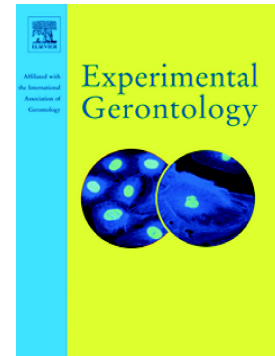


Developing and validating intrinsic capacity scores in early adulthood and midlife: 15-year follow-up from the Northern Finland Birth Cohort 1966

Leah Hillari, Georgina M. Hosang, Rozenn Nedelec, Alex Lewin, Marjo-Riitta Järvelin



PII: S0531-5565(26)00319-0

DOI: <https://doi.org/10.1016/j.exger.2026.113340>

Reference: EXG 113340

To appear in: *Experimental Gerontology*

Received date: 12 March 2026

Revised date: 9 September 2026

Accepted date: 23 September 2026

Please cite this article as: L. Hillari, G.M. Hosang, R. Nedelec, et al., Developing and validating intrinsic capacity scores in early adulthood and midlife: 15-year follow-up from the Northern Finland Birth Cohort 1966, *Experimental Gerontology* (2026), <https://doi.org/10.1016/j.exger.2026.113340>

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Developing and validating intrinsic capacity scores in early adulthood and midlife: 15-year follow-up from the Northern Finland Birth Cohort 1966

Authors

Leah Hillari, MSc ^{a*}, Georgina M. Hosang, PhD ^{bc}, Rozenn Nedelec, PhD ^d, Alex Lewin, PhD ^e, Marjo-Riitta Järvelin, MD, PhD ^{adefgh}

Author Affiliations

^a Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, W12 0BZ, United Kingdom

^b Centre for Psychiatry & Mental Health, Wolfson Institute of Population Health, Faculty of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, United Kingdom

^c Department of Psychology, Kingston University, Kingston upon Thames, KT1 2EE, United Kingdom

^d Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, 90014, Finland

^e Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, United Kingdom

^f Unit of Primary Health Care, Oulu University Hospital, Oulu, 90220, Finland

^g Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, UB8 3PH, United Kingdom

^h Medical Research Council Centre for Environment and Health, School of Public Health, Imperial College London, London, W12 0BZ, United Kingdom

***Corresponding author**

Leah Hillari

Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, 86 Wood Lane, London, W12 0BZ, United Kingdom

Email: l.hillari22@imperial.ac.uk; **Tel:** +44 7932 305088

Journal Pre-proof

Abstract

Background: Intrinsic capacity (IC) - the aggregate of physical and mental capacities - has been studied in older adults, but research on its profile and prognostic value earlier in adulthood is scarce. Existing IC measurement approaches also lack theoretical alignment, psychometric evaluation, and validation. We therefore developed and validated IC scores in adulthood using a theory-aligned formative-reflective approach.

Methods: IC was assessed at ages 31 and 46 in 8,756 participants from the Northern Finland Birth Cohort 1966 using 33 variables across psychological, cognitive-sensory, locomotor, and vitality domains. Domain scores were derived using unidimensional confirmatory factor analysis and aggregated to composite scores. Construct validity of scores was examined through stability, change and known-groups gradients. Predictive validity was tested using covariate-adjusted logistic regression models of associations between 31-year IC and incident mental disorder, cardiovascular disease, type 2 diabetes, functional limitations, and low well-being at age 46.

Results: Domain models demonstrated acceptable fit and reliability. Mean composite IC was moderate-to-high at age 31 (67 on a 0–100 scale) but declined significantly by age 46 (61), with the steepest reductions in locomotor capacity. Lower composite IC was observed in groups with socioeconomic disadvantage, unhealthy behaviours, and health conditions. Higher 31-year composite IC predicted lower odds of all outcomes at age 46 (odds ratios per 10-point increase: 0.69–0.82, all $p < 0.001$).

Conclusions: Composite IC declined from early adulthood to midlife, and predicted key physical, mental, and functional outcomes. Formative-reflective approaches and earlier IC assessment may strengthen prevention-oriented healthy ageing strategies.

Keywords: Intrinsic Capacity, Life course, Functional ability, Healthy ageing, Validation, Longitudinal, Northern Finland Birth Cohort 1966

1. Introduction

Global population ageing is accelerating¹. This shift is accompanied by rising burdens of non-communicable disease and disability, with substantial healthcare, societal, and economic consequences^{1,2}. Yet poor health and functional decline are not inevitable consequences of ageing. Early intervention and supportive environments can prevent disease and disability and enable individuals to maintain high levels of function and autonomy into advanced age¹. This perspective underpins the World Health Organization (WHO)'s framework of Healthy Ageing, which shifts emphasis from disease to the preservation of function, autonomy, and well-being during ageing¹. In this framework, functional ability underpins well-being and the capacity to continue engagement in valued activities during ageing. Functional ability, in turn, arises from the interaction between an individual's intrinsic capacity (IC) - the totality of available physical and mental capacities - and the surrounding environment¹. IC is subdivided into five core domains: cognitive, psychological, sensory, locomotor and vitality capacity (see *Section 2.2.1* for details)³.

A growing body of research has established IC as a sensitive marker of physical and mental reserve, and a robust predictor of disability, functional decline, frailty, hospitalisation, mortality and poor quality of life in older adults (≥ 65 years)^{4,5}. Emerging evidence further suggests that lower IC is predictive of incident cardiovascular disease⁶, type 2 diabetes⁷, and diagnosed depressive disorders⁸.

Despite mounting evidence for its value as a marker of healthy ageing and predictor of health outcomes, IC is not the only multidimensional approach to assessing healthy ageing. Another widely used approach is frailty assessment, which focuses on vulnerability to stressors arising from reduced physiological reserve. This vulnerability is commonly operationalised using the frailty index, which quantifies the accumulation of symptoms, diseases, impairments and disabilities during ageing^{9,10}. By contrast, IC focuses on the physical and mental capacities

an individual can draw upon to maintain functional ability during ageing ¹. Although low IC and frailty overlap, they can identify different groups ¹¹ and are increasingly regarded as distinct but complementary constructs ¹⁰. Frailty is particularly useful for identifying heightened vulnerability and complex care needs in older populations, whereas IC supports earlier assessment, monitoring and prevention in younger populations ^{10,12}. Consistent with this role, IC has been shown to provide additional prognostic information beyond conventional frailty measures for dementia and mortality outcomes ¹³. IC may also be considered as an antecedent of frailty, as lower IC has predicted incident frailty ¹⁴ and impairments in IC domains have predicted transitions from non-frailty to frailty ¹⁵. Together, these findings support IC as a complementary approach for identifying decline in capacity and reserve earlier in the pathway towards frailty and other adverse health outcomes. Beyond frailty-based measures, several other multidimensional approaches have been developed to conceptualise and assess healthy ageing. These approaches are described using overlapping terms, including healthy, successful, active and productive ageing, and vary substantially in their definitions, constituent domains and measurement methods ^{16–18}. This conceptual and methodological heterogeneity limits the comparability and synthesis of findings across studies. The WHO Healthy Ageing framework helps address this limitation by providing an evidence-informed reference structure for defining and assessing healthy ageing ^{10,18}. Within this framework, IC was delineated through extensive evidence synthesis and expert consultation ³, with its five domains broadly covering the essential individual-level components of healthy ageing identified in epidemiological reviews ¹⁶. Given its evidence-based, comprehensive and function-oriented structure, IC is particularly well suited to monitoring the capacities that underpin healthy ageing in younger adult populations. Despite these strengths, important empirical and methodological gaps remain in the development and application of IC measures. First, although IC is conceived as a dynamic

attribute that changes across the life course ¹, most studies have focused on older adults ¹⁹ and assessed IC at a single timepoint in old age ⁵. Recent work has started to model IC trajectories in older cohorts ^{20,21}, but longitudinal studies starting in young adulthood are largely absent. Consequently, the level, variability, and trajectory of IC during periods when most capacities are near their peak remain poorly characterised ^{22,23}. Understanding IC in early adulthood, its evolution across midlife, and its prognostic value for later-life health is therefore critical, as this period may represent a key window for prevention-oriented interventions to promote healthy ageing.

Second, another central limitation of the IC literature concerns substantial inconsistency in the construction of domain and composite scores, typically without explicit alignment to IC's measurement model, that is, the assumed relationship between the construct and its observed indicators ^{24,25}. IC is defined as “the composite of all physical and mental capacities”, a definition which naturally implies a formative measurement model at the composite level, whereby overall capacity is formed by distinct domains, each contributing a unique aspect of capacity ^{26,27}. A recent review supports this view, arguing that IC domains lack a shared physiological basis, are not necessarily correlated, and are not conceptually interchangeable ²⁴. Within IC domains, however, indicators are expected to reflect a common latent construct and share variance. This points to a reflective measurement model at the domain level, in which latent domains, such as locomotor capacity, give rise to their observed indicators, including gait and balance measures. Together these considerations support a dual-level structure of IC: reflective measurement at the domain level and a formative aggregation across domains at the composite level ^{27,28}.

Despite its central importance for score construction, IC's measurement model has rarely been specified in prior work ²⁴. Existing scoring approaches can be broadly classified into two approaches: additive approaches that average or sum domain scores ^{29,30}, and single-

factor latent models ^{22,23} that load all indicators across domains onto a “global IC” factor. The latter implicitly assume a reflective composite, treating indicators across domains as interchangeable manifestations of a single capacity trait. This assumption is difficult to justify given the conceptual and biological distinctiveness of IC domains and their divergent ageing trajectories ^{24,31}.

Additive indices are more consistent with a formative conceptualisation of composite IC but commonly apply summation at both the composite and domain levels. Summation at the domain levels overlooks the potential reflective nature of domains, implicitly assuming equal indicator weighting and sufficient shared variance among indicators to reflect a common underlying construct. These assumptions are rarely tested, and equal weighting is particularly questionable given that indicators often contribute differentially to their respective domain. Moreover, domain-level summation precludes formal testing of domain structure and coherence, reliability, validity and longitudinal measurement invariance. Together, these limitations call for a dual-level strategy that aligns scoring with theory, respecting the formative nature of composite IC while providing psychometric insights in its reflective domains.

1.1 Study rationale and aims

To address key gaps in IC research, including the limited focus on younger age groups and inconsistencies in IC operationalisation, this study examined IC longitudinally in early adulthood and midlife in a large, population-based birth cohort.

The aims of this study were threefold: (i) to develop domain and composite IC scores at ages 31 and 46 using a dual-level measurement strategy, with reflective modelling of domains and formative aggregation at the composite level; (ii) to evaluate the psychometric robustness and construct validity of the derived IC scores, including domain-level model fit and reliability, temporal stability and change, and known-groups gradients across socioeconomic,

behavioural, and health characteristics; (iii) to examine the predictive validity of IC scores for subsequent physical, mental, and functional health outcomes.

Together, these analyses extend IC research beyond older populations to early adulthood and midlife while advancing a theory-aligned approach to IC measurement and validation.

2. Methods

2.1 Study design and sample

Data were drawn from the Northern Finland Birth Cohort 1966 (NFBC1966), a prospective, population-based birth cohort, including 12,231 births (96% of all births in the two northernmost provinces of Finland in 1966)^{32,33}. Pregnant women with an expected delivery date in 1966 were recruited on average around the 16th gestational week. The offspring have subsequently been followed up at birth and ages 1, 14, 31, and 46 years^{32,33}.

The present study used questionnaire and clinical data collected at ages 31 (1997–98) and 46 years (2012–14). At age 31, a subset of cohort members living in Northern Finland and capital areas were invited for clinical assessment, while all living cohort members with a known address were invited at age 46. In total, 8,690 participants returned questionnaires and 6,007 attended clinical examinations at age 31, while 7,146 participants returned questionnaires and 5,832 attended clinical examinations at age 46^{32,33}. A total of 4,016 participants completed at least one questionnaire and attended the clinical examination at both adult follow-ups³³.

For the present analyses, the imputation base was restricted to participants who contributed any study data (questionnaire and/or clinical) at age 31, age 46, or both, yielding 8,998 eligible participants. Of these, 2,333 contributed data at age 31 only, 720 at age 46 only, and 5,945 at both follow-ups. After imputation, 242 participants who had not fasted before blood

collection at age 31 were excluded to ensure valid assessment of metabolic markers, resulting in a full analytical sample of 8,756 participants. No corresponding exclusion was required at age 46 because blood sampling was restricted to fasted participants.

The full analytical sample of 8,756 was used for the construction of IC scores and most validation analyses. The only validation analyses that did not use the full analytical sample were the predictive-validity analyses of incident mental disorder, cardiovascular disease and type 2 diabetes. For these analyses, participants with the respective condition already present at age 31 were excluded to minimise reverse causation. Consequently, the outcome-specific samples were $n = 8,010$ – $8,054$ for mental disorder, $n = 8,479$ – $8,525$ for cardiovascular disease and $n = 8,630$ – $8,652$ for type 2 diabetes across the multiply imputed datasets.

Among the 12,058 live-born NFBC1966 cohort members, cumulative mortality was 3.5% by the 31-year follow-up, increasing to 5.6% by the 46-year follow-up. Among the 11,640 cohort members alive at age 31, 2.2% died during the approximately 15-year interval between the 31- and 46-year follow-ups. Participant flow and analysis-specific samples are summarised in *Figure 1*.

2.2 Variables

2.2.1 Intrinsic capacity variables at age 31 and 46

IC was operationalised using the WHO five-domain framework with a total of 33 variables assessed at ages 31 and 46¹. Variables were selected based on WHO recommendations and availability at both ages (see *eTable 1*, Supplemental Materials for the variable selection process). Only one cognitive indicator and two sensory indicators, each with a limited number of response categories, were available at both timepoints. Treating the single cognitive item and the two sensory items as separate domains would have prevented a consistent factor-based scoring approach across domains, including its advantages for psychometric evaluation. The separate cognitive and sensory scores also had limited

granularity and marked ceiling effects, which could reduce sensitivity to between-person and longitudinal differences. We therefore combined cognition, vision and hearing in a single cognitive-sensory factor model. This was conceptually supported by the shared neurophysiological basis of cognitive and sensory functioning³⁴ and empirically supported by moderate-to-strong inter-item correlations, strong standardised factor loadings, and acceptable-to-excellent reliability. Combining the three indicators produced a more differentiated score with a greater range of possible values and reduced the disproportionate influence of any single indicator. Because this operationalisation departs from the WHO five-domain structure, we additionally examined the robustness of the findings in sensitivity analyses that retained cognitive and sensory capacity as separate domains (see *Section 2.3.4* for details). The following variables were used for IC score construction (see *eTable 2*, Supplemental Materials for measurement details):

Psychological capacity: Sleep, life satisfaction, life orientation (dispositional optimism versus pessimism), depression and anxiety symptoms.

Cognitive-sensory capacity: Self-rated cognitive function, vision and hearing.

Locomotor capacity: Self-rated mobility and physical performance capacity.

Vitality capacity is commonly defined as an underlying physiological driver of IC, subdivided into three components: neuromuscular function, energy and metabolism, and immune and stress response³⁵.

- Neuromuscular function: Isometric handgrip strength, cardiorespiratory fitness, and self-reported respiratory function.
- Energy and metabolism: Body mass index, waist-to-hip ratio, isometric back muscle endurance, blood-derived metabolic markers (fasting insulin, glucose, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglycerides), self-reported energy levels and appetite loss.

- Immune and stress response: Systolic blood pressure and diastolic blood pressure, inflammation via high-sensitivity C-reactive protein, self-perceived immune status.

2.2.2 Variables for intrinsic capacity score validation

Known-groups validity, a form of construct validity, refers to the ability of a measure to distinguish between groups expected *a priori* to differ in the underlying construct^{36,37}. In the present study, known-groups validity was assessed at age 31 using variables that have consistently been linked to IC: sociodemographic characteristics (sex, education, occupation, gross household income), health behaviours (smoking, alcohol intake, physical activity), and health status (physician-diagnosed or treated mental disorders, cardiovascular disease, diabetes, self-rated well-being, and functional limitations)^{4,38}. Given the well documented links between IC and adverse health outcomes⁴, predictive validity was tested at age 46 using self-reported physician-diagnosed or treated mental disorders, cardiovascular disease, and type 2 diabetes at age 46. Self-rated well-being and functional limitations were additionally included, which are the two primary outcomes of IC within the WHO Healthy Ageing framework^{1,5}. Given their well-established associations with IC and adverse health outcomes³⁸, the sociodemographic and behavioural variables described above were included as confounders in predictive validity analyses. Full measurement details of all variables used for IC score validation are provided in *eTable 3*, Supplemental Materials.

2.3 Analytical approach

2.3.1 Data preparation and imputation procedure

All analyses were conducted in R (version 4.5.2)³⁹. Statistical significance was set at $p \leq 0.05$. p -values are reported to three significant digits, with values < 0.001 shown in scientific notation in tables and summarised in text as $p < 0.001$. Values below numerical precision are reported as bounds.

Continuous variables were standardised to a 0–100 scale using the percent-of-maximum-possible (POMP) method⁴⁰. POMP was applied separately for females and males to adjust for sex-related differences in their distributions^{28,41}. Variables were coded so that higher scores reflected worse capacity. Continuous variables with notable skewness were log- or Box-Cox-transformed.

Missing data were substantial: 84.5% of participants had incomplete data, with cardiorespiratory fitness showing the highest missingness (48.5%). Data were not missing completely at random (MCAR) (Little's test $p < 0.001$). We addressed missing data using Multivariate Imputation by Chained Equations⁴², including all IC variables, predictors, outcomes, and relevant auxiliary variables⁴³. Predictive mean matching was used for continuous variables, logistic regression models for binary variables, and proportional odds models for ordinal variables⁴⁴. Fifty imputed datasets were generated to reflect the proportion of missingness in cardiorespiratory fitness⁴⁴. Ten iterations per dataset were used, with trace plots confirming convergence. Complete-case analyses yielded similar results (not reported here).

2.3.2 Construction of domain-specific and composite intrinsic capacity scores

Analyses testing factor-analytic assumptions, data suitability and sampling adequacy are detailed in *eText 1* and *eTable 4*, Supplemental Materials. Relationships among the IC indicators within each domain are presented in correlation matrices (*eFigures 1–4*, Supplemental Materials). Domain-specific IC scores were derived from unidimensional confirmatory factor analysis (CFA) at each timepoint. Domains with continuous variables only were modelled using the robust maximum likelihood method and domains including ordinal variables were estimated using the weighted least squares mean and variance adjusted (WLSMV) method⁴⁵.

Model fit was assessed using the comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA), with conventional thresholds applied (adequate fit: CFI/TLI ≥ 0.90 , RMSEA ≤ 0.08 ; good fit: CFI/TLI ≥ 0.95 , RMSEA ≤ 0.06)³⁷. CFA results were averaged across imputed datasets. Where unidimensional models showed poor fit, exploratory factor analysis was performed to explore potential multidimensional structures.

Variables with statistically significant and practically meaningful factor loadings ($\geq |0.30|$) were retained⁴⁶. The statistical significance of factor loadings was estimated at an alpha level of 0.05 using Norman and Streiner's⁴⁷ formula $\lambda \geq 3.92 / \sqrt{(N - 2)}$, where λ represents the minimum absolute factor loading value required for statistical significance. Standardised factor loadings linking each retained IC indicator to its corresponding domain factor are presented in *eFigures 5–8*, Supplemental Materials. Reliability of CFA models was evaluated using McDonald's omega (ω). ω values were averaged across imputations.

Temporal measurement invariance across ages 31 and 46 was tested using multi-group CFA (*eText 2*, *eTable 5*, Supplemental Materials). As invariance was not consistently supported for all domains, we imposed the pooled factor loadings from the 31-year models onto the 46-year models. This ensured that factor scores were placed on a common metric and directly comparable across time⁴⁸. Factor scores were extracted separately within each imputed dataset, reverse-coded for interpretability so that higher values indicated greater capacity, and rescaled using the POMP method to 0–100 scores, where one point represents 1% of the observed capacity range. Composite IC scores were computed as the arithmetic mean of the domain scores within each imputed dataset. Descriptive statistics were averaged across imputations, while inferential analyses were conducted within each imputed dataset and pooled as described below.

To characterise the extent of overlap and distinctiveness among the domains contributing to composite IC, relationships among the domain-specific scores were examined. Pearson correlations between domain scores were calculated separately at ages 31 and 46 within each imputed dataset. Correlations were Fisher- z transformed, pooled across imputations using Rubin's rules and back-transformed for presentation. Spearman rank correlations were additionally performed for the cognitive-sensory scores as they showed some deviation from normality.

2.3.3 Validation of intrinsic capacity scores

Construct validity of IC scores was examined through temporal stability, change and known-groups gradients. Temporal stability of domain-specific and composite IC scores between ages 31 and 46 was assessed using Pearson's correlations (r) and Spearman's rank-order correlations (ρ) when parametric assumptions were violated. Correlations were estimated within each imputed dataset, Fisher- z transformed, pooled using Rubin's rules, and back-transformed to the original scale.

Change in IC scores from age 31 to 46 years was tested using paired t -tests pooled across imputations following Rubin's rules. Mean scores, standard deviations and Cohen's d_z (standardised mean change) were averaged across imputations. For domains with non-normal paired differences, Wilcoxon signed-rank tests were additionally performed; Hodges–Lehmann (HL) median differences and 95% confidence intervals (CIs) were summarised across imputations (Rubin's rules are not defined for rank-based tests), and p -values were combined using Fisher's method. Age-related declines in IC domains were considered evidence of temporal construct validity.

Known-groups validity was evaluated at 31 years by examining univariate associations between IC scores and known sociodemographic, behavioural and health groups^{4,38}. Group differences were tested using Welch's independent t -tests for binary variables and linear

models with omnibus Wald tests for variables with ≥ 3 categories. For significant omnibus tests, pairwise contrasts were examined using Welch's independent t-tests and Holm adjustment. All estimates were pooled across imputations using Rubin's rules, and descriptive statistics were averaged across imputations.

Predictive validity of IC scores was examined using logistic regression models estimating associations between IC scores at age 31 and odds of incident mental disorders,

cardiovascular disease, type 2 diabetes, well-being, and functional limitations at age 46.

Logistic regression models were fitted within each imputed dataset and pooled using Rubin's rules. Models were adjusted sequentially: unadjusted (Model 1); adjusted for sex (Model 2); additionally for education, occupation, and income (Model 3); and further for smoking, alcohol use, and physical activity (Model 4). To aid interpretation, odds ratios (ORs) and 95% CIs are presented per 10-point increase in IC on the 0–100 POMP scale.

2.3.4 Sensitivity analyses

Two sets of sensitivity analyses were conducted. First, we assessed whether the findings were robust to retaining cognition and sensory capacity as separate domains rather than combining them into a single cognitive-sensory score. Within each imputed dataset, the three-category cognitive function item was reverse-coded and POMP-transformed to a 0–100 scale, with higher values indicating greater capacity. Vision and hearing were similarly reverse-coded and POMP-transformed separately to 0–100 scales and then averaged with equal weight to derive a sensory capacity score. An alternative five-domain composite was calculated as the arithmetic mean of cognitive, sensory, psychological, locomotor and vitality capacity within each imputed dataset, consistent with the primary composite. Agreement between the primary and alternative composite scores was assessed using Pearson and Spearman correlations.

Validation analyses of temporal stability, longitudinal change and fully adjusted predictive validity were repeated for the alternative composite and the separate cognitive and sensory

domain scores using the same procedures as in the primary analyses. As an additional sensitivity analysis of known-groups validity, we examined whether combining cognition and sensory capacity had obscured domain-specific educational differences. The separately constructed cognitive and sensory scores were therefore compared by educational level at age 31 using the same procedure as for the primary cognitive–sensory score.

Second, sex-specific sensitivity analyses examined whether longitudinal changes in IC and associations between IC and health outcomes differed between females and males. Mean IC levels and changes from age 31 to 46 were estimated separately by sex using the same procedures as in the primary analyses, and differences in mean change between females and males were estimated within each imputed dataset and pooled using Rubin’s rules. Sex modification of associations between IC and health outcomes was assessed by adding an IC-by-sex interaction term to each fully adjusted regression model. Sex-specific ORs per 10-point increase in IC and 95% CIs were derived from the interaction models using the same covariates, outcome-specific samples and baseline disease exclusions as in the primary analyses.

2.3.5 Attrition analyses

Potential attrition bias was assessed by comparing participants lost to the 46-year follow-up to those who remained in the study, and participants with incomplete IC data to those with complete IC data. Comparisons were made on key sociodemographic, behavioural, and health variables measured at baseline (31 years) using independent t-tests for continuous variables and Pearson’s chi-square or Fisher’s exact tests for categorical variables.

3. Results

3.1 Construction of domain-specific and composite intrinsic capacity scores

3.1.1 Model fit and reliability

Factor-analytic assumptions were satisfied for all domains after appropriate corrections (see *eText 1* and *eTable 4*, Supplemental Materials). The correlation matrices for each domain showed adequate shared inter-item variance for factor analysis with most absolute correlations ≥ 0.30 (*eFigures 1–4*, Supplemental Materials).

Cognitive-sensory models yielded perfect fit indices, which are not informative given the limited number of variables. Model fit was adequate-to-good for the psychological and locomotor domains at both timepoints, whereas fit was poorer for the vitality domain. Complete-case analyses yielded similarly poorer fit for the vitality models at both ages, suggesting that the lower fit was not primarily an artefact of multiple imputation. All pooled fit statistics for the unidimensional CFA models by domain and timepoint are presented in *eTable 6*, Supplemental Materials. Pooled McDonald's ω values ranged from 0.606 to 0.921 across models and timepoints, with the vitality domains demonstrating the lowest values. Detailed reliability results are described in *eText 3*, Supplemental Materials.

We theorised that the poorer fit and lower ω values for the vitality domains may be attributable to a more complex multidimensional structure underlying vitality-related variables. Indeed, exploratory factor analysis on the vitality domains showed good fit for a four-factor model at 31 years and a three-factor model at 46 years (*eText 4* and *eTable 7*, Supplemental Materials). One-factor models were imposed for the vitality domains to ensure comparable model structure across domains and time.

3.1.2 Factor structure and score distribution

Most indicators loaded meaningfully $\geq |0.30|$ and significantly on their respective domain factors at both ages (*eFigures 5–8*, Supplemental Materials). In the vitality domain, several indicators did not meet loading criteria and were excluded from the final models (age 31:

glucose, energy, respiration, appetite loss, perceived immune status; age 46: appetite loss, handgrip strength, energy). The final vitality models comprised 10 variables per timepoint. Correlations between domains were generally positive but small at age 31 ($r = 0.001$ – 0.272) and small-to-moderate at age 46 ($r = 0.086$ – 0.471) (*eFigure 9*, Supplemental Materials). At age 31, the strongest association was between psychological and locomotor capacity ($r = 0.272$), whereas psychological and vitality capacity were essentially unrelated ($r = 0.001$). The associations of vitality with psychological and cognitive-sensory capacity were not statistically significant. At age 46, the strongest association was between cognitive-sensory and locomotor capacity ($r = 0.471$), and the weakest was again between psychological and vitality capacity ($r = 0.086$). All associations were statistically significant. Spearman rank correlations for the cognitive-sensory scores yielded a comparable pattern of associations. Overall, these modest inter-domain correlations were consistent with the domains representing related but conceptually distinct components and with their treatment as separate contributors to a formative composite score.

Most extracted domain and composite scores were approximately normally distributed. The cognitive-sensory domains and composite IC scores exhibited slight skew and excess kurtosis but remained well within acceptable limits for parametric testing (± 2 for skewness; ± 7 for excess kurtosis). Graphs showing the distribution of IC scores can be found in *eFigures 10–11*, Supplemental Materials and pooled skewness and excess kurtosis values in *eTable 8*, Supplemental Materials.

3.2 Validation of intrinsic capacity scores

3.2.1 Construct validity – Temporal stability in intrinsic capacity

Pearson correlations between ages 31 and 46 were highest for vitality ($r = 0.586$), followed by locomotor ($r = 0.351$) and psychological capacity ($r = 0.306$), with lower stability for cognitive–sensory capacity ($r = 0.241$). Composite IC showed moderate stability ($r = 0.365$)

(for scatterplots of correlations see *eFigure 12*, Supplemental Materials). Spearman rank correlations for non-normally distributed domains yielded comparable results (composite IC: $\rho = 0.374$; cognitive–sensory IC: $\rho = 0.253$). All correlations were statistically significant (all $p < 0.001$).

3.2.2 Construct validity – Change in intrinsic capacity

Paired-samples t-tests demonstrated significant declines in mean cognitive–sensory, locomotor, and composite IC from 31 to 46 years, while psychological capacity increased significantly (all $p < 0.001$) (*Table 1*). Vitality capacity was, on average, 4.17 points lower at age 46, but this change was not statistically significant ($p = 0.158$). The largest mean decline was observed in locomotor capacity ($\Delta = -10.84$), and composite IC decreased by 5.70 points. Wilcoxon signed-rank tests for non-normally distributed domains corroborated these findings (cognitive-sensory IC: HL median change -9.83 , [95% CI $-14.54, -5.08$], Fisher-combined $p < 0.001$; composite IC: HL median change -5.64 , [95% CI $-9.16, -0.54$], Fisher-combined $p < 0.001$).

Table 1. Pooled mean changes in intrinsic capacity between ages 31 and 46 ($n = 8,756$).

Domain	Mean ₃₁ (SD)	Mean ₄₆ (SD)	Δ (95% CI)	t (df)	p -value	Cohen's d_z
Cog.-sensory	78.78 (31.70)	73.53 (28.16)	-5.25 [$-7.41, -3.10$]	-4.87 (65.3)	7.52×10^{-6}	-0.14
Psychological	60.55 (20.93)	64.07 (20.73)	3.52 [$2.44, 4.59$]	6.51 (83.9)	5.27×10^{-9}	0.14
Locomotor	73.40 (21.41)	62.56 (24.00)	-10.84 [$-11.81, -9.86$]	-22.1 (106)	1.70×10^{-41}	-0.42
Vitality	64.99 (13.43)	60.82 (14.48)	-4.17 [$-10.00, 1.67$]	-1.44 (49.2)	0.158	-0.33
Composite IC	66.96 (16.34)	61.26 (17.59)	-5.70 [$-8.80, -2.60$]	-3.69 (50.8)	5.40×10^{-4}	-0.30

Note. Mean scores and SDs at ages 31 and 46 were averaged across imputations. Within-person mean changes (Δ) and 95% CIs were pooled using Rubin's rules. Cohen's d_z represents the standardised mean change and was averaged across imputations. Positive Δ values indicate capacity increase from age 31 to 46, while negative values indicate capacity decline.

Abbreviations: Cog.-sensory = Cognitive-sensory; SD = standard deviation; CI = confidence interval; df = degrees of freedom; IC = intrinsic capacity.

3.2.3 Construct validity – Known-groups gradients in intrinsic capacity

The final analytical sample of 8,756 individuals comprised 4,480 females (51.2%) and 4,276 males (48.8%) (*Table 2*). Mean IC scores varied significantly across sociodemographic, behavioural, and health-related groups (*Table 2*; for post hoc comparison see *eTable 9*, Supplemental Materials). IC showed clear socioeconomic patterning: participants with higher education, income, and occupational status had higher composite and domain scores compared to those with lower education, income, and lower-status occupations. However, two exceptions emerged: cognitive-sensory IC did not differ by education, and vitality IC did not differ by income.

A health-status gradient was also evident: compared with participants who had functional limitations, low well-being, a mental disorder, diabetes, or cardiovascular disease, those without these conditions had higher composite and domain scores. Exceptions were vitality IC by cardiovascular disease and psychological IC by diabetes, which showed no difference. IC also differed according to different lifestyle patterns. Compared with those not meeting WHO physical activity guidelines, participants who met the guidelines had higher composite and domain scores (except for cognitive-sensory scores). Compared with former and current smokers, never smokers scored higher across domains and composite IC. For alcohol intake, IC differences were more nuanced. Generally, compared with moderate- or high-risk drinkers, abstainers and low-risk drinkers had higher composite and domain IC. Sex differences were domain-specific: men scored higher in psychological and locomotor capacity, whereas women scored higher than men in the cognitive-sensory and vitality domains. However, the composite score did not differ by sex.

Table 2. Pooled mean intrinsic capacity scores by sociodemographic, behavioural and health subgroups at age 31 ($n = 8,756$).

Variable category	All, n (%)	Mean (SD), <i>p</i> -value				
		Composite IC	Cognitive-sensory	Psychological	Locomotor	Vitality
Sex						
Male	4,276 (48.8)	67.28 (16.88), 0.224	77.33 (32.62), 0.004	62.71 (20.95), 3.69×10^{-18}	79.48 (19.91), 2.31×10^{-83}	59.22 (12.21), 8.53×10^{-21}
Female	4,480 (51.2)	66.66 (15.70)	80.16 (30.70)	58.48 (20.70)	67.60 (21.18)	70.50 (11.65)
Education						
Basic or less	5,183 (59.2)	65.25 (16.69), 2.86×10^{-16}	78.07 (31.91), 0.099	59.67 (21.26), 1.13×10^{-5}	71.40 (21.94), 2.09×10^{-22}	62.92 (13.40), 1.18×10^{-19}
Matric./Higher	3,573 (40.8)	69.45 (15.37)	79.80 (31.32)	61.83 (20.38)	76.29 (20.28)	67.99 (12.28)
Occupation						
Upper-level	1,583 (18.1)	71.45 (14.68), 9.48×10^{-31}	80.15 (31.68), 1.85×10^{-6}	65.45 (19.74), 5.95×10^{-52}	79.84 (18.62), 1.63×10^{-42}	67.05 (12.25), 5.20×10^{-22}
Lower/entrepreneur	3,185 (36.4)	68.40 (15.27)	81.15 (30.18)	62.11 (19.95)	72.66 (20.87)	66.54 (12.80)
Manual/farming	2,630 (30.0)	65.58 (16.22)	77.48 (31.97)	59.74 (20.84)	73.50 (21.04)	62.45 (13.26)
Not working	1,358 (15.5)	61.04 (18.36)	74.14 (33.84)	52.76 (22.37)	67.44 (24.26)	63.88 (14.08)
Household income						
1st quartile (lowest)	2,334 (26.7)	63.77 (17.08), 5.77×10^{-12}	78.68 (30.78), 0.019	54.14 (21.84), 1.13×10^{-43}	69.65 (22.73), 7.34×10^{-18}	64.73 (13.78), 0.081
2nd quartile	2,127 (24.3)	67.64 (15.72)	80.79 (30.04)	61.03 (20.65)	73.53 (21.05)	64.59 (13.15)
3rd quartile	2,158 (24.6)	67.86 (15.55)	79.50 (30.93)	62.17 (19.81)	74.20 (20.83)	64.81 (12.98)
4th quartile (highest)	2,137 (24.4)	68.88 (16.21)	76.16 (34.64)	65.45 (19.54)	76.55 (20.23)	65.86 (12.75)
Smoking status						
Never	3,111 (35.5)	69.36 (16.01), 4.31×10^{-20}	80.27 (31.15), 0.012	63.73 (20.14), 2.05×10^{-30}	75.30 (21.67), 1.99×10^{-9}	66.32 (13.02), 7.39×10^{-13}
Former	1,949 (22.3)	67.78 (15.57)	79.46 (31.01)	61.44 (20.25)	73.61 (21.03)	65.90 (12.86)
Current	3,696 (42.2)	64.51 (16.56)	77.16 (32.39)	57.41 (21.48)	71.68 (21.26)	63.38 (13.34)
Alcohol intake						
No alcohol	794 (9.06)	64.45 (18.14), 1.11×10^{-12}	76.87 (32.81), 0.005	61.75 (20.97), 4.02×10^{-25}	66.09 (25.51), 1.05×10^{-14}	64.72 (13.07), 0.004
WHO Low-risk	7,402 (84.5)	67.63 (15.96)	79.39 (31.46)	61.17 (20.65)	74.19 (20.82)	65.17 (13.05)
WHO Moderate-risk	357 (4.08)	63.32 (16.37)	74.75 (32.15)	53.11 (21.64)	73.59 (20.68)	64.27 (14.90)
WHO High-risk	203 (2.32)	58.90 (16.65)	70.89 (32.55)	46.66 (22.30)	72.69 (21.00)	60.93 (14.88)

Variable category	All, n (%)	Mean (SD), <i>p</i> -value				
		Composite IC	Cognitive-sensory	Psychological	Locomotor	Vitality
Physical activity						
< 600 MET-min/week	3,973 (45.4)	63.55 (16.46), 1.37 × 10 ⁻³²	77.85 (31.82), 0.084	58.34 (21.0), 6.46 × 10 ⁻¹⁵	67.29 (21.76), 9.02 × 10 ⁻⁸⁸	62.98 (13.54), 1.70 × 10 ⁻¹⁶
≥ 600 MET-min/week	4,783 (54.6)	69.80 (15.60)	79.54 (31.55)	62.39 (20.7)	78.47 (19.73)	66.66 (12.66)
Functional limitations						
No (0)	6,909 (78.9)	71.74 (13.04), 1.06 × 10 ⁻³²	90.54 (18.90), 4.52 × 10 ⁻³⁵	62.37 (20.2), 8.71 × 10 ⁻¹⁷	75.16 (20.24), 3.79 × 10 ⁻¹¹	65.37 (12.98), 0.033
Yes (1)	1,847 (21.1)	49.12 (14.77)	34.82 (31.10)	53.74 (22.3)	66.78 (24.19)	63.58 (13.87)
Well-being						
High (0)	6,454 (73.7)	70.18 (14.89), 4.37 × 10 ⁻⁸	83.58 (28.50), 0.009	64.45 (19.6), 7.52 × 10 ⁻³⁶	74.65 (20.68), 2.89 × 10 ⁻⁹	65.64 (12.98), 1.09 × 10 ⁻⁵
Low (1)	2,302 (26.3)	58.10 (16.81)	65.86 (35.18)	49.59 (20.7)	69.91 (22.99)	63.16 (13.61)
Diabetes						
No (0)	8,643 (98.7)	67.10 (16.23), 2.43 × 10 ⁻⁷	78.96 (31.61), 0.002	60.59 (20.9), 0.173	73.52 (21.37), 1.17 × 10 ⁻⁴	65.11 (13.11), 7.04 × 10 ⁻⁷
Yes (1)	113 (1.29)	56.30 (17.34)	64.70 (33.99)	57.62 (21.0)	64.37 (22.48)	55.91 (15.98)
CVD						
No (0)	8,506 (97.10)	67.16 (16.13), 2.05 × 10 ⁻⁴	78.99 (31.49), 0.049	60.78 (20.9), 8.41 × 10 ⁻⁷	73.58 (21.27), 0.001	65.02 (13.18), 0.411
Yes (1)	250 (2.86)	60.35 (20.14)	71.58 (36.68)	52.87 (22.0)	67.31 (24.99)	64.16 (13.70)
Mental disorder						
No (0)	8,035 (91.80)	68.30 (15.42), 2.92 × 10 ⁻³⁴	80.33 (30.76), 1.91 × 10 ⁻¹⁸	62.42 (20.0), 5.37 × 10 ⁻¹²⁸	74.24 (20.89), 6.11 × 10 ⁻¹⁸	65.14 (13.06), 0.008
Yes (1)	721 (8.24)	52.07 (18.24)	61.51 (36.29)	39.79 (20.2)	64.03 (24.66)	63.34 (14.49)

Note. Mean IC scores and SDs at age 31 were averaged across imputations. *p*-values reflect univariate group comparisons pooled across imputations using Rubin's rules. Differences between binary groups were tested using Welch's independent t-tests; differences across variables with three or more categories were tested using linear models with omnibus Wald tests.

Abbreviations: IC = intrinsic capacity; SD = standard deviation; CVD = cardiovascular disease; WHO = World Health Organization; Matric. = matriculation examination; MET-min/week = metabolic equivalent of task minutes per week.

3.2.4 Predictive validity – Associations between intrinsic capacity and health outcomes

In fully adjusted logistic regression models (*Table 3* and *Figure 2*), each 10-point increase in composite IC at age 31 was associated with lower odds of functional limitations, low well-being, mental disorder, cardiovascular disease and type 2 diabetes at age 46 (ORs = 0.694–0.819; all $p < 0.001$). By domain, cognitive-sensory, locomotor, and psychological domains were each inversely associated with all outcomes (all $p < 0.05$), except psychological capacity with cardiovascular disease. The vitality domain predicted only type 2 diabetes and cardiovascular disease after covariate adjustment. Unadjusted and partially adjusted models can be found in *eTables 10A-10E*, Supplemental Materials.

Table 3. Fully adjusted associations between intrinsic capacity at age 31 and health outcomes at age 46.

Variable	Functional limitations		Low well-being		Mental disorder		Cardiovascular disease		Type 2 diabetes	
	OR [95% CI]	<i>p</i> -value	OR [95% CI]	<i>p</i> -value	OR [95% CI]	<i>p</i> -value	OR [95% CI]	<i>p</i> -value	OR [95% CI]	<i>p</i> -value
Composite IC	0.764 [0.718–0.813]	1.88 × 10 ⁻¹³	0.803 [0.763–0.845]	2.44 × 10 ⁻¹⁴	0.819 [0.770–0.871]	1.52 × 10 ⁻⁹	0.737 [0.645–0.843]	1.25 × 10 ⁻⁵	0.694 [0.631–0.764]	1.25 × 10 ⁻¹²
Cognitive-sensory	0.886 [0.862–0.910]	3.21 × 10 ⁻¹⁴	0.947 [0.922–0.973]	1.22 × 10 ⁻⁴	0.957 [0.925–0.991]	0.013	0.927 [0.863–0.996]	0.040	0.944 [0.896–0.993]	0.027
Psychological	0.884 [0.850–0.920]	8.08 × 10 ⁻⁹	0.783 [0.753–0.815]	2.70 × 10 ⁻²⁴	0.780 [0.745–0.817]	7.41 × 10 ⁻²²	0.926 [0.829–1.034]	0.170	0.908 [0.840–0.982]	0.016
Locomotor	0.903 [0.875–0.933]	1.36 × 10 ⁻⁹	0.922 [0.892–0.953]	2.31 × 10 ⁻⁶	0.947 [0.909–0.987]	0.010	0.816 [0.738–0.902]	8.19 × 10 ⁻⁵	0.811 [0.754–0.873]	3.93 × 10 ⁻⁸
Vitality	0.960 [0.904–1.020]	0.183	1.003 [0.945–1.064]	0.926	1.022 [0.944–1.107]	0.587	0.628 [0.515–0.767]	1.10 × 10 ⁻⁵	0.436 [0.377–0.504]	2.37 × 10 ⁻²¹

Note. Values are ORs with 95% CIs from Rubin-pooled logistic regression models estimating associations between IC scores at age 31 and health outcomes at age 46.

Models were adjusted for sex, socioeconomic indicators (education, occupation, gross household income), and health behaviours (smoking status, alcohol intake, and physical activity). ORs correspond to a 10-point increase in IC on a 0–100 percent-of-maximum-possible scale. Participants with the relevant condition at baseline (age 31) were excluded from analyses of incident disease and outcome-specific analytic samples were $n = 8,010$ – $8,054$ for mental disorder, $n = 8,479$ – $8,525$ for cardiovascular disease, and $n = 8,630$ – $8,652$ for type 2 diabetes.

Abbreviations: OR = odds ratio; CI = confidence interval; IC = intrinsic capacity.

3.3 Sensitivity analyses

3.3.1 Separate cognitive and sensory domains

The alternative five-domain composite showed very high agreement with the primary composite at age 31 (Pearson $r = 0.930$; Spearman $\rho = 0.977$, $p < 0.001$) and age 46 ($r = 0.966$; $\rho = 0.982$, $p < 0.001$). Temporal stability from age 31 to 46 was relatively low for the separate cognitive ($r = 0.216$) and sensory ($r = 0.209$) scores, similar to the combined cognitive-sensory score ($r = 0.241$). Temporal stability was slightly lower for the alternative five-domain composite ($r = 0.297$) than for the primary composite ($r = 0.365$), although it remained statistically significant. Spearman correlations yielded similar results (*eTable 11*, Supplemental Materials).

Cognitive and sensory capacity showed small mean reductions that did not reach statistical significance. In contrast, the alternative composite declined significantly by 4.66 points ($p < 0.001$), broadly consistent with the 5.70-point decline observed for the primary composite. This finding was corroborated by the Wilcoxon rank-based analysis (HL median change = -4.71 , Fisher-combined $p < 0.001$) (*eTable 12*, Supplemental Materials).

The alternative five-domain composite predicted all five outcomes in fully adjusted models (ORs per 10-point increase = 0.802 – 0.890 , all $p < 0.001$), paralleling the findings for the primary composite. However, the separate cognitive and sensory scores showed more selective associations than the combined cognitive–sensory domain, which predicted all five outcomes. Cognitive capacity predicted functional limitations, low well-being and type 2 diabetes, while sensory capacity predicted functional limitations and low well-being only (*eTable 13*, Supplemental Materials). Overall, these findings support the robustness of the primary composite results to retaining cognitive and sensory capacity as separate domains, while indicating more outcome-specific associations for the separate observed scores.

Neither the separate cognitive score (mean difference for higher versus lower education: -0.12 points, 95% CI -2.56 to 2.33 ; $p = 0.925$) nor the sensory score (mean difference: 1.34 points, 95% CI -0.73 to 3.41 ; $p = 0.202$) differed significantly by education at age 31. This indicated that the absence of an educational difference in the combined cognitive-sensory score was not attributable to opposing domain-specific patterns.

3.3.2 Sex-related sensitivity analyses

Sex-stratified analyses indicated different patterns of capacity change from age 31 to 46. Composite IC declined in both females ($\Delta = -3.71$, $p = 0.018$) and males ($\Delta = -7.78$, $p < 0.001$), with a 4.07-point greater decline in males. Declines in cognitive-sensory and locomotor capacity were also greater in males, whereas vitality declined more in females. Psychological capacity increased in females but remained stable in males. All formal sex differences in change were statistically significant (all $p < 0.001$), and Wilcoxon rank-based analyses corroborated the direction of cognitive-sensory and composite IC changes in both sexes (*eTable 14*, Supplemental Materials).

In predictive-validity analyses, sex-specific associations were broadly comparable. For composite IC, fully adjusted ORs per 10-point increase ranged from 0.692 to 0.809 in females and from 0.695 to 0.831 in males across the five outcomes. None of the 25 IC-score-by-sex interactions was statistically significant (interaction $p = 0.098$ – 0.958), providing no evidence that predictive associations differed by sex (*eTable 15*, Supplemental Materials).

3.4 Attrition analyses

Participants lost to the 46-year follow-up and those with incomplete IC data tended to be male, less healthy, engage in more health-risk behaviours and have lower socioeconomic status, compared to individuals that continued to participate until 46 years and to those with complete IC data. Full results from attrition analyses are provided in *eTables 16–18*, Supplemental Materials.

4. Discussion

4.1 Summary of key findings

We developed and validated domain and composite IC scores in a large population-based birth cohort (NFBC1966) at ages 31 and 46 years. IC scores demonstrated acceptable psychometric properties, with domain models showing adequate fit, meaningful factor loadings, and satisfactory reliability; lower fit for vitality reflected its multidimensional nature. Temporal stability was moderate for most IC scores but lower for cognitive–sensory capacity. Composite IC and cognitive–sensory, locomotor and vitality capacities declined from early adulthood to midlife, consistent with expected age-related changes ^{22,23,49}, although the decline in vitality capacity was not statistically significant. Psychological capacity increased significantly. IC scores also demonstrated known-groups validity across socioeconomic, behavioural, and health characteristics. Predictive validity was evidenced by associations between higher IC at age 31 and reduced odds of cardiovascular disease, type 2 diabetes, mental disorder, functional limitations, and low well-being at age 46.

4.2 Intrinsic capacity levels and change from early adulthood to midlife

We observed moderate-to-high average IC levels at age 31 (67 on the 0–100 POMP scale), consistent with evidence that most capacities are near their peak in early adulthood ^{22,23}. From this stage onwards, composite IC declined modestly but significantly over 15 years, extending prior evidence of capacity decline in older populations across shorter follow-up periods (one-to-six years) ⁵⁰. The extent of capacity decline varied across domains: locomotor capacity showed the largest significant decline, followed by a smaller significant decline in cognitive–sensory capacity. In contrast, the separate cognitive and sensory scores showed only small, non-significant reductions, and the decline in vitality capacity was also not statistically significant. In older adults, declines in cognitive capacity are most frequently reported alongside psychological, sensory and/or locomotor impairment ⁵⁰. Together, this

pattern suggests that substantial locomotor losses - often described as triggers of a disabling cascade⁵¹ - may emerge earlier, while cognitive decline increasingly co-occurs with other capacity losses later in life. Our findings therefore point to locomotor capacity as an important early target for interventions aimed at preserving capacity across domains. The rationale for targeting locomotor capacity earlier in adulthood would have been strengthened by validation against additional clinically relevant locomotor outcomes, such as falls, frailty and sarcopenia. Nevertheless, locomotor capacity predicted later functional limitations as well as mental and cardiometabolic outcomes in the present study, supporting its potential relevance as an intervention target for maintaining functional ability, well-being and physical health during ageing.

Psychological capacity was the only domain to show a significant increase from early adulthood into midlife, consistent with evidence of selective gains in psychological IC⁵². Such improvements in well-being are well documented across adulthood and may reflect enhanced emotion regulation, coping skills and greater focus on positive experiences⁵³.

Notably, the divergent trajectory of psychological capacity would have been obscured by a composite-only view, highlighting the importance of domain-specific analyses.

Sex-related sensitivity analyses further showed that changes in IC differed between females and males. Declines in composite, cognitive-sensory and locomotor capacity were greater in males, whereas vitality declined more in females and the increase in psychological capacity was confined to females. These patterns should be interpreted in the context of sex differences already present at age 31. For example, males had higher baseline locomotor and psychological capacity, whereas females had higher baseline vitality. The magnitude of subsequent change may therefore partly reflect differences in starting levels and the scope for further improvement or decline.

4.3 Known-groups validity – Socioeconomic and health gradients in intrinsic capacity

Despite moderate-to-high average IC at age 31, between-person variability was substantial and persisted into midlife (coefficients of variation $\approx 21\text{--}40\%$). This variability can partly be explained by sociodemographic and health-related characteristics. Replicating earlier work^{4,38}, we found higher IC in participants with higher income, education and occupational status, healthier behaviours, and who were free of chronic conditions, compared to their more disadvantaged or ill counterparts. Interestingly, composite IC did not differ by sex as commonly reported in prior research^{49,54}. However, domain-specific differences were evident: women had higher cognitive-sensory and vitality capacity, whereas men showed higher locomotor and psychological capacity. Similar domain-level patterns have been reported in older adults, although higher vitality capacity is typically found in older men, not women^{52,54}. The discrepancy may reflect age-dependent patterns in vitality capacity: young women may show higher vitality in young adulthood, with steeper declines in later life due to menopause-related hormonal and cardiometabolic changes⁵⁵. Consistent with this possibility, our sex-specific sensitivity analyses showed that vitality capacity declined significantly in females between ages 31 and 46, whereas it changed little in males and the change was not statistically significant. Further research should directly investigate the contribution of reproductive, hormonal, cardiometabolic and psychosocial factors to sex differences in vitality changes across midlife.

The absence of a significant educational difference in cognitive-sensory capacity at age 31 was somewhat unexpected, particularly given established associations between education and cognitive functioning. Combining cognition with sensory capacity might have attenuated a cognition-specific educational gradient; however, sensitivity analyses showed that neither the separate cognitive nor sensory score differed significantly by education, arguing against

cancellation of opposing domain-specific associations. The finding may instead reflect the limited granularity of the single self-reported cognitive item and two self-reported sensory items, together with restricted variability and ceiling effects in this relatively young cohort. More differentiated objective assessments may be more sensitive to education-related variation in cognitive and sensory functioning and should be included in future research.

4.4 Predictive validity of intrinsic capacity – Associations with health

outcomes

Higher IC at age 31 was associated with lower odds of incident cardiovascular disease, type 2 diabetes and mental disorder at age 46, extending emerging evidence from older adults on reduced physical and mental disease risk^{6–8} to much earlier life stages. Higher IC was also associated with fewer functional limitations and higher well-being, corroborating the WHO framework of Healthy Ageing and prior research in older adults^{4,5}. Although the effect sizes were modest, each 10-point increase in composite IC corresponded to 18%–31% lower odds across the five outcomes, and the associations persisted after adjustment for socioeconomic and behavioural factors. The ongoing NFBC1966 follow-up will enable future analyses of whether IC measured at age 31 predicts health and functional outcomes over a substantially longer period and whether IC assessed at age 46 provides additional or stronger prognostic information. Although locomotor capacity predicted functional limitations, future validation against other locomotor-related outcomes such as falls, frailty or sarcopenia would be needed for clarifying the clinical consequences of the marked locomotor decline observed from early adulthood to midlife.

Domain-specific analyses added nuanced predictive information. Vitality capacity predicted physical outcomes only, while the other domains predicted both physical and mental/functional outcomes. These patterns are conceptually coherent: vitality capacity is a physiological construct chiefly tied to physical disease processes³⁵, while the other domains

may shape both physical and mental health via direct effects and indirect lifestyle/psychosocial pathways ⁵⁶. Nevertheless, the predictive relevance of vitality may broaden in older adulthood as physiological decline becomes more pronounced and increasingly affects functioning and mental well-being.

Taken together, these findings support composite IC as a summary indicator of overall capacity, while highlighting the added value of domain scores for outcome-specific risk stratification and tailored prevention. No IC-by-sex interaction was statistically significant in the predictive models, providing no evidence that the associations between IC and subsequent health outcomes differed between females and males.

4.5 Measurement and operationalisation of intrinsic capacity

Most IC studies have not explicitly aligned score construction with IC's measurement model, despite evidence supporting a formative composite and reflective domains ^{27,28,57}. To address this, we implemented a dual-level approach that treats composite IC as a formative aggregate of distinct domains, while using reflective factor analysis to derive domain scores.

This dual-level approach offers several advantages. At the domain level, factor analysis accounts for differential indicator weighting based on factor loadings and enables psychometric evaluation of domain structure to inform variable selection and model refinement. It also allows formal testing of temporal measurement invariance, a prerequisite for valid comparison of change across time ⁵⁸. At the composite level, aggregating domain scores avoids imposing a single reflective factor across all heterogeneous domains, a practice that conflicts with the formative nature of composite IC but is common in prior research ^{22,23}.

The limited cognitive and sensory assessments required a pragmatic operational decision. Combining cognition, vision and hearing permitted a common factor modelling approach and produced a more differentiated score than the separate observed measures. Nevertheless, it departs from the WHO five-domain structure. Sensitivity analyses retaining cognition and

sensory capacity as separate domains yielded an alternative composite that was highly correlated with the primary composite, declined longitudinally and predicted all five outcomes, indicating that the principal composite findings were robust to the alternative domain definition. However, unlike the combined cognitive–sensory score, which declined significantly and predicted all five outcomes, the separate cognitive and sensory scores showed small non-significant reductions and more selective predictive associations. Both predicted functional limitations and low well-being, while cognitive capacity additionally predicted type 2 diabetes. These differences may reflect genuine domain-specific associations but may also arise from the limited granularity and marked ceiling effects of the one-item cognitive and two-item sensory scores. Overall, the findings support the combined cognitive–sensory score as a pragmatic primary operationalisation while highlighting the need for more comprehensive domain-specific assessment.

To the best of our knowledge, only two studies have operationalised IC using a dual-formative-reflective measurement framework^{27,28}. Qi and colleagues²⁸ implemented a measurement approach closely resembling ours with the main difference being that they added partial least squares structural equation modelling (PLS-SEM) to confirm domain indicators. Their scores showed known-groups gradients and predictive validity for subsequent functional limitations, paralleling our findings. However, unlike our approach, their study did not report formal longitudinal measurement-invariance testing, leaving uncertainty about whether observed longitudinal changes reflect true changes or measurement artefacts. Nevertheless, PLS-SEM-based approaches offer valuable opportunities for refining domain weights and higher-order composite IC structures.

Our domain-specific analyses were particularly informative for vitality, the least well-defined IC domain⁵⁹. Factor analysis confirmed a multidimensional structure at both ages, encompassing metabolic, neuromuscular, and immune-related indicators as proposed by the

WHO working group on vitality capacity³⁵. However, indicators did not cluster cleanly into proposed subcomponents, and factor structure shifted between ages 31 and 46. For instance, metabolic variables were strongly correlated with neuromuscular and immune/stress indicators and loaded onto different factors at both ages. These findings suggest that vitality comprises distinct subdomains that only partly align with current conceptualisations and that appear to change over time, underscoring the need for further refinement of vitality measurement across the life course. Because the present study included only two adult timepoints, further longitudinal research across younger, middle-aged and older adulthood is needed to determine how vitality measurement changes across the life course. Such studies should formally assess measurement invariance to distinguish true changes in capacity from changes in domain composition or indicator weighting.

4.6 Strengths and limitations

The study has several strengths. It addresses a major gap in longitudinal evidence on IC in young adulthood and midlife. IC was assessed comprehensively across all five domains using a multimodal battery of performance tests, clinical biomarkers, and validated questionnaires aligned with WHO guidance. This design improved measurement precision and minimised bias, relative to the predominantly self-report-based and retrospective studies in older populations¹⁹. Bias was further mitigated through multiple imputation, addressing limitations of complete-case approaches typically used in prior IC research¹⁹. Another strength is the dual-level scoring approach, explicitly aligned with IC's measurement model and conceptual definition and supported by our extensive validation analyses.

Several limitations should be acknowledged. First, cognition and sensory capacity were combined into a single latent cognitive-sensory domain due to the limited number of available variables. While this approach was supported by moderate-to-strong inter-item correlations and shared neurophysiological underpinnings³⁴, it departs from the WHO

framework, which treats these domains separately. Nevertheless, all IC domains were assessed, whereas many studies omit at least one domain, most commonly sensory capacity¹⁹. Moreover, sensitivity analyses using an alternative composite that retained cognitive and sensory capacity as separate domains supported the robustness of the primary findings.

Second, full temporal measurement invariance was not achieved in all domains. Although some authors consider comparison of factor means acceptable under partial invariance⁵⁸, we opted for the conservative approach of enforcing measurement invariance by constraining equal factor loadings across time. This approach helped ensure that observed longitudinal changes in IC reflected true changes in capacity rather than changes in indicator weighting across time.

Third, vitality capacity exhibited a multidimensional structure but was modelled as a single factor to maintain cross-domain and longitudinal comparability, likely minimising measurement precision. Future work should test whether modelling vitality as a multidimensional construct improves validity and predictive performance.

Fourth, changes in treatment, health behaviours and health status during follow-up may have influenced the observed predictive associations. Future longitudinal analyses should examine whether these factors mediate or modify the relationship between early-adulthood IC and later outcomes.

Fifth, generalisability may be limited by the cohort's ethnically homogeneous, high-resource Northern Finnish setting and by the historical timing of data collection. Participants were born in 1966 and assessed at ages 31 and 46 in 1997–1998 and 2012–2014, respectively. Secular changes in education, employment, living conditions, healthcare, health behaviours and cardiometabolic risk profiles may therefore limit direct generalisation of the observed IC levels and associations to contemporary younger adults. Replication in more recent birth cohorts and in populations with greater ethnic, socioeconomic and geographical diversity is

warranted. Nevertheless, the population-based design and the limited existing evidence on IC in Nordic populations make the study an important contribution of regional and life-course evidence.

Sixth, selective participation and loss to follow-up may have introduced selection bias. At inception, the NFBC1966 included approximately 96% of births in the source population, reflecting very high population coverage and limiting selection bias at initial recruitment^{32,33}.

However, as is common in long-running longitudinal cohort studies, participation declined over subsequent follow-ups. The NFBC1966 cohort profile shows that participation in the adult follow-ups was socially patterned, with higher participation among women and individuals with higher socioeconomic position and education³³. Our own attrition analyses similarly showed that participants lost to the 46-year follow-up or with incomplete IC data tended to be male and to have less favourable socioeconomic, behavioural and health characteristics than those retained in the study (*eTables 16–18*, Supplemental Materials).

Multiple imputation was used to reduce bias arising from incomplete data but cannot fully account for selective non-participation, attrition or data that are missing not at random.

Residual selection bias therefore cannot be excluded and may have resulted in some underrepresentation of individuals with poorer health and greater socioeconomic disadvantage.

Seventh, sex- and gender-related considerations warrant caution. Our primary aim was to derive a single set of domain and composite IC scores on a common metric for use across the cohort. We therefore did not estimate separate sex-specific scores. Instead, we accounted for sex differences during score construction by POMP-standardising relevant indicators separately for females and males. Moreover, we examined sex differences in the key validation analyses by comparing baseline IC levels and longitudinal changes between females and males and by testing whether sex modified the predictive associations between

IC and subsequent health outcomes. It remains possible that the measurement properties or domain structure of IC differ by sex. Future studies specifically designed to address this question should formally assess sex-based measurement invariance and, where data permit, gender-related heterogeneity to strengthen inference and support sex- and gender-sensitive applications of IC.

Finally, individual-level mortality data were not available within the approved analytical dataset, preventing evaluation of all-cause mortality as an additional indicator of the clinical relevance of IC. Given the relatively young age of the cohort and the low proportion of deaths between ages 31 and 46 (2.2%), evaluation of mortality would have been constrained at this stage. Longer follow-up and accumulation of additional events will enable more robust assessment of associations between adult IC and mortality.

5. Conclusions

We developed and validated domain-specific and composite IC scores in young and middle-aged adults using a formative-reflective approach that is aligned with IC's conceptual definition and measurement model. IC was already informative in early adulthood: composite IC declined significantly from the early 30s and predicted major physical, mental, and functional outcomes 15 years later. These findings extend IC research beyond geriatric populations, positioning IC as a robust marker of physical and mental reserve with prognostic value well before old age. Assessing IC in young adulthood could strengthen early surveillance, risk stratification, and primary prevention. Socioeconomic, behavioural and health gradients highlighted early inequalities in capacity and identified vulnerable groups that may benefit most from capacity-building interventions.

Changes in capacity varied across domains: cognitive-sensory and locomotor capacities declined significantly, vitality capacity showed a non-significant reduction, and psychological capacity increased significantly. This pattern suggests opportunities for early prevention of decline as well as enhancement of capacities amenable to improvement. Moreover, all capacities predicted both physical and mental/functional outcomes except vitality capacity, which was selectively associated with physical disease, supporting domain- and outcome-specific intervention strategies. Future research should refine domain measurement, particularly vitality, replicate our formative-reflective approach across settings, and identify modifiable life-course determinants of divergent IC profiles. Longer-term follow-up studies should compare the prognostic value of IC assessed at multiple stages of adulthood and test whether improving IC earlier in life translates into healthier ageing profiles.

Declarations

Ethics approval and consent to participate

All participants provided informed written consent before study entry, and ethical approval for the NFBC1966 was received from the University of Oulu Ethical Committee and the Ethical Committee of the Northern Ostrobothnia Hospital District 94/2011 (12.12.2011). The study was conducted in accordance with the Declaration of Helsinki.

Data statement

NFBC1966 data are available from the University of Oulu, Infrastructure for Population Studies. Permission to use the data can be applied for research purposes via an electronic material request portal. In the use of data, we follow the European Union General Data Protection Regulation (679/2016) and the Finnish Data Protection Act. The use of personal data is based on a cohort participant's written informed consent in their latest follow-up study, which may cause limitations to its use. Please contact the NFBC project centre (NFBCprojectcenter@oulu.fi) and visit the cohort website (www.oulu.fi/nfbc) for more information.

Conflict of interests

The authors declare that they have no conflict of interests.

Author contributions

LH: Conceptualisation; data curation; formal analysis; investigation; methodology; project administration; software; validation; visualisation; writing (original draft), writing (review & editing).

GH: Conceptualisation; investigation; methodology; project administration; supervision; validation; writing (review & editing).

RN: Conceptualisation; investigation; methodology; project administration; supervision; validation; writing (review & editing).

AL: Conceptualisation; investigation; methodology; project administration; software; supervision; validation; writing (review & editing).

MRJ: Conceptualisation; funding acquisition; investigation; methodology; project administration; supervision; validation; writing (review & editing).

All authors read and approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. LH was supported by a doctoral studentship from the London Interdisciplinary Social Science Doctoral Training Partnership funded by the Economic and Social Research Council (ESRC) at Imperial College London. MRJ was supported by the British Academy (grant number RaR/100571). MRJ and RN were supported by the European Union's Horizon 2020 research and innovation programme (LongITools) (grant number 874739). AL, MRJ, and RN were supported by the European Union's Horizon 2020 research and innovation programme (DynaHEALTH) (grant number 633595). RN was supported by the European Union's Horizon 2020 research and innovation programme (OBCT) (grant number 101080250).

The Northern Finland Birth Cohort 1966 (NFBC1966) 31-year follow-up study was supported by the University of Oulu (grant number 65354); Oulu University Hospital (grant numbers 2/97, 8/97); the Ministry of Health and Social Affairs (grant numbers 23/251/97, 160/97, 190/97); the National Institute for Health and Welfare, Helsinki (grant number 54121); and the Regional Institute of Occupational Health, Oulu, Finland (grant numbers 50621, 54231). The NFBC1966 46-year follow-up study was supported by the University of Oulu (grant number 24000692); Oulu University Hospital (grant number 24301140); and the

European Regional Development Fund (grant number 539/2010 A31592). The funding bodies had no role in the study design; data collection, analysis, or interpretation; manuscript writing; or the decision to submit the manuscript for publication.

Acknowledgements

We wish to thank all cohort members, researchers and NFBC project centre personnel who participated in the NFBC data collections. The developers and funders of NFBC1966 bear no responsibility for the analyses or interpretations presented in this article. We extend our deepest gratitude to Jotheeswaran Amuthavalli Thiyagarajan (Epidemiologist, WHO Ageing and Health Unit) for his expert guidance on the conceptualisation and operationalisation of IC.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (version 5.2; OpenAI) to support identification of relevant studies, assist with code debugging, improve clarity and grammar in the initial manuscript draft, and help format the manuscript in accordance with journal submission guidelines. The AI tool was not used for carrying out substantive research tasks, including content generation, data collection, data analysis, interpretation of results, or the formulation of conclusions. All analyses, interpretations, and final decisions regarding the content of the manuscript were performed by the authors, who take full responsibility for the accuracy and integrity of the work.

References

1. World Health Organization. World report on ageing and health. 2015. Accessed January 6, 2026. <https://www.who.int/publications/i/item/9789241565042>
2. Atella V, Piano Mortari A, Kopinska J, et al. Trends in age-related disease burden and healthcare utilization. *Aging Cell*. 2019;18(1):e12861. doi:10.1111/accel.12861
3. 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(12):1653-1660. doi:10.1093/gerona/gly011
4. Yang Y, Ma G, Wei S, et al. Adverse outcomes of intrinsic capacity in older adults: A scoping review. *Arch Gerontol Geriatr*. 2024;120:105335. doi:10.1016/j.archger.2024.105335
5. Sánchez-Sánchez JL, Lu WH, Gallardo-Gómez D, et al. Association of intrinsic capacity with functional decline and mortality in older adults: a systematic review and meta-analysis of longitudinal studies. *Lancet Healthy Longev*. 2024;5(7):e480-e492. doi:10.1016/S2666-7568(24)00092-8
6. Ramírez-Vélez R, Iriarte-Fernández M, Santafé G, et al. Association of intrinsic capacity with incidence and mortality of cardiovascular disease: Prospective study in UK Biobank. *J Cachexia Sarcopenia Muscle*. 2023;14(5):2054-2063. doi:10.1002/jcsm.13283
7. Gao WX, Liu J, Wang J, et al. Association of intrinsic capacity with incident type 2 diabetes mellitus in older Chinese: Guangzhou Biobank Cohort Study. *Arch Gerontol Geriatr*. 2025;129:105687. doi:10.1016/j.archger.2024.105687
8. Ramírez-Vélez R, Borda MG, Sáez de Asteasu ML, Izquierdo M. Intrinsic capacity as a predictor of depression onset in middle-aged and older adults: Insights from the UK Biobank. *J Affect Disord*. 2025;388:119590. doi:10.1016/j.jad.2025.119590
9. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8:24. doi:10.1186/1471-2318-8-24
10. Cesari M, Canevelli M, Amuthavalli Thiyagarajan J, et al. Frailty and intrinsic capacity: integrating complementary concepts to promote healthy ageing and transformation of care. *Age Ageing*. 2026;55(5):afag138. doi:10.1093/ageing/afag138
11. Qi Y, van Schoor NM, Schaap LA, Hoogendijk EO. Disentangling the overlap between frailty and intrinsic capacity in older adults. *Exp Gerontol*. 2026;213:113006. doi:10.1016/j.exger.2025.113006
12. Belloni G, Cesari M. Frailty and intrinsic capacity: two distinct but related constructs. *Front Med (Lausanne)*. 2019;6:133. doi:10.3389/fmed.2019.00133
13. Numbers K, Samtani S, Hanewald K, Cheung J, Sachdev PS, Brodaty H. Intrinsic capacity predictors of dementia and mortality in the Sydney Memory and Ageing Study. *Int J Geriatr Psychiatry*. 2025;40(10):e70156. doi:10.1002/gps.70156
14. Yu R, Leung J, Leung G, Woo J. Towards healthy ageing: using the concept of intrinsic capacity in frailty prevention. *J Nutr Health Aging*. 2022;26(1):30-36. doi:10.1007/s12603-021-1715-2

15. Liu S, Kang L, Liu XH, et al. Trajectory and correlation of intrinsic capacity and frailty in a Beijing elderly community. *Front Med (Lausanne)*. 2021;8:751586. doi:10.3389/fmed.2021.751586
16. Lu W, Pikhart H, Sacker A. Domains and measurements of healthy aging in epidemiological studies: a review. *Gerontologist*. 2019;59(4):e294-e310. doi:10.1093/geront/gny029
17. Silva N, Rajado AT, Esteves F, et al. Measuring healthy ageing: current and future tools. *Biogerontology*. 2023;24(6):845-866. doi:10.1007/s10522-023-10041-2
18. Piriú AA, Bufali MV, Cappellaro G, Compagni A, Torbica A. Conceptualisation and measurement of healthy ageing: insights from a systematic literature review. *Soc Sci Med*. 2025;374:117933. doi:10.1016/j.socscimed.2025.117933
19. Liang Y, Shang S, Gao Y, et al. Measurements of intrinsic capacity in older adults: A scoping review and quality assessment. *J Am Med Dir Assoc*. 2023;24(3):267-276.e2. doi:10.1016/j.jamda.2022.09.011
20. Zhou Y, Wang G, Li J, et al. Trajectory of intrinsic capacity among community-dwelling older adults in China: The China Health and Retirement Longitudinal Study. *Arch Gerontol Geriatr*. 2024;124:105452. doi:10.1016/j.archger.2024.105452
21. Salinas-Rodríguez A, Fernández-Niño JA, Rivera-Almaraz A, Manrique-Espinoza B. Intrinsic capacity trajectories and socioeconomic inequalities in health: the contributions of wealth, education, gender, and ethnicity. *Int J Equity Health*. 2024;23:48. doi:10.1186/s12939-024-02136-0
22. Beyene MB, Visvanathan R, Ahmed M, Benyamin B, Beard JR, Amare AT. Development and validation of an intrinsic capacity score in the UK Biobank study. *Maturitas*. 2024;185:107976. doi:10.1016/j.maturitas.2024.107976
23. 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(11):e026119. doi:10.1136/bmjopen-2018-026119
24. Koivunen K, Schaap LA, Hoogendijk EO, Schoonmade LJ, Huisman M, van Schoor NM. Exploring the conceptual framework and measurement model of intrinsic capacity defined by the World Health Organization: A scoping review. *Ageing Res Rev*. 2022;80:101685. doi:10.1016/j.arr.2022.101685
25. Beard JR, Chen M. Intrinsic Capacity as a composite outcome measure: Opportunities and challenges. *J Nutr Health Aging*. 2023;27(6):398-400. doi:10.1007/s12603-023-1923-z
26. Koivunen K, Hoogendijk EO, Schaap LA, Huisman M, Heymans MW, van Schoor NM. Development and validation of an intrinsic capacity composite score in the Longitudinal Aging Study Amsterdam: a formative approach. *Aging Clin Exp Res*. 2023;35(4):815-825. doi:10.1007/s40520-023-02366-2
27. Rodríguez-Laso Á, García-García FJ, Rodríguez-Mañas L. The ICOPE intrinsic capacity screening tool: Measurement structure and predictive validity of dependence and hospitalization. *J Nutr Health Aging*. 2023;27(10):808-816. doi:10.1007/s12603-023-1985-y

28. Qi Y, Schaap LA, Schalet BD, et al. The development of intrinsic capacity measures for longitudinal research: The Longitudinal Aging Study Amsterdam. *Exp Gerontol*. 2024;197:112599. doi:10.1016/j.exger.2024.112599
29. Locquet M, Sanchez-Rodriguez D, Bruyère O, et al. Intrinsic capacity defined using four domains and mortality risk: A 5-year follow-up of the SarcoPhAge cohort. *J Nutr Health Aging*. 2022;26(1):23-29. doi:10.1007/s12603-021-1702-7
30. Gutiérrez-Robledo LM, García-Chanes RE, Pérez-Zepeda MU. Allostatic load as a biological substrate to intrinsic capacity: A secondary analysis of CRELES. *J Nutr Health Aging*. 2019;23(9):788-795. doi:10.1007/s12603-019-1251-5
31. Hoogendijk EO, Dent E, Koivunen K. Intrinsic capacity: an under-researched concept in geriatrics. *Age Ageing*. 2023;52(10):afad183. doi:10.1093/ageing/afad183
32. University of Oulu. Northern Finland Birth Cohort 1966. 2020. Accessed January 6, 2026. <http://urn.fi/urn:nbn:fi:att:bc1e5408-980e-4a62-b899-43bec3755243>
33. Nordström T, Miettunen J, Auvinen J, et al. Cohort Profile: 46 years of follow-up of the Northern Finland Birth Cohort 1966 (NFBC1966). *Int J Epidemiol*. 2021;50(6):1786-1787. doi:10.1093/ije/dyab109
34. Baltes PB, Lindenberger U. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychol Aging*. 1997;12(1):12-21. doi:10.1037/0882-7974.12.1.12
35. Bautmans I, Knoop V, Amuthavalli Thiagarajan J, et al. WHO working definition of vitality capacity for healthy longevity monitoring. *Lancet Healthy Longev*. 2022;3(11):e789-e796. doi:10.1016/s2666-7568(22)00200-8
36. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22. doi:10.1186/1471-2288-10-22
37. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Front Public Health*. 2018;6:149. doi:10.3389/fpubh.2018.00149
38. Huang ZT, Lai ETC, Luo Y, Woo J. Social determinants of intrinsic capacity: A systematic review of observational studies. *Ageing Res Rev*. 2024;95:102239. doi:10.1016/j.arr.2024.102239
39. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2024. Accessed January 13, 2026. <https://www.R-project.org/>
40. Cohen P, Cohen J, Aiken LS, West SG. The problem of units and the circumstance for POMP. *Multivariate Behav Res*. 1999;34(3):315-346. doi:10.1207/s15327906mbr3403_2
41. Stolz E, Mayerl H, Freidl W, Roller-Wirnsberger R, Gill TM. Intrinsic capacity predicts negative health outcomes in older adults. *J Gerontol A Biol Sci Med Sci*. 2022;77(1):101-105. doi:10.1093/gerona/glab279
42. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67. doi:10.18637/jss.v045.i03
43. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi:10.1002/sim.4067

44. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: A tutorial on multiple imputation. *Can J Cardiol*. 2021;37(9):1322-1331. doi:10.1016/j.cjca.2020.11.010
45. Li CH. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods*. 2016;48(3):936-949. doi:10.3758/s13428-015-0619-7
46. Watkins MW. Exploratory factor analysis: a guide to best practice. *J Black Psychol*. 2018;44(3):219-246. doi:10.1177/0095798418771807
47. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th ed. Oxford, UK: Oxford University Press; 2015. doi:10.1093/med/9780199685219.001.0001
48. Widaman KF, Ferrer E, Conger RD. Factorial invariance within longitudinal structural equation models: measuring the same construct across time. *Child Dev Perspect*. 2010;4(1):10-18. doi:10.1111/j.1750-8606.2009.00110.x
49. Lu WH, Rolland Y, Guyonnet S, de Souto Barreto P, Vellas B. Reference centiles for intrinsic capacity throughout adulthood and their association with clinical outcomes: a cross-sectional analysis from the INSPIRE-T cohort. *Nat Aging*. 2023;3(12):1521-1528. doi:10.1038/s43587-023-00522-x
50. Wang Y, Tang N, Shao M, Song J, Su Q, Gao Y. Transitions and trajectories in intrinsic capacity states over time: a systematic review. *BMC Geriatr*. 2025;25:407. doi:10.1186/s12877-025-05938-1
51. Gonzalez-Bautista E, Llibre-Guerra JJ, Sosa AL, et al. Exploring the natural history of intrinsic capacity impairments: longitudinal patterns in the 10/66 study. *Age Ageing*. 2023;52:afad137. doi:10.1093/ageing/afad137
52. Bernal MC, Batista E, Martínez-Ballesté A, Solanas A. A functional approach to model intrinsic capacity in ageing trajectories. *Sci Rep*. 2025;15:7878. doi:10.1038/s41598-025-92271-7
53. Carstensen LL, Turan B, Scheibe S, et al. Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychol Aging*. 2011;26(1):21-33. doi:10.1037/a0021285
54. Chen M, Hanewald K, Si Y, Gu Y, Beard JR. Intrinsic capacity across 15 countries in the Survey of Health, Aging, and Retirement in Europe. *JAMA Netw Open*. 2025;8(5):e259792. doi:10.1001/jamanetworkopen.2025.9792
55. Nappi RE, Chedraui P, Lambrinoudaki I, Simoncini T. Menopause: a cardiometabolic transition. *Lancet Diabetes Endocrinol*. 2022;10(6):442-456. doi:10.1016/S2213-8587(22)00076-6
56. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. *Lancet*. 2015;385(9968):640-648. doi:10.1016/S0140-6736(13)61489-0
57. Gonzalez-Bautista E, Beard JR. The challenge of measuring intrinsic capacity. *J Nutr Health Aging*. 2023;27:806-807. doi:10.1007/s12603-023-2012-z
58. Yang F, Su Q, Ran Q, Hu Y. Longitudinal change of intrinsic capacity and associated factors in older Chinese adults: China Health and Retirement Longitudinal Study. *J Nutr Health Aging*. 2024;28(5):100214. doi:10.1016/j.jnha.2024.100214

59. Chew J, Lee J, Hernandez HHC, Munro YL, Lim CL, Lim WS. The vitality domain of intrinsic capacity: A scoping review of conceptual frameworks and measurements. *J Frailty Aging*. 2025;14(4):100058. doi:10.1016/j.tjfa.2025.100058

Journal Pre-proof

Figure Captions

Figure 1. Sample flowchart of the Northern Finland Birth Cohort 1966 at ages 31 and 46.

Note. Participant selection in the NFBC1966 at the 31-year (1997–1998) and 46-year (2012–2014) follow-ups, showing the derivation of the imputation base, full analytical sample and analysis-specific samples. The full analytical sample was used for IC score construction and most validation analyses. The analysis-specific subsamples were used only for predictive-validity analyses of incident mental disorder, cardiovascular disease and type 2 diabetes, for which participants with the respective condition already present at age 31 were excluded. Counts for target populations, questionnaire returns and clinical assessment attendance were obtained from the NFBC1966 reference paper ³³.

Abbreviations: NFBC1966 = Northern Finland Birth Cohort 1966; IC = intrinsic capacity.

Alt text: Flowchart showing NFBC1966 participation at ages 31 and 46 and derivation of the imputation base, full analytical sample and outcome-specific analytical samples.

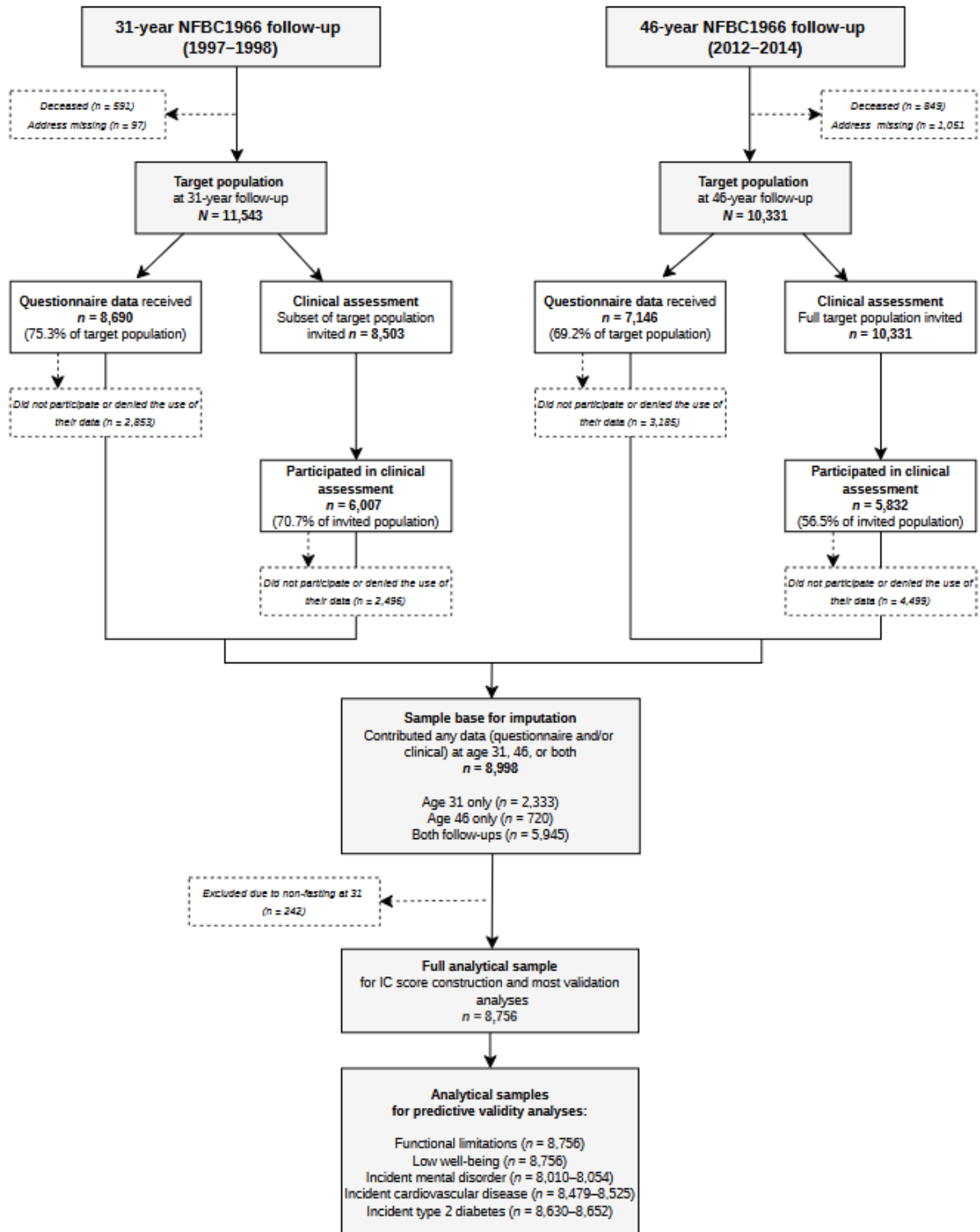
Figure 2. Forest plots for fully adjusted associations between intrinsic capacity at age 31 and health outcomes at age 46.

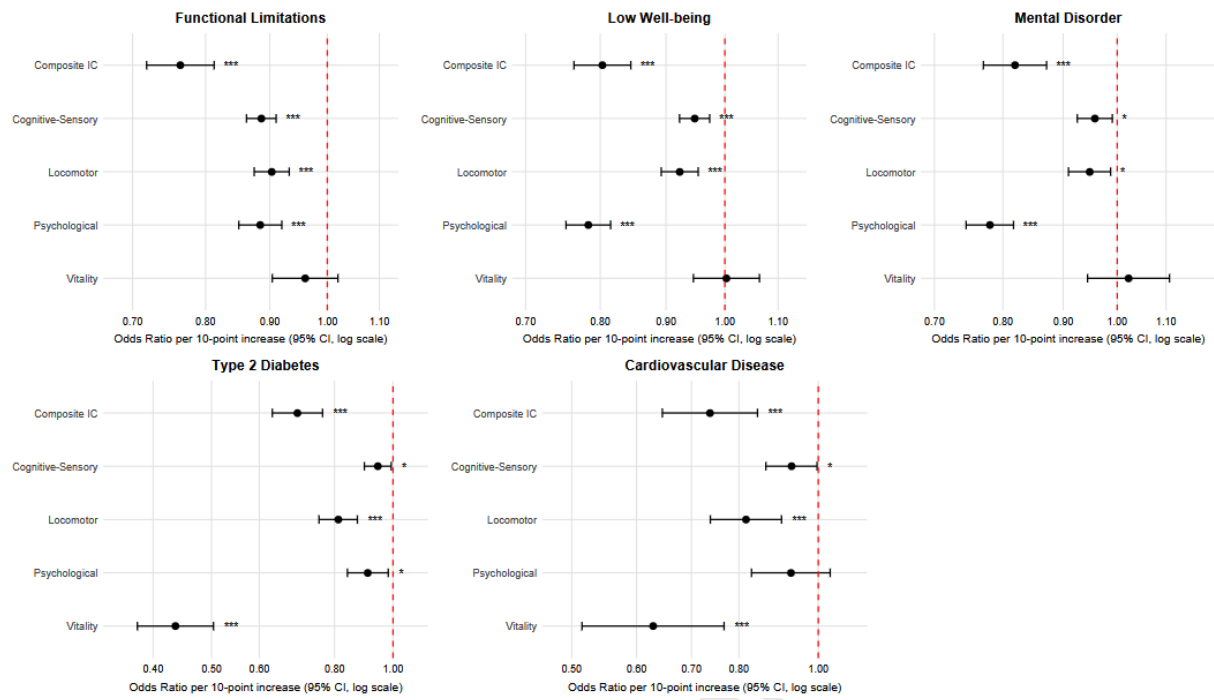
Note. Forest plots show ORs and 95% CIs from Rubin-pooled logistic regression models for IC scores at age 31 in relation to five health outcomes at age 46. Models were adjusted for sex, socioeconomic indicators (education, occupation, and household income), and health behaviours (smoking status, alcohol intake, and physical activity). Points show ORs per 10-point increase in IC on a 0–100 percent-of-maximum-possible scale; horizontal bars present 95% CIs; the dashed vertical line denotes OR = 1 (no association). Participants with the relevant condition at baseline (age 31) were excluded from analyses of incident disease outcomes and outcome-specific analytic samples were $n = 8,010$ – $8,054$ for mental disorder, $n = 8,479$ – $8,525$ for cardiovascular disease, and $n = 8,630$ – $8,652$ for type 2 diabetes.

Abbreviations: IC = intrinsic capacity; CI = confidence interval; ORs = odds ratio.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Alt text: Forest plots of adjusted associations between intrinsic capacity at age 31 and five functional, mental and cardiometabolic outcomes at age 46.





Highlights

- Intrinsic capacity scores built using reflective domains and formative composites.
- Scores validated longitudinally at ages 31 and 46 in a population birth cohort.
- Composite capacity declined from the early 30s to midlife; locomotor declined most.
- Higher composite capacity predicted better physical, mental and functional health.
- Lower capacity was linked to disadvantage, unhealthy lifestyle and ill health.