Episodic Memory and Verbal Fluency Tasks: Normative Data from Nine Nationally Representative Samples

Elvira Lara[1,2,3,*] , Marta Miret[1,2,3], Albert Sanchez-Niubo[3,4], Josep Maria Haro[3,4], Seppo Koskinen[5], Matilde Leonardi[6], Beata Tobiasz-Adamczyk[7], Somnath Chatterji[8] and José Luis Ayuso-Mateos[1,2,3]

[1]Department of Psychiatry, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria Princesa (IIS-Princesa), Madrid, Spain
[2]Department of Psychiatry, Universidad Autónoma de Madrid, Madrid, Spain
[3]Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, Instituto de Salud Carlos III, Madrid, Spain
[4]Research, Innovation and Teaching Unit, Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain
[5]Population Health Research Unit, National Institute for Health and Welfare, Helsinki, Finland
[6]Neurology, Public health, Disability Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
[7]Department of Medical Sociology, Jagiellonian University Medical College, Krakow, Poland
[8]Department of Information, Evidence and Research, World Health Organization, Geneva, Switzerland

Abstract

Objective: This study aims to generate country-specific norms for two episodic memory tasks and a verbal fluency test among middle-aged and older adults using nationally representative data from nine low-, middle-, and high-income countries. **Method:** Data from nine countries in Africa, Asia, Europe, and Latin America were analyzed ($n = 42,116$; aged 50 years or older). Episodic memory was assessed with the word list memory (three trials of immediate recall) and word list recall (delayed recall). Verbal fluency was measured through the animal naming task. Multiple linear regression models with country-specific adjustments for gender, age, education, and residential area were carried out. **Results:** Both age and education showed high influence on test performance (i.e. lower cognitive performance with increasing age and decreasing years of education, respectively), while the effect of sex and residential area on cognitive function was neither homogeneous across countries nor across cognitive tasks. **Conclusions:** Our study provided sex-, age-, education-, and residential area-specific regression-based norms that were obtained from one of the largest normative study worldwide on verbal recall and fluency tests to date. Findings derived from this study will be especially useful for clinicians and researchers based at countries where cognitive norms are limited.

Keywords: Cognition, Assessment, Norms/Normative studies, Neuropsychology, Elderly/Aging, Population-based study

INTRODUCTION

According to the projections of the United Nations (2019), the number of individuals aged 65 or over is expected to more than double, reaching 1.5 billion by 2050. The share of persons above 80 years is expected to grow even faster. Within this context, people from low- and middle-income countries will account for most of the growth over the coming decades. These population estimates will have profound social and economic implications worldwide. For instance, this demographic transition will be accompanied by an increased burden of neuropsychiatric disorders such as dementia. Today, around 50 million people live with dementia, and this figure is projected to triple by 2050 (Prince et al., 2015). This steep increase will again be more pronounced among low- and middle-income countries, where limited epidemiological research on both pathological and non-pathological aging is yet available (World Health Organization, 2017).

Neuropsychological assessment is a basic tool to describe the cognitive performance of individuals and to identify both age-related cognitive changes and cognitive decline that are not a normal part of aging. Furthermore, it helps to evaluate improvements or further deterioration in cognitive performance and predict cognitive trajectories. As the world's population continues to grow, so will the demand for neuropsychological assessment. As a result, the development of appropriate neuropsychological normative data is essential for descriptive accuracy and diagnosis. Of relevant need are those tests of

verbal recall and verbal fluency, which are typically administered in protocols aimed at identifying subjects at higher risk of cognitive decline (Nyberg & Pudas, 2019). Particularly, poor performance on tasks measuring episodic memory has been consistently suggested to be a common manifestation in individuals with Alzheimer's Disease (AD), even at the very early stages (Backman, Jones, Berger, Jonsson Laukka, & Small, 2005; Weissberger et al., 2017). On the other hand, verbal fluency tests assess a diverse combination of cognitive domains (i.e. semantic memory, executive function, language, and processing speed) and present high discriminant validity for different types of neurodegenerative diseases, including AD, vascular dementia, and frontotemporal dementia (Jones, Laukka, & Bäckman, 2006; Van den Berg, Jiskoot, Grosveld, van Swieten, & Papma, 2017).

Previous research has provided neuropsychological norms for various settings and age-groups (Alenius et al., 2019; Beeri et al., 2006; Cavaco et al., 2013; Contador et al., 2016; Costa et al., 2014; Guruje et al., 1995; Hänninen et al., 2010; Kosmidis, Vlahou, Panagiotaki, & Kiosseoglou, 2004; Machado et al., 2009; Mejía-Arango, Wong, & Michaels-Obregón, 2015; Mitrushina, Boone, & Razani, 2005; Olabarrieta-Landa et al., 2015; Sosa et al., 2009; Vogel, Stokholm, & Jorgensen, 2019; Yang et al., 2012). For example, Sosa et al. (2009) showed normative values by sex, age, and education for four components of the 10/66 Dementia Research Group cognitive test battery for 13,649 individuals aged 65+ years in five Latin American countries, China, and India. In the same line, Olabarrieta et al. (2015) generated norms for several verbal fluency tests across 11 Latin America countries with country-specific adjustments for gender, age, and education in 3977 adults from 18 to 95 years of age. Although many widely used cognitive measures have been validated and standardized for evaluation of individuals across the life span, norms from some groups of individuals with diverse socioeconomic and cultural characteristics, especially among those countries where the population aging will be concentrated, are scarce or non-existent. Moreover, the majority of earlier research used convenience and/or small samples (Beeri et al., 2006; Cavaco et al., 2013; Costa et al., 2014; Guruje et al., 1995; Hänninen et al., 2010; Kosmidis et al., 2004; Machado et al., 2009; Mejía-Arango et al., 2015; Olabarrieta-Landa et al., 2015; Sosa et al., 2009; Vogel et al., 2019; Yang et al., 2012) rather than nationally or community-based representative samples (Abbott et al., 2019; Alenius et al., 2019; Contador et al., 2016; Kenny et al., 2013). Numerous studies have showed the sociodemographic influence on neuropsychological instruments based on methodological approaches that rely on artificial categorization. However, normative values derived from regression modeling provide more precise adjustment to the subject's cognitive performance and can assist in early identification of cognitive impairment in middle- and older-aged adults (Crawford, Garthwaite, Denham, & Chelune, 2012; Guardia-Olmos, Pero-Cebollero, Rivera, & Arango-Lasprilla, 2015). In order to overcome the above-mentioned limitations, the present study aims to generate country-specific norms for two tests of episodic memory and a verbal fluency task among middle- and older-aged adults using nationally representative data from nine countries.

METHODS

The Survey

Data from the Collaborative Research on Aging in Europe (COURAGE in Europe) [http://www.courageineurope.eu/] and the WHO Study on Global AGEing and Adult Health (SAGE) [https://www.who.int/healthinfo/sage/en/] studies were analyzed. The COURAGE in Europe project was undertaken between 2011 and 2012 in Finland, Poland, and Spain, while the SAGE study was conducted in China, Ghana, India, Mexico, Russia, and South Africa between 2007 and 2010. These six countries were considered low- and middle-income at the time of the survey, whereas the countries included in the COURAGE study were high-income countries based on the World Bank classification (World Bank, 2011).

Potential respondents were selected by a stratified, multistage, clustered area probability design to generate nationally representative samples. Distinct strata were used according to country characteristics, including geographical areas (e.g. regions, provinces, communities, states, districts), population size, residential area (i.e. urban/rural), and race (Garin et al., 2016). A list of household occupants was created for the final sampling units, from which one participant was randomly selected following age quotas (18–49 and 50+ years). In the SAGE study, however, individuals from 50+ households were all invited to participate (Kowal et al., 2012). Sampling weights were then generated to account for the population distribution obtained from the National Institute of Statistics or the United Nations Statistical Division for COURAGE in Europe and SAGE, respectively.

Overall, the sample comprised 53,269 non-institutionalized adults aged 18+ years. Individual response rates were as follows: 51% (Mexico), 53% (Finland), 67% (Poland), 68% (India), 70% (Spain), 77% (South Africa), 80% (Ghana), 83% (Russia), and 93% (China). Both surveys followed equivalent similar protocol to collect comparable, reliable, and valid data on health, lifestyle habits, and well-being outcomes. Standardized physical examinations and a neuropsychological test battery assessment were also performed. The protocol was translated from English into the local languages according to the WHO guidelines for translation and adaptation of instruments (World Health Organization, 2019). Face-to-face structured interviews were completed by trained personnel using a computer-assisted personal interview (CAPI) or a paper and pencil format. Quality control procedures were carried out during the fieldwork. If an individual was not able to undertake the survey because of severe (cognitive or physical) limitations, a shorter version of the questionnaire was administered to a proxy. Ethical approvals were obtained from the WHO Ethical Review Committee and the appropriate local ethics research review boards. All participants provided written informed consent.

The present study was focused on participants aged 50 years or older. Individuals who participated in the survey *via* a proxy respondent, whose information on cognition was lacking, were further excluded, yielding a final analytical sample of 42,116 individuals.

Measures

Cognitive performance

Cognitive function was measured by means of three performance tests assessing different domains [three elements from the CERAD (Consortium to Establish a Registry for Alzheimer's Disease) neuropsychological battery] (Morris, Heyman, & Mohs, 1989). The tests from the CERAD consisted of a word list memory (three trials of immediate recall), a word list recall (delayed recall), and a verbal fluency task. Word list memory assessed learning ability and episodic memory. In the COURAGE study, respondents were presented with 10 cards with unrelated common nouns to read aloud and remember. As for the SAGE project, the examiner read out the list of 10 words. Word list memory and word list recall measured the ability to recall the nouns given in the word list immediately and after presentation of a distractor task, respectively. Scores were the total number of words correctly recalled across the three learning trials (range 0–30) and after a brief delay (range 0–10). Higher values correspond to better performance. Individuals with missing values in any of the three trials of immediate recall were removed (less than 0.4%). The animal naming task assessed verbal fluency and production, semantic memory, and executive function. Participants were encouraged to name as many different animals as possible within a 60-s period. These include different breeds, gender, and generation-specific names. Repetitions and proper names (e.g. Lassie) were not credited. If the individual gave no answers within 15 s, the interviewer provided prompts or repeated basic instructions. In the same line, participants who stopped before the assigned time elapsed were encouraged to continue. Scores are the total number of valid answers. Higher scores indicate better performance.

Statistical Analyses

Descriptive analyses on the basic sociodemographic and clinical characteristics (i.e. sex, age, education, marital status, residential area, and medical conditions) were performed for each country. In order to obtain the expected score for each cognitive test with country-specific adjustments for sex, age, education, and residential area, multiple linear regression analyses were carried out. Regression-based methods play a useful role in neuropsychological research and clinical practice and produce methodologically robust outputs (Crawford et al., 2012). This approach is suitable for the comparison of an individual's obtained score to those predicted from large-scale and demographically corrected normative data.

The assumptions of the linear regressions were first statistically and visually tested. Models were built for each country independently using the raw scores of each cognitive test. Sex (0 = women; 1 = men), age (in years), education (in years), and residential area (0 = urban; 1 = rural) were included as the predictor variables. Results from all multiple linear regression models are presented as intercept, unstandardized regression coefficients (B) with its corresponding 95% confidence intervals (CIs), square root of the mean square residual (root MSE), and adjusted R^2. Complete-case analysis was performed. The level of statistical significance was set at $p < .05$. Sample weighting and the complex study design were considered in these analyses to adjust for the population structure of the respective countries. Data analyses were performed using Stata 15.1 (Stata Corp LP, College Station, Texas).

To calculate demographically adjusted z scores and percentiles from linear regression models, the following formula was then computed:

$$Z \text{ score} = (Y_i - Y'_i)/\text{Root MSE}$$

$$\begin{aligned} Y'_i = \beta_0 &+ (\beta_{\text{sex}} \times \text{Sex}_i) + (\beta_{\text{age}} \times \text{Age}_i) \\ &+ (\beta_{\text{education}} \times \text{Education}_i) \\ &+ (\beta_{\text{residential area}} \times \text{Residential area}_i) \end{aligned}$$

where:

- Y_i = participant's real score
- Y'_i = participant's expected score
- Root MSE = squared root of the mean square residual (i.e. standard deviation of the error term)
- β_0 = intercept
- β_{sex} = regression coefficient for woman
- Sex_i = participant's real sex
- β_{age} = regression coefficient for age
- Age_i = participant's real age
- $\beta_{\text{education}}$ = regression coefficient for years of education
- Education_i = participant's real years of education
- $\beta_{\text{residential area}}$ = regression coefficient for urban area
- $\text{Residential area}_i$ = participant's real residential area

Z scores can be positive or negative numbers. A positive Z score can be interpreted as a better cognitive performance than expected. On the contrary, a negative Z score would indicate that the individual showed worse cognitive performance than could be expected taking into account the subject's sex, age, years of education, and residential area. More precisely, Z scores ≤ -1.0 usually denote "borderline cognitive impairment," ≤ -1.5 would indicate (non-clinical) "mild cognitive impairment," while a Z score ≤ -2.0 generally represent "significant cognitive impairment" (Abbott et al., 2019; Schinka et al., 2010). A standard method for creating point estimates of the percentile ranks was then used in keeping with Crawford, Garthwaite, and Slick (2009). Inspired by previous initiatives (Cavaco et al., 2013; Crawford et al., 2009; Larouche et al., 2016), we have developed a user-friendly Libreoffice Calc® file for clinicians and researchers to facilitate the interpretation of the subject's cognitive

performance. This open source software can be downloaded for free at https://www.libreoffice.org/. The professional is requested to introduce the individual's main characteristics and the real score for the specific cognitive test. The program will then provide the expected score for this individual, his/her *z* score and its corresponding point estimate of the percentile rank. This file is available in the Supplementary Material.

RESULTS

Sample Characteristics

Among our analytical sample of 42,116 individuals, the mean age was 62.8 years (SD = 9.5), and there were more women than men (52.5% *vs.* 47.5%). Table 1 provides a summary of the sociodemographic and clinical characteristics of the study sample by country. The level of education showed significant variability across countries, with the highest percentage of participants with completed secondary or tertiary education found in Finland (82.7%), Poland (75.0%), and Russia (92.5%). The percentage of respondents married or in a relationship was higher in China (85.1%) and India (76.9%). Likewise, the Chinese, Indian, and Ghanaian samples showed the largest proportion of individuals living in rural areas. Overall, an important proportion of participants had hypertension (>60%, except for India), while Mexico and Spain had the highest prevalence of diabetes and depression. A proportion ranging from 3.7 to 6.5 was found for stroke.

Sociodemographic Effects on Cognitive Performance

Tables 2–4 show the results from the linear regression analyses. Assumptions for regression analyses were fulfilled for each model. The effect of sex, age, years of education, and residential area were explored for each cognitive test across all sites independently.

The influence of sex and residential area to the prediction of the cognitive performance was not uniform. For example, regression analysis showed significant differences between men and women from Poland in all cognitive tasks. On the contrary, cognitive performance was not associated with sex in any of the cognitive measures among the individuals living in Mexico or South Africa. Increasing age was significantly and consistently associated with lower cognitive scores across all sites and cognitive measures (*B* coefficients from −0.02 to −0.22). Similarly, higher years of education was significantly associated with higher cognitive scores in all tasks and countries. Overall, these sociodemographic variables accounted from 0.03 (South Africa for the verbal fluency task) to 0.30 (Finland and Russia for the immediate recall task) of variance in cognitive performance.

Normative Data

The previous tables would be used for the adjustment of the cognitive scores to generate normative data by country, sex, age, education, and residential area of the individual. As an illustration, assume that you have evaluated a Spanish man of 65 years living in a rural area who completed 9 years of formal education and generated 15 animal names in the verbal fluency task. According to the aforementioned equation, the individual expected score would be

$$\text{Score}_{\text{predicted}} = 27.40 + (1.31 \times 1) + (-0.20 \times 65) + (0.32 \times 9)$$

The adjusted predicted score for this subject is 18.59. Next, to calculate the individual's *Z* score, the predicted score is subtracted from the observed score and then divided by the root MSE, namely (15–18.59)/5.97. This *z* score of −0.60 equates to approximately the 27th percentile, representing a cognitive performance in the low average for his specific reference group, albeit clinically normal. In other words, around 27% of the Spanish men of 65 years and 9 years of formal education would name ≤15 words on this semantic fluency task. Please refer to the supplementary material in order to generate demographically adjusted norms using the Libreoffice Calc® file without the need for manual calculation.

DISCUSSION

Normative values for various widely used neuropsychological instruments were presented. While cognitive norms are increasingly available across the entire adult age range, respondents are commonly recruited following a purposive sampling method. To the best of our knowledge, this is the first and largest study to date to report normative data among nationally representative samples of low-, middle-, and high-income countries. The study design and procedure together with the regression-based approach extend the validity of our findings.

Although our objective was not to compare mean scores of cognitive tests across countries, but to report normative data for each test and country separately, we observed that Ghana, India, China, and South Africa consistently showed lower scores across all cognitive tasks.

Numerous studies have examined the effect of sociodemographic variables on cognitive performance. Our findings are in line with other previous research reporting that both age and education have a strong impact on cognitive results, albeit with different magnitude effects (Alenius et al., 2019; Beeri et al., 2006; Brody, Kramarow, Taylor, & McGuire, 2019; Contador et al., 2016; Ganguli et al., 2010; Hankee et al., 2016; Melikyan et al., 2019; Olabarrieta-Landa et al., 2015; Sosa et al., 2009; Van Der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006). It is possible that

Table 1. Baseline demographic and clinical characteristics of the study sample by country

Survey	COURAGE study ($n = 7987$)			SAGE study ($n = 34,129$)					
Country	Finland ($n = 1452$)	Poland ($n = 2910$)	Spain ($n = 3625$)	China ($n = 13,175$)	Ghana ($n = 4305$)	India ($n = 6560$)	Mexico ($n = 2313$)	Russia ($n = 3938$)	South Africa ($n = 3838$)
Sex									
Male	46.0	43.5	46.4	49.8	52.5	51.0	46.8	38.9	44.1
Female	54.0	56.5	53.6	50.2	47.5	49.0	53.2	61.1	55.9
Age (mean; SD)	65.0 (10.0)	64.2 (9.9)	66.3 (10.4)	62.6 (9.0)	64.4 (10.7)	61.5 (8.9)	63.0 (10.6)	63.9 (10.41)	61.6 (9.5)
50–59 years	33.3	38.3	32.2	44.9	39.7	48.6	48.1	45.2	49.9
60–69 years	35.4	33.0	28.2	31.9	27.5	30.9	25.6	24.6	30.6
70–79 years	20.1	19.9	29.3	18.6	23.1	16.0	17.7	21.8	14.0
80+ years	11.2	8.8	10.3	4.6	9.7	4.5	8.6	8.4	5.5
Education[a] (mean; SD)	11.2 (4.0)	11.3 (3.5)	9.9 (6.1)	7.2 (3.6)	9.2 (4.1)	7.6 (4.1)	6.1 (4.2)	11.2 (3.4)	8.0 (3.9)
No/basic education	1.3	2.4	32.3	42.0	64.3	61.3	55.6	2.0	49.1
Primary education	15.9	22.6	31.3	21.0	11.0	14.8	24.0	5.6	22.4
Secondary education	56.8	59.3	25.6	32.5	21.1	18.8	12.3	74.2	22.8
Tertiary education	25.8	15.7	10.8	4.5	3.6	5.1	8.1	18.2	5.7
Marital status									
Single	8.5	7.9	8.5	1.1	1.3	0.7	7.0	2.7	14.3
Married or in a relationship	64.7	68.1	62.4	85.1	59.3	76.9	73.0	58.3	55.9
Separated, divorced, or widowed	26.8	24.0	29.1	13.8	39.4	22.4	20.0	39.0	29.8
Residential area[b]									
Urban	75.8	69.0	83.9	47.3	41.1	28.9	78.8	72.7	64.9
Rural	24.2	31.0	16.1	52.7	58.9	71.1	21.2	27.3	35.1
Medical conditions[c]									
Depression	15.4	13.5	26.4	1.3	8.0	16.0	16.9	5.6	4.8
Diabetes	12.9	13.6	16.3	6.6	3.8	6.9	17.6	7.0	9.2
Hypertension	66.9	67.7	64.2	63.7	62.0	39.9	62.9	73.5	79.6
Stroke	4.9	5.8	5.3	3.8	3.7	3.7	5.6	6.5	4.9

Values are percentages for each category unless otherwise indicated. Some percentages are based on an incomplete sample because of missing data [less than 1%, except for level of education in Mexico (4%) and South Africa (15%)].

Abbreviation: SD = Standard deviation.

[a]Level of education followed the International Standard Classification of Education (ISCED-11). No/basic education was defined as illiterate, no formal education received but can read and write, or incomplete primary schooling.

[b]An urban area includes towns, cities, and metropolitan areas that have been legally declared as being urban. A rural area includes commercial farms, small settlements, rural villages, and other areas which are further away from towns and cities.

[c]Medical conditions were ascertained through the use of combined criteria (except for diabetes) including self-reported diagnosis, lifetime symptoms, and blood pressure measurement. For a detailed description please refer to Garin et al. (2016).

education contributes to cognitive performance by altering how individuals process cognitive tasks (Stern et al., 2018) and further slowing cognitive decline (Hankee et al., 2016). On the contrary, Seblova, Berggren, and Lovden (2020), who conducted a systematic review and meta-analysis of observational longitudinal cohort studies, have recently revealed that length of formal education was not associated with changes in cognitive performance. As suggested by Luck et al. (2018), it is also plausible that the effect of education on cognition is explained by reverse causality. That is, higher cognitive abilities may facilitate individuals to attain higher levels of formal education. However, this contrasts with the findings of a meta-analysis providing evidence for causal effects of education on a wide variety of cognitive abilities across the life span (Ritchie & Tucker-Drob, 2018). Educational attainment may also be related with engagement in healthy lifestyles and health care utilization that would ultimately affect cognitive performance. Beyond the impact of age and education on cognitive function, it is also plausible to explain differences in tests scores by cohort effects. Previous studies have reported that younger cohorts outperformed earlier generations for multiple cognitive domains

Table 2. Multiple linear regression models for immediate recall

	Finland	Poland	Spain	China	Ghana	India	Mexico	Russia	South Africa
	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)
Intercept	26.96***	22.24***	25.88***	23.25***	20.69***	21.59***	25.36***	29.76***	21.37***
	(24.93, 29.00)	(19.87, 24.60)	(24.47, 27.28)	(21.83, 24.67)	(18.90, 22.48)	(19.97, 23.22)	(20.68, 30.03)	(26.46, 33.05)	(19.06, 23.67)
Sex (ref. women)	−1.76***	−1.07***	−0.28	−0.41***	−0.01	−0.32	−0.60	−0.86*	0.12
	(−2.09, −1.44)	(−1.63, −0.51)	(−0.64, 0.07)	(−0.65, −0.17)	(−0.39, 0.37)	(−0.77, 0.13)	(−1.42, 0.22)	(−1.65, −0.06)	(−0.44, 0.68)
Age (in years)	−0.14***	−0.15***	−0.20***	−0.12***	−0.07***	−0.10***	−0.15***	−0.22***	−0.10***
	(−0.17, −0.12)	(−0.18, −0.12)	(−0.22, −0.18)	(−0.14, −0.10)	(−0.09, −0.04)	(−0.12, −0.08)	(−0.21, −0.08)	(−0.26, −0.18)	(−0.14, −0.06)
Education (in years)	0.31***	0.34***	0.22***	0.30***	0.15***	0.26***	0.21***	0.28***	0.30***
	(0.24, 0.38)	(0.23, 0.44)	(0.17, 0.27)	(0.24, 0.36)	(0.09, 0.20)	(0.21, 0.32)	(0.09, 0.32)	(0.19, 0.36)	(0.21, 0.38)
Residential area (ref. urban)	0.06	−0.49	−0.49	−0.29	0.20	0.08	−0.60	1.26	0.92*
	(−0.51, 0.64)	(−1.14, 0.15)	(−1.19, 0.20)	(−0.77, 0.19)	(−0.34, 0.75)	(−0.69, 0.85)	(−2.28, 1.07)	(−0.76, 3.29)	(0.15, 1.69)
Root MSE	3.61	5.31	4.43	4.35	3.97	3.67	3.46	4.38	4.44
Adjusted R^2	0.30	0.16	0.29	0.11	0.04	0.13	0.21	0.30	0.10

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
Abbreviations: CI = Confidence interval; Root MSE = Square root of the mean square residual.
Some analyses are based on an incomplete sample because of missing data in the cognitive test (less than 5% except for Finland 7.4%; Mexico 5.7%).

Table 3. Multiple linear regression models for delayed recall

	Finland	Poland	Spain	China	Ghana	India	Mexico	Russia	South Africa
	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)	B (95%CI)
Intercept	10.24***	7.82***	9.68***	8.09***	6.64***	6.75***	10.30***	9.52***	7.67***
	(9.11, 11.38)	(6.81, 8.82)	(9.00, 10.36)	(7.51, 8.67)	(5.82, 7.45)	(5.78, 7.71)	(8.77, 11.82)	(7.52, 11.52)	(6.43, 8.90)
Sex (ref. women)	−0.74***	−0.62***	−0.36***	−0.09	0.09	−0.38**	−0.24	−0.32	0.14
	(−0.93, −0.55)	(−0.84, −0.41)	(−0.52, −0.21)	(−0.21, 0.03)	(−0.13, 0.31)	(−0.63, −0.13)	(−0.87, 0.39)	(−0.65, 0.01)	(−0.16, 0.44)
Age (in years)	−0.07***	−0.06***	−0.09***	−0.06***	−0.04***	−0.04***	−0.09***	−0.09***	−0.05***
	(−0.08, −0.06)	(−0.07, −0.05)	(−0.10, −0.08)	(−0.07, −0.05)	(−0.05, −0.03)	(−0.05, −0.03)	(−0.11, −0.06)	(−0.11, −0.07)	(−0.07, −0.03)
Education (in years)	0.13***	0.13***	0.07***	0.12***	0.07***	0.09***	0.07**	0.10**	0.10***
	(0.10, 0.16)	(0.08, 0.17)	(0.06, 0.09)	(0.10, 0.14)	(0.04, 0.10)	(0.07, 0.11)	(0.03, 0.12)	(0.04, 0.17)	(0.06, 0.14)
Residential area (ref. urban)	−0.02	−0.26	−0.35*	−0.11	0.04	0.04	−0.65	0.60	0.18
	(−0.28, 0.24)	(−0.54, 0.03)	(−0.66, −0.04)	(−0.34, 0.12)	(−0.28, 0.35)	(−0.46, 0.54)	(−1.39, 0.10)	(−0.32, 1.52)	(−0.20, 0.56)
Root MSE	1.78	2.21	1.91	2.01	2.11	1.83	1.99	2.20	2.14
Adjusted R^2	0.27	0.16	0.27	0.10	0.04	0.07	0.20	0.21	0.07

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
Abbreviations: CI = Confidence interval; Root MSE = Square root of the mean square residual.
Some analyses are based on an incomplete sample because of missing data in the cognitive test (less than 5% except for Finland 7.4%; Mexico 5.7).

Table 4. Multiple linear regression models for verbal fluency

	Finland B (95% CI)	Poland B (95% CI)	Spain B (95% CI)	China B (95% CI)	Ghana B (95% CI)	India B (95% CI)	Mexico B (95% CI)	Russia B (95% CI)	South Africa B (95% CI)
Intercept	28.47*** (25.21, 31.72)	21.43*** (17.92, 24.95)	27.40*** (25.48, 29.32)	16.12*** (14.80, 17.43)	17.68*** (15.37, 19.99)	11.80*** (10.34, 13.26)	20.55*** (16.17, 24.92)	19.95*** (14.74, 25.15)	12.16*** (10.44, 13.88)
Sex (ref. women)	−0.10 (−0.86, 0.65)	1.27*** (0.60, 1.94)	1.31*** (0.73, 1.89)	0.66*** (0.40, 0.93)	0.60* (0.05, 1.15)	0.51* (0.06, 0.95)	1.05 (−0.13, 2.22)	0.90 (−0.70, 2.50)	0.43 (−0.13, 0.98)
Age (in years)	−0.16*** (−0.21, −0.12)	−0.16*** (−0.19, −0.12)	−0.20*** (−0.23, −0.18)	−0.07*** (−0.09, −0.05)	−0.09*** (−0.12, −0.06)	−0.02* (−0.05, −0.00)	−0.10** (−0.17, −0.04)	−0.19*** (−0.25, −0.12)	−0.05*** (−0.08, −0.02)
Education (in years)	0.51*** (0.41, 0.61)	0.57*** (0.43, 0.70)	0.32*** (0.26, 0.37)	0.24*** (0.18, 0.30)	0.14** (0.05, 0.22)	0.15*** (0.09, 0.22)	0.25*** (0.13, 0.38)	0.43*** (0.26, 0.60)	0.10* (0.01, 0.18)
Residential area (ref. urban)	−0.06 (−1.03, 0.91)	−0.50 (−1.32, 0.31)	−0.02 (−0.88, 0.85)	−0.83** (−1.46, −0.21)	1.53*** (0.65, 2.41)	−0.61 (−1.41, 0.18)	−1.80 (−3.78, 0.17)	2.94 (−2.50, 8.39)	0.74* (0.11, 1.38)
Root MSE	6.62	7.07	5.97	4.60	5.35	3.47	4.13	7.02	3.74
Adjusted R^2	0.18	0.16	0.25	0.07	0.05	0.06	0.18	0.17	0.03

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
Abbreviations: CI = Confidence interval; Root MSE = Square root of the mean square residual.
Some analyses are based on an incomplete sample because of missing data in the cognitive test (less than 5% except for Finland 7.0%; Mexico 7.4%; Russia 11.3%; South Africa 5.7%).

(Dickinson & Hiscock, 2011; Karlsson, Thorvaldsson, Skoog, Gudmundsson, & Johansson, 2015), with both measures of fluid and crystallized intelligence being vulnerable to these effects (Cornelis et al., 2019; Pietschnig, Voracek, & Formann, 2010). The exposure to analogous social and environmental circumstances for those being born at a specific place and moment in time will presumably influence cognitive abilities, while the progressive improvement of education with the passing years is thought to play a key role in determining cognitive performance (Schaie, Willis, & Pennak, 2005). In view of the above, normative data become less valid over the years, as the use of outdated norms may overestimate an individual's cognitive performance when compared with an older cohort. In this regard, our work, in addition to using nationally representative samples, has the advantage of providing normative values from recent data. It has to be mentioned, though, that observed changes in cognition may be partially explained by cohort effects given the cross-sectional nature of our study. There is vast literature emphasizing the importance of including sex to produce appropriate norms for neuropsychological tests. However, the effect of sex in this and other studies seems controversial (Luck et al., 2018; Melikyan et al., 2019; Olabarrieta-Landa et al., 2015; Sosa et al., 2009; Yang et al., 2012). Our findings showed a significant advantage for women in measures of episodic memory in some countries, whereas men performed better at the animal naming task. Finally, the residential area appeared to have a weaker influence on cognitive performance. Aspects of socioeconomic and sociocultural status could possibly influence cognitive scores (Contador et al., 2016). Furthermore, differences in the study population (e.g. age range, country of origin), study setting (e.g. clinically based vs. population-based), and measurement approaches (e.g. mean/standard deviations, T scores, percentiles, Z scores) may explain the observed heterogeneity.

The applicability of population-based norms within specific subpopulations is gaining greater attention because of the unstoppable globalization. Previous studies have suggested that neuropsychological batteries may be biased against, for example, ethnic minority groups as these tests may not reflect the same underlying construct with respect to the population for which the norms were developed. In this regard, measurement invariance is important to determine whether, under different conditions, assessment procedures yield equivalence of a latent cognitive structure (Barnes et al., 2016; Wicherts, 2016). Even if we have used nationally representative samples to generate cognitive norms for each country, measurement invariance has not been tested in the present study. Therefore, we caution the readers regarding the appropriateness of unitary norms for specific subgroups or country comparisons.

Tests of verbal recall and fluency are almost universally used and are typically included in standard protocols of cognitive tests. They can be administered in both clinical and research settings, and they may be an appropriate option for some individuals (e.g. those with sensory limitations) who might otherwise not be suitable for neuropsychological assessment. These tasks are quick and easy to administer and

have value beyond their usefulness as reference measures of cognitive testing, being sensitive to many types of brain disorders (Backman et al., 2005; Jones et al., 2006; Van den Berg et al., 2017; Weissberger et al., 2017). It is important to ensure reliability and validity of cognitive testing while taking into account the influence of socioeconomic and cultural factors. This increased sensibility would not only facilitate early detection but could also assist to improve the effectiveness of interventions.

Strengths and Limitations

The strengths of the present study include the use of nationally representative samples from very culturally and economically diverse sites. Moreover, this study included a regression-based approach to obtain neuropsychological tests norms, instead of previous common methods (e.g., standardized Z scores or percentiles based on means and SD derived from arbitrary categories), which constitutes a convenient contribution to the applied clinical research field (Crawford et al., 2012). Furthermore, all interviewers were involved in a course prior to the survey to enable them learning about the procedures to administrate performance tests in an accurate, clear, and comparable way. Some limitations should also be taken into account when interpreting our study results. First, despite our best attempt to exclude individuals with cognitive limitations, it is possible that some respondents presenting cognitive impairment or even mild dementia may have been included in our analyses. However, we made use of family-reported dementia diagnosis, and interviewers were trained to identify individuals with severe cognitive function who were unable to undertake the interview. Second, some of our regression analyses are based on an incomplete sample because of missing data (a missing rate of 5% or less in most cases). Statistical methods to handle missing data were not performed due to the nature of missing data, as the majority of cases arose from individuals who were not capable to complete the tasks. Indeed, individuals who did not respond to one or more of the cognitive tests were more likely to be older, have lower years of education, and present higher disability (data not shown). Third, as it can be seen from the heterogeneous R^2, there should other variables that may account for cognitive performance that have not been considered when generating normative data.

CONCLUSIONS

Neuropsychological assessment constitutes a key component for cognitive testing, diagnosis, and disease progression monitoring. In order to assist in the accurate identification and follow-up of the (risk of) cognitive dysfunction, demographically adjusted norms are necessary. Our study provided sex-, age-, education-, and residential area-specific regression-based norms that were obtained from one of the largest normative studies worldwide on verbal recall and fluency tests to date. Both age and education were associated with test performance, while the effect of sex and residential area on cognitive function was not homogeneous across countries neither cognitive tasks. Findings derived from this study will be especially useful for clinicians and researchers based at countries where normative data are limited.

CONFLICT OF INTEREST

The authors have nothing to disclose.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S1355617720000582

REFERENCES

Abbott, R.A., Skirrow, C., Jokisch, M., Timmers, M., Streffer, J., van Nueten, L., ... Weimar, C. (2019). Normative data from linear and nonlinear quantile regression in CANTAB: Cognition in mid-to-late life in an epidemiological sample. *Alzheimer's & Dementia*, *11*, 36–44. doi: 10.1016/j.dadm.2018.10.007

Alenius, M., Koskinen, S., Hallikainen, I., Ngandu, T., Lipsanen, J., Sainio, P., ... Hanninen, T. (2019). Cognitive performance among cognitively healthy adults aged 30–100 years. *Dementia and Geriatric Cognitive Disorders Extra*, *9*(1), 11–23. doi: 10.1159/000495657

Backman, L., Jones, S., Berger, A.K., Jonsson Laukka, E., & Small, B.J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, *19*(4), 12. doi: 10.1037/0894-4105.19.4.520.supp

Barnes, L.L., Yumoto, F., Capuano, A., Wilson, R.S., Bennett, D.A., & Tractenberg, R.E. (2016). Examination of the factor structure of a global cognitive function battery across race and time. *Journal of the International Neuropsychological Society*, *22*(1), 66–75. doi: 10.1017/S1355617715001113

Beeri, M.S., Schmeidler, J., Sano, M., Wang, J., Lally, R., Grossman, H., & Silverman, J.M. (2006). Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*, *67*(6), 1006–1010. doi: 10.1212/01.wnl.0000237548.15734.cd

Brody, D.J., Kramarow, E.A., Taylor, C.A., & McGuire, L.C. (2019). Cognitive performance in adults aged 60 and over: National health and nutrition examination survey, 2011–2014. *National Health Statistics Reports*, *126*, 23.

Cavaco, S., Goncalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., ... Teixeira-Pinto, A. (2013). Semantic fluency and phonemic fluency: Regression-based norms for the Portuguese population. *Archives of Clinical Neuropsychology*, *28*(3), 262–271. doi: 10.1093/arclin/act001

Contador, I., Almondes, K., Fernandez-Calvo, B., Boycheva, E., Puertas-Martin, V., Benito-Leon, J., & Bermejo-Pareja, F. (2016). Semantic Verbal Fluency: Normative Data in Older Spanish Adults From NEDICES Population-Based Cohort. *Archives of Clinical Neuropsychology*, *31*, 9. doi: 10.1093/arclin/acw071

Cornelis, M.C., Wang, Y., Holland, T., Agarwal, P., Weintraub, S., & Morris, M.C. (2019). Age and cognitive decline in the UK Biobank. *PLoS ONE*, *14*(3), e0213948. doi: 10.1371/journal.pone.0213948

Costa, A., Bagoj, E., Monaco, M., Zabberoni, S., De Rosa, S., Papantonio, A.M., ... Carlesimo, G.A. (2014). Standardization and normative data obtained in the Italian population for a new verbal fluency instrument, the phonemic/semantic alternate fluency test. *Neurological Sciences*, *35*(3), 365–372. doi: 10.1007/s10072-013-1520-8

Crawford, J.R., Garthwaite, P.H., Denham, A.K., & Chelune, G.J. (2012). Using regression equations built from summary data in the psychological assessment of the individual case: Extension to multiple regression. *Psychological Assessment*, *24*(4), 4. doi: 10.1037/a0027699

Crawford, J.R., Garthwaite, P.H., & Slick, D.J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, *23*(7), 1173–1195. doi: 10.1080/13854040902795018

Dickinson, M.D., & Hiscock, M. (2011). The Flynn effect in neuropsychological assessment. *Applied Neuropsychology*, *18*(2), 136–142. doi: 10.1080/09084282.2010.547785

Ganguli, M., Snitz, B.E., Lee, C.W., Vanderbilt, J., Saxton, J.A., & Chang, C.C. (2010). Age and education effects and norms on a cognitive test battery from a population-based cohort: The Monongahela-Youghiogheny Healthy Aging Team. *Aging & Mental Health*, *14*(1), 100–107. doi: 10.1080/13607860903071014

Garin, N., Koyanagi, A., Chatterji, S., Tyrovolas, S., Olaya, B., Leonardi, M., ... Haro, J.M. (2016). Global multimorbidity patterns: A cross-sectional, population-based, multi-country study. *The Journals of Gerontology Series A Biological Sciences and Medical Sciences*, *71*(2), 205–214. doi: 10.1093/gerona/glv128

Guardia-Olmos, J., Pero-Cebollero, M., Rivera, D., & Arango-Lasprilla, J.C. (2015). Methodology for the development of normative data for ten Spanish-language neuropsychological tests in eleven Latin American countries. *NeuroRehabilitation*, *37*(4), 493–499. doi: 10.3233/NRE-151277

Guruje, O., Unverzargt, F.W., Osuntokun, B.O., Hendrie, H.C., Baiyewu, O., Ogunniyi, A., & Hali, K.S. (1995). The CERAD neuropsychological test battery: Norms from a Yoruba-speaking Nigerian sample. *West African Journal of Medicine*, *14*(1), 5.

Hankee, L.D., Preis, S.R., Piers, R.J., Beiser, A.S., Devine, S.A., Liu, Y., ... Au, R. (2016). Population normative data for the CERAD word list and Victoria stroop test in younger- and middle-aged adults: Cross-sectional analyses from the Framingham heart study. *Experimental Aging Research*, *42*(4), 315–328. doi: 10.1080/0361073X.2016.1191838

Hänninen, T., Pulliainen, V., Sotaniemi, M., Hokkanen, L., Salo, J., Hietanen, M., ... Erkinjuntti, T. (2010). Early detection of cognitive changes in memory diseases: New cut-off scores for the Finnish version of CERAD neuropsychological battery. *Doudecim*, *126*(17), 9.

Jones, S., Laukka, E.J., & Bäckman, L. (2006). Differential verbal fluency deficits in the preclinical stages of Alzheimer's disease and vascular dementia. *Cortex*, *42*(3), 347–355. doi: 10.1016/s0010-9452(08)70361-7

Karlsson, P., Thorvaldsson, V., Skoog, I., Gudmundsson, P., & Johansson, B. (2015). Birth cohort differences in fluid cognition in old age: Comparisons of trends in levels and change trajectories over 30 years in three population-based samples. *Psychology and Aging*, *30*(1), 83–94. doi: 10.1037/a0038643

Kenny, R.A., Coen, R.F., Frewen, J., Donoghue, O.A., Cronin, H., & Savva, G.M. (2013). Normative values of cognitive and physical function in older adults: Findings from the Irish Longitudinal Study on Ageing. *Journal of the American Geriatrics Society*, *61*(Suppl 2), S279–S290. doi: 10.1111/jgs.12195

Kosmidis, M.H., Vlahou, C.H., Panagiotaki, P., & Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society*, *10*(2), 164–172. doi: 10.1017/S1355617704102014

Kowal, P., Chatterji, S., Naidoo, N., Biritwum, R., Fan, W., Lopez Ridaura, R., ... Collaborators, S. (2012). Data resource profile: The World Health Organization Study on global AGEing and adult health (SAGE). *International Journal of Epidemiology*, *41*(6), 1639–1649. doi: 10.1093/ije/dys210

Larouche, E., Tremblay, M.P., Potvin, O., Laforest, S., Bergeron, D., Laforce, R., ... Hudon, C. (2016). Normative data for the montreal cognitive assessment in middle-aged and elderly Quebec-French people. *Archives of Clinical Neuropsychology*, *31*, 8. doi: 10.1093/arclin/acw076

Luck, T., Pabst, A., Rodriguez, F.S., Schroeter, M.L., Witte, V., Hinz, A., ... Riedel-Heller, S.G. (2018). Age-, sex-, and education-specific norms for an extended CERAD Neuropsychological Assessment Battery – Results from the population-based LIFE-Adult-Study. *Neuropsychology*, *32*(4), 461–475. doi: 10.1037/neu0000440

Machado, T.H., Charchat Fichman, H., Lucas Santos, E., Koenig, A.M., Amaral Carvalho, V., Santos Fernandes, C., ... Paulo Caramelli, P. (2009). Normative data for healthy elderly on the phonemic verbal fluency task – FAS. *Dementia & Neuropsychologia*, *3*(1), 6. doi: 10.1590/S1980-57642009DN30100011

Mejía-Arango, S., Wong, R., & Michaels-Obregón, A. (2015). Normative and standardized data for cognitive measures in the Mexican Health and Aging Study. *Salud Pública de México*, *57*(1), 7. doi: 10.21149/spm.v57s1.7594

Melikyan, Z.A., Corrada, M.M., Dick, M.B., Whittle, C., Paganini-Hill, A., & Kawas, C.H. (2019). Neuropsychological test norms in cognitively intact oldest-old. *Journal of the International Neuropsychological Society*, 25(5), 530–545. doi: 10.1017/S1355617719000122

Mitrushina, M.N., Boone, B.B., & Razani, J. (2005). *Handbook of normative data for neuropsychological assessment.* Cambridge: Oxford University Press.

Morris, J., Heyman, A., & Mohs, R. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I: Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39, 6.

Nyberg, L., & Pudas, S. (2019). Successful memory aging. *Annual Review of Psychology*, 70, 219–243. doi: 10.1146/annurev-psych-010418-103052

Olabarrieta-Landa, L., Rivera, D., Galarza-Del-Angel, J., Garza, M.T., Saracho, C.P., Rodriguez, W., ... Arango-Lasprilla, J.C. (2015). Verbal fluency tests: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, 37(4), 515–561. doi: 10.3233/NRE-151279

Pietschnig, J., Voracek, M., & Formann, A.K. (2010). Pervasiveness of the IQ rise: A cross-temporal meta-analysis. *PLoS ONE*, 5(12). doi: 10.1371/journal.pone.0014406

Prince, M., Wimo, A., Guerchet, M., Ali, G.C., Wu, Y., & Prina, M. (2015). World Alzheimer Report 2015. The global impact of dementia: An analysis of prevalence, incidence, cost and trends.

Ritchie, S.J., & Tucker-Drob, E.M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, 29(8), 1358–1369. doi: 10.1177/0956797618774253

Schaie, K.W., Willis, S.L., & Pennak, S. (2005). An historical framework for cohort differences in intelligence. *Research in Human Development*, 2(1), 21.

Schinka, J.A., Loewenstein, D.A., Raj, A., Schoenberg, M.R., Banko, J.L., Potter, H., & Duara, R. (2010). Defining mild cognitive impairment: Impact of varying decision criteria on neuropsychological diagnostic frequencies and correlates. *The American Journal of Geriatric Psychiatry*, 18(8), 684–691. doi: 10.1097/JGP.0b013e3181e56d5a

Seblova, D., Berggren, R., & Lovden, M. (2020). Education and age-related decline in cognitive performance: Systematic review and meta-analysis of longitudinal cohort studies. *Ageing Research Reviews*, 58, 101005. doi: 10.1016/j.arr.2019.101005

Sosa, A.L., Albanese, E., Prince, M., Acosta, D., Ferri, C.P., Guerra, M., ... Stewart, R. (2009). Population normative data for the 10/66 Dementia Research Group cognitive test battery from Latin America, India and China: A cross-sectional survey. *BMC Neurology*, 9, 48. doi: 10.1186/1471-2377-9-48

Stern, Y., Arenaza-Urquijo, E.M., Bartres-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., ... Conceptual Frameworks, W. (2018). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*, S1552–S5260. doi: 10.1016/j.jalz.2018.07.219

United Nations, Department of Economic and Social Affairs, & Population Division. (2019). *World population prospects 2019: Highlights.*

Van den Berg, E., Jiskoot, L.C., Grosveld, M.J.H., van Swieten, J.C., & Papma, J.M. (2017). Qualitative assessment of verbal fluency performance in frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders*, 44(1–2), 35–44. doi: 10.1159/000477538

Van Der Elst, W., Van Boxtel, M.P.J., Van Breukelen, G.P.J., & Jolles, J. (2006). Normative data for the animal, profession and letter M naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *Journal of the International Neuropsychological Society*, 12, 10. doi: 10.10170S1355617706060115

Vogel, A., Stokholm, J., & Jorgensen, K. (2019). Normative data for eight verbal fluency measures in older Danish adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology, and Cognition*, 1–11. doi: 10.1080/13825585.2019.1593935

Weissberger, G.H., Strong, J.V., Stefanidis, K.B., Summers, M.J., Bondi, M.W., & Stricker, N.H. (2017). Diagnostic accuracy of memory measures in Alzheimer's dementia and mild cognitive impairment: A systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 354–388. doi: 10.1007/s11065-017-9360-6

Wicherts, J.M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, 30(7), 1006–1016. doi: 10.1080/13854046.2016.1205136

World Bank. (2011). The world by income and region. Retrieved from https://www.worldbank.org/

World Health Organization. (2017). *Global action plan on the public health response to dementia 2017–2025.* Geneva.

World Health Organization. (2019). Process of translation and adaptation of instruments. Retrieved from https://www.who.int/substance_abuse/research_tools/translation/en/

Yang, L., Unverzagt, F.W., Jin, Y., Hendrie, H.C., Liang, C., Hall, K.S., ... Gao, S. (2012). Normative data for neuropsychological tests in a rural elderly Chinese cohort. *The Clinical Neuropsychologist*, 26(4), 641–653. doi: 10.1080/13854046.2012.666266