 41 items related to intrinsic capacity and functional ability, covering the biopsychosocial aspects of health and functioning usually found in general population surveys.<sup>33</sup> All items were assessed across studies and successfully harmonized in at least three studies. Study-specific variables were harmonized into dichotomous items expressing the presence or absence of difficulties. Continuous variables were dichotomized in the first quartile, indicating the presence of difficulties. The harmonization process of each item can be found at [Supplementary Table S1](#), available as [Supplementary data](#) at *IJE* online.

### Data selection

The sample size used to construct the scale included all individuals above 18 years of age. We selected all individuals with available data on at least one of the 41 items and used their earliest observed assessment.

### Statistical analysis

We developed the healthy ageing scale using item response theory (IRT) models.<sup>34</sup> We chose the two-parameter logistic IRT model, where the probability of endorsing a response category is modelled as a function of two item parameters, item discrimination and item difficulty, and a person parameter. To test the adequacy of the model as a measurement scale, its fit was assessed using the Root Mean Square Error of Approximation (RMSEA; good fit <0.06), the Comparative Fit Index (CFI; good fit >0.95) and the Tucker-Lewis index (TLI; good fit >0.95).<sup>35</sup> The initial estimation of item and person parameters (calibration) was obtained by applying full information maximum

likelihood estimation on the sample matrix of response patterns. The score of each individual was calculated using the expected *a posteriori* estimation method. Maximum score reliability and model marginal reliability were assessed.<sup>36</sup>

As each study had a different subset of available items, differences between the scores on the scale and scores on study-specific scales were assessed using intraclass correlation coefficients (ICC). High ICC values indicate that the scoring system is stable to obtain same person scores despite different item subsets.

To establish the homogeneity of the scale across studies, a logistic regression framework was used to detect differential item functioning (DIF), which indicates whether items are measured in the same way for all studies.<sup>37</sup> If a study had any items exhibiting DIF, specific parameters were estimated applying the overall model in the study-specific sample. To rescale specific parameters to the full sample, scaling parameters were linearly moved using the Stocking-Lord equating approach.<sup>38</sup> Thus, the procedure took into account study-specific IRT parameters for items with DIF by equating study-specific parameters to the main scale using a test characteristic curve equating procedure.<sup>39</sup> The resulting scores were converted to T-scores with a mean of 50 and a standard deviation of 10. These constituted the final healthy ageing scores of the sample based on the main scale.

To check concurrent validity of the scale, weighted linear trends of the T-score means by age groups were assessed for each study. Second, a multiple linear regression model was carried out to investigate whether the T-scores were associated with known risk factors for poor health in adulthood. The following variables were included, which were all self-reported by the participant: sociodemographic variables such as country, sex, age, year of birth, education and wealth; lifestyles such as smoking and obesity; most frequent physician-diagnosed diseases such as high blood pressure, cardiovascular disease, respiratory disease, diabetes, joint disorders; one frequent disorder such as depression that was based on the psychometric instrument of each study; and one important social environment such as loneliness which was based in some studies on a symptom of the depression instrument and in others on questions about the feeling of loneliness. Variables with too many missing structural values were avoided because some studies were not found or were not valid to be harmonized. Predictive validity was supported using the harmonized age at death variable in those studies with follow-up and provided by each study. Kaplan-Meier survival estimators of T-scores categorized into four groups (less than or equal to 40, 41–50, 51–60 and 61 or higher) and time from participant assessment to death

were plotted. Finally, to check convergent validity, each country's T-score means were compared with two established indicators for health and wealth: the Healthy Life Expectancy (HALE) at birth and the Gross Domestic Product (GPD) per capita, both from the year 2016 in each country.<sup>40</sup> Potential sex, age and birth cohort effects were previously removed by calculating the difference between T-scores and predicted T-scores from a linear regression adjusted by sex, age, year of birth and their interactions.

All analyses were conducted with the statistical software R and appended as [Supplementary materials](#), available as [Supplementary data](#) at IJE online.

## Results

The sample size was 343 915 individuals, 55% were female and the median age was 60 years (see [Table 1](#)). This sample size was 98.9% of the total eligible participants who had information on at least one item. Of the remaining 1.1% (3978 individuals), 60% were from the SAGE study. Studies that provided the largest sample sizes were SHARE (30.8%), SAGE (12.3%) and HRS (10.6%). Years of interview were mostly between 2000 and 2015 except in ALSA and HRS, which started in the early 90s.

In HRS, new cohorts were refreshed over the years. Medians of years of birth were between 1940 and 1950 except in the studies 10/66, ALSA and HRS, which included individuals who were born earlier.

The IRT model converged successfully with an excellent fit (RMSEA = 0.03, TLI = 0.99 and CFI = 0.99). The IRT parameter estimates showed that daily activity items had the highest values for discrimination and cognitive items had the lowest values (see [Table 2](#)).

T-scores were computed for all the individuals, with high scores indicating healthier ageing. The T-scores range was from 12 to 69, left-skewed with a mean of 50.2 and a standard deviation of 10. The model had maximum reliability of 0.975 at the T-score 35.3, with reliability over 0.90 from T-scores 23.2 to 48.5, and a model marginal reliability of 0.83. ICCs between the T-scores for the main scale and each study-specific subset of items were higher than 0.89 (see [Table S1](#)).

Items with DIF were found in three or fewer studies, except for the item 'energy' that presented DIF in 6 studies. On the other hand, ENRICA, HRS, and MHAS were the only studies without any items with DIF. All others exhibited from 1 to 8 items with DIF. 10/66, ALSA and SAGE were the studies with the highest proportion of items with DIF (see [Table S1](#)).

**Table 1** Descriptive statistics of sex, age, years of interview and birth by study of participants included in the construction of the healthy ageing scale

Studies	Sample size		Females %	Age			Years of interview			Years of birth		
	N	%		min	med	max	min	med	max	min	med	max
10/66	16886	4.9	62	65	73	110	2001	2005	2010	1896	1932	1944
ALSA	2087	0.6	49	64	78	103	1992	1992	1993	1889	1914	1927
CHARLS	20273	5.9	52	19	57	101	2011	2011	2013	1910	1954	1993
COURAGE	10780	3.1	57	18	60	104	2011	2011	2012	1903	1951	1994
ELSA	17984	5.2	55	19	59	94	2002	2002	2015	1908	1944	1987
ENRICA	2519	0.7	53	60	67	93	2008	2009	2010	1915	1941	1950
HAPIEE	26664	7.8	53	44	58	75	2002	2004	2008	1932	1945	1962
HRS	36320	10.6	56	18	56	103	1992	1994	2013	1890	1938	1992
H2000-11	8417	2.4	55	30	49	101	2000	2000	2012	1900	1951	1981
JSTAR	7105	2.1	52	46	63	77	2007	2007	2011	1930	1945	1964
KLOSA	10 254	3.0	57	45	61	105	2006	2006	2006	1901	1945	1961
LASI	1413	0.4	56	21	53	102	2010	2010	2010	1907	1956	1989
MHAS	19 848	5.8	44	18	57	114	2001	2001	2012	1895	1946	1992
SAGE	42 268	12.3	57	18	58	114	2007	2007	2010	1893	1949	1991
SHARE	105 829	30.8	56	22	62	104	2004	2011	2013	1900	1946	1991
TILDA	8463	2.5	56	49	62	82	2010	2010	2012	1930	1948	1961
All	343 915	100	55	18	60	114	1992	2007	2015	1889	1945	1994

10/66, 10/66 Dementia Research Group Population-Based Cohort Study; ALSA, Australian Longitudinal Study of Ageing; CHARLS, China Health and Retirement Longitudinal Study; COURAGE, Collaborative Research on Ageing in Europe; ELSA, English Longitudinal Study of Ageing; ENRICA, Seniors-ENRICA; HAPIEE, Health, Alcohol and Psychosocial factors In Eastern Europe study; HRS, Health and Retirement Study; H2000-11, Health 2000/2011 study; JSTAR, Japanese Study of Aging and Retirement; KLOSA, Korean Longitudinal Study of Ageing; LASI, Longitudinal Aging Study in India; MHAS, Mexican Health and Aging Study; SAGE, Study on Global Ageing and Adult Health; SHARE, Survey of Health, Ageing and Retirement in 20 countries from Europe; TILDA, Irish Longitudinal study on Ageing; min, minimum; med, medium; max, maximum.



**Table 2** List of domains and items related to functioning and health and their parameter estimates from the item response theory (IRT) model

Domains	Presence or absence of difficulties	IRT parameter estimates and standard errors				
		Discrimination		Difficulty		
Cognition	Memory	0.8332	(0.0059)	-0.5914	(0.0065)	
	Immediate recall <sup>a</sup>	0.6325	(0.0049)	-0.9731	(0.0092)	
	Delayed recall <sup>a</sup>	0.6714	(0.0051)	-1.2130	(0.0100)	
	Verbal fluency <sup>a</sup>	0.6080	(0.0057)	-1.6981	(0.0160)	
	Orientation in time	0.8449	(0.0078)	-1.7307	(0.0155)	
	Processing speed <sup>a</sup>	0.5912	(0.0144)	-1.9291	(0.0453)	
	Numeracy <sup>a</sup>	1.0404	(0.0118)	-1.9586	(0.0203)	
	Psychology symptoms	Sleeping	0.8334	(0.0052)	-0.5605	(0.0057)
Vitality	Experiences some degree of pain	1.0616	(0.0059)	-0.1463	(0.0042)	
	Having high level of energy	0.9119	(0.0054)	-0.5781	(0.0054)	
Sensory functions	Urinary incontinence	1.0969	(0.0111)	-2.3546	(0.0196)	
	Near vision	0.9438	(0.0061)	-0.9104	(0.0069)	
	Far vision	1.2639	(0.0075)	-1.1091	(0.0062)	
	Eyesight using glasses or lens as usual	0.9421	(0.0086)	-1.5122	(0.0126)	
	Hearing in general	0.8212	(0.0067)	-1.9818	(0.0151)	
	Hearing in a conversation	0.8426	(0.0107)	-2.2121	(0.0262)	
	Locomotion/mobility	Stooping, kneeling or crouching	2.4717	(0.0120)	-0.5059	(0.0029)
		Lifting or carrying weights	2.7130	(0.0134)	-0.7834	(0.0031)
Climbing stairs		2.7327	(0.0137)	-0.7940	(0.0031)	
Getting up from sitting down		2.4166	(0.0125)	-0.8256	(0.0035)	
Walking by yourself and without any equipment		3.1335	(0.0161)	-0.8763	(0.0030)	
Pulling or pushing large objects		3.1691	(0.0202)	-0.8848	(0.0036)	
Sitting for long periods		2.0455	(0.0114)	-1.1322	(0.0047)	
Reaching or extending arms		2.1929	(0.0129)	-1.4727	(0.0054)	
Walking speed <sup>a</sup>		0.8995	(0.0111)	-1.6364	(0.0160)	
Dizziness when walking on a level surface		1.3230	(0.0128)	-1.7363	(0.0142)	
Activities of daily living	Picking up things with fingers	2.3139	(0.0156)	-1.8427	(0.0068)	
	Getting in or out of bed	3.4954	(0.0239)	-1.4949	(0.0044)	
	Bathing or showering	3.5997	(0.0253)	-1.5996	(0.0045)	
	Getting dressed	2.6935	(0.0180)	-1.6404	(0.0055)	
	Moving around the home	3.4746	(0.0271)	-1.6976	(0.0053)	
	Using the toilet	3.5668	(0.0277)	-1.7960	(0.0054)	
	Eating	3.0627	(0.0255)	-2.0989	(0.0073)	
	Instrumental activities of daily living	Doing housework	3.1663	(0.0185)	-1.0530	(0.0035)
Shopping for groceries		4.3188	(0.0376)	-1.3521	(0.0045)	
Getting out of the house		3.4506	(0.0377)	-1.3939	(0.0056)	
Difficulties in preparing meals		3.8589	(0.0360)	-1.5886	(0.0058)	
Using a map		1.6814	(0.0149)	-1.6329	(0.0108)	
Managing money, bills or expenses		2.4045	(0.0266)	-1.8498	(0.0107)	
Taking medications		3.0022	(0.0311)	-2.0162	(0.0096)	
Making telephone calls	2.9271	(0.0332)	-2.0392	(0.0108)		

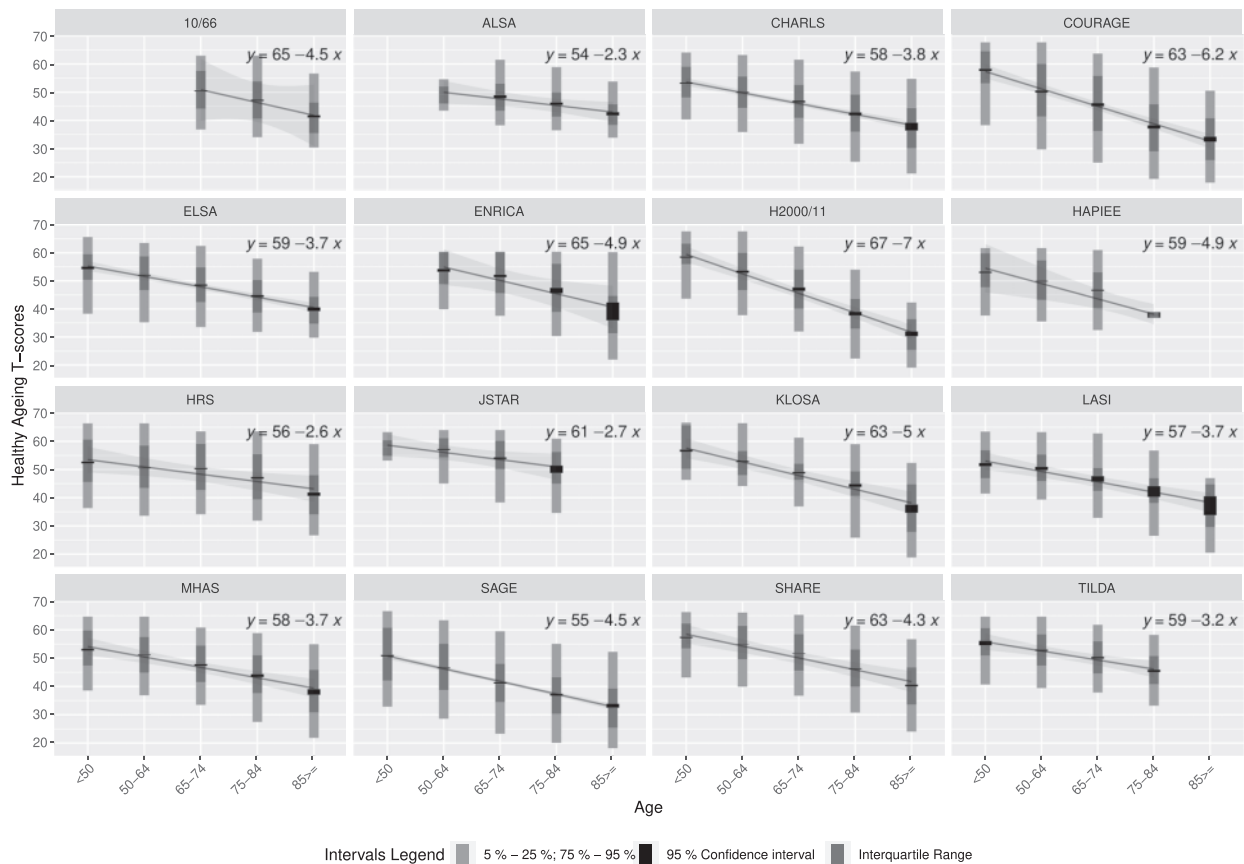
<sup>a</sup>Item was dichotomized in the first quartile indicating presence of difficulties.

In the linear transformation of the equating approach, additive parameter estimates ranged from -0.15 to 0.11, and the multiplicative parameter estimates ranged from 0.90 to 1.11. Studies without items with DIF had the same scores (see Table S1).

In all studies, we observed that T-scores were lower in each older group (see Figure 1). COURAGE and H2000/11

had the highest decreasing slopes, -6.2 and -7.0, respectively, which indicate the number of T-score units that decreased in each older group.

Regarding the results of the multiple linear regression, we found that males and/or individuals with higher educational level, greater wealth and never smoking had higher T-scores. In contrast, obesity, arterial hypertension,



**Figure 1** Trends of healthy ageing T-scores by age groups in each study

depression, physical diseases and loneliness were associated with lower T-scores (see [Table 3](#)).

The studies H2000/11, MHAS and TILDA did not provide mortality information for reasons of confidentiality, and LASI and SAGE had available only one wave. In the Kaplan-Meier estimations, mortality risk was higher for each lower T-score group throughout the observed time period. The group with the lowest T-scores had a 50% survival probability in 10 years and for the other groups it was in at least 20 years (see [Figure 2](#)).

Graphical representations of T-scores by country show that cohorts from Switzerland (mean of 56.5), Japan (55.6) and Denmark (55.4) had the highest T-scores (see [Figure 3](#)). In contrast, cohorts from Ghana (40.4), India (40.7) and Russia (42.7) had the lowest T-scores. Correlations between T-score means by country and ecological country indicators were 0.81 with HALE and 0.58 with GDP (see [Figure 4](#)).

## Discussion

In this study, we developed a scale for measuring healthy ageing comprising 41 items of intrinsic capacity and functional ability, by employing harmonized data from 16

international ageing cohorts. The IRT model resulted in excellent reliability ( $>0.90$ ) in T-scores from 23.5 to 48.5, with marginal reliability of 0.83, rendering the model adequate for group comparisons. Concurrent validity of the scale with sociodemographic, life and health factors, and predictive validity with mortality, have shown that this scale corresponds well with health status and could be potentially useful for conducting international ageing research.

We found that the scale was related to HALE, which is known to differ between countries. Regarding GDP, countries' well-being is known to be a determinant of health.<sup>41</sup> The fact that the scale is sensitive to differences in GDP between countries indicates that it is a potential health outcome useful for aggregate comparisons. However, these comparisons are perhaps most interesting when examining what is left behind. For example, why do countries with a high HALE, such as Spain or South Korea, or with a high GDP, such as the USA or Ireland, show worse ageing outcomes than countries such as Denmark or Peru? Such observations provide the basis for further studies and thought-provoking hypotheses. Unfortunately, they are far beyond the scope of this study.

Several scales are currently available for measuring specific aspects of health and ageing, albeit none concomitantly

**Table 3** Multiple linear regression between the T-scores of the healthy ageing scale and sociodemographic, life and health factors, adjusted by country<sup>a</sup>

Variables	Mean (SD <sup>b</sup> ) or %	Standardized coefficients	95% confidence interval
Age	62 (12)	-4.25	(-4.36, -4.15)
Year of birth	1944 (13)	-1.65	(-1.76, -1.54)
Age x year of birth	..	0.30	(0.28, 0.31)
Sex (reference: females)	55	0	..
Males	44	0.83	(0.81, 0.86)
Missing	1	-0.10	(-0.13, -0.07)
Education (reference: less than primary)	13	0	..
Primary	22	0.25	(0.21, 0.30)
Secondary	43	1.04	(0.99, 1.09)
Tertiary	15	1.27	(1.23, 1.31)
Missing	7	-0.42	(-0.46, -0.39)
Wealth <sup>c</sup> (reference: 1st quintile = less wealthy)	18	0	..
2nd quintile	16	0.18	(0.15, 0.21)
3rd quintile	17	0.34	(0.31, 0.37)
4th quintile	17	0.51	(0.48, 0.54)
5th quintile	18	0.75	(0.71, 0.78)
Missing	14	0.53	(0.49, 0.58)
Smoking (reference: never)	49	0	..
Past	23	-0.14	(-0.17, -0.11)
Currently	20	-0.33	(-0.36, -0.30)
Missing	8	0.24	(0.20, 0.28)
Obesity (reference: no)	66	0	..
Yes	17	-0.66	(-0.69, -0.64)
Missing	17	0.68	(0.63, 0.72)
Arterial hypertension (reference: no)	54	0	..
Yes	36	-0.33	(-0.35, -0.30)
Missing	10	-0.35	(-0.39, -0.30)
Depression (reference: no)	70	0	..
Yes	19	-2.05	(-2.08, -2.02)
Missing	11	-1.45	(-1.48, -1.41)
CVD <sup>d</sup> (reference: no)	83	0	..
Yes	14	-1.23	(-1.26, -1.20)
Missing	3	0.33	(0.29, 0.36)
Respiratory disease <sup>e</sup> (reference: no)	89	0	..
Yes	9	-0.76	(-0.79, -0.74)
Missing	2	-0.08	(-0.12, -0.04)
Diabetes (reference: no)	66	0	..
Yes	11	-0.52	(-0.55, -0.49)
Missing	23	0.52	(0.48, 0.56)
Joint disorders <sup>f</sup> (reference: no)	69	0	..
Yes	22	-1.70	(-1.73, -1.67)
Missing	9	-0.60	(-0.66, -0.55)
Loneliness (reference: no)	57	0	..
Yes	17	-0.55	(-0.58, -0.52)
Missing	26	-0.004	(-0.044, 0.036)
Adjusted R-squared <sup>g</sup> = 0.42			

<sup>a</sup>Variable 'country' is categorical with 38 countries.

<sup>b</sup>SD: standard deviation.

<sup>c</sup>Wealth: quantiles of household incomes and asset information from participants within their country cohort.

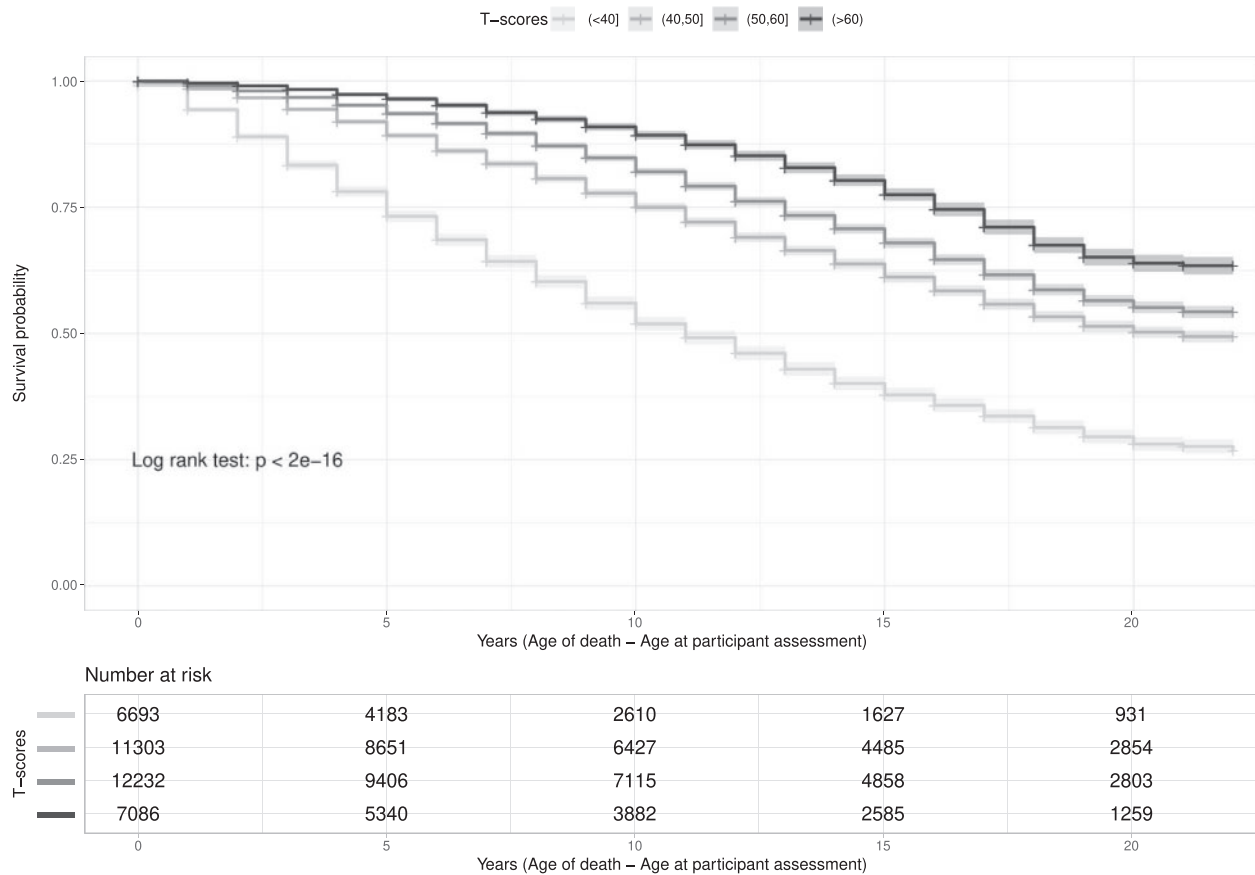
<sup>d</sup>At least one of the following cardiovascular diseases (CVD): angina, stroke, myocardial infarction, heart attack, coronary heart disease, congestive heart failure, heart murmur, valvular disease, cerebral vascular disease.

<sup>e</sup>At least one of the following respiratory diseases: asthma, chronic obstructive pulmonary disease, bronchitis, emphysema.

<sup>f</sup>At least one of the following joint disorders: arthritis, rheumatism, osteoarthritis.

<sup>g</sup>R-squared refers to the coefficient of determination.





**Figure 2** Kaplan-Meier curves of healthy ageing T-scores categorized into four groups (<40, 40–49, 50–59, >60) and time from participant assessment to death

including the comprehensive assessment of intrinsic capacity and functional ability. For example, the WHODAS 2.0 scale has been widely used to assess individual disability at population levels or in clinical practice.<sup>42</sup> However, it focuses on the individual's functioning in interaction with the social environment, thus assessing the functional ability of a person rather than intrinsic capacity. Similarly, the Active Ageing Index, which covers diverse aspects of active and healthy ageing particularly in European populations, focuses on individuals' functional ability as well as interaction with social and labour environments, rather than on intrinsic capacity.<sup>8</sup> On the other hand, a recent composite ageing measure arising from the ELSA study included only intrinsic capacity.<sup>9</sup> The latter showed predictive capacity for measuring an individual's functioning, thus separating the concepts of intrinsic capacity and functional ability. Hence, the available scales are limited to specific populations and separate or merge some aspects of intrinsic capacity and functional ability with other domains used to describe ageing: for instance, physiological and physical health, personal perception and social environment, among others. It must be underlined that most existing cohort studies were not designed to measure intrinsic capacity and

functional ability separately. For example, self-reported vision problems might reflect both intrinsic capacity (those who do not use glasses) and functional ability (those who use glasses). Therefore, it is more appropriate to incorporate all measures related to intrinsic capacity and functional ability, to capture healthy ageing when using existing cohort data.

The ATHLOS scale was based on the IRT modelling framework that allows for estimating item parameters that are independent of person scores in the sample from which they are obtained.<sup>34</sup> This entails that individual scores from the IRT measurement scale are comparable between individuals from different studies responding to different item subsets. Moreover, even though measurement biases can occur, IRT modelling allows testing and adjusting the effect of potential confounders on individual responses, such as specific effects of cohorts, gender and cultural factors. Using modelling differential item functioning, it is possible to obtain directly comparable measures across groups.<sup>43,44</sup> IRT also provides the means to scores in a universal reference, yielding the possibility of rescaling individual scores obtained in a group to an arbitrary reference scale of choice.<sup>43,45</sup>

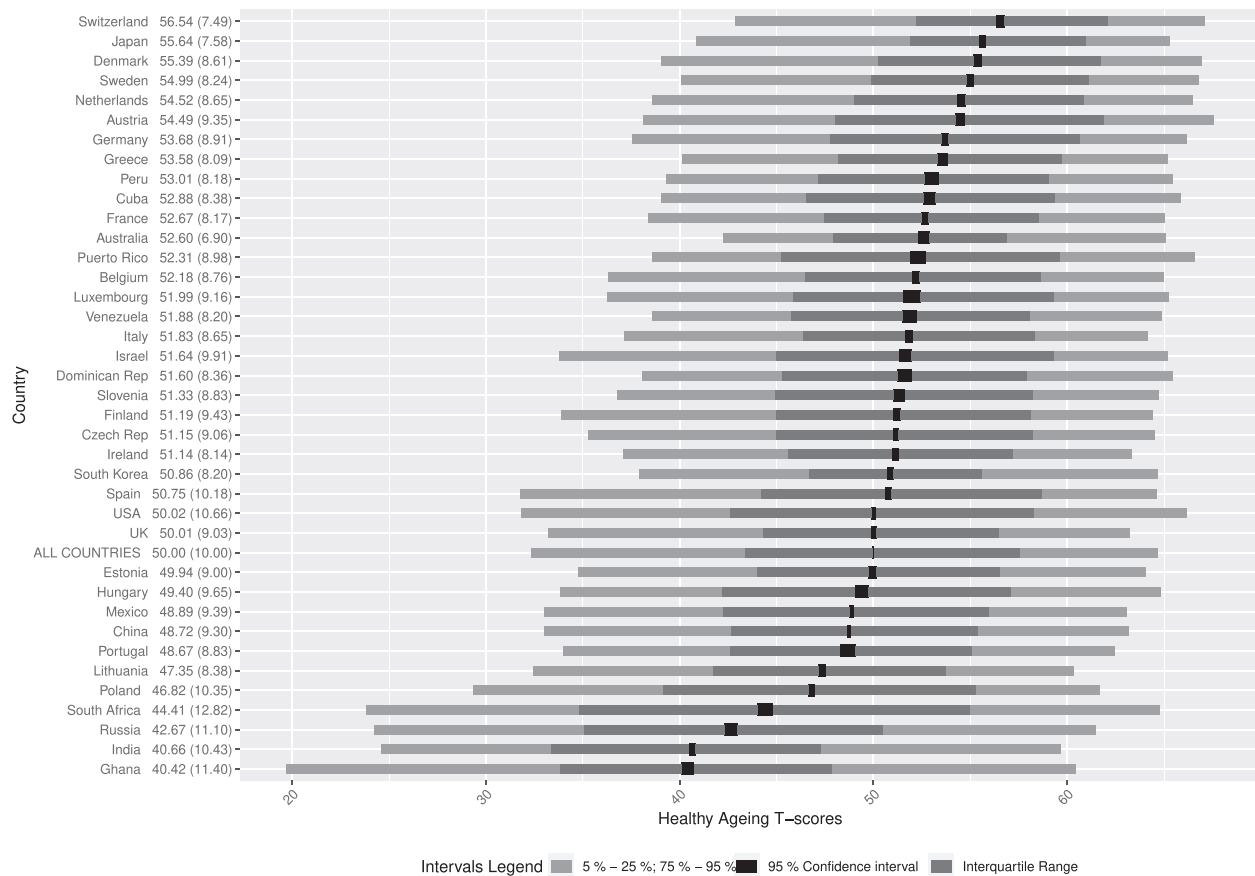
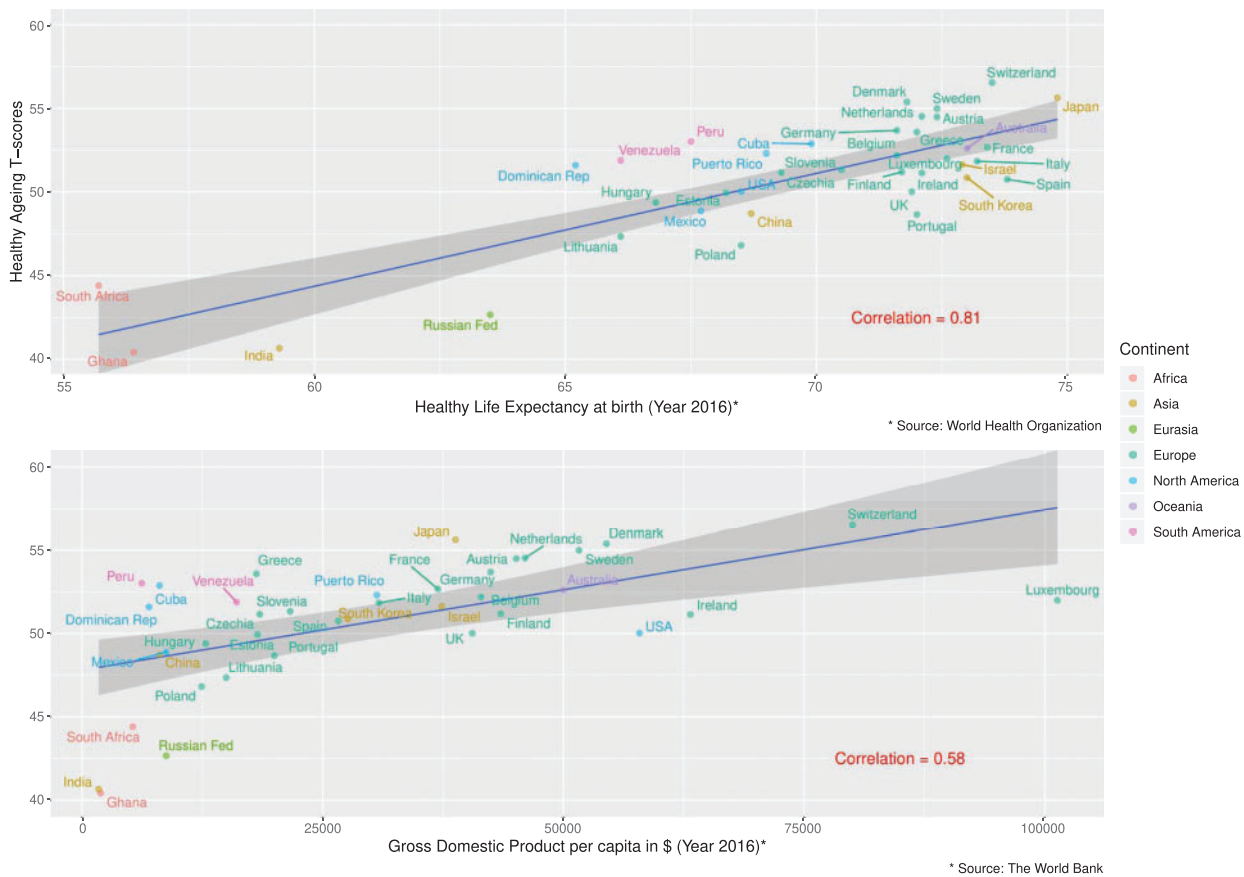


Figure 3 Distributions of healthy ageing T-scores by country

Methodological attempts to create the scale with similar subsets of items were successfully conducted by using only some specific studies within the ATHLOS project.<sup>46–49</sup> However, the integration of multiple independent samples in the ATHLOS dataset had the challenge of obtaining and harmonizing a large number of intrinsic capacity and functional ability items that are difficult to collect with individual samples. Having information about many items makes it more feasible to capture the diverse heterogeneity of the individual’s healthy ageing. Moreover, the IRT model overcomes the potential obstacle of the presence of incomplete data, wherein some items serve as anchors for equating responses in different studies.<sup>50</sup> Scores of individuals from smaller subsets of items should obtain the score estimates with the full subset of items, but with higher measurement error. Nevertheless, the ICC that was conducted showed that the scoring with different subsets of items was well correlated to scores from the main scale. Furthermore, in the case of choosing a subset of items for a new study, priority should be given to those with the highest IRT discrimination parameters, such as ADL/IADL (activities of daily living, instrumental activities of daily living) items. Using these items, it would be possible to establish a

minimum set of information for measuring the ageing status of individuals.

Our results must be interpreted taking into account that observational data can introduce selection and information biases. The data come from the general population and therefore the scale may under-represent people with greater dependency living in nursing homes or other facilities, or with greater cognitive impairment. These biases can be aggravated when integrating data from studies with different sampling designs, questionnaires and ways of asking for the same information. In addition, the harmonization process had to address multiple issues of heterogeneity across studies. Items had to be dichotomized to include the maximum number of items per study. For instance, similar questions with different levels required choosing the cut-off point for dichotomization. Harmonizing questions addressing the same type of difficulty (e.g. sleep was restless versus difficulty sleeping) was more difficult. For this reason, potential sources of DIF are usually studied by subgroups as a means of detecting heterogeneity when obtaining an item. Future research should address other potential sources of DIF such as sex and age, even though in preliminary analyses there were already very few items with DIF.



**Figure 4** Healthy ageing T-score means by country compared with the indicators 'healthy life expectancy' and 'gross domestic product', respectively

From the list of 41 items covering the biopsychosocial aspects of health and functioning, others may be missing such as lung function, grip strength or even biomarkers of immune function. These items, which may already be quite correlated with the items included, are usually less common and would therefore add more complexity to the scale. Other items more related to depressive/anxious mental health symptoms, such as feelings of sadness or loneliness, were considered candidates for inclusion in the scale. However, in carrying out the analyses, these elements introduced new dimensions related to psychopathology. As it would have been more difficult to interpret a single measure of ageing, we chose to exclude these items from the scale, leaving them aside for further research for the diagnosis of affective disorders in ageing studies.

Healthy ageing scores have become useful tools in daily clinical practice for patient prognosis, as well as for the development of future public health strategies with the globally rapid pace of ageing. It is a fact that accuracy is the cornerstone of this kind of score, so the wide use of a scale like this in populations with divergent ethnic, genetic, social and cultural characteristics, and hence variable risk factors, could lead to a specific variability in the prediction

of healthy ageing. However, through our methodological approach, the accuracy and the validity of the healthy ageing estimation models, using multinational populations and using different variables, represents an important topic in the field of healthy ageing and is the first step towards understanding the complex process of ageing.

We believe that the development of this scale will make it possible to move forward the epidemiological research of healthy ageing. This single scale can then be used across studies conducted internationally. When applied to longitudinal data, these scores may provide reliable measures of healthy ageing scores over time. In this way, the ATHLOS scale can identify patterns of healthy ageing trajectories and their determinants, and critical points in time when changes in trajectories are induced, enabling the design and implementation of timely clinical and public health interventions to optimize and promote healthy ageing.

All the studies included in the ATHLOS project gave permission for the secondary use of their data by the ATHLOS Consortium. Data may be shared on request to the corresponding author with permission of the ATHLOS consortium.

## Supplementary Data

Supplementary data are available at *IJE* online

## Funding

This work was supported by the 5-year Ageing Trajectories of Health: Longitudinal Opportunities and Synergies (ATHLOS) project. The ATHLOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635316. See [Supplementary material](#) for more information about funding for each study.

## Acknowledgements

The authors thank the ATHLOS Consortium for useful discussions, and gratefully acknowledge the funding of institutions and the work of people who carried out the studies and provided data for this paper. See [Supplementary material](#) for acknowledgements for each study. A.R. is supported by a grant from the Italian Ministry of Health (Ricerca Corrente, Fondazione Istituto Neurologico C. Besta, Linea 4-Outcome Research: dagli Indicatori alle Raccomandazioni Cliniche).

## Author Contributions

A.S.-N. led the work on harmonizing and integrating the ATHLOS database, carried out the statistical analyses, interpreted the results and drafted the paper. C.G.F. guided the statistical analyses, interpreted the results and critically reviewed the paper. I.G.-V. participated in the database management and the statistical analyses and interpreted the results. Y.-T.W., C.D., J.dF., A.dT. and F.F.C. participated in the study design, database management and statistical support and critically reviewed the paper. H.A., I.B.-M., E.C., L.E.-C., E.G.-E., B.M.-M., A.R. and S.T. participated in the database management and critically reviewed the paper. M.Prina, J.L.A.-M., M.B., S.C., M.L., S.K., B.O., A.P., M.Prince, F.R.-A., A.T. and B.T.-A. participated in the acquisition of data and in the study design and critically reviewed the paper. D.P., J.B., I.K., W.S., S.S. and J.M.H. participated in the study design and critically reviewed the paper. All authors gave final approval of the version to be published.

## Conflict of Interest

None declared.

## References

1. United Nations Department of Economic and Social Affairs. World Population Prospects 2019: *Highlights*. *Multimedia Library*. 2019. . [https://population.un.org/wpp/Publications/Files/WPP2019\\_Highlights.pdf](https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf) (24 October 2019, date last accessed).
2. World Health Organization. World Report on Ageing and Health. 2015. [www.who.int](http://www.who.int) (18 July 2018, date last accessed).
3. Cosco TD, Prina AM, Perales J, Stephan BCM, Brayne C. Operational definitions of successful aging: a systematic review. *Int Psychogeriatr* 2014;**26**:373–81.

4. Lu W, Pikhart H, Sacker A. Domains and measurements of healthy aging in epidemiological studies: a review. *Gerontologist* 2019;**59**:e294–310.
5. Michel J-P, Sadana R. 'Healthy Aging' concepts and measures. *J Am Med Dir Assoc* 2017;**18**:460–64.
6. Beard JR, Officer A, De Carvalho IA *et al*. The World Report on Ageing and Health: a policy framework for healthy ageing. *Lancet* 2016;**387**:2145–54.
7. Michel J-P, Graf C, Ecarnot F. Individual healthy aging indices, measurements and scores. *Ageing Clin Exp Res* 2019;**31**: 1719–25.
8. Zaidi A, Gasior K, Zolyomi E, Schmidt A, Rodrigues R, Marin B. Measuring active and healthy ageing in Europe. *J Eur Soc Policy* 2017;**27**:138–57.
9. Beard JR, Jotheeswaran AT, Cesari M, Araujo de Carvalho I. The structure and predictive value of intrinsic capacity in a longitudinal study of ageing. *BMJ Open* 2019;**9**:e026119.
10. Cesari M, De Carvalho IA, Thiyagarajan JA *et al*. Evidence for the domains supporting the construct of intrinsic capacity. *J Gerontol A Biol Sci Med Sci* 2018;**73**:1653–60.
11. Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychol Methods* 2009;**14**: 150–64.
12. Hu P, Lee J. *Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Chronic Medical Conditions*. Santa Monica, CA: RAND Corporation, 2012.
13. Kaye J, Hofer SM. Integrative analysis of longitudinal studies on aging and dementia (IALSA). *Imov Aging* 2017;**1**:1275.
14. Minicuci N, Naidoo N, Corso B, Rocco I, Chatterji S, Kowal P. Data Resource Profile: Cross-national and cross-study sociodemographic and health-related harmonized domains from SAGE plus CHARLS, ELSA, HRS, LASI and SHARE (SAGE+ Wave 2). *Int J Epidemiol* 2019;**48**:14.
15. Sanchez-Niubo A, Egea-Cortés L, Olaya B *et al.*; ATHLOS Consortium. Cohort Profile: The Ageing Trajectories of Health – Longitudinal Opportunities and Synergies (ATHLOS) project. *Int J Epidemiol* 2019;**48**:1052–53.
16. Prina AM, Acosta D, Acosta I *et al*. Cohort Profile: The 10/66 Study. *Int J Epidemiol* 2017;**46**:406.
17. Luszcz MA, Giles LC, Anstey KJ, Browne-Yung KC, Walker RA, Windsor TD. Cohort Profile: The Australian Longitudinal Study of Ageing (ALSA). *Int J Epidemiol* 2016;**45**:1054–63.
18. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). *Int J Epidemiol* 2014;**43**:61–68.
19. Leonardi M, Chatterji S, Koskinen S *et al.*; on behalf of COURAGE in Europe Project's Consortium. Determinants of health and disability in ageing population: the COURAGE in Europe Project (collaborative research on ageing in Europe). *Clin Psychol Psychother* 2014;**21**:193–98.
20. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort Profile: The English Longitudinal Study of Ageing. *Int J Epidemiol* 2013;**42**: 1640–48.
21. Rodríguez-Artalejo F, Graciani A, Guallar-Castillón P *et al*. Rationale and methods of the study on nutrition and cardiovascular risk in Spain (ENRICA). *Rev Española Cardiol* 2011;**64**: 876–82.

22. Peasey A, Bobak M, Kubinova R *et al.* Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: Rationale and design of the HAPIEE study. *BMC Public Health* 2006;6:255.
23. Koskinen S. *Health 2000 and 2011 Surveys—THL Biobank*. Helsinki, Finland: National Institute for Health and Welfare. 2011. <https://thl.fi/fi/web/thl-biobank/for-researchers/sample-collections/health-2000-and-2011-surveys> (20 April 2020, date last accessed).
24. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR, Weir DR. Cohort Profile: The Health and Retirement Study (HRS). *Int J Epidemiol* 2014;43:576–85.
25. Hidehiko I, Satoshi S, Hideki H. *JSTAR First Results 2009 Report*. Tokyo, Japan: Research Institute of Economy, Trade and Industry (RIETI). 2009. <https://www.rieti.go.jp/jp/publications/dp/09e047.pdf> (9 November 2015, date last accessed).
26. Park JH, Lim S, Lim JY *et al.* An overview of the Korean Longitudinal Study on Health and Aging. *Psychiatry Investig* 2007;4:84–95.
27. Arokiasamy P, Bloom DE, Feeney KC, Ozolins M. *Longitudinal Aging Study in India: Vision, Design, Implementation, and Some Early Results*. 2011. <https://lasi.hsph.harvard.edu/publications/longitudinal-aging-study-india-vision-design-implementation-and-some-early-results> (22 June 2018, date last accessed).
28. Wong R, Michaels-Obregon A, Palloni A. Cohort Profile: The Mexican Health and Aging Study (MHAS). *Int J Epidemiol* 2017;46:e2.
29. Kowal P, Chatterji S, Naidoo N *et al.*; the SAGE Collaborators. Data Resource Profile: The World Health Organization Study on global AGEing and adult health (SAGE). *Int J Epidemiol* 2012;41:1639–49.
30. Börsch-Supan A, Brandt M, Hunkler C *et al.* Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol* 2013;42:992–1001.
31. Whelan BJ, Savva GM. Design and methodology of the Irish Longitudinal Study on Ageing. *J Am Geriatr Soc* 2013;61: S265–68.
32. Fortier I, Raina P, Van den Heuvel ER *et al.* Maelstrom Research guidelines for rigorous retrospective data harmonization. *Int J Epidemiol* 2017;46:103–05.
33. Fortier I, Doiron D, Little J *et al.* Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies. *Int J Epidemiol* 2011;40: 1314–28.
34. Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. In: Lord FM, Novick MR (eds). *Statistical Theories of Mental Test Scores*. Boston, MA: Addison-Wesley Reading, 1968, pp. 397–479.
35. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 1999;6:1–55.
36. Thissen D, Wainer H (eds). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.
37. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. *J Educ Measurement* 1990;27:361–70.
38. Kolen MJ, Brennan RL. *Test Equating, Scaling, and Linking: Methods and Practices*. 3rd edn. New York, NY: Springer, 2014.
39. Stocking ML, Lord FM. Developing a common metric in item response theory. *Appl Psychol Meas* 1983;7:201–10.
40. World Health Organization. *Global Health Observatory Data Repository*. <http://apps.who.int/gho/data> (25 November 2019, date last accessed).
41. Marmot M. Social determinants of health inequalities. *Lancet* 2005;365:1099–1104.
42. Federici S, Bracalenti M, Meloni F, Luciano JV. World Health Organization disability assessment schedule 2.0: an international systematic review. *Disabil Rehabil* 2017;39:2347–80.
43. Curran PJ, Hussong AM, Cai L *et al.* Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Dev Psychol* 2008;44:365–80.
44. Curran PJ, Hussong AM. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol Methods* 2009;14: 81–100.
45. Heuvel ER, Griffith LE, Sohel N, Fortier I, Muniz-Terrera G, Raina P. Latent variable models for harmonization of test scores: a case study on memory. *Biom J* 2020;62:34–52. <https://doi.org/10.1002/bimj.201800146>.
46. Caballero FF, Soulis G, Engchuan W *et al.* Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: The ATHLOS project. *Sci Rep* 2017;7:43955.
47. De La Fuente J, Caballero FF, Sanchez-Niubo A *et al.* Determinants of health trajectories in England and the United States: an approach to identify different patterns of healthy ageing. *J Gerontol A Biol Sci Med Sci* 2018;73:1512–18.
48. Daskalopoulou C, Koukounari A, Ayuso-Mateos J, Prince M, Prina A. Associations of lifestyle behaviour and healthy ageing in five Latin American and the Caribbean Countries. A 10/66 population-based cohort study. *Nutrients* 2018;10:1593.
49. Daskalopoulou C, Koukounari A, Wu YT *et al.* Healthy ageing trajectories and lifestyle behaviour: the Mexican Health and Aging Study. *Sci Rep* 2019. doi: 10.1038/s41598-019-47238-w.
50. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;38: 28–42.