MUNICH RESEARCH INSTITUTE FOR THE
ECONOMICS OF AGING AND SHARE
ANALYSES (MEA-SHARE GGMBH)

DISCUSSION PAPER

RELEASE NOTE TO PRELIMINARY SHARE-HCAP CLASSIFICATION OF MILD AND SEVERE COGNITIVE IMPAIRMENT IN FIVE EUROPEAN COUNTRIES

AXEL BÖRSCH-SUPAN, SALIMA DOUHOU, IRENE FERNÁNDEZ &
MARCELA C. OTERO

04-2024

Munich Research Institute for the Economics of Aging and SHARE Analyses (MEA-SHARE gGmbH)

Leopoldstrasse 139
80804 Munich, Germany
Phone +49 89 4111205-151

WWW.MEA-SHARE.EU

Release note to preliminary SHARE-HCAP classification of mild and severe cognitive impairment in five European countries

Version 0.1

Axel Börsch-Supan, Salima Douhou, Irene Fernández, Marcela C. Otero

17 May 2024

1. Introduction

This note describes the release of a first and preliminary measurement of the cognitive status of selected SHARE respondents in five European countries. It is part of the Survey of Health, Aging, and Retirement in Europe's Harmonized Cognitive Assessment Protocol (SHARE-HCAP) project that is supported by the US National Institute of Aging under grant R01 AG056329. SHARE-HCAP is a sub-study of SHARE, the largest social science longitudinal study in Europe (Börsch-Supan et al., 2013), and aims to assess cognitive status in five countries: Czech Republic, Denmark, France, Germany, and Italy.

HCAP stands for the Harmonized Cognitive Assessment Protocol that has been developed by the US Health and Retirement Study (Weir et al., 2014) in order to harmonize the measurement of cognition in a global network of sister studies spanning North and South America, Asia, and Africa (Langa et al., 2020). SHARE-HCAP was developed to address the need for a standardized assessment of cognitive impairment in the European context.

SHARE-HCAP is comprised of an in-depth battery of cognitive tests that assess several cognitive domains, including memory, executive functioning, language and fluency, visuospatial skills, and orientation to time and place. SHARE-HCAP also includes an interview conducted with a family member or friend that assesses informant-reported cognitive functioning and ability to perform activities of daily living. Results from SHARE-HCAP can be linked with economic, health, and social data from the core SHARE study, and global data from other HCAP sister studies.

In this preliminary release, SHARE provides a first estimate of the cognitive status measured as three categories: normal, mild cognitive impairment and severe cognitive impairment possibly caused by the presence of Alzheimer's disease (AD) or AD-related dementias (ADRD) (Albert et al., 2011, McKhann et al., 2011, Manly et al., 2022).

This note describes the methodology that has been used to compute this first estimate of cognitive status and its shortcomings. It is structured as follows. Section 2 describes the selection of test items. Section 3 summarizes the main outcomes of the data collection. Section 4 presents the confirmatory factor analysis that was used to condense the items enumerated in Section 2. Section 5 details the classification algorithm that produces the three-category cognitive status. Sections 6 and 7 conclude with several caveats for users of this preliminary estimate. We first note that we detected a substantial extent of heterogeneity among countries which puts some doubt on the measurement invariance across countries. Second, we note that using the estimated cognitive status as if it were an exogenously given explanatory variable creates several econometric challenges.

2. Item selection

The SHARE-HCAP substudy was initiated in 2017. It consisted of in-depth cognitive testing of the respondent, including other respondent data, and an informant interview data (Douhou et al., 2024). All tests were based on the Harmonized Cognitive Assessment Protocol (HCAP) from the U.S. Health and Retirement Study (HRS), which has been successfully adapted for samples in England, Mexico, India, China, and South Africa. The adaptation process for the five European countries (Czech Republic, Denmark, France, Germany, and Italy) included consultation with the SHARE-HCAP Project Advisory Board, local cognition experts and native speakers within each SHARE country team.

SHARE-HCAP cognitive tests were divided into five broad domains representing memory, executive functioning, visuospatial skills, language and fluency, and orientation. These domains were selected based on prior theoretical and empirical work and are widely accepted categories of cognitive functioning (Lezak, 2004).

Tables 1 and 2 show the tests included in the respondent and informant interviews of SHARE-HCAP:

Table 1: Respondent tests of SHARE-HCAP
MMSE
HRS TICS (3 items: Object naming; naming president)
CERAD Word List – Recall: Immediate and delayed, Recognition
Semantic Fluency (Animal Naming)
Symbol cancellation test
Timed Backward Counting Task
Brief Community Screening Instrument for Dementia (CSI-D; 4 items)
Story recall – immediate, delayed and recognition
CERAD Constructional Praxis – immediate and delayed
Symbol Digit Modalities Test (SDMT)
HRS Number Series
Raven’s Standard Progressive Matrices
Trail Making Test (Part A and Part B)

Table 2: SHARE-HCAP informant report items
Background information
Jorm IQCODE
Blessed part 2
HRS Activities
10/66 Dementia Research Group Informant Questionnaire (4 items)
CSI-D Cognitive Activities
Blessed part 1

The orientation domain consisted of two scores comprised of 11 items. One score was the sum of 10 items assessing orientation to time and place selected from the Mini-Mental State Exam (MMSE; Folstein, 1975). The second score came from an item that asked the name of the country’s current primary political leader (e.g., prime minister, president), which was adapted from the Telephone Interview for Cognitive Status (TICS; Brandt, Spencer, & Folstein, 1988).

The memory domain consisted of 11 scores from the immediate, delayed, and recognition recall of a 10-word list from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; Morris et al., 1989), and the immediate, delayed, and recall of a logical memory test from the Wechsler Logical Memory Scale Fourth Edition (WMS-IV; Wechsler, 2009) and the East Boston Memory Test (i.e., Brave Man story) (Scherr et al., 1988). Additionally, the delayed recall of shapes was included and taken from the CERAD constructional praxis test (Morris et al., 1989; Rosen et al., 1984).

The executive functioning domain consisted of 8 scores, including performance on the Raven's Standard Progressive Matrices (Raven, 1981, 1989, 2000) as adapted by HRS, Trail Making Test Parts A and B (TMT; Reitan, 1992), symbol cancellation test (Mesulam, 1985), Symbol Digit Modalities Test (SDMT; Smith, 1982), backward counting (Agrigoroaei & Lachman, 2011), HRS Number Series (Fisher, McArdle, McCammon, Sonnega, & Weir, 2013), attention and calculation from MMSE.

The language and fluency domain consisted of 4 scores, including performance on animal naming (Woodcock et al., 2001; Schrank & Flanagan, 2003; Weir et al., 2014), naming described objects from the TICS, and naming common objects from Community Screening Instrument for Dementia (CSI-D; Hall et al., 1993; Hall et al., 2000). Additionally, several items from the MMSE were selected including object naming, writing a sentence, repeating a phrase, following a three-step oral command, and reading and following written instructions.

The visuospatial domain consisted of two scores representing performance on the CERAD constructional praxis copy task, and the MMSE drawing task.

3. Outcomes of the data collection

SHARE-HCAP's sampling strategy has been inspired by the strategy used in the HCAP study of the English Longitudinal Study of Aging (ELSA-HCAP) (Cadar et al., 2021). Briefly, respondents hypothesized to be at greater risk of cognitive impairment based on prior performance on the word recall test in the core SHARE in an earlier wave were oversampled to ensure adequate number of respondents with mild to severe cognitive impairment. For more details, see Douhou et al. (2024).

Response rates by risk group in the corresponding earlier wave are shown in Table 3:

Table 3: Overall response rates and by risk group.

At risk of CI	Severe CI	Mild CI	Normal	Overall
Czechia	60.9%	75.6%	74.9%	72.4%
Denmark	67.4%	76.5%	73.5%	73.2%
France	61.3%	69.1%	75.9%	72.2%
Germany	67.4%	70.3%	78.7%	74.5%
Italy	67.9%	82.7%	83.3%	79.3%

The resulting SHARE-HCAP data comprises data from 2687 respondents in the Czech Republic, Denmark, France, Germany and Italy (Czech n = 502; Danish n = 573, French n = 528, German n = 547, and Italian n = 537). Pooled across these five countries, respondents were on average 75.4 years old and primarily female (54.5%). 63.4% completed secondary education as assessed by the International Standard Classification of Education (ISCED; United Nations Educational, Scientific and Cultural Organization, 2003).

Table 4: Number of observations per indicator.

Indicator	Observations	% Missing
CERAD word list immediate recall	2683	0.15 %
MMSE word list, immediate recall	2681	0.22 %
WMS-IV logical memory immediate recall	2664	0.86 %
Brave Man immediate recall	2679	0.30 %
CERAD word delayed recall	2656	1.15 %
WMS-IV logical memory delayed recall	2645	1.56 %
MMSE word list, delayed recall	2676	0.41 %
CERAD constructional praxis delayed recall	2587	3.72 %
Brave Man delayed recall	2645	1.56 %
CERAD word list recognition	2650	1.38 %
WMS-IV logical memory recognition	2098	21.92 %
Raven's Progressive Matrices	2609	2.90 %
Trail Making Test part B	2349	12.58 %
HRS Number Series	2273	15.41 %
SDMT	2431	9.53 %
Trail Making Test part A	2510	6.59 %
MMSE Attention	2684	0.11 %
Backward counting	2609	2.90 %
Symbol Cancellation test	2514	6.44 %
CERAD Constructional praxis, copy	2627	2.23 %
MMSE drawing	2599	3.28 %
MMSE Orientation	2685	0.07 %
TICS: naming president	2465	8.26 %
Animal fluency	2674	0.48 %
TICS: object naming	2684	0.11 %
MMSE Naming	2684	0.11 %
CSI-D Naming	2679	0.30 %

In general, item nonresponse is not a major concern in this study, as can be seen in Table 4, which lists the number of observations per indicator and the percentage missing across the

SHARE-HCAP countries. The few indicators that stand out are the WMS-IV logical memory recognition (21.92%), the HRS Number Series (15.41%) and TMT part B (12.58%). In all three indicators, Italy accounts for roughly at least 40% of the missings per indicator. Recognition of story points followed right after the delayed recall of the two stories in SHARE-HCAP, which may point to respondent fatigue. The HRS Number Series was administered after WMS-IV logical memory recognition, and we see that a majority of respondents, who did not do the recognition test, also did not respond to the numeracy test. The Trail Making Test Part B is the last test in the SHARE-HCAP battery. Probably due to respondent fatigue, the percentage missing is relatively high for this test.

No imputation was used in the results presented in this note. We plan to use several imputation methods in our future work.

4. Confirmatory factor analyses

Confirmatory factor analysis models were used to examine the factor structure of SHARE-HCAP's cognitive test battery. Based on theory and prior empirical work, unidimensional models representing the cognitive domains of orientation, visuospatial skills, memory, executive functioning, and language & fluency were tested on the full sample. If fit was not adequate, models were re-specified based on modification indices and theoretical and/or methodological considerations. Details are provided in the paper by Otero et al. (forthcoming).

The cognitive status provided in the preliminary release is based on unidimensional models applied to the full SHARE-HCAP sample. There was one exception for the orientation and visuospatial domains, which were estimated as two correlated domains within a two-factor model given that the orientation and visuospatial were measured by two items each and therefore did not satisfy model identification requirements (Kline, 2011). Factor structures are shown in Figures 1-4.

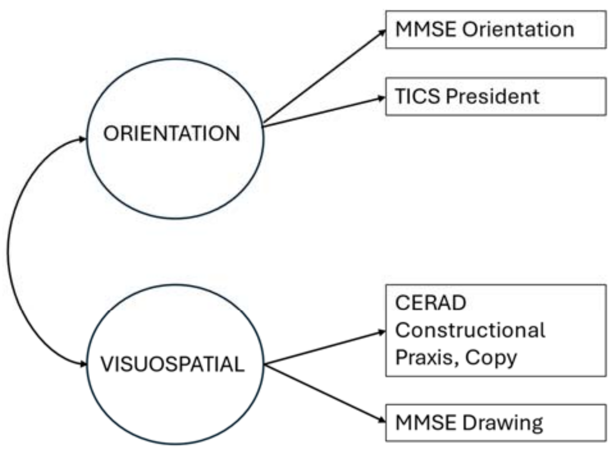


Figure 1. Diagram of correlated two-factor model for orientation and visuospatial domains.

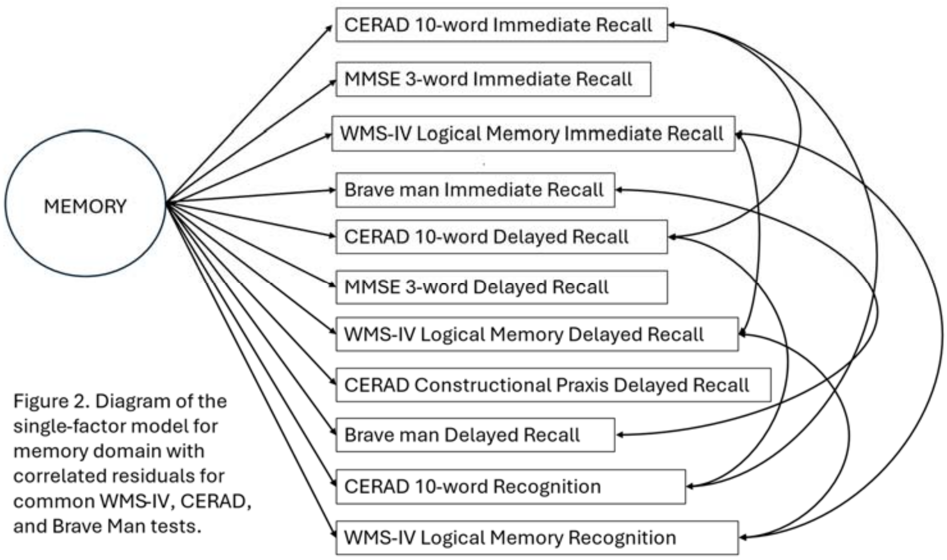


Figure 2. Diagram of the single-factor model for memory domain with correlated residuals for common WMS-IV, CERAD, and Brave Man tests.

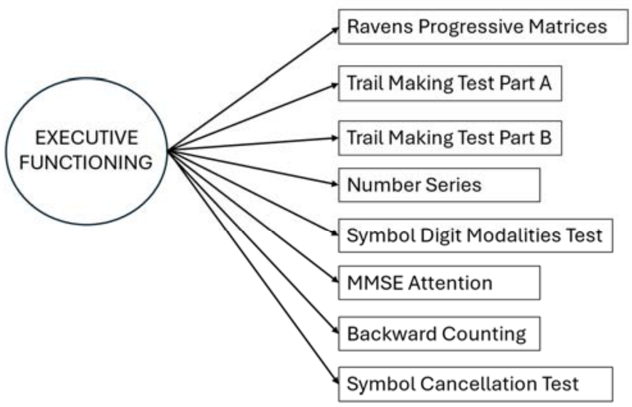


Figure 3. Diagram of the single-factor model for executive functioning domain.

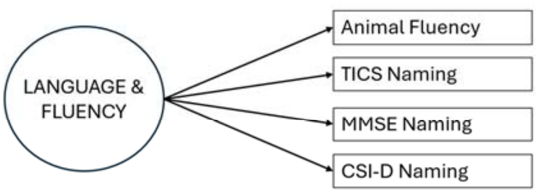


Figure 4. Diagram of the single-factor model for language and fluency domain.

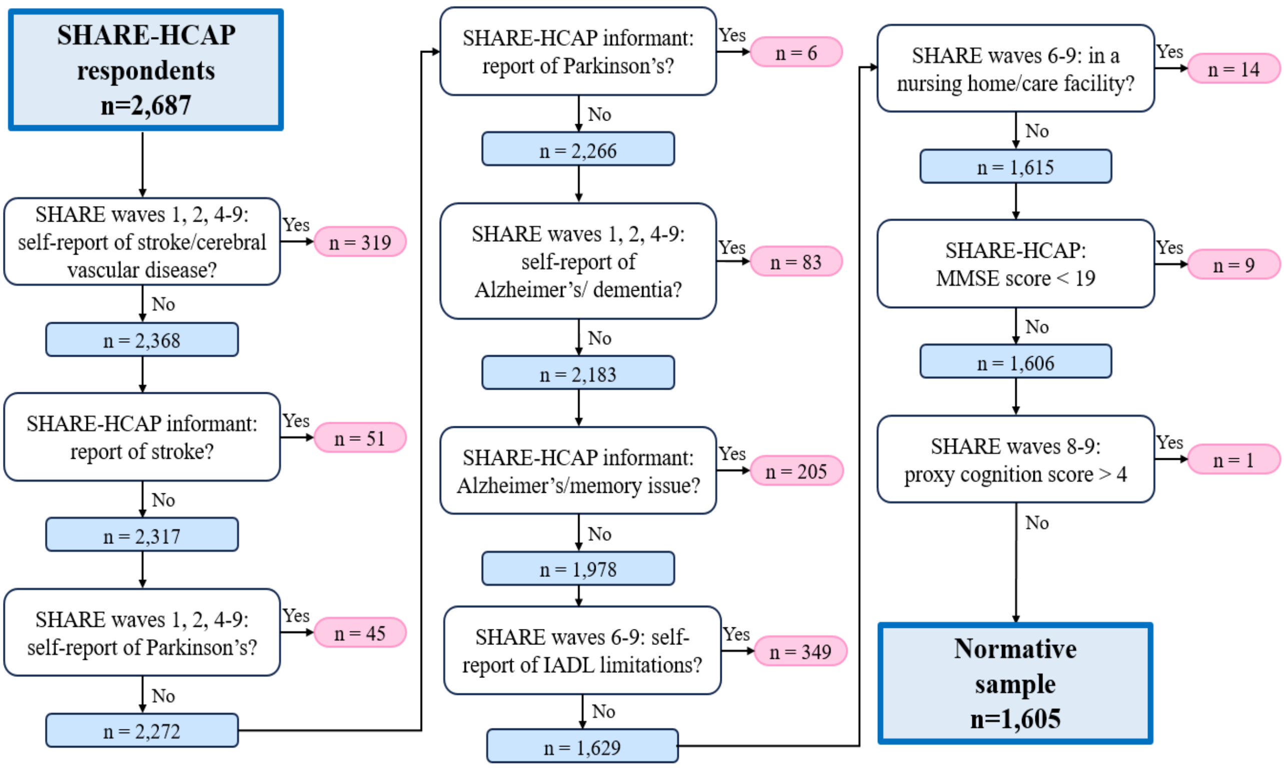
Final model fits ranged from perfect (orientation and visuospatial domains) to adequate (memory, executive functioning, language & fluency) per Hu and Bentler's (1999) criteria for Comparative Fit Index (CFI), the Root Mean Squared Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR).

5. Classification

To compute the cognitive classification scores provided in this first and preliminary release (normal, MCI, SCI), we follow the approach that has been described in Manly et al. (2022). It has three steps. First, we selected a normative sample from the SHARE-HCAP sample. The exclusion criteria were based on conditions that have been found to be related to pathological cognitive ageing. Specifically, individuals with neurodegenerative disease, stroke, or significant cognitive or functional impairment were excluded from the normative sample.

Figure 5 presents a more detailed flowchart with the criteria used for the selection of the normative sample from the SHARE-HCAP sample. Respondents excluded from the normative sample fulfilled one or more of the criteria for which we relied on information reported by a friend/family member of the respondent ("informant") and self-reported information in earlier core waves of SHARE. The criteria were inspired from Manly et al (2022) and discussions with our Project Advisory Board. The final normative sample includes 1,605 respondents from the SHARE-HCAP sample.

Figure 5: Flowchart of selection of normative sample

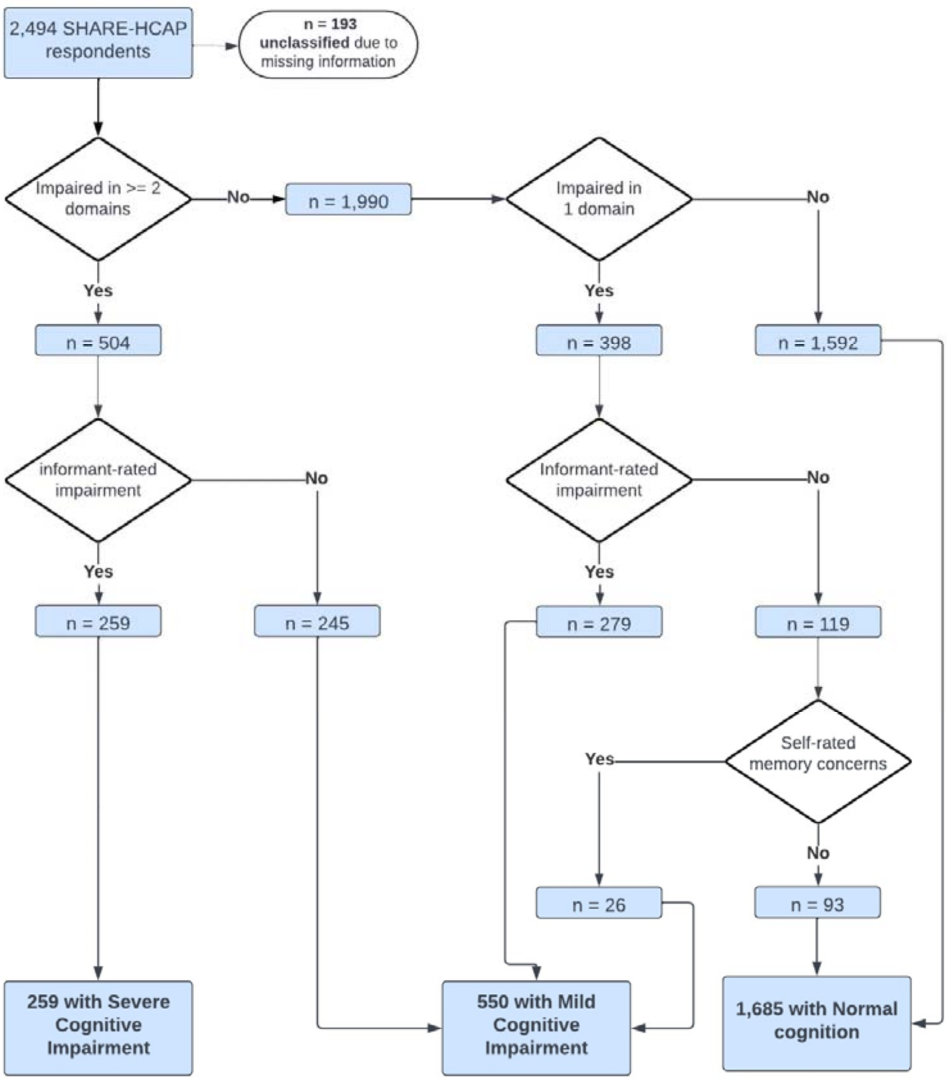


Second, we use factor analysis to derive factor score estimates in each of the five domains based on the unidimensional models described in the previous section. The factor score estimates within the normative sample were rank normalized and then a (linear) regression adjustment on the rank-based normalized scores was performed using age, gender, education and country of residence. Using the results from the regression, estimates of expected performance were generated for every combination of age, gender, education and country for the full SHARE-HCAP sample. For this, an adjusted score for every respondent was generated, which was standardized by dividing the standard error of estimate from the regression model. Following the Manly et al. approach, these were then placed on a T-score distribution (mean = 50, standard deviation =10) and rounded to the nearest integer.

A T-score of 50 indicates that a person is performing at the average level expected for an individual who is considered free from severe cognitive impairment, taking into account their age, gender, education, and country of residence. And a T-score below 35 suggests that a person performs 1.5 standard deviations below the mean, when compared to individuals with similar demographic characteristics who are considered free from severe cognitive impairment.

Third, we use a deterministic algorithm to classify respondents in three classes of cognition: normal, mild cognitive impairment and severe cognitive impairment. The classification algorithm is an exact replication of the one used by Manly et al. as shown in Figure 6 below.

Figure 6: Flowchart of SHARE-HCAP Classification Algorithm



193 observations could not be classified because essential information was missing. The distribution of their factor score estimate in each of the domains, as per CFA in section 4, resembles that of respondents that were classified as MCI.

6. Results

Table 5 shows some descriptive statistics by country, age and education, measured by the international standard code of education (ISCED). Italy and the Czech Republic stand out with a markedly higher shares of respondents categorized as MCI and SCI. The age and education gradients show the expected pattern.

These basic patterns are confirmed in a multivariate ordered probit regression (Table 6).

Table 5: Descriptive statistics (percentages)

Country	normal	MCI	SCI	Total (N*obs*)
Germany	74.76	20.04	5.20	100 (*519*)
Italy	58.51	22.75	18.74	100 (*523*)
France	73.13	18.96	7.92	100 (*480*)
Denmark	75.67	17.43	6.90	100 (*522*)
Czechia	54.44	32.22	13.33	100 (*450*)

Age	normal	MCI	SCI	all
65-69	26.23	20.00	6.95	22.85
70-74	24.93	23.45	12.36	23.3
75-79	20.71	18.91	22.78	20.53
80-84	15.61	20.00	24.32	17.48
85+	12.52	17.64	33.59	15.84
Total	100	100	100	100

Education	normal	MCI	SCI	all
Primary	32.34	39.27	55.60	36.29
Secondary	38.46	39.45	30.89	37.89
Tertiary	29.20	21.27	13.51	25.82
Total	100	100	100	100

Note: unweighted data

Table 6: Ordered probit regression of cognitive status

	Coef.	z-stat
Age	0.032	9.27
Female	-0.171	-3.32
ISCED	-0.072	-2.49
Czech Rep.	0.485	5.90
Denmark	0.004	0.05
France	-0.024	-0.27
Italy	0.369	4.12
Nobs	2,494	
Pseudo R2	0.052	
LR chi2(7)	213.1	

7. Heterogeneity across the five countries

Users are advised to be careful with the interpretation of this first and preliminary measurement of cognitive status in SHARE-HCAP. Very generally, the cognitive status provided is not a diagnose of mild or severe cognitive impairment of a responding individual but only an estimate that suffers from false positives as well as false negatives.

Furthermore, the estimation of the cognitive status has been based on the pooled data without distinguishing differences across the five countries due to the relatively small sample size in each country, i.e., we used the same model for the confirmatory factor analysis and the same thresholds in the categorization algorithm for all five SHARE-HCAP countries.

We checked for heterogeneity across the five countries in three ways. First, we ran Chow-type tests using OLS regressions of the indicators on the relevant factor score estimate (Bayesian plausible values) and a set of socio-demographic variables (age, age squared, ISCED, and gender) with 'country' as the grouping variable. The null hypothesis tested is whether the five country-specific intercepts and slopes are equal to those obtained in the pooled sample. Table 7 reports the p-values from the regressions. As can be seen, this null hypothesis is rejected for almost all indicators.

Table 7: Chow test p-values for equality of country-specific and pooled results.

Indicator	Constraint on intercepts		Constraint on Slopes		Constraint on intercepts and slopes	
CERAD word list immediate recall	0.26		0.00	reject	0.00	reject
MMSE word list, immediate recall	0.80		0.00	reject	0.00	reject
WMS-IV logical memory immediate recall	0.11		0.22		0.02	reject
Brave Man immediate recall	0.02	reject	0.00	reject	0.00	reject
CERAD word delayed recall	0.00	reject	0.01	reject	0.00	reject
WMS-IV logical memory delayed recall	0.17		0.00	reject	0.00	reject
MMSE word list, delayed recall	0.76		0.00	reject	0.00	reject
CERAD constructional praxis delayed recall	0.21		0.29		0.36	
Brave Man delayed recall	0.00	reject	0.00	reject	0.00	reject
CERAD word list recognition	0.53		0.00	reject	0.00	reject
WMS-IV logical memory recognition	0.85		0.00	reject	0.00	reject
Raven's Progressive Matrices	0.40		0.00	reject	0.00	reject
Trail Making Test part B	0.58		0.01	reject	0.04	reject
HRS Number Series	0.61		0.53		0.53	
SDMT	0.01	reject	0.00	reject	0.00	reject
Trail Making Test part A	0.10		0.00	reject	0.00	reject
MMSE Attention	0.01	reject	0.00	reject	0.00	reject
Backward counting	0.81		0.04	reject	0.09	
Symbol Cancellation test	0.36		0.01	reject	0.00	reject
CERAD Constructional praxis, copy	0.12		0.26		0.12	
MMSE drawing	0.26		0.58		0.43	
MMSE Orientation	0.61		0.00	reject	0.00	reject
TICS: naming president	0.44		0.00	reject	0.00	reject
Animal fluency	0.22		0.00	reject	0.00	reject
TICS: object naming	0.42		0.00	reject	0.00	reject
MMSE Naming	0.01	reject	0.00	reject	0.00	reject
CSI-D Naming	0.13		0.00	reject	0.00	reject

Second, we estimated five country-specific unidimensional factor analytical models in addition to the model for the pooled data. Using the same methodology as in Table 5, we find an even higher number of rejections.

A third way to assess whether the five countries are sufficiently similar to be described by a common factor analytical model is to use principal components analysis. We compared the coefficients associated with the first five principal components in the pooled sample with those coefficients that were estimated separately for each country. Except for Denmark, in all other countries at least two coefficients deviated by at least two standard deviations from coefficients derived from the pooled sample.

These results provide strong indications that there is significantly more heterogeneity across the five countries than assumed by the pooled model presented in the previous section work. Meaningful group comparisons – e.g. countries – require the establishment of measurement invariance of the instrument. Users should therefore restrict their analyses to the pooled data and abstain from country-specific analyses. Users should also abstain from extrapolating these cognitive-status estimates to any other country in the SHARE data since it is most likely that the heterogeneity is not confined to the country included in SHARE-HCAP.

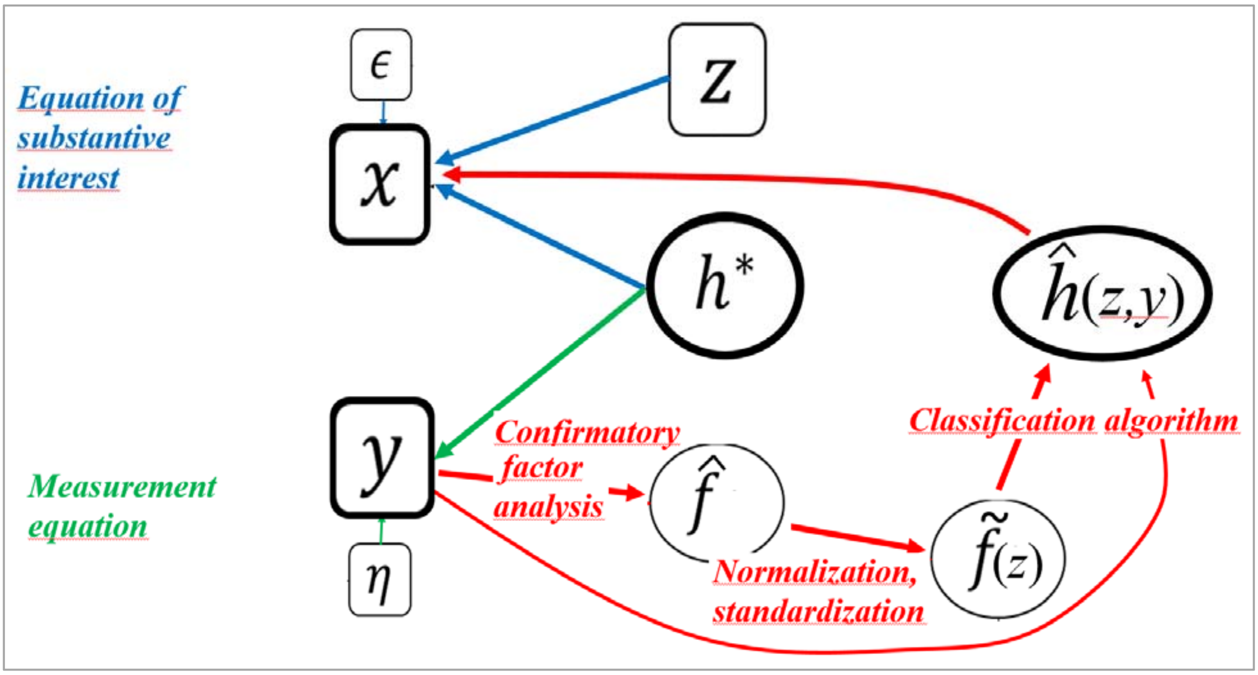
Further work is necessary to compute stable factor structures in each of the five countries. This is the main reason to stress that the cognitive status estimates provided in this release are very preliminary.

We also note that the thresholds in the classification algorithm (see Figure 6) are those used in HRS-HCAP. However, these thresholds may not be optimal for the five European countries. Better sensitivity and specificity may be found with thresholds that are tailored to the European context.

8. Independence assumptions needed to use SHARE-HCAP cognitive status as explanatory variable

In addition to the failure of the homogeneity assumption underlying the pooled factor analytical model, users are advised that strong independence assumptions are necessary to apply the estimated cognitive scores as explanatory variables in a regression that has some socio-economic variable as outcome variable. Figure 7 shows the general set-up of such regressions:

Figure 7: Typical application



There is first an equation of substantive interest, which measures the influence of cognition h^* (assumed to be a scalar such that it can be converted into the three categories normal, MCI and SCI) and other characteristics $\boldsymbol{z}$ (a vector with dimension M) on some outcome variable *x* (assumed to be a scalar), which varies across countries *i* and respondents *n*:

$$x_{in} = \beta\, h_{in}^* + \boldsymbol{\gamma}'\boldsymbol{z_{in}} + \epsilon_{in} \qquad \text{Equation of substantive interest}$$

For example, we want to understand the long-term care arrangement (*x* is the choice between nursing home, lives with children, home care) as a function of the individual's cognition h^* and other characteristics *z* such as age, gender, co-morbidities etc. β and γ may be common across countries or vary by country.

True cognition h^* is latent (hence the asterisk) and is estimated by the outcome $\hat{h}(\boldsymbol{z},\boldsymbol{y})$ of the deterministic classification algorithm

$$\hat{h}_{in}(\boldsymbol{z},\boldsymbol{y}) = j \ \text{ if } \left[\tilde{\boldsymbol{f}}_{\boldsymbol{in}}(\boldsymbol{z}) > \boldsymbol{\theta}_{\boldsymbol{i}}^{\boldsymbol{j}} \,\&\, \boldsymbol{y}_{\boldsymbol{in}} > \boldsymbol{\omega}_{\boldsymbol{i}}^{\boldsymbol{j}}\right] \qquad \text{Classification algorithm}$$

where $j =$ normal, MCI or SCI.

The classification algorithm is a sequence of conditions that involve observed indicators $\boldsymbol{y}$ (i.e., items of the SHARE-HCAP battery, a vector of dimension K) and a vector of the estimated factor scores $\hat{\boldsymbol{f}}$ that is produced by the confirmatory factor analysis and then standardized for age, gender, etc. (i.e., variables in $\boldsymbol{z}$) to generate $\tilde{\boldsymbol{f}}(\boldsymbol{z})$. Referring to Figures 1-4, we have assumed five factors (orientation, visuospatial skills, memory, executive functioning, and language & fluency) such that $\hat{\boldsymbol{f}}$ and $\tilde{\boldsymbol{f}}(\boldsymbol{z})$ are vectors with dimension 5. Each classification step imposes vectors of thresholds $\boldsymbol{\theta}_{\boldsymbol{i}}^{\boldsymbol{j}}$ and $\boldsymbol{\omega}_{\boldsymbol{i}}^{\boldsymbol{j}}$ on the standardized factor scores $\tilde{\boldsymbol{f}}(\boldsymbol{z})$ and indicators $\boldsymbol{y}$ as shown in Figure 6. The thresholds are specific for each cognition category *j* and may vary by country *i* (although they do not for our preliminary estimates).

The factor scores are derived from the confirmatory factor analysis that has been described in Section 3 and relate the latent cognition h^* to the observed indicators $\boldsymbol{y}$:

$$\boldsymbol{y}_{in} = \boldsymbol{\delta}' h_{in}^* + \boldsymbol{\eta}_{in} \qquad \text{Measurement equation}$$

While the classification algorithm is deterministic, both the equation of interest and the measurement equation are stochastic with error components ϵ_{in} and η_{in} that may vary across countries *i* and respondents *n*. The loadings δ may be common across countries (as we have done in our preliminary analysis) or vary by country (as we will do in our future work due to the indications in the previous section).

Replacing h^* by $\hat{h}(\boldsymbol{z},\boldsymbol{y})$ in the equation of substantive interest, as it is common practice, creates several statistical problems. First, while $\hat{h}(\boldsymbol{z},\boldsymbol{y})$ is an estimated variable via the confirmatory factor analysis and therefore has its own variability, users tend to treat $\hat{h}(\boldsymbol{z},\boldsymbol{y})$ as deterministic and ignore the additional variability in their assessment of statistical significance. Second, a regression of the equation of substantive interest will only generate unbiased results if the stochastic terms ε and η are independent from each other. This is not the case if similar unobserved components influence both the type of care and the indicators that measure cognition, a most likely case. Users should be aware of these statistical problems. They can be solved by approaches that combine the estimation of both the equation of interest and the measurement equation, e.g., by using MIMIC models with correlated error structures and suitable instruments.

Users should use the appropriate weights provided by SHARE (cross-sectional or longitudinal). However, in spite of the relatively high response rates among eligible SHARE respondents (Table 3), there is additional sample selectivity in the SHARE-HCAP sample. Future releases will provide additional weights to account for this.

Acknowledgments

Substantial funding for the collection and analysis of the SHARE-HCAP data was granted by the US National Institute on Aging (R01 AG056329). We thank John Phillips, Minki Chatterji and Jon King for their enduring support. The EU-Commission's contribution to SHARE-HCAP through H2020 (SHAREDEV3, No. 676536) is gratefully acknowledged. Data collection was also supported by national sources.

SHARE and SHARE-HCAP protocols were approved by the Ethics Committee of the Max Planck Institute in Germany.

The implementation of SHARE-HCAP has been made possible with the support of many people. Special thanks go to the members of the Project Advisory Board of SHARE-HCAP, Beatrice Baaba Tawiah, Annette Scherpenzeel, the country teams of the Czech Republic, Denmark, France, Germany, and Italy, and the staff at SHARE Central for their tremendous support throughout.

References

Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3): 270-279. doi:10.1016/j.jalz.2011.03.008

Agrigoroaei, S., & Lachman, M. E. (2011). Cognitive functioning in midlife and old age: Combined effects of psychosocial behavioral factors.

Brandt, J., Spencer, M. and Folstein, M. (1988). The Telephone Interview for Cognitive Status. Neuropsychiatry, Neuopsychology, and Behavioral Neurology, 1, 111-117.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). International journal of epidemiology, 42(4), 992-1001.

Cadar, D., Abell, J., Matthews, F. E., Brayne, C., Batty, G. D., Llewellyn, D. J., & Steptoe, A. (2021). Cohort Profile Update: The Harmonised Cognitive Assessment Protocol Sub-study of the English Longitudinal Study of Ageing (ELSA-HCAP). *International journal of epidemiology*, *50*(3), 725–726i. https://doi.org/10.1093/ije/dyaa227

Douhou, S., Otero, M., Pettinicchi, Y., Bergmann, M., de Luca, G., Quezada, M., Rieckmann, A., & Börsch-Supan, A. (2024). Cognition and Ageing: Introducing SHARE-HCAP. In M. Bergmann, M. Wagner, & A. Börsch-Supan (Eds.), *SHARE Wave 9 Methodology: From the SHARE Corona Survey 2 to the SHARE Main Wave 9 Interview* (p. 171-174). Munich: SHARE-ERIC.

Fisher, Gwenith G., John J. McArdle, Ryan J. McCammon, Amanda Sonnega, and David R. Weir. New Measures of Fluid Intelligence in the HRS (2013).

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

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Hall, K. S., Hendrie, H. C., Brittain, H. M., Norton, J. A., et al. (1993). The development of a dementia screening interview in two distinct languages. *International Journal of Methods in Psychiatric Research, 3*(1), 1–28.

Hall, K. S., Gao, S., Emsley, C. L., Ogunniyi, A. O., Morgan, O., & Hendrie, H. C. (2000). Community screening interview for dementia (CSI 'D'); performance in five disparate study sites. *International journal of geriatric psychiatry*, *15*(6), 521–531. https://doi.org/10.1002/1099-1166(200006)15:6<521::aid-gps182>3.0.co;2-f

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jones, R. N., Manly, J. J., Langa, K. M., Ryan, L. H., Levine, D. A., McCammon, R., & Weir, D. (2024). Factor structure of the Harmonized Cognitive Assessment Protocol neuropsychological battery in the Health and Retirement Study. Journal of the International Neuropsychological Society, 30(1), 47–55. doi:10.1017/S135561772300019X

Jorm, A. F. (2004). The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): A review. In International Psychogeriatrics (Vol. 16, Issue 3, pp. 275–293). https://doi.org/10.1017/S1041610204000390

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling. Third Edition*. The Guildford Press.

Langa, K. M., Ryan, L. H., McCammon, R. J., Jones, R. N., Manly, J. J., Levine, D. A., Sonnega, A., Farron, M., & Weir, D. R. (2020). The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. Neuroepidemiology, 54(1), 64–74. https://doi.org/10.1159/000503004

Lezak, M. D. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Manly, J. J., Jones, R. N., Langa, K. M., Ryan, L. H., Levine, D. A., McCammon, R., Heeringa, S. G., & Weir, D. (2022). Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project. JAMA neurology, 79(12), 1242–1249. https://doi.org/10.1001/jamaneurol.2022.3543

McKhann G.M., Knopman D.S., Chertkow H., et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3): 263-269. doi:10.1016/j.jalz.2011.03.005

Mesulam, M. M. (Ed.). (1985). Contemporary neurology series, volume 26. Philadelphia: Davis.

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellits, E.D., Clark, C., and the CERAD investigators (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology, 39, 1159-1165.

Muthén, L. K. & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eight Edition.* Los Angeles, CA: Muthén & Muthén.

Otero, M., Fernández, I., Douhou, S., Börsch -Supan, A. (forthcoming): Factor structure of the Survey of Health Aging and Retirement in Europe, Harmonized Cognitive Assessment Protocol. Discussion paper, Munich Research Institute for the Economics of Aging and SHARE Analyses

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. Cognitive Psychology, 41, 1 – 48.

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. Journal of Educational Measurement, 26, 1–16.

Raven, J. (1981). Manual for Raven's Progressive Matrices and Vocabulary Scales. Research supplement no. 1: The 1979 British standardization of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, together with comparative data from earlier studies in the UK, US, Canada, Germany, and Ireland. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Reitan, R.M. (1992). Trail Making Test: Manual for Administration and Scoring. Tucson, AZ: Reitan Neuropsychological Laboratory.

Rosen, W.G., Mohs, R.C., & Davis, K.L. (1984). A new rating scale for Alzheimer's disease. American Journal of Psychiatry, 141, 1356-1364.

Scherr, P. A., Albert, M. S., Funkenstein, H. H., Cook, N. R., Hennekens, C. H., Branch, L. G., White, L. R., Taylor, J. O., & Evans, D. A. (1988). Correlates of cognitive function in an elderly community population. *American journal of epidemiology*, *128*(5), 1084–1101. https://doi.org/10.1093/oxfordjournals.aje.a115051

Schrank, F. A., and Flanagan, D. P. (2003). WJIII clinical use and interpretation: Scientist-practitioner perspectives: Academic Press.

Smith, A. (1982). Symbol Digits Modalities Test. Los Angeles: Western Psychological Services.

UNESCO United Nations Educational, Scientific and Cultural Organization. (2003). International Standard Classification of Education, ISCED 1997. In: Hoffmeyer-Zlotnik, J.H.P., Wolf, C. (eds) Advances in Cross-National Comparison. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9186-7_10

Wechsler, D. (2009). Wechsler Memory Scale–Fourth Edition (WMS–IV) technical and interpretive manual. San Antonio, TX: Pearson.

Weir, D., McCammon, R., Ryan, L., & Langa, K. (2014). Cognitive Test Selection for the Harmonized Cognitive Assessment Protocol (HCAP). Ann Arbor, Michigan: Institute for Social Research, University of Michigan; 2014., October. http://hrsonline.isr.umich.edu/sitedocs/userg/HCAP_testselection.pdf

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishing.