Author(s): Qi, Yuwei; Schaap, Laura A.; Schalet, Benjamin D.; Hoogendijk, Emiel O.; Deeg, Dorly J.H.; Visser, Marjolein; Koivunen, Kaisa; Huisman, Martijn; van Schoor, Natasja M.

Title: The development of intrinsic capacity measures for longitudinal research : The Longitudinal Aging Study Amsterdam

Year: 2024

Version: Published version

Please cite the original version:

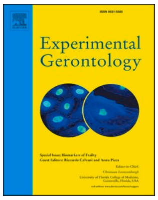
Qi, Y., Schaap, L. A., Schalet, B. D., Hoogendijk, E. O., Deeg, D. J., Visser, M., Koivunen, K., Huisman, M., & van Schoor, N. M. (2024). The development of intrinsic capacity measures for longitudinal research : The Longitudinal Aging Study Amsterdam. Experimental Gerontology, 197, Article 112599. https://doi.org/10.1016/j.exger.2024.112599

Experimental Gerontology 197 (2024) 112599

Contents lists available at ScienceDirect

Experimental Gerontology

journal homepage: www.elsevier.com/locate/expgero



The development of intrinsic capacity measures for longitudinal research: The Longitudinal Aging Study Amsterdam

Yuwei Qi [a,*], Laura A. Schaap [b,c], Benjamin D. Schalet [e,f], Emiel O. Hoogendijk [a,b,g], Dorly J.H. Deeg [a], Marjolein Visser [d], Kaisa Koivunen [h], Martijn Huisman [a,b,i], Natasja M. van Schoor [a,b]

[a] Amsterdam UMC location Vrije Universiteit Amsterdam, Epidemiology and Data Science, the Netherlands
[b] Aging and Later Life, Amsterdam Public Health Research Institute, the Netherlands
[c] Department of Health Sciences, Faculty of Science, Amsterdam Public Health Research Institute, Amsterdam Movement Sciences, Vrije Universiteit Amsterdam, the Netherlands
[d] Department of Health Sciences, Faculty of Science, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, the Netherlands
[e] Amsterdam University Medical Centers, Department of Epidemiology and Data Science, the Netherlands
[f] Amsterdam Public Health, Methodology, Amsterdam, The Netherlands
[g] Department of General Practice, Amsterdam UMC Location Vrije Universiteit Amsterdam, the Netherlands
[h] Faculty of Sport and Health Sciences and Gerontology Research Center, University of Jyväskylä, Jyväskylä, Finland
[i] Department of Sociology, Vrije Universiteit Amsterdam, the Netherlands

ARTICLE INFO

Section Editor: Michael Drey

Keywords:
Intrinsic capacity
Longitudinal measures
Ageing
Formative model

ABSTRACT

Background: The World Health Organization has introduced the construct of intrinsic capacity (IC) as an important component of healthy ageing and overall well-being in older adults The present study aimed to develop domain-specific and composite IC scores and to validate these scores by examining their longitudinal relation with functioning.

Methods: We used prospective data on participants aged 57 to over 90 years, with a 10-year follow-up, from the Longitudinal Aging Study Amsterdam, an ongoing cohort study of older Dutch men and women Using a formative, stepwise approach, we identified indicators across the different domains of IC, i.e. vitality, sensory, cognition, psychology, and locomotion, using a combination of unidimensional factor analyses and Partial Least Squares Structural Equation Modelling (PLS-SEM). Next, domain-specific and composite IC scores were generated, and the construct validity (score across age groups) and criterion validity (relationship with change in functional limitations) were assessed.

Results: The multiple unidimensional factor analyses and PLS-SEM identified a total of 18 indicators, covering the five domains of IC. The mean composite IC score was 70.9 (SD = 0.9) in men and 69.7 (0.8) in women. The domain-specific and composite IC scores all showed good construct validity, with known-group validation results indicating age-related declines. A higher composite IC score was associated with less functional limitations over time (B = 0.20, 95%CI [0.19, 0.22]).

Conclusion: The developed domain-specific IC scores and the composite IC score effectively discriminated age-related declines in IC. Additionally, the composite IC score was longitudinally associated with functional limitations. By creating this comprehensive and reliable tool for tracking IC, we aim to provide valuable insights into the dynamics of ageing and support more effective strategies for promoting health and well-being throughout later life. These scores establish a foundation for future research to track longitudinal changes across various IC domains and relate these changes to key age-related outcomes.

---

* Corresponding author.
E-mail address: y.qi1@amsterdamumc.nl (Y. Qi).

https://doi.org/10.1016/j.exger.2024.112599
Received 2 August 2024; Received in revised form 13 September 2024; Accepted 26 September 2024

1. Introduction

The World Health Organization (WHO) defines healthy ageing as "the process of developing and maintaining the functional ability that enables well-being in older age" (Beard et al., 2016). Functional ability is determined by the intrinsic capacity (IC) of individuals and by the environment they live in. IC is therefore crucial for understanding and promoting healthy ageing. However, developing measures for longitudinal research on IC presents various hurdles.

One major challenge in developing measures of IC lies in the complexities of its measurement model. The current consensus identifies five pivotal domains—locomotion, cognition, vitality, psychology, and sensory—as key components of IC, which refers to the composite of physical and mental capacities an individual can draw on (Koivunen et al., 2022; George et al., 2021; World Health Organization, 2017). These five domains have been used in WHO's reports on healthy ageing and have also been employed in various other studies concerning IC ((World Health Organization, 2019a; Ma et al., 2020; Rodríguez-Laso et al., 2023; Yan Wang et al., 2024). Nonetheless, it is important to note that research on IC is still evolving, and the identification of these five domains might not be definitive. IC transforms the idea of "healthy ageing" from a disease-focused to a function-focused perspective. Researchers have increasingly recognized the value of IC in measuring individual capacities and its connection with various aspects of functional ability in the context of ageing. The construct has been successfully validated and empirically examined in different cohorts (Beard et al., 2019; Beard et al., 2022; Si et al., 2023). Yet, the measurement model that guides the operationalization of IC in practice has not been explicitly defined in these studies. In a recent scoping review by our group, it has been suggested that IC should be examined as a formative construct (Koivunen et al., 2022). In a formative measurement model, the construct is formed by the combination of its domains, each representing a different facet that collectively defines the construct (Bollen and Diamantopoulos, 2017). Specifically, the formative approach means that individuals' total IC is constituted by their capacities regarding vitality, sensory, cognition, psychology and locomotion (Koivunen et al., 2022; Koivunen et al., 2023; Cesari et al., 2018). As a result, each domain contributes to the overall construct of IC, and changes in any domain can affect the total IC. This is in contrast to situations where a concept corresponds to a reflective measurement model, which means that changes in the overall construct would be reflected in changes across all individual indicators (Bollen and Diamantopoulos, 2017). Given this formative nature of IC, which encompasses five distinct domains, selecting indicators for composite measures should ideally be done at the domain-specific level to indicate the contributions of each domain (Fleuren et al., 2018).

Another challenge to measures of IC has been the plethora of ways to select data for each specific domain of IC (Diaz and Banerjee, 2023). Since the introduction of IC, extensive research has been conducted to measure and validate this construct (Beyene et al., 2024; Sanchez-Niubo et al., 2020). Our group has developed a cross-sectional composite IC score using prediction modelling (Koivunen et al., 2023). The developed IC score included seven indicators covering all five domains of the IC. Recently, more guidelines have become available to better define IC, as a collection of articles have provided recommendations for items to include in standardized questionnaires and identified measurement areas in need of further research (Diaz and Banerjee, 2023). For instance, it is recommended that nutritional assessment should be considered in the vitality domain to capture the individual's capacity to maintain homeostasis (Cesari et al., 2022). These advancements in understanding IC's components suggested the importance of revisiting and refining our measurement approaches to ensure they capture the construct's full breadth and depth. Additionally, prioritising measures on capacities over (solely) deficits promotes a more holistic and positive approach to understanding healthy ageing.

Establishing a reliable and consistent measure for longitudinal studies to monitoring IC is crucial, as emphasized by the ICOPE model, to effectively monitor and understand changes in individual capacities over time (World Health Organization, 2019a). Most earlier studies on development of measures on IC have been cross-sectional, with the exception of Beyene et al. (Beyene et al., 2024), where IC measure was developed and validated using longitudinal data repositories. The current study aimed to operationalize the concept of IC into standardized measures by developing scores for the five domains and an overall composite of IC across four measurement waves. We established a formative measurement model, maintaining the five domains as distinct dimensions. Indicators were identified for each domain individually, rather than assuming that all five domains stem from overall IC as in reflective models. Finally, the construct validity and longitudinal criterion validity of these scores were tested by examining their relationship with functioning over time.

2. Methods

2.1. Study sample

The present study is part of the ongoing IMPROve (Intrinsic Capacity Maintenance for Promoting Healthy Ageing) study, which seeks to operationalize and enhance the application of the WHO's framework on healthy aging in research and practice. For this, we used data from the Longitudinal Aging Study Amsterdam (LASA), which is an ongoing longitudinal study based on a nationally representative sample of the Dutch older population (Hoogendijk et al., 2020). Briefly, a random sample was drawn from population registers from eleven municipalities in the Netherlands. The LASA study started in 1992/93 consisting of 3107 participants aged 55–85 years. Since then, data are collected approximately every 3 years with a face-to-face general interview and a medical interview, which also includes performance tests in the homes of the respondents. In 2002/03, a second cohort of participants aged 55–64 years was added using the same sampling frame as the original cohort. The study was approved by the Ethical Review Board of the VU University Medical Center. All participants signed an informed consent before participating in the study. The LASA study protocol was conducted in accordance with the Declaration of Helsinki and received approval from the medical ethics committee of the VU University Medical Centre (IRB numbers: 92/138, 2002/141).

For the current study, data from these two cohorts was combined. We used their measurements in respectively 1995/96 and in 2005/06 as the baseline measures for this particular study. These measurement cycles were treated as baseline measurement because not all relevant indicators for operationalizing IC were available at the first measurement cycle of these two cohorts in 1992/1993 and 2002/2003, respectively. For the first cohort, follow-up measurements were performed in 1998/99, 2001/02, and 2005/06. For the second cohort, follow-up measurements were performed 2008/09, 2011/12, and 2015/16, respectively. At baseline, the total number of participants for cohort 1 was 2545 and for cohort 2 was 908. With the combined cohorts, two study samples were used: an indicator selection sample and a score construction sample. The indicator selection sample included individuals who participated in both the general and medical interviews at baseline as some potential indicators were collected in the general interview and others in the medical interview (Fig. 1). This sample consisted of 2333 participants: 1509 from the first cohort and 824 from the second cohort.

For the score construction sample, we used a larger sample of 3246 participants at baseline, aged 55 and over, including 2372 from the first cohort and 874 from the second cohort (see Fig. 2). This sample is larger in comparison to the indicator selection sample, as some missings on the indicators were allowed when constructing the domain-specific scores (see section IC scores construction), but not during the selection of indicators for the domains. All participants completed the general interview, but not all underwent the medical interview. Some participants who did not participate at one wave did participate in later waves. This

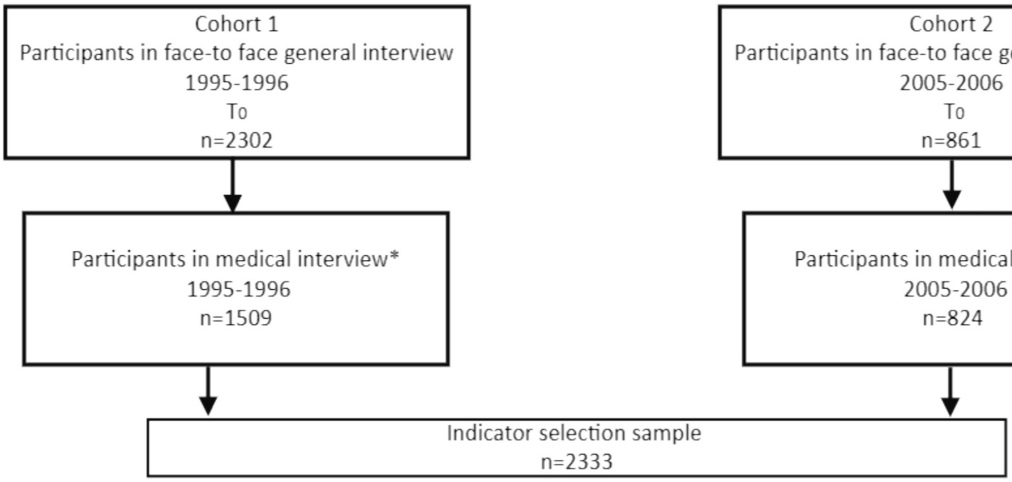


Fig. 1. Flowchart of the indicator selection sample
*In 1995–1996, only people born before 1931 were asked to participate in the medical interview.

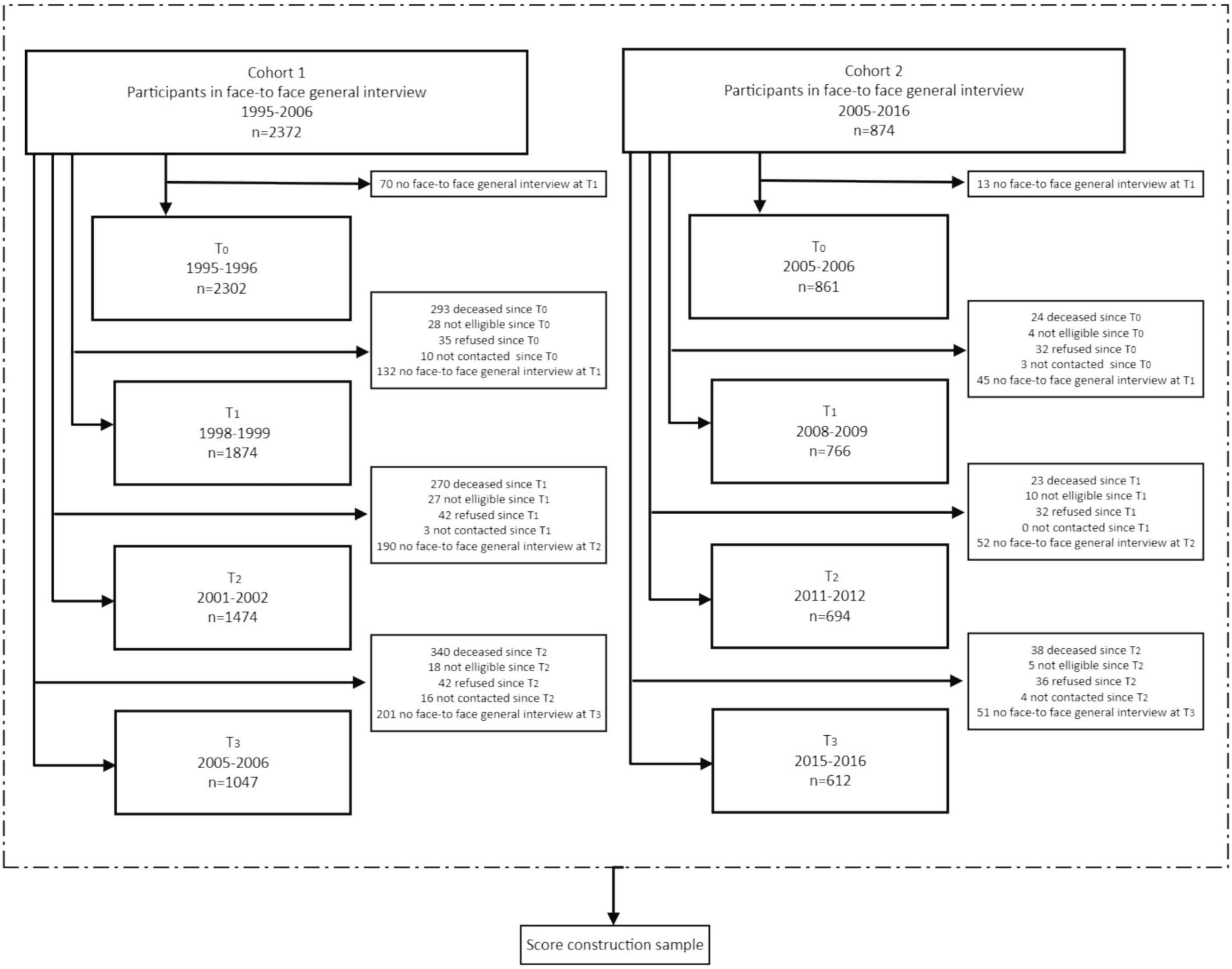


Fig. 2. Flowchart of the score construction sample.

score construction sample was also used to evaluate whether the IC scores consistently and accurately differentiate individuals across different age groups.

2.2. Steps of IC scores development and validation

Using the five-domain structure of IC, which has been employed in the ICOPE screening tool and other studies (Koivunen et al., 2022; George et al., 2021; World Health Organization, 2019a), we followed a stepwise procedure to create the domain-specific IC scores and the composite IC scores.

2.2.1. Indicator selection

First, we identified candidate indicators of IC based on our past work (Koivunen et al., 2022; Koivunen et al., 2023), literature search, and expert opinions. These indicators covered the five domains of IC (Beard et al., 2019): vitality, sensory, cognition, psychology, and locomotion. Second, we verified the presence of each potential indicator in the LASA dataset to ensure the feasibility of developing longitudinal measures. A total of 22 indicators measured at baseline were considered. The inclusion of indicators under the vitality domain was guided by the working definition of vitality as proposed by the WHO working group and their consensus regarding potential markers to measure vitality capacity. Hand grip strength (George et al., 2021; Beard et al., 2019; Koivunen et al., 2023; Gutiérrez-Robledo et al., 2021; Gutiérrez-Robledo et al., 2019), peak flow (George et al., 2021; Beard et al., 2022; van Schoor et al., 2012), calf circumference (Sanchez-Rodriguez et al., 2023), appetite (Gaussens et al., 2023), sleep quality, and self-reported weight change (Gaussens et al., 2023) were considered to construct the *vitality* domain. These indicators cover several important attributes of vitality. Hand grip strength and peak flow measure neuromuscular function. Calf circumference measures body composition (Bautmans et al., 2022). Appetite was chosen as age-related physiological changes, such as decreased ghrelin release in the stomach, are known to increase feelings of fullness and reduce appetite (Cox et al., 2020). Self-reported weight loss assesses the nutritional aspect and sleep quality evaluates energy levels. Sensory function includes vision and hearing (George et al., 2021).Self-rated items on hearing in a conversation, being able to use a normal telephone, near vision, and far vision were considered to construct the *sensory* domain (Gutiérrez-Robledo et al., 2019). General cognitive functioning (López-Ortiz et al., 2022), information processing speed (Koivunen et al., 2023), and episodic memory (Koivunen et al., 2023) were considered to construct the *cognition* domain. These performance-based measures are commonly used as key indicators within the cognition domain (George et al., 2021). Anxiety symptoms (López-Ortiz et al., 2022)and depressive symptoms (López-Ortiz et al., 2022) have been commonly used as indicators for the psychology domain. It is also important to question whether the absence of anxiety or depressive symptoms fully captures the spectrum of psychological capacity, particularly on the positive end. Research suggests that resources related to a sense of control and the ability to mentally adapt to adversities can be preserved or even enhanced through growth, experiences, and learning during ageing (Wister and Cosco, 2020; Charles and Carstensen, 2010). Therefore, we have also included mastery (Golino et al., 2020), self-efficacy (Koivunen et al., 2023), and self-esteem (Astrone et al., 2022) as these capacities may be essential in compensating for physiological losses. Walking speed (George et al., 2021; Chen et al., 2023), chair rise test (Beard et al., 2019), standing balance (George et al., 2021), and cardigan test were considered to construct the *locomotion* domain. These are commonly used performance-based measures for assessing locomotion (George et al., 2021). We have also explored the potential cross-domain relevance of grip strength and walking speed, assessing whether grip strength should be considered as indicator for *locomotion* (George et al., 2021) and whether walking speed should be considered as indicator for *vitality* (Zhao et al., 2021). Currently, it is not clear how to best measure nutritional status as an indicator of vitality. Therefore, we have additionally tested the utility of Body Mass Index (BMI) and weight change as indicators. The detailed descriptions of how these variables were measured are provided in Appendix 1.

Second, in the indicator selection sample, unidimensional factor analyses and Partial Least Squares Structural Equation Modelling (PLS-SEM) were used to select indicators to construct domain-specific IC scores. Using the indicator selection sample, we first rescaled all the candidate indicators of IC using the percent of maximum possible ("POMP") method (Cohen et al., 2013; Cohen et al., 1999), which was calculated by linearly transforming each participant's raw score into a percentage of the maximum total score of the measure in the sample. The rescale was stratified by sex. After rescaling, all indicators ranged from 0 (low capacity) to 100 (high capacity). To facilitate any possible direct comparison with other studies, we reported POMP units in the present study. A POMP score of 70, for example, indicates that the score is 70 % of the maximum possible value for the measurement in question. Next, separate unidimensional factor analyses were performed on each of the five domains of IC. Indicators with loading >0.40 were selected for further analysis (Clark and Watson, 2019). This threshold was established to ensure that only indicators with a substantial relationship to the underlying domain were included. The multiple unidimensional factor analyses were performed using the "psych" package with the "fa" function in R programming for statistical computing version 4.1.2 (Revelle and Revelle, 2015).

Following the initial selection process, selected indicators were incorporated into a PLS-SEM model. PLS-SEM is the preferred approach when formatively specified constructs are concerned (Hair Jr et al., 2021a). The structural model estimated the relationships between latent constructs (i.e. the five IC domains), with correlations between the five domains being estimated The measurement model specifies the relationships between each domain as latent constructs and their corresponding observed indicators (Fig. 3). For this, indicator correlation weights represent each indicator's relative importance to the construct and indicator loading represents the absolute contribution of an indicator to its construct (Lohmöller, 2013) (Hair Jr et al., 2021b). It is also recommended to consider the absolute contribution of a formative indicator to the construct, which is determined by the formative indicator's loading (Cenfetelli and Bassellier, 2009). In general, indicator loadings of 0.50 (Cenfetelli and Bassellier, 2009) and higher suggest the indicator makes a sufficient absolute contribution to forming the construct, even if it lacks a significant relative contribution (Hair Jr et al., 2021b). Significance of the indicator weights was based on 10,000 bootstrap samples (Hair Jr et al., 2021a). The PLS-SEM model was estimated using SEMinR package in R programming for statistical computing version 4.1.2 (Ray et al., 2021).

2.2.2. IC scores construction

In the score construction sample, the domain-specific IC scores and composite IC scores were constructed by the indicators with statistically significant weights in the PLS-SEM model. At each measurement wave (T_0 through T_3), a mean score (domain-specific score) was first calculated for each domain using the selected indicators. We applied the general rule that the mean domain-specific score was calculated when 50 % or more of the indicators were present (Fairclough and Cella, 1996). Subsequently, the composite IC score was computed by averaging the five domain-specific scores, but only if all five domain-specific scores were present. All computed scores are reported in POMP units, ranging from 0 to 100. Descriptive statistics were calculated for both the indicator selection sample and the score construction sample.

2.2.3. IC scores validation

First, descriptive statistics were calculated for both the indicator selection sample and the score construction sample. In the score construction sample, we tested the construct validity of both the domain-specific and composite IC scores. Additionally, we tested the criterion validity of the composite IC score in relation to functional limitations. Construct validity was tested using the known-groups' validity (Mokkink et al., 2010), this involved assessing whether the scores could effectively distinguish between groups known to differ in IC, such as different age groups. Based on the hypothesis that IC decreases with age, we compared the IC scores across these groups. For this, data that were originally structured according to measurement wave was restructured to represent each IC domain at 3-year age intervals, covering intervals ranging from 57 to 59 years to 90+ years. Scatter plots were generated using ggplot2 in R to visualize the relationship between age categories and the main score of each domain.

Criterion validity was assessed by examining the association between composite IC scores and functional limitations over four measurement waves using a linear mixed-effects model (LMM). Functional limitation was assessed using a scale based on six items that measured limitations in performing certain activities. Response categories ranged from 1 to 5, and were summed to create a scale ranging from 6 to 30. A higher score indicates less functional limitation (Eekhoff et al., 2019). The detailed description of the six items can be found in Appendix 1. The LMM model was formulated with functional limitation as dependent variable and IC scores as primary independent variable. Both functional limitation and IC scores were measured at four time points, from T_0 to T_3. The model included a fixed effect of composite IC score for each wave and a random intercept to account for between-subject variability. Age and sex were included as covariates. The model was fitted using the restricted maximum likelihood (REML) estimation method to provide unbiased variance component estimates. The Wald test was used to assess the significance of fixed effects, and confidence intervals were calculated for model parameters. Model diagnostics were performed to evaluate the assumptions of homoscedasticity and normality of residuals. The LMM model was estimated using the lme4 package for R programming for statistical computing version 4.1.2 (Bates, 2010).

3. Results

The average age of participants in the indicator selection sample was 71.3 years, with a standard deviation (SD) of 8.3 years. Among the participants, 51.8 % were female (Table 1). The score construction sample had a mean age of 70.0 years (SD = 8.6) with 53.2 % female participants. The relatively large number of missing values for calf circumference was due to the introduction of this measurement during the year of data collection, resulting in the first participants not being measured.

3.1. Indicator selection

Table 2 shows the results of the unidimensional factor analyses in which each IC domain was considered individually in relation to their corresponding candidate indicators. Based on standardized loadings from the pattern matrix, three variables—grip strength (0.62), peak flow (0.71), and calf circumference (0.48)—demonstrated loadings above the 0.40 threshold, indicating a significant contribution to the vitality domain. For sensory, cognition and psychology, all candidate indicators showed loadings higher than the threshold. For locomotion, only the balance test showed a loading below the threshold. Based on these results, a total of 18 candidate (in bold) indicators with loading above 0.40 were selected to fit in the PLS-SEM model.

Table 1

Sample characteristics of the indicator selection sample ($n = 2333$) and the score construction sample ($n = 3246$) at T_0, as Percentage of Possible Maximum (POMP) units.

	Indicator selection sample		Score construction sample	
	Mean (SD)/ Percentage	N. valid cases	Mean (SD)/ Percentage	N. valid cases
Age	71.3(8.3)	2333	70.0(8.6)	3246
Sex(female)	51.8 %	2333	53.2 %	3246
Vitality				
Hand grip strength	53.1 (14.8)	2279	53.1 (14.8)	2279
Peak flow	49.8 (17.6)	2115	49.8 (17.6)	2130
Calf circumference	42.3 (17.3)	1547	42.0 (16.8)	1972
Appetite	93.6 (19.2)	2286	93.8 (18.8)	3084
Sleep quality	67.9 (23.6)	2185	68.5 (23.4)	2844
Weight change	75.1 (36.2)	2322	75.1 (36.2)	2322
Sensory				
Hearing in a conversation	83.5 (27.4)	2326	84.9 (26.3)	3115
Use normal telephone	96.3 (15.9)	2326	96.6 (15.3)	3115
Near vision	88.3 (22.8)	2331	88.0 (22.9)	3122
Far vision	93.2 (18.9)	2324	93.2 (19.0)	3115
Cognition				
General cognitive functioning	87.1 (13.3)	2329	89.2 (11.6)	3153
Information processing speed	48.2 (18.0)	2217	48.2 (18.0)	2217
Episodic memory	56.5 (17.9)	2283	56.5 (17.9)	2283
Psychology				
Anxiety symptoms	85.0 (17.4)	2284	85.8 (16.7)	3080
Depressive symptoms	83.0 (15.8)	2277	84.2 (15.3)	3072
Mastery	60.5 (17.9)	2243	62.9 (17.1)	3025
Self-efficacy	55.4 (13.8)	2257	55.5 (13.9)	3040
Self esteem	67.6 (15.4)	2263	68.9 (14.9)	3047
Locomotion				
Walking speed	87.9 (7.3)	2238	88.1 (7.2)	3015
Chair rise test	81.1 (10.7)	2103	84.9 (7.3)	2823
Balance test	79.1 (38.5)	2248	80.6 (37.3)	3020
Cardigan test	84.7 (9.3)	2281	85.0 (9.2)	3080

3.2. Indicator selection-additional analyses

We have additionally evaluated the utility of Body Mass Index (BMI) alongside weight change as indicators of vitality by assigning scores that reflect increasing levels of capacity. Specifically, BMI was categorized from 1 to 4, with 1 representing underweight (BMI <18.5), 2 indicating obesity (BMI higher than 30), 3 for overweight (BMI between 25 and 29.9), and 4 corresponding to normal weight (BMI between 18.5 and 24.9). Weight change was similarly scored: 1 for involuntary weight loss, 2 for weight gain due to any reason, 3 for voluntary weight loss, and 4 for no changes in weight. Neither of these variables demonstrated acceptable correlations with vitality. In unidimensional factor analysis, BMI correlated at 0.12 after adjustment for overlaps with other indicators, falling well below our threshold of 0.40. Weight change correlated at only 0.05 after adjustment for overlaps with other indicators, also falling below the threshold.

Table 3 shows the indicator weights and their corresponding 95 % bootstrap confidence intervals for each IC domain. If a confidence

Table 2

Factor loadings of candidate IC indicators: separate unidimensional factor analysis.

	Vitality	*Sensory*	*Cognition*	*Psychology*	*Locomotion*
Hand grip strength	**0.62**				
Peak flow	**0.71**				
Calf circumference	**0.48**				
Appetite	0.28				
Sleep quality	0.19				
Weight change	0.05				
Hearing in a conversation		**0.52**			
Use normal telephone		**0.53**			
Near vision		**0.50**			
Far vision		**0.43**			
General cognitive functioning			**0.68**		
Information processing speed			**0.77**		
Episodic memory			**0.70**		
Anxiety symptoms				**0.65**	
Depressive symptoms				**0.75**	
Mastery				**0.72**	
Self-efficacy				**0.57**	
Self esteem				**0.66**	
Walking speed					**0.78**
Chair rise test					**0.68**
Balance test					0.19
Cardigan test					**0.62**

interval does not include the value zero, the weight can be considered statistically significant, and the indicator can be retained (Hair Jr et al., 2021a). The analysis of indicator weights concludes the evaluation of the formative measurement models. In the current model, all indicators showed significant weights corresponding to their domains, therefore all indicators were retained (Cenfetelli and Bassellier, 2009).

Fig. 3 visualizes the structural model together with indicator loadings of each IC indicator. For the structural model, bivariate correlations were estimated in regard to the relationship between the five domains. The five domains did not exhibit high correlations between each other (range 0.03 to 0.46). For the current model, only calf circumference showed loading that was lower than 0.50, with other loadings ranging from 0.60 to 0.88. However, as the weight of calf circumference was statistically significant, we have retained calf circumference as indicator for vitality.

3.3. *IC scores validation*

We computed five domain-specific IC scores and one composite IC score for each measurement wave, using the statistically significant indicators from the PLS-SEM model. Descriptive statistics, baseline characteristics, IC domain scores and composite scores for the score construction sample at each measurement wave are available in Appendix 2.

Construct validity was assessed using the known-groups' validity by testing whether the scores could effectively distinguish between different age groups. Fig. 4 presents the domain-specific IC scores and composite IC scores across different age groups. Each point represents the mean score for a specific age group of each domain-specific score. The x-axis represents the age groups, and the y-axis depicts the mean domain-specific scores. The vertical lines extending above and below each mean score represent the standard deviation above and below the mean score for each age group. Overall, we observed a clear trend of lower IC with increasing age across all domains. The average vitality score was lower in higher age groups, starting at an average of 58 (SD 10) for the youngest group and dropping to 36 (SD 11) for those aged 90 or older. The average sensory and locomotion scores also followed a consistent decline up to the 90 years or older group, with increased variability in these scores observed with age. The cognition and psychology domain both showed similar trends of decline with ageing. The variability in the cognition scores remained rather consistent in comparison to other domains. The composite IC score also showed a steady decline from 78 to 60, with relatively small variability.

The mixed effects linear regression model with functional limitations as outcome achieved a R^2 of 0.76. Fixed effects revealed significant predictors including composite IC score ($B = 0.20$, $p < 0.001$), age ($B = -0.10$, $p < 0.001$), and sex ($B = -1.32$, $p < 0.001$), each statistically significant with Satterthwaite degrees of freedom adjustments. Random effects showed substantial variability attributed to individual respondents ($ICC = 0.67$) and minimal variability across the four measurement waves ($ICC = 0.01$). Overall, the mixed effects linear regression model indicates that higher IC scores were related to less functional limitations, as evidenced by the upward slope of the regression line. The figure showing the relationship between composite IC score and functional limitation can be found in Appendix 4.

4. Discussion

The current study aimed to operationalize the concept of IC into standardized longitudinal measures by adopting a formative measurement model and utilizing data from the Longitudinal Aging Study

Table 3

Measurement model of the five IC domains: Indicator weights and bootstrap 95 % confidence interval, relative importance of one indicator to its domain based on results from PLS-SEM model.

	Vitality	*Sensory*	*Cognition*	*Psychology*	*Locomotion*
Hand grip strength	**0.65 [0.62, 0.69]**				
Peak flow	**0.50 [0.47, 0.53]**				
Calf circumference	**0.08 [0.02, 0.14]**				
Hearing in a conversation		**0.44 [0.40, 0.49]**			
Use normal telephone		**0.32 [0.27, 0.37]**			
Near vision		**0.41 [0.37, 0.46]**			
Far vision		**0.33 [0.28, 0.38]**			
General cognitive functioning			**0.39 [0.36, 0.42]**		
Information processing speed			**0.45 [0.43, 0.48]**		
Episodic memory			**0.41 [0.39, 0.43]**		
Anxiety symptoms				**0.10 [0.06, 0.14]**	
Depressive symptoms				**0.34 [0.30, 0.37]**	
Mastery				**0.35 [0.32, 0.39]**	
Self-efficacy				**0.37 [0.33, 0.40]**	
Self esteem				**0.16 [0.12, 0.19]**	
Walking speed					**0.61 [0.55, 0.66]**
Chair rise test					**0.26 [0.21, 0.31]**
Cardigan test					**0.38 [0.33, 0.43]**

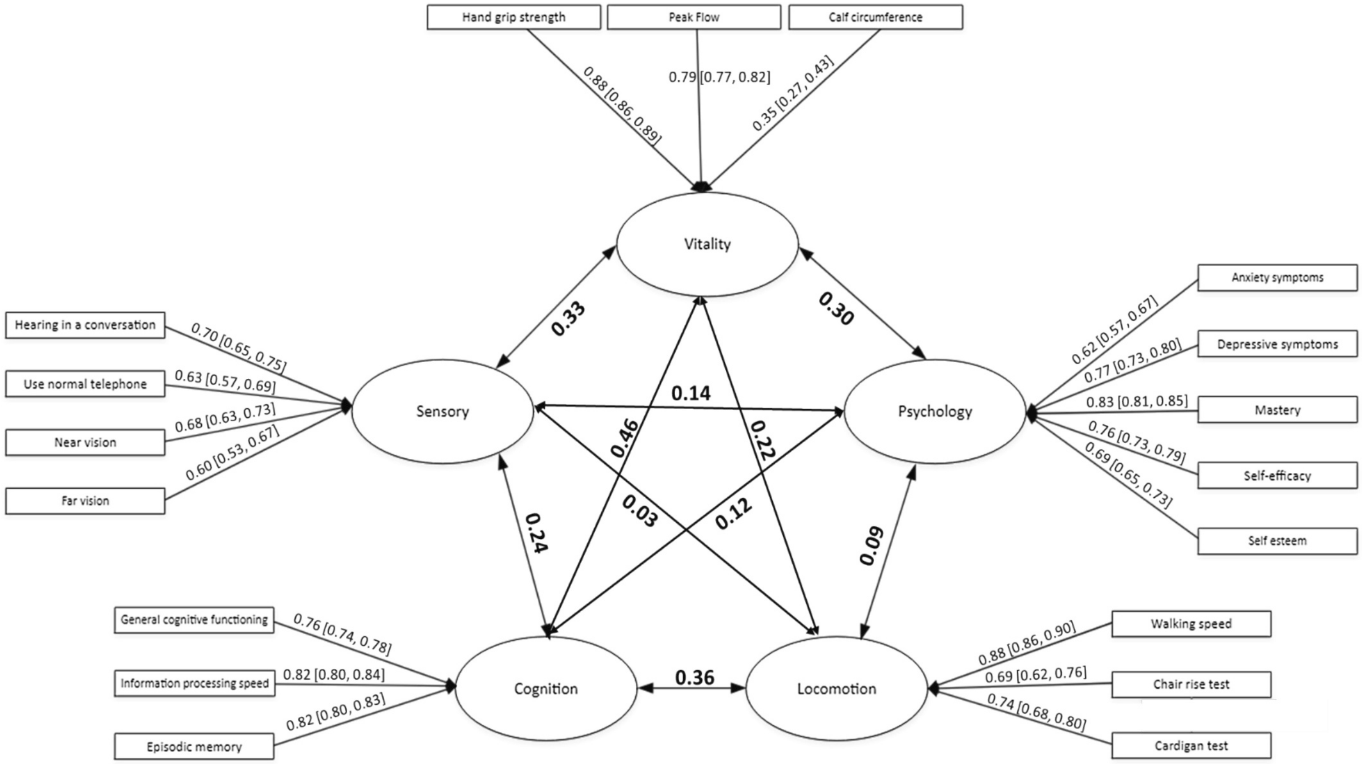


Fig. 3. Structural model of the five domains of IC with results of the path analysis.

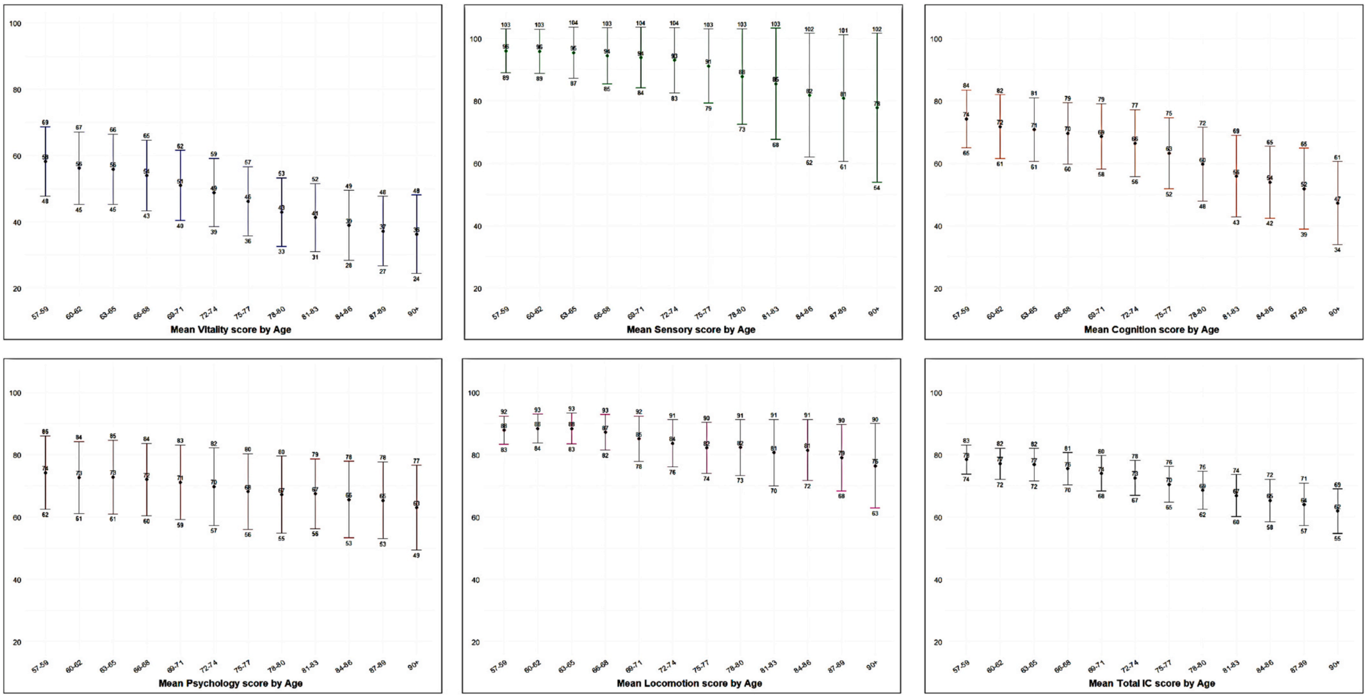


Fig. 4. Age group comparison in domain-specific IC scores and composite IC scores to test known groups' validity, as Percentage of Possible Maximum (POMP) units.

Amsterdam. We have identified 18 indicators covering the five domains of IC. The developed domain-specific IC scores and composite IC score demonstrated good validity and indicated age-related declines. Higher composite IC scores were associated with fewer functional limitations over time.

Guided by the WHO's definition of IC, our study used the five domains that are commonly recognized in the literature (World Health Organization, 2019a; Rodríguez-Laso et al., 2023; World Health Organization, 2019b). Our objective was not to explore the conceptualisation of IC, but rather to operationalize the domains of IC using existing data sources and framework. Building on our previous work, we assumed a formative measurement model for IC. The formative approach is rarely acknowledged in the development of IC measures, even though IC is recognized as a multidimensional construct (Koivunen et al., 2023; Cesari et al., 2018; Nascimento et al., 2023) that could be more appropriately operationalized with composite measures rather than reflective scales. We demonstrated an approach for developing both domain-specific and composite IC measures using formative constructs. A formative model suggests that IC is an operationalization of a multidimensional construct, summarizing various conceptually distinct domains (World Health Organization, 2019a). This operationalization aligns with the findings from the network analysis by Koivunen et al., which indicates that IC should not be conceptualized as stemming from a single, general trait (Lohmöller, 2013). The recognition of IC as a formative construct has important implications for research. It necessitates the use of appropriate statistical techniques that can accurately capture the different dimensions of this construct, moving beyond techniques that assume an overall latent variable model. This perspective also influences how interventions are designed and implemented. Since IC is not derived from a single trait but from multiple independent domains, interventions can be more precisely tailored to target specific areas of deficiency within an individual's intrinsic capacity profile, focusing on achieving specific outcomes.

At the domain level, we considered a total of 22 candidate indicators, which spanned the five IC domains. While the selection of candidate indicators for the sensory, cognition, and psychology domains was relatively straightforward, identifying appropriate indicators for the vitality and locomotion domains proved more challenging. The current working definition of vitality suggests that it represents a physiological state, arising from either normal or accelerated biological ageing processes (Bautmans et al., 2022). This state is the outcome of interactions among multiple physiological systems and is manifested in various bodily functions such as energy and metabolism, neuromuscular function, and stress response capabilities (Beard et al., 2019; Bautmans et al., 2022). We have considered hand grip strength, peak flow, calf circumference, appetite, sleep quality, and weight as indicators for vitality. Among these, hand grip strength, peak flow, and calf circumference were selected to construct the domain score for vitality. Hand grip strength is often used to measure vitality (Beard et al., 2019; Beard et al., 2022; Aliberti et al., 2022) as it captures the vital sign of physiological reserve and biological age (Granic et al., 2017; Sayer and Kirkwood, 2015; Lu et al., 2023). Peak flow, similarly, has been utilized in studies as an indicator of vitality (Gutiérrez-Robledo et al., 2019; van Schoor et al., 2024). Although not as commonly employed, calf circumference has demonstrated its relevance as a significant marker of nutritional status (Bonnefoy et al., 2002; Cruz-Jentoft et al., 2018). Together, these three indicators can possibly provide a comprehensive measure of vitality.

Among the indicators not selected for the vitality domain, weight change exhibited the lowest loading. We evaluated the utility of BMI and weight change as indicators of vitality, but none of these variables were selected as indicator due to non-significant correlations and weights. This finding suggests that, within our sample, BMI and weight change may not effectively capture the essence of vitality, particularly in relation to metabolism. This could be due to these indicators not adequately reflecting the same physiological and metabolic components of vitality or that the operational definitions used did not capture the necessary aspects of these measures. This finding suggests the need for further exploration into measures that might more accurately reflect the metabolism and stress response aspects of the vitality domain. The broader issue highlighted by our results is that the operational definitions of vitality domain is still being developed (Bautmans et al., 2022). The empirical meaning of a construct might differ from its intended meaning based on the chosen indicators. This risk exists for both formative models and reflective models and ties into the issue of indicator validity (Fleuren et al., 2018). We may also need to extend this consideration to many other constructs, which should always be considered when interpreting research findings.

Hand grip strength has commonly been used as an indicator for vitality, reflecting overall muscle strength and physiological resilience. However, given muscle strength's importance in activities requiring bodily stabilization and support, hand grip strength might also be pertinent to locomotion (Nayasista et al., 2022). The additional unidimensional factor analyses (Appendix 3) showed that hand grip strength had a loading of 0.46 in the locomotion domain, which is lower than its loading in the vitality domain (0.62), reaffirming its primary association with vitality. Additionally, walking speed, which results from the interplay of an individual's physiological capabilities and their perceptions of environmental conditions and task demands, has been advocated as sign of vitality (Middleton et al., 2015). Our analyses found that walking speed had a loading of 0.27 in the vitality domain, which is lower than its loading in the locomotion domain. This finding confirms walking speed as indicator for the locomotion domain within the IC framework.

Finally, we assessed the construct validity of both the developed domain-specific IC scores and the composite IC scores. Unlike reflective models, where factor analysis can confirm construct validity, formative measurement models do not derive meaningful interpretive value from statistically calculating internal consistency, nor do they confirm a single underlying latent variable (Fleuren et al., 2018). This is due to formative measurement models not assuming unidimensionality and, as such, requiring more intricate methods for assessing validity, such as construct and criterion validity (Fleuren et al., 2018). Our findings indicate a consistent decline with age across the individual domains of vitality, sensory, locomotion, cognition, and psychology, as well as in the overall IC score. This uniform decline suggests the robust construct validity of our IC scores. Collectively, these results validate that the IC scores effectively mirror the gradual decline of IC, a fundamental aspect of ageing. It also suggests that focusing only on the aggregated overall IC score across multiple domains might lead to the loss of information about the different contributing aspects (Koivunen et al., 2023). While our composite score is a comprehensive representation of IC and has practical advantages, the multidimensional nature of the composite IC score also represents a potential limitation. It is for example possible that individual domains will change independently of each other. These changes might be “hidden” from the researchers view if only the composite IC is examined. Therefore, we recommend that researchers always compute and report both domain-level scores along with the composite IC scores. It is worth mentioning that most studies, including ours, have used functional abilities-related measures to validate the developed IC measure (Beard et al., 2019; Beyene et al., 2024; Gutiérrez-Robledo et al., 2021; Salinas-Rodríguez et al., 2022). The WHO's healthy ageing model suggests that an individual's level of IC significantly influences their functional ability in interaction with the surrounding environment (Belloni and Cesari, 2019). Following this idea, we considered functional decline as a likely result of reduced IC and used it to validate the developed composite IC score. However, IC is a unique construct and thus one suitable gold standard does not exist yet (Hoogendijk et al., 2023).

We used POMP scores to construct each domain-specific score and subsequently averaged these to obtain the composite IC score. POMP scores are akin to summed scores in that they are not derived from a factor model, Item Response Theory (IRT), or latent variable model with parameters derived from specific samples (Cohen et al., 2013; Cohen et al., 1999; Mendenhall et al., 2009). The choice to use summed POMP scores is grounded in their practicality and ease of application. One significant concern with relying on factor loadings to construct measures is the variability of factor loadings across different samples. Widaman and Revelle note that (Widaman and Revelle, 2023), factor loadings used in sample-based estimation of factor scores can vary across different samples and may deviate from population values, leading to fluctuations in factor scoring weights and introducing additional uncertainty from one sample to another, as well as between samples and the population (MacCallum and Tucker, 1991). This variability can lead to inconsistencies in how scores are calculated across different studies or populations, making it less straightforward to compare results and draw meaningful conclusions. The practicality of summed scores is well-documented and aligns with the goals of our study, which is to make measures of IC available using existing resources. As noted by Sijtsma et al. (Sijtsma et al., 2024), summed scores derive their value from their ability to predict practically relevant events and behaviours. This approach ensures that our findings are directly applicable and relevant in real-world contexts. While our composite score is a comprehensive representation of IC and has practical advantages, the multidimensional nature of the composite IC score also represents a potential limitation. It is for example possible that individual domains will change independently of each other. These changes might be “hidden” from the researchers view if only the composite IC is examined. Therefore, we recommend that researchers always compute and report both domain-level scores along with the composite IC scores.

One strength of our study lies in the comprehensive assessment of IC. Our assessment considers all five domains that are generally considered critical components of IC (Beard et al., 2019; Cesari et al., 2018). Another strength of our study includes a formative conceptual framework and measurement model of IC, which provided the foundation for developing our scores. Considering the different implications of formative versus reflective measurement models (Bollen and Diamantopoulos, 2017; Fleuren et al., 2018), the distinction between reflective and formative measurement is not merely conceptual nitpicking. For instance, following the formative approach, interventions aiming to improve the overall IC can target individual domains as they “cause” the IC (Bollen and Diamantopoulos, 2017). Moreover, we benefited from a large, nationally representative sample of older adults in the Netherlands, utilizing a longitudinal study design. This design primarily incorporated performance-based and continuous measurements across various domains of functioning, enhancing the robustness and applicability of our findings to real-world settings. Being able to monitor these capacities continuously at various time points provides a more detailed and effective analysis than methods relying on simplistic categorical measures of late-life events, such as mortality. Importantly, this setting aligns closely with the concept of capacity, with continuous and longitudinal measures on IC.

Our study also has several limitations that should be noted. Firstly, some missing data for candidate indicators of IC at baseline, along with participant loss during follow-up, reduced our sample size. This may have caused our study sample to represent a slightly healthier subset of the target population. Consequently, selective dropout might have led to a slight underestimation of functional decline during the follow-up period. In our study, participants who did not continue with face-to-face interviews after baseline were older (mean age 73.9 years, SD 8.9) compared to those who remained (mean age 70.8 years, SD 8.6). They also had a slightly lower baseline functional limitation score (mean 26.0 SD 5.5) than those in the score construction sample (mean 27.2, SD 4.6). Additionally, our cohort included few participants aged 90 years or older, which might have led to an underestimation of functional decline due to selective dropout. LASA employed oversampling of older people and men to maintain sufficient numbers in these age and sex groups. However, despite these efforts, attrition may still introduce bias in estimating longitudinal relationships, particularly in relation to functional decline. Future studies should aim to examine the external validity of the constructed IC scores enhance generalizability. Another limitation is the absence of objective measures for vision and hearing in our data. Reliance on self-assessments may not provide as precise or standardized information about sensory capacities compared to performance-based tests. However, self-reported data can also offer valuable insights into daily functioning and the personal experiences of older adults. Furthermore, the limited response options available in our sensory indicators may not fully capture the entire spectrum of sensory function. Moreover, we were unable to include validated assessment of nutritional status under the vitality domain. Valid measures of undernutrition, such as using the Global Leadership Initiative on Malnutrition (GLIM)consensus definition (Cederholm et al., 2019), in future research could provide a more comprehensive evaluation of the vitality domain (World Health Organization, 2019b). Additionally, we did not fully explore the relationships between the domains of intrinsic capacity. Vitality, for example, has been suggested as the underlying physiological determinant of intrinsic capacity and is therefore related to the other domains of intrinsic capacity (Bautmans et al., 2022). For instance, the ability to move from one place to another is part of locomotion capacity; however, sufficient muscular strength is necessary to complete functional tasks (e.g., rising from a chair) and is also linked to neuromuscular function, which is a component of vitality capacity (World Health Organization, 2021). Similarly, other domains may both influence and be influenced by one another. Future research is needed to study the temporal order of the different domains. Furthermore, we did not investigate the measurement invariance of the five domains of IC over time. Future research should confirm that our developed scores hold equivalent meaning across different populations and over time. Lastly, as our study focused on the exploration of how to operationalize IC and the five domains using a formative model, it is possible that future research may need to account for new consensus and emerging properties of IC as the understanding of this construct evolves. Our approach describes one possible method of developing measures of IC using data from existing longitudinal studies, recognizing that the construct of IC itself is still being developed and refined.

Taken together, our findings suggest that the developed IC scores capture a comprehensive range of information related to an individual's physical and mental capacities. Our findings have a number of implications. Firstly, our analytical framework could enable researchers to identify and address the underlying causes of changes in these capacities. Moreover, measurable trajectories of capacity could serve as valuable research outcomes. Algorithms could be devised to process this information and outline capacity trajectories, informing self-management, clinical practice, and further research. Utilizing these trajectories as research outcomes could also enhance the comparison of intervention impacts across various conditions. As medicine moves towards more personalized and precise approaches, better data are needed on how different subpopulations respond to specific interventions. Stratifying by intrinsic capacity could help identify which groups might benefit most from particular interventions, offering a more relevant categorization than by chronological age or comorbidity alone (Gore et al., 2018). Additionally, our findings support reevaluating how ageing and functional capacity are assessed in older adults. By viewing IC as a formative, multidimensional construct, we gain a deeper understanding of the ageing process, acknowledging that declines in functional capacity do not occur uniformly across all domains. This insight paves the way for more personalized healthcare strategies that cater to individual needs rather than applying a one-size-fits-all approach.

5. Conclusion

IC covers a broad range of health characteristics, encompassing both physiological and psychological changes associated with ageing. Monitoring its decline might offer insights into trajectories of functioning and may guide prevention strategies. The findings from this study provide a foundational starting point for building an evidence base aimed at enhancing our understanding of healthy ageing from a holistic, capacity-based perspective.

Funding

The authors acknowledge ZonMW for their support (grant number: 09120012110024). The Longitudinal Aging Study Amsterdam (LASA) is largely supported by a grant from the Netherlands Ministry of Health, Welfare and Sports, Directorate of Long-Term Care.

Statement of human participants and/or animals

The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent

All participants of the LASA study provided written informed consent.

Author statement

Yuwei Qi, Laura A. Schaap, Benjamin D. Schalet, Emiel O. Hoogendijk, Dorly J.H. Deeg, Marjolein Visser, Kaisa Koivunen, Martijn Huisman, Natasja M. van Schoor: Conceptualization, Methodology; **Yuwei Qi, Benjamin D. Schalet:** Formal analysis; the first draft of the manuscript was written by **Yuwei Qi** and all authors commented on previous versions of the manuscript.

CRediT authorship contribution statement

Yuwei Qi: Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Laura A. Schaap:** Writing – review & editing, Conceptualization. **Benjamin D. Schalet:** Writing – review & editing, Formal analysis. **Emiel O. Hoogendijk:** Writing – review & editing, Conceptualization. **Dorly J.H. Deeg:** Writing – review & editing, Conceptualization. **Marjolein Visser:** Writing – review & editing, Conceptualization. **Kaisa Koivunen:** Writing – review & editing, Conceptualization. **Martijn Huisman:** Writing – review & editing, Conceptualization. **Nastasja M. van Schoor:** Writing – review & editing, Conceptualization.

Conflict of interest

The authors declare that they do not have competing interests.

Data availability

The dataset generated during the current study are not publicly available due to confidentiality, but the data underlying the results presented in this study are available from the Longitudinal Aging Study Amsterdam (LASA). Data of LASA may be requested for research purposes. More information on data requests can be found on the LASA website: www.lasa-vu.nl.

Appendix 1. Considered indicators for measurement of IC and the six items measuring functional limitation

Vitality

Hand grip strength, calf circumference, peak flow, appetite, and sleep quality were considered to construct the *vitality* domain. Hand grip strength was assessed with a strain-gauged dynamometer (Takei TKK 5001; Takei Scientific Instruments Co. Ltd., Tokyo, Japan) (Stel et al., 2004). Participants performed two maximum forced trials with both hands in a standing position with the arm along the body. The total hand grip score was calculated by summing and dividing by two the maximum values of the right and left hands.

Peak expiratory flow rate (PEFR, also known as peak flow) is defined as a person's maximum speed of expiration (van Schoor et al., 2012). PEFR primarily reflects large airway flow and depends on the amount of airway obstruction, the voluntary effort and muscular strength of the participant. For the measurements, PEFR was measured using the Mini-Wright peak flow meter. Participants were asked to take a maximum inspiration and to breath out with maximum effort into the peak flow meter. The highest score of three measurements was used in the analyses (van Schoor et al., 2012; Cook et al., 1989).

Calf circumference is a simple anthropometric measure that is highly correlated with muscle mass (Wijnhoven, 2010). The calf circumference was measured to the nearest 0.001 m on the left leg with the participant standing straight, feet 20 cm apart, body weight equally distributed on both feet and at the level of the widest circumference of the calf.

Information on appetite was obtained using the second question from the Dutch translation of the Centre for Epidemiologic Studies Depression scale (CES—D) (Radloff, 1977): "In the past week, I did not feel like eating, my appetite was poor". A score ranging from 1 (no problems with appetite) to 4 (poor appetite) was used in the analyses.

A self-administered questionnaire was used to assessed sleep quality. The presence and frequency of sleep disturbances was measured using three categorical questions regarding having difficulties with sleep onset, sleep continuity, and early morning awakening. Questions were formulated as follows: 'Do you experience difficulties falling asleep?', 'Do you experience interruptions in your sleep?', and 'Do you wake up too early?'. Response options to each of the three questions were 'almost never', 'sometimes', 'frequently', and 'almost always'. Response options were assigned scores of one, two, three, and four, respectively. These scores were summed to compute a scale ranging from 3 (no sleep disturbances) to 12 (many sleep disturbances) (van der Linden et al., 2023).

Information on weight change pertained to the reason of the short-term weight change in the past 6 months. A score was constructed to categorize weight change into four distinct levels,: 1 (low capacity), characterized by involuntary weight loss; 2 weight gain for any reason; 3 voluntary weight loss; and 4 (high capacity) no change in weight. Voluntary weight change can be due to diet or physical activity, while involuntary weight change may be the result of disease, social factors, or unknown reasons.

Sensory

Self-rated items of vision and hearing were considered to construct the sensory domain. Vision was assessed with two items: “Can you read the normal, small print in the newspaper?”, and “Can you recognize someone's face from a distance of 4 m?”. Hearing was assessed with two items: “Can you follow a conversation in a group of three or four persons?”, and “Can you use a normal telephone?” All items could be given on a scale from 1 (no difficulty) to 4 (much difficulty).

Cognition

Cognitive functioning information processing speed, and episodic memory were considered to construct the *cognition* domain. Episodic memory was measured with a 15 Words Test (15WT), which was a Dutch version of the Auditory Verbal Learning Test (Dik et al., 2000; Rey, 1958). In the test, participants were instructed to learn 15 one-syllable nouns, which were read aloud by the interviewer. The same word list was repeated in three trials, and after each trial, the participant was asked to recall as many words as possible. The maximum number of correctly remembered words was used in the analyses, which measured immediate memory.

Information processing speed was assessed by a Coding task, which was an adjusted version of the Alphabet Coding Task, a letter substitution task (Maylor, 1996). The Coding task is not pure a measure of information processing speed but also reflects a global measure of intellectual functioning (Bouma et al., 1996). In the assessment, two rows of characters were shown; each character in the upper row belonged to a character in the lower row. In the test itself, the upper row contained characters and the lower row was empty. The participants were instructed to complete as many two-character combinations as possible by naming the corresponding character. The total assessment consisted of three trials of 1 min, and the score of each trial was defined as the number of completed combinations, irrespective of the number of wrong answers. The mean score for the three trials was used in the analyses (range 1.0 to 42.7, higher score indicated better intellectual functioning).

General cognitive functioning was measured with the Mini-Mental State Examination (MMSE), which is a test consisting of 23 items representing various domains of cognitive functioning: orientation (Folstein et al., 1975). The score ranges from 0 to 30, with a higher score indicating better cognitive capacity.

Psychology

Anxiety symptoms, depressive symptoms, mastery, self-efficacy, and self-esteem were considered to construct the *psychology* domain. Anxiety was measured with the anxiety subscale of the Hospital Anxiety Depression Scale (HADS-A) consisting of seven items (Spinhoven et al., 1997). In the adaptation for LASA, the response options ranged from 1 (rarely or never) to 4 (mostly or always), and the sum score ranged from 0 to 21, with higher scores indicating higher anxiety (de Beurs et al., 1999). Depressive symptoms were assessed with the Center for Epidemiologic Studies Depression Scale (CES—D) scale (Radloff, 1977). The scale consists of 20 items measuring depressive symptoms experienced in the past week. The total score ranges from 0 to 60, with higher scores indicating more severe symptoms. In the present study, the second item which measured appetite was excluded.

Mastery refers to a sense of being in control of events and ongoing situations and was measured with the five-item version of the seven-item Pearlin Mastery Scale (Pearlin and Schooler, 1978; Deeg and Huisman, 2010). The total score ranged from 5 to 25, with higher scores indicating a higher sense of mastery. Self-efficacy, which is the belief of a person in their own ability to organize and execute certain behaviours, was measured with a 12-item version of the General Self-Efficacy Scale (GSES-12) (Bosscher and Smit, 1998), which total score ranged from 12 to 60, with higher scores indicating higher self-efficacy. Self-esteem reflects a person's overall evaluation or appraisal of their own worth. Self-esteem was measured by an adapted version of the Rosenberg Self-esteem scale (Rosenberg, 1965). A score ranging from 0 to 4 was used in the analyses, with higher scores indicating higher levels of self-esteem.

Locomotion

Walking speed, chair rise test, standing balance, and cardigan test were considered to construct the *locomotion* domain. Walking speed was measured as the time (seconds) needed to walk 3 m, turn around, and then walk back 3 m as fast as possible. A value of 0 was assigned if the participant was in a wheelchair or physically incapable of performing the test. In the chair rise test, participants folded their arms across the chest, and the time to perform five sit-to-stand rises was measured in seconds. A value of 0 was assigned if the participant could not stand up without using their hands or if they did not complete all five rises. Standing balance was measured with feet in the tandem position for a maximum of 10s. A value of 0 was assigned if the participant was physically incapable of performing the test. For testing the ability to put on and take off a cardigan, the time required to put on and take off a cardigan, which was brought in by the interviewer, was scored. A value of 0 was assigned if the participant could not complete the test or completed only with help.

Functional limitation

Participants were asked to which degree they experienced difficulties in performing the following activities of daily living: going up and down the stairs, getting (un-) dressed, sitting down and rising from a chair, cutting one's own toenails, walking 400 m, and using own or public transportation (Stel et al., 2004). Response categories included “yes, without difficulty; yes, with some difficulty; yes, with much difficulty; only with help; no, I cannot” with corresponding scores ranging from 1 to 5. A functional limitation score (range 6–30) was computed by summing the scores of the individual questions.

Appendix 2. Descriptive statistics of the score construction sample together with IC domain scores and composite scores per measurement wave ($n = 3246$).

	Mean (SD) Female	N.Valid cases	Mean (SD) Male	N.Valid cases
T0 (1995/96, 2005/06)				
Age	70.1 (8.6)	1726	70.0 (8.6)	1520
Vitality	45.4 (11.5)	1187	52.8 (13.3)	1099
Sensory	90.6 (14.5)	1657	90.8 (12.9)	1467
Cognition	67.0 (13.3)	1191	62.2 (11.8)	1104
Psychology	70.2 (11.9)	1604	72.9 (11.5)	1445
Locomotion	89.5 (4.2)	1577	82.0 (6.1)	1423
Total	72.8 (7.1)	1121	72.6 (7.0)	1048
T1 (1998/99, 2008/09)				
Age	71.8 (8.2)	1539	72.4(8.2)	1278
Vitality	43.7 (9.5)	1134	54.1 (12.9)	949
Sensory	92.3 (12.9)	1418	93.0 (11.3)	1187
Cognition	65.3 (12.1)	1138	63.3 (12.9)	959
Psychology	68.6 (12.3)	1314	71.8 (11.6)	1149
Locomotion	90.4 (4.6)	1299	84.7 (5.9)	1135
Total	72.2 (6.5)	1050	73.6 (7.1)	900
T2 (2001/02, 2011/12)				
Age	74.6 (7.8)	1352	73.4 (7.3)	1058
Vitality	49.6 (11.1)	1048	53.4 (11.8)	899
Sensory	91.5 (14.4)	1171	92.5 (10.3)	959
Cognition	67.9 (12.8)	1051	65.9 (11.2)	902
Psychology	68.5 (12.4)	1093	72.7 (11.6)	921
Locomotion	81.8 (7.5)	1077	91.8 (3.6)	916
Total	72.6 (6.8)	976	75.7 (6.5)	860
T3 (2005/06, 2015/16)				
Age	75.3 (5.6)	1089	76.3 (6.2)	822
Vitality	45.1 (10.7)	826	48.4 (10.1)	686
Sensory	91.5 (13.2)	861	91.9 (11.9)	700
Cognition	64.7 (12.8)	852	64.6 (13.1)	691
Psychology	66.2 (13.1)	839	69.5 (12.7)	684
Locomotion	78.2 (7.5)	832	73.3 (7.9)	689
Total	69.6 (6.7)	757	70.1 (6.4)	649

Appendix 3. Factor loadings of hand grip strength and walking speed, results from separate unidimensional models of Vitality and Locomotion.

	Vitality
Hand grip strength	0.72
Peak flow	0.68
Calf circumference	0.41
Calf circumference	0.28
Appetite	0.18
Sleep quality	0.05
Walking speed	0.27

	Locomotion
Walking speed	0.85
Chair rise test	0.60
Balance test	0.29
Cardigan test	0.58
Hand grip strength	0.38

Appendix 4. Relationship between composite IC score and functional limitation in a mixed effects linear regression model*

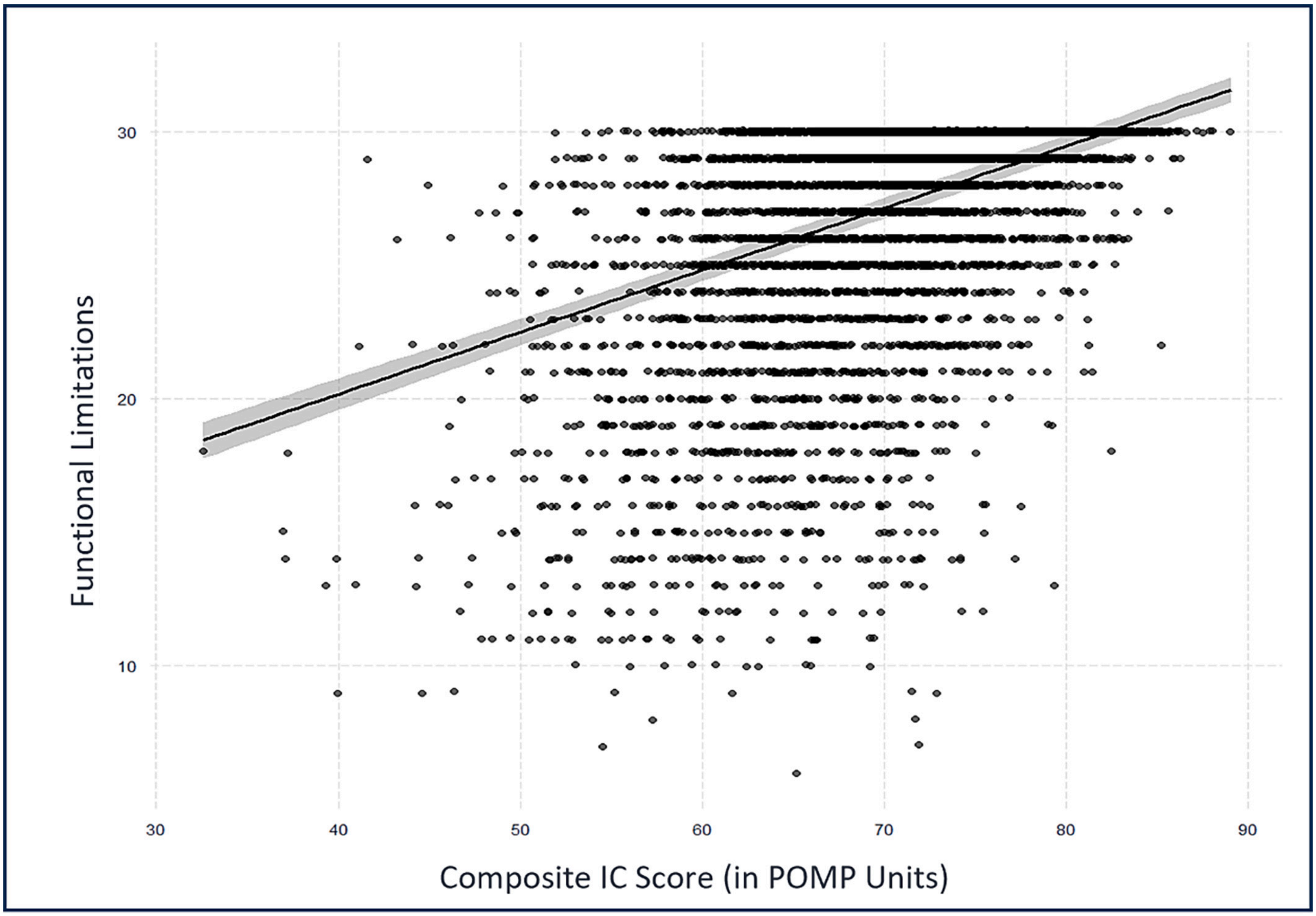


Fig. A4. * The x-axis represents the composite IC score (as POMP units), ranging from low capacity to high capacity while the y-axis displays the functional limitations, ranging from high limitation to low limitation (6 to 30). Each data point, slightly jittered for clarity, represents individual observations across the range of total IC. The shaded area surrounding the regression line denotes the 95 % confidence interval.

References

Aliberti, M.J.R., Bertola, L., Szlejf, C., Oliveira, D., Piovezan, R.D., Cesari, M., et al., 2022. Validating intrinsic capacity to measure healthy aging in an upper middle-income country: findings from the ELSI-Brazil. Lancet Reg Health Am. 12, 100284.

Astrone, P., Perracini, M.R., Martin, F.C., Marsh, D.R., Cesari, M., 2022. The potential of assessment based on the WHO framework of intrinsic capacity in fragility fracture prevention. Aging Clin. Exp. Res. 34 (11), 2635–2643.

Bates, D.M., 2010. lme4: Mixed-Effects Modeling with R. Springer.

Bautmans, I., Knoop, V., Amuthavalli Thiyagarajan, J., Maier, A.B., Beard, J.R., Freiberger, E., et al., 2022. WHO working definition of vitality capacity for healthy longevity monitoring. Lancet Healthy Longev. 3 (11) e789-e96.

Beard, J.R., Jotheeswaran, A.T., Cesari, M., Araujo de Carvalho, I., 2019. The structure and predictive value of intrinsic capacity in a longitudinal study of ageing. BMJ Open 9 (11), e026119.

Beard, J.R., Officer, A., de Carvalho, I.A., Sadana, R., Pot, A.M., Michel, J.-P., et al., 2016. The world report on ageing and health: a policy framework for healthy ageing. Lancet 387 (10033), 2145–2154.

Beard, J.R., Si, Y., Liu, Z., Chenoweth, L., Hanewald, K., 2022. Intrinsic capacity: validation of a new WHO concept for healthy aging in a longitudinal Chinese study. J. Gerontol. A Biol. Sci. Med. Sci. 77 (1), 94–100.

Belloni, G., Cesari, M., 2019. Frailty and intrinsic capacity: two distinct but related constructs. Front Med (Lausanne). 6, 133.

Beyene, M.B., Visvanathan, R., Ahmed, M., Benyamin, B., Beard, J.R., Amare, A.T., 2024. Development and validation of an intrinsic capacity score in the UK biobank study. Maturitas 185, 107976.

Bollen, K.A., Diamantopoulos, A., 2017. In defense of causal-formative indicators: a minority report. Psychol. Methods 22 (3), 581–596.

Bonnefoy, M., Jauffret, M., Kostka, T., Jusot, J., 2002. Usefulness of calf circumference measurement in assessing the nutritional state of hospitalized elderly people. Gerontology 48 (3), 162–169.

Bosscher, R.J., Smit, J.H., 1998. Confirmatory factor analysis of the general self-efficacy scale. Behav. Res. Ther. 36 (3), 339–343.

Bouma JM, Mulder JL, Lindeboom J. Neuropsychologische diagnostiek: handboek: Swets & Zeitlinger; 1996.

Cederholm, T., Jensen, G.L., Correia, M., Gonzalez, M.C., Fukushima, R., Higashiguchi, T., et al., 2019. GLIM criteria for the diagnosis of malnutrition - a consensus report from the global clinical nutrition community. Clin. Nutr. 38 (1), 1–9.

Cenfetelli, R.T., Bassellier, G., 2009. Interpretation of formative measurement in information systems research. MIS Q. 689–707.

Cesari, M., Araujo de Carvalho, I., Amuthavalli Thiyagarajan, J., Cooper, C., Martin, F.C., Reginster, J.-Y., et al., 2018. Evidence for the domains supporting the construct of intrinsic capacity. The Journals of Gerontology: Series A. 73 (12), 1653–1660.

Cesari, M., Sadana, R., Sumi, Y., Amuthavalli Thiyagarajan, J., Banerjee, A., 2022. What is intrinsic capacity and why should nutrition be included in the vitality domain? The Journals of Gerontology: Series A. 77 (1), 91–93.

Charles, S.T., Carstensen, L.L., 2010. Social and emotional aging. Annu. Rev. Psychol. 61 (1), 383–409.

Chen, Y.-J., Kukreti, S., Yang, H.-L., Liu, C.-C., Yeh, Y.-C., Fung, X.C.C., et al., 2023. Psychometric properties of instruments assessing intrinsic capacity: a systematic review. Asian Journal of Social Health and Behavior. 6 (4).

Clark, L.A., Watson, D., 2019. Constructing validity: new developments in creating objective measuring instruments. Psychol. Assess. 31 (12), 1412–1427.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Routledge.

Cohen, P., Cohen, J., Aiken, L.S., West, S.G., 1999. The problem of units and the circumstance for POMP. Multivar. Behav. Res. 34 (3), 315–346.

Cook, N.R., Evans, D.A., Scherr, P.A., Speizer, F.E., Vedal, S., Branch, L.G., et al., 1989. Peak expiratory flow rate in an elderly population. Am. J. Epidemiol. 130 (1), 66–78.

Cox, N.J., Morrison, L., Ibrahim, K., Robinson, S.M., Sayer, A.A., Roberts, H.C., 2020. New horizons in appetite and the anorexia of ageing. Age Ageing 49 (4), 526–534.

Cruz-Jentoft, A.J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., et al., 2018. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 48 (1), 16–31.

de Beurs, E., Beekman, A.T., van Balkom, A.J., Deeg, D.J., van Dyck, R., van Tilburg, W., 1999. Consequences of anxiety in older persons: its effect on disability, well-being and use of health services. Psychol. Med. 29 (3), 583–593.

Deeg, D.J.H., Huisman, M., 2010. Cohort differences in 3-year adaptation to health problems among Dutch middle-aged, 1992–1995 and 2002–2005. Eur. J. Ageing 7 (3), 157–165.

Diaz, T., Banerjee, A., 2023. Welcome to the supplement on measurement of healthy ageing. Age Ageing 52(Supplement_4):iv1-iv2.

Dik, M.G., Jonker, C., Bouter, L.M., Geerlings, M.I., van Kamp, G.J., Deeg, D.J., 2000. APOE-epsilon4 is associated with memory decline in cognitively impaired elderly. Neurology 54 (7), 1492–1497.

Eekhoff, E.M.W., van Schoor, N.M., Biedermann, J.S., Oosterwerff, M.M., de Jongh, R., Bravenboer, N., et al., 2019. Relative importance of four functional measures as predictors of 15-year mortality in the older Dutch population. BMC Geriatr. 19 (1), 92.

Fairclough, D.L., Cella, D.F., 1996. Functional assessment of Cancer therapy (FACT-G): non-response to individual questions. Qual. Life Res. 5 (3), 321–329.

Fleuren, B.P.I., van Amelsvoort, L.G.P.M., Zijlstra, F.R.H., de Grip, A., Kant, I., 2018. Handling the reflective-formative measurement conundrum: a practical illustration based on sustainable employability. J. Clin. Epidemiol. 103, 71–81.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12 (3), 189–198.

Gaussens, L., González-Bautista, E., Bonnefoy, M., Briand, M., Tavassoli, N., De Souto, Barreto P., et al., 2023. Associations between vitality/nutrition and the other domains of intrinsic capacity based on data from the INSPIRE ICOPE-care program. Nutrients 15 (7), 1567.

George, P.P., Lun, P., Ong, S.P., Lim, W.S., 2021. A rapid review of the measurement of intrinsic capacity in older adults. J. Nutr. Health Aging 25 (6), 774–782.

Golino, H., Thiyagarajan, J.A., Sadana, R., Teles, M., Christensen, A.P., Boker, S.M., 2020. Investigating the Broad Domains of Intrinsic Capacity, Functional Ability and Environment: An Exploratory Graph Analysis Approach for Improving Analytical Methodologies for Measuring Healthy Aging.

Gore, P.G., Kingston, A., Johnson, G.R., Kirkwood, T.B.L., Jagger, C., 2018. New horizons in the compression of functional decline. Age Ageing 47 (6), 764–768.

Granic, A., Davies, K., Jagger, C., R MD, Kirkwood TBL, Sayer AA., 2017. Initial level and rate of change in grip strength predict all-cause mortality in very old adults. Age Ageing 46 (6), 970–976.

Gutiérrez-Robledo, L.M., García-Chanes, R., Pérez-Zepeda, M., 2019. Allostatic load as a biological substrate to intrinsic capacity: a secondary analysis of CRELES. The Journal of nutrition, health and aging. 23 (9), 788–795.

Gutiérrez-Robledo, L.M., García-Chanes, R.E., González-Bautista, E., Rosas-Carrasco, O., 2021. Validation of two intrinsic capacity scales and its relationship with frailty and other outcomes in Mexican community-dwelling older adults. J. Nutr. Health Aging 25 (1), 33–40.

Hair Jr., J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S., 2021a. Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook. Springer Nature.

Hair Jr., J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S., et al., 2021b. Evaluation of Formative Measurement Models. A Workbook, Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R, pp. 91–113.

Hoogendijk, E.O., Deeg, D.J.H., de Breij, S., Klokgieters, S.S., Kok, A.A.L., Stringa, N., et al., 2020. The longitudinal aging study Amsterdam: cohort update 2019 and additional data collections. Eur. J. Epidemiol. 35 (1), 61–74.

Hoogendijk, E.O., Dent, E., Koivunen, K., 2023. Intrinsic capacity: an under-researched concept in geriatrics. Age Ageing 52 (10).

Koivunen, K., Hoogendijk, E.O., Schaap, L.A., Huisman, M., Heymans, M.W., van Schoor, N.M., 2023. Development and validation of an intrinsic capacity composite score in the longitudinal aging study Amsterdam: a formative approach. Aging Clin. Exp. Res. 35 (4), 815–825.

Koivunen, K., Schaap, L.A., Hoogendijk, E.O., Schoonmade, L.J., Huisman, M., van Schoor, N.M., 2022. Exploring the conceptual framework and measurement model of intrinsic capacity defined by the World Health Organization: a scoping review. Ageing Res. Rev. 80, 101685.

Lohmöller, J.-B., 2013. Latent Variable Path Modeling with Partial Least Squares. Springer Science & Business Media.

López-Ortiz, S., Lista, S., Peñín-Grandes, S., Pinto-Fraga, J., Valenzuela, P.L., Nisticò, R., et al., 2022. Defining and assessing intrinsic capacity in older people: a systematic review and a proposed scoring system. Ageing Res. Rev. 79, 101640.

Lu, W.-H., González-Bautista, E., Guyonnet, S., Martinez, L.O., Lucas, A., Parini, A., et al., 2023. Investigating three ways of measuring the intrinsic capacity domain of vitality: nutritional status, handgrip strength and ageing biomarkers. Age Ageing 52 (7).

Ma, L., Chhetri, J.K., Zhang, Y., Liu, P., Chen, Y., Li, Y., Chan, P., 2020. Integrated care for older people screening tool for measuring intrinsic capacity: preliminary findings from ICOPE pilot in China. Front. Med. 7, 576079.

MacCallum, R.C., Tucker, L.R., 1991. Representing sources of error in the common-factor model: implications for theory and practice. Psychol. Bull. 109 (3), 502.

Maylor, E.A., 1996. Age-related impairment in an event-based prospective-memory task. Psychol. Aging 11 (1), 74–78.

Mendenhall, A.N., Fristad, M.A., Early, T.J., 2009. Factors influencing service utilization and mood symptom severity in children with mood disorders: effects of multifamily psychoeducation groups (MFPGs). J. Consult. Clin. Psychol. 77 (3), 463–473.

Middleton, A., Fritz, S.L., Lusardi, M., 2015. Walking speed: the functional vital sign. J. Aging Phys. Act. 23 (2), 314–322.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., et al., 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63 (7), 737–745.

Nascimento, L.M.D., Cruz, T., Silva, J., Silva, L.P., Inácio, B.B., Sadamitsu, C.M.O., Scheicher, M.E., 2023. Use of intrinsic capacity domains as a screening tool in public health. Int. J. Environ. Res. Public Health 20 (5).

Nayasista, A.H., Tinduh, D., Alit Pawana, I.P., Mei Wulan, S.M., Utomo, D.N., Soenarnatalina, M., 2022. Effect of combined locomotor training and aerobic exercise on increasing handgrip strength in elderly with locomotive syndrome: a randomised controlled trial. Ann. Med. Surg. 78, 103800.

Pearlin, L.I., Schooler, C., 1978. The structure of coping. J. Health Soc. Behav. 2-21.

Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1 (3), 385–401.

Ray S, Danks N, Calero Valdez A. SEMinR: Domain-specific language for building, estimating, and visualizing structural equation models in R. Estimating, and Visualizing Structural Equation Models in R (August 6, 2021). 2021.

Revelle, W., Revelle, M.W., 2015. Package 'psych'. The comprehensive R archive network. 337 (338).

Rey, A., 1958. L'examen clinique en psychologie [The clinical examination in psychology.].. Presses Universitaries De France, Oxford, England, p. 222.

Rodríguez-Laso, Á., García-García, F.J., Rodríguez-Mañas, L., 2023. The ICOPE intrinsic capacity screening tool: measurement structure and predictive validity of dependence and hospitalization. J. Nutr. Health Aging 27 (10), 808–816.

Rosenberg, M., 1965. Society and the Adolescent Self-Image. Princeton University Press.

Salinas-Rodríguez, A., González-Bautista, E., Rivera-Almaraz, A., Manrique-Espinoza, B., 2022. Longitudinal trajectories of intrinsic capacity and their association with quality of life and disability. Maturitas 161, 49–54.

Sanchez-Niubo, A., Forero, C.G., Wu, Y.-T., Giné-Vázquez, I., Prina, M., De La Fuente, J., et al., 2020. Development of a common scale for measuring healthy ageing across the world: results from the ATHLOS consortium. Int. J. Epidemiol. 50 (3), 880–892.

Sanchez-Rodriguez, D., Demonceau, C., Bruyère, O., Cavalier, E., Reginster, J.Y., Beaudart, C., 2023. Intrinsic capacity and risk of death: focus on the impact of using different diagnostic criteria for the nutritional domain. Maturitas 176, 107817.

Sayer, A.A., Kirkwood, T.B., 2015. Grip strength and mortality: a biomarker of ageing? Lancet 386 (9990), 226–227.

Si, Y., Hanewald, K., Chen, S., Li, B., Bateman, H., Beard, J.R., 2023. Life-course inequalities in intrinsic capacity and healthy ageing. China. Bull World Health Organ. 101 (5), 307-16c.

Sijtsma, K., Ellis, J.L., Borsboom, D., 2024. Recognize the value of the sum score, Psychometrics' greatest accomplishment. Psychometrika 89 (1), 84–117.

Spinhoven, P., Ormel, J., Sloekers, P., Kempen, G., Speckens, A.E., van Hemert, A.M., 1997. A validation study of the hospital anxiety and depression scale (HADS) in different groups of Dutch subjects. Psychol. Med. 27 (2), 363–370.

Stel, V.S., Pluijm, S.M.F., Deeg, D.J.H., Smit, J.H., Bouter, L.M., Lips, P., 2004. Functional limitations and poor physical performance as independent risk factors for self-reported fractures in older persons. Osteoporos. Int. 15 (9), 742–750.

van der Linden, M., Olthof, M.R., Wijnhoven, H.A.H., 2023. The association between caffeine consumption from coffee and tea and sleep health in male and female older adults: a cross-sectional study. Nutrients 16 (1).

van Schoor, N.M., de Jongh, R.T., Daniels, J.M., Heymans, M.W., Deeg, D.J., Lips, P., 2012. Peak expiratory flow rate shows a gender-specific association with vitamin D deficiency. J. Clin. Endocrinol. Metab. 97 (6), 2164–2171.

van Schoor, N.M., de Jongh, R.T., Lips, P., Deeg, D.J., Kok, A.A., 2024. Long-term trajectories of peak expiratory flow rate in older men and women show linear decline mainly determined by baseline levels. Aging Clin. Exp. Res. 36 (1), 93.

Widaman, K.F., Revelle, W., 2023. Thinking thrice about sum scores, and then some more about measurement and analysis. Behav. Res. Methods 55 (2), 788–806.

Wijnhoven HAH, van Bokhorst-de van der Schueren MAE, Heymans MW, de Vet HCW, Kruizenga HM, Twisk JW, Visser M. Low Mid-Upper Arm Circumference, Calf circumference, and body mass index and mortality in older persons. The Journals of Gerontology: Series A. 2010;65A(10):1107–14.

Wister, A.V., Cosco, T.D., 2020. Introduction: Perspectives of Resilience and Aging. Emerging Science and Future Possibilities, Resilience and Aging, pp. 1–14.

World Health Organization. WHO clinical consortium on healthy ageing: topic focus: frailty and intrinsic capacity: report of consortium meeting, 1–2 December 2016 in Geneva, Switzerland. Geneva: World Health Organization; 2017 2017. Contract No.: WHO/FWC/ALC/17.2.

World Health Organization, 2019a. Integrated Care for Older People (ICOPE): Guidance for Person-Centred Assessment and Pathways in Primary Care. World Health Organization.

World Health Organization. WHO Clinical Consortium on Healthy Ageing 2018: Report of Consortium Meeting, 11–12 December 2018, Geneva, Switzerland. Geneva: World Health Organization; 2019 2019b. (Contract No.: WHO/FWC/ALC/19.2).

World Health Organization, 2021. Decade of Healthy Ageing: Baseline Report. World Health Organization.

Yan Wang, N., Liu, X., Kong, X., Sumi, Y., Chhetri, J.K., Hu, L., et al., 2024. Implementation and impact of the World Health Organization integrated care for older people (ICOPE) program in China: a randomised controlled trial. Age Ageing 53(1):afad249.

Zhao, J., Chhetri, J.K., Chang, Y., Zheng, Z., Ma, L., Chan, P., 2021. Intrinsic capacity vs. multimorbidity: a function-centered construct predicts disability better than a disease-based approach in a community-dwelling older population cohort. Front. Med. 8, 753295.