Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research

Emma Nichols[1], Derek K Ng[1], Shablina Hayat[2], Kenneth M Langa[3,4,5], Jinkook Lee[6,7], Andrew Steptoe[2], Jennifer A Deal[1], Alden L Gross[1]

1. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
2. Institute of Epidemiology and Health Care, University College London, London, UK
3. Department of Internal Medicine, University of Michigan, Ann Arbor, MI
4. Institute for Social Research, University of Michigan, Ann Arbor, MI
5. VA Center for Clinical Management Research, University of Michigan, Ann Arbor, MI
6. Department of Economics, University of Southern California, Los Angeles, CA
7. Center for Economic and Social Research, University of California, Los Angeles, CA

Corresponding Author: Emma Nichols, enicho30@jh.edu, 615 N Wolfe St. Baltimore MD, 21205

Keywords: measurement, cognition, dementia, aging, cross-national comparisons, global health

Color figures: 1 (Figure 2)

Word count: 3499/3500

Data elements: 5 (2 Tables, 3 Figures)

Abstract

INTRODUCTION: Most cognitive assessments have been developed in high-income countries but are used in diverse contexts. Differences in culture and context may affect performance of cognitive items.

METHODS*:* We used the Harmonized Cognitive Assessment Protocol surveys in the US, Mexico, India, England, and South Africa (combined N=11,364) to quantify associations across countries between cognitive items and cognitive impairment status using age- and sex-adjusted logistic regression.

RESULTS*:* Associations were stronger in the US (Median Odds Ratio [OR] across items=0.17) and England (Median OR=0.19), compared to South Africa (Median OR=0.23), India (Median OR=0.29), and Mexico (Median OR=0.28). Items assessing memory (e.g. delayed recall tasks) had the most consistent associations of the largest magnitudes across contexts.

DISCUSSION*:* Transporting cognitive items among countries and cultures warrants caution. Our results can guide the design of future instruments by identifying items that performed well either in individual contexts or across the range of contexts considered.

Word count: 149/150

1. Background

It is expected that 71% of individuals living with dementia will reside in low- and middle-income countries by 2050 [1]. However, most dementia research conducted to-date has taken place in high-income countries [2]. Research in diverse geographic settings can inform our understanding of the distribution of disease burden, raise awareness of dementia in contexts where this may be lacking, and can guide policy decisions, resource allocation, and public health planning efforts. Cross-national research can also identify differences in the effects of modifiable risk factors, informing targeted prevention efforts. Furthermore, comparisons across countries with wider ranges of risk factor profiles and larger variation in the causes and consequences of dementia may lead to new findings on modifiable risk factors or disease progression.

Recently, there has been increased attention on cross-national research focused on dementia and cognitive aging, spearheaded by large coordinated efforts such as the 10/66 Studies or the Harmonized Cognitive Assessment Protocol (HCAP) surveys [3,4]. The HCAP surveys represent one of the largest efforts to-date to conduct comparable population-representative studies on dementia and cognitive aging across geographic contexts [4].

Despite these efforts to conduct research in diverse geographical contexts, there is little available evidence to guide the design and implementation of cross-national studies on dementia. For example, the HCAP surveys leveraged evidence from the Aging, Demographics and Memory Study and the Religious Orders Study and Memory and Aging Project, two US-based cohorts, to guide selection of survey questions (items) on cognition for inclusion in the cross-national HCAP battery [4–6]. Other dementia studies in diverse settings, including in central Africa, Brazil, and China, have based item selection on expert opinion or prior work in other low-income settings without context-specific validation studies or other quantitative evidence [7–9].

However, demographic and cultural factors, such as language of test administration, sex/gender (sex), urbanicity, and race/ethnicity can impact performance on cognitive test items, holding underlying cognitive ability constant [10–13]. Many of these factors vary across geographies. Therefore, it is necessary to closely consider the utility of survey items selected for cross-national research; standardization of instruments may not be enough for valid and comparable measurement.

This study aims to provide concrete guidance for dementia measurement in future cross-national efforts through the evaluation of items on cognitive functioning for use in measuring and classifying dementia using the HCAP surveys. We will quantify differences and similarities across countries in associations between cognitive impairment and items on cognitive functioning to evaluate the utility of items for future research.

2. Methods

2.1 Methods Overview

The analytic plan had two main components: 1) Classification of cognitive impairment, and 2) Evaluation of associations between cognitive impairment and items on cognition (Figure 1). Step 1 was required because HCAP studies did not include clinical evaluations for formal dementia diagnoses. Therefore, we used an actuarial neuropsychological norms approach to define impairment; this approach conceptualizes impairment as a discrepancy between cognitive performance and demographically-adjusted norms [14].

2.2 Harmonized Cognitive Assessment Protocol (HCAP) Surveys

The HCAP series aimed to assess cognitive aging and dementia cross-nationally in sub-samples from the larger Health and Retirement Study International Partner Surveys (HRS IPS). The HRS IPS surveys used multistage probability sampling to generate nationally representative (with the exception of South Africa) samples of adults in private households [15–18]. The South African HRS IPS is instead representative of the rural sub-district of Agincourt [19]. HCAP sub-samples in the US and Mexico randomly sampled eligible participants, whereas the other studies oversampled those with low levels of cognition. We used data from the baseline HCAP wave in the US [4], Mexico [20], India [21], England [22], and South Africa [19]. Informed consent was obtained for all participants. Sample sizes ranged from 4096 in India to 606 in South Africa (combined N = 11,364). We excluded individuals with missing data on covariates (age, sex, education, race/ethnicity in the US) or high levels of missingness in cognitive testing (greater than 50% missingness leading to poor reliability of scores in all cognitive domains), resulting in a final sample size of 11,250 (excluded N=62 [US], 18 [England], 46 [South Africa], 1 [India], 56 [Mexico]) (details in the Appendix A).

2.3 Cognitive Measures

Table 1 describes the full list of cognitive items and compares their inclusion across studies. While collaborative efforts sought to ensure the highest possible concordance, some adaptations were necessary to accommodate different languages, cultures, and levels of numeracy and literacy [4]. Items on memory had the highest overlap among studies, followed by items on orientation and language. Items on executive functioning had the least overlap. Assessments of visuospatial functioning were brief, but included at least one item in all studies.

2.4 Sociodemographic and health questions

We considered sociodemographic factors in HCAP studies based on cultural relevance and data availability. In the US, we considered race and ethnicity. In India and Mexico we considered rurality, and in India and South Africa we also used literacy status. In all countries we dichotomized educational attainment based on the distribution in each study. To evaluate depressive symptomology, we considered all items administered from the Center for Epidemiologic Studies – Depression scale in each study [23]. Details on definitions of these variables are in Appendix A. Finally, we used information on informant-reported stroke, Alzheimer's disease, and memory problems from all HCAP studies with the exception of Mexico due to a lack of data availability. We additionally considered self-reported stroke and heart attack from the prior HRS IPS wave in all studies.

2.5 Step 1: Classification of Cognitive Impairment

We used an actuarial neuropsychological norms approach to classify cognitive impairment. This approach has three steps: (1) quantify cognitive functioning by cognitive domain; (2) define a normative sample of individuals unlikely to develop cognitive impairment; and (3) within basic demographic categories, compare cognitive scores between the normative sample and individual participants to define impairment. Previous work used similar methodology within the Mexico HCAP sample [24]. This process was completed independently within each HCAP study.

To quantify cognitive functioning by domain in each study, we used confirmatory factor analysis (CFA) models [25]. We estimated models for orientation, executive functioning, memory, and language. We were unable to estimate visuospatial functioning as two studies included only 1 item assessing visuospatial functioning.

We used information on functional limitations and self-reported health to define a cognitively robust group in each study. Using multivariable linear regression, we estimated normative cognitive scores within demographic categories from data on participants from the cognitively robust group. Norms were estimated separately in each country. Demographic categories included an individual's age, sex, and educational attainment (dichotomized). We further stratified norms by race and ethnicity in the US, rurality in India and Mexico, and literacy in India and South Africa due to relevance of these additional characteristics in each setting. To compare cognitive scores from the normative sample to scores from participants in the broader study samples within demographic categories, we calculated residual scores, which represent the difference between an individual's cognitive performance and their expected cognitive performance based on demographic characteristics. Individuals were defined as impaired if they had a residual score less than 1.5 standard deviations from demographically-corrected norms in any cognitive domain [13]. Individuals with missing scores on all cognitive domains were excluded (N=36 across all studies). Details on CFA models, definitions of the cognitively robust group, and the calculation of residual scores are in Appendix A.

2.6 Step 2: Description of data and evaluation of associations between cognitive impairment and items on cognition and functional limitations

We characterized HCAP samples using descriptive statistics. We then assessed patterns of missing data and quantified variability of responses to binary cognitive items (the proportion answering items correctly).

For our primary analysis, we used weighted multiple logistic regression (details on survey weights in Appendix A) controlling for age and sex to quantify associations between cognitive impairment and each item on cognition. To ensure effect sizes were comparable between binary and continuous items, we divided all non-binary items by 2 times the item's standard deviation [26]. Because each item on cognitive functioning also contributed to the classification of cognitive impairment, we used an iterative approach to avoid circularity. Specifically, to estimate the association between each cognitive item and cognitive impairment we re-calculated the classification of cognitive impairment (including re-estimating CFA models and re-calculating demographically adjusted norms) leaving out data on the item of interest. While this procedure does lead to 64 different sets of classifications (one for each item of interest), differences between classification sets were minimal (details in Appendix A). When there were fewer than 5 individuals in any given item response category and impairment status combination, we did not estimate odds ratios due to model instability (details in Appendix A). To make direct comparisons of the effect sizes between different HCAP studies, we subtracted parameters on the log scale. We assumed additive variance for normally distributed parameter estimates to calculate the variance of differences. To summarize effect sizes either across items or countries, we used the median as a measure of central tendency to prevent outliers from having undue influence. We used histograms of estimated odds ratios to inspect differences in the distribution of associations across countries.

2.7 Sensitivity Analyses

The US and England studies included individuals 65 years and older; younger participants were included in other countries. Therefore, we conducted a sensitivity analysis where we subset data to individuals 65 and older across all studies to ensure observed differences were not due to differences in age distributions of studies.

To test the sensitivity of results to the use of the neuropsychological norms approach for classification, we repeated primary analyses using latent class analysis as an alternative strategy for classification [27] (details in Appendix A).

3. Results

3.1 Descriptive statistics

The mean age was higher in the US (75.8, SD=7.5) and England (75.9, SD=7.1) in comparison to South Africa (69.2, SD=11.1), India (69.0, SD=7.6), and Mexico (68.1, SD=9.0) (Table 1). Educational attainment was highest in the United Sates (28.7% with post-secondary education), and in England (13.0% with post-secondary education). In comparison, in South Africa, India, and Mexico most participants had either no education or primary education only.

3.2 Missingness for items on cognition

Missingness was less than 10% for almost all items in the US and England, with the exceptions of the HRS Number Series in the US and the Trail-Making Test Part B in England (Appendix A figure S4). Higher levels of missingness were observed in a larger number of items in Mexico (4 items), India (12 items), and South Africa (9 items). In South Africa and India, items on executive functioning had the highest levels of missingness, with 68% missingness on the Trail-Making Test Part B and 54% missingness on the Symbol Digit Modalities Test in South Africa and 44% missingness on the Serial 7s test in India.

3.3 Associations for items on cognition

High performance (good or correct scores) on all cognitive items was negatively associated with cognitive impairment across all locations. However, there was substantial heterogeneity in the strength of the associations observed (Figure 2).

Memory. Across all settings, some of the items with the most consistently large associations with cognitive impairment tested memory performance, including the CERAD immediate sum of 3 trials (Median Odds Ratio [OR]=0.09; Range=0.07–0.17), the CERAD word list delay (Median OR=0.12; Range=0.09–0.20), and the logical memory delayed task (Median OR=0.13; Range=0.13–0.16).

198 *Language.* A number of items assessing language had low variability (most individuals answered
199 correctly), suggesting that these items may only help in classifying a small number of individuals
200 (Appendix A Figure S6). Additionally, several items, including the following instructions, do with a
201 hammer, and naming the prime minister/president items showed notable variation in estimated
202 associations between countries. For example, the do with a hammer item from the Community
203 Screening Instrument – Dementia (CSID) battery had a substantially stronger association with cognition
204 in Mexico (OR=0.14; 95% Confidence Interval [CI] 0.08–0.25) as compared to the US (OR=0.53; 0.41–
205 0.70) or India (OR=0.44; 0.38–0.51). Of language items administered, the animal fluency task showed
206 the most consistently strong relationship with cognitive impairment across each HCAP study (Median
207 OR=0.19; Range = 0.14–0.32).

208 *Executive functioning.* Of items measuring executive functioning, only letter or symbol cancellation was
209 administered across all HCAP studies, and it showed a fairly strong and consistent association with
210 cognitive impairment in all locations (Median OR=0.17; Range = 0.11–0.49); associations were weakest
211 in England (OR=0.49; 0.37–0.65) and South Africa (OR=0.38; 0.22–0.63). While the Symbol Digit
212 Modalities Test was not administered in India, it also showed robust associations with cognitive
213 impairment across the remaining countries (Median OR=0.18; Range=0.10–0.28). A number of items
214 were administered in only one or two studies. The Token test and Problem solving test were only
215 administered in India, but showed the strongest associations with cognitive functioning (Token test
216 OR=0.16, 0.13–0.19; Problem solving test OR=0.15, 0.12–0.18) of executive functioning items
217 administered in the India HCAP study.

218 *Orientation.* Similarly to items on language, the majority of orientation items had low variability (most
219 individuals answered correctly), indicating these items may only help in classifying a small proportion of
220 individuals (Appendix A Figure S6). Due to differences in the administration of orientation items across
221 studies as well as low numbers of incorrect responses, which led to model instability and suppressed

estimates, there were no orientation items with associations for all studies (Appendix A Figure S5). Across the four samples evaluated (variability was too low in England to estimate an odds ratio), the item assessing the day of the week had the strongest and most consistent associations with cognitive impairment (Median OR=0.19; Range=0.14–0.29).

3.4 Overall patterns and sensitivity analyses

Across all items, associations between cognitive impairment and survey items were stronger in the US (Median OR [Inter-Quartile Range=IQR]=0.17 [0.13–0.32]) and England (Median OR [IQR]=0.19 [0.13–0.25]), as compared to South Africa (Median OR [IQR]=0.23 [0.18–0.35]), India (Median OR [IQR]=0.29 [0.22–0.33]), and Mexico (Median OR [IQR]=0.28 [0.15–0.35]), although associations were meaningfully strong in all studies. These differences can additionally be visualized in terms of shifts in the distributions of estimated odds ratios between countries (Figure 3).

Subsetting to individuals over 65 had minimal effects on comparisons (Appendix A Figure S9). Results from latent class analysis also broadly replicated the pattern of findings from primary analyses (Appendix A Figures S7-S8).

4. Discussion

This study evaluated patterns in associations between cognitive impairment and items assessing cognition across countries. We found substantial variability across HCAP studies, although the magnitude of variation was different across items. The observed heterogeneity suggests that the performance of items for classification purposes is not consistent across settings. In general, associations between cognitive impairment and items on cognition were strongest in the US and England, as compared to South Africa, India, and Mexico. Many items in the HCAP battery were developed in high-income settings [28–31]. Associations between responses to these items and cognitive impairment may be somewhat weaker, to varying degrees, in other contexts.

Despite overall patterns, some cognitive items showed strong to moderate associations with cognitive impairment across all studies and should be recommended for use in future cross-national research. In particular, a number of memory items (CERAD immediate and delayed recall, and logical memory delayed recall) as well as the animal fluency task and the orientation item on naming the day of the week showed consistently strong associations with cognitive impairment in each study. Other items performed well in specific settings, such as the item on naming a hammer, which had a stronger association with cognitive impairment in Mexico compared to other contexts. Such items should be considered for use in settings they perform well in, but may not be optimal candidates for cross-national comparisons.

Differences in item performance may be due to differences in cultural contexts and educational attainment of participants in different HCAP studies. Prior work on the Hindi version of the Mini Mental State Examination for use in Ballabgarh, India found that participants did not keep track of years and were often not attuned to geographic location beyond the boundaries of their village, which affected

performance of items on orientation to time and place [32]. In this study, we also found weaker associations between items on orientation and cognitive impairment in India.

Prior work on cognition in Cree-speaking natives in Canada found that items involving calculations or numeracy requirements were challenging to implement due to low levels of educational attainment [33]. The Mexico and India HCAP studies did not administer many of the executive functioning/attention items included in the US and England studies due to concerns about education and numeracy. The South Africa HCAP study did administer items requiring numeracy, but we found high levels of missingness and weak associations with cognitive impairment in some of these items, including Trail-Making Test parts A & B and the Backwards counting test. Based on these convergent findings, we would not recommend the use of executive functioning tests with strong numeracy requirements for cross-national research. The symbol or letter cancellation task does not have such requirements and had strong to moderate associations with cognitive impairment across studies, indicating this item may be a better choice for cross-national research. Additionally, the two executive functioning items added to the India HCAP survey to assess executive functioning performed well compared to other executive functioning items in this setting. Future work should explore whether these items perform well in other low numeracy settings and in cross-national research.

This study leveraged large population-representative samples; minimal sample exclusions and use of sampling weights help ensure that findings are relevant to broader populations. However, the size and scale of the HCAP studies made the administration of gold-standard clinician-based diagnoses of dementia cost-prohibitive [34]. Instead, we used a neuropsychological norms approach to classifying cognitive impairment and assumed that normative samples across countries represented comparably healthy groups. The neuropsychological norms approach has been shown to result in fewer false positives compared to conventional criteria for mild cognitive impairment, and is highly correlated with Alzheimer's disease biomarkers [14,35,36]. We also conducted a sensitivity analysis using latent class

analysis as an alternative classification method, and found overall patterns remained consistent. Our classification of cognitive impairment likely captured more mild forms of impairment as compared to a dementia diagnosis and we did not require deficits in functional limitations. Despite differences, the measurement of cognitive impairment is critical to the measurement of dementia, therefore conclusions regarding the measurement of cognitive impairment will apply to the measurement of dementia as well. In our primary analyses, we were unable to incorporate uncertainty from the estimation of cognitive impairment in logistic regression models; instead we treated impairment status as fixed, in line with other studies relying on algorithmic classifications [37,38]. Classification uncertainty was taken into consideration in secondary analyses using latent class analysis, which yielded similar inferences.

Additionally, due to data availability constraints, our analysis is limited to data from 5 countries, 2 of which were similar high-income contexts (US and England). However, our results can provide important insights into the measurement of dementia in the specific contexts examined, and broader patterns highlight differences between measurement in high-income contexts (US and England) compared to other settings (Mexico, South Africa, India). Future research should seek to incorporate new HCAP data from additional countries, as these are released.

We focused on one metric of item quality, the association between cognitive impairment and specific items, but other information will likely impact item selection in future studies. Considerations surrounding the magnitude of missing data, the variability of binary items, and comprehensive content coverage across cognitive domains will also be important. Furthermore, while this study evaluated the utility of items on cognition for classification purposes, other uses exist (e.g. for the evaluation of specific cognitive subdomains), which may be considered.

The diversity of studies using different methods for decisions on the selection and inclusion of survey items has led to extreme heterogeneity across the literature, with a recent systematic review finding

over 230 different diagnostic procedures used in 237 studies for the assessment of dementia prevalence or incidence [2]. This variation highlights the lack of consensus on the best way to measure dementia across settings. Our results highlight the challenges in conducting cross-national research, which likely contribute to observed heterogeneity in measurement within the field.

Despite these challenges, we identified items on cognition which had strong associations with cognitive impairment either across settings or in individual HCAP studies. Items that performed consistently well across settings may be useful in future cross-national research, and can potentially be leveraged to allow post-hoc statistical harmonization efforts using item response theory methods [39]. Items that had strong associations with cognitive impairment in specific locations should be considered for use in those locations to improve measurement quality in a given study. Some differences in assessments of cognition can and should persist in cross-national research due to differences in culture and context. However, our results can guide the selection of a common set of items for use in cross-national research and can help standardize assessments across new epidemiologic studies on cognitive aging and dementia.

Acknowledgements: The Health and Retirement Study: Harmonized Cognitive Assessment Protocol survey is sponsored by the National Institute on Aging (grant number U01 AG058499) and is conducted by the University of Michigan. The MHAS (Mexican Health and Aging Study) is partly sponsored by the National Institutes of Health/National Institute on Aging (grant number NIH R01 AG018016) in the US and the Instituto Nacional de Estadística y Geografía (INEGI) in Mexico. Data files and documentation are public use and available at www.MHASweb.org. The LASI-DAD project is funded by grants R01 AG051125, RF1 AG055273, and U01 AG064948 from the National Institute on Aging and the National Institutes of Health. The HCAP substudy in ELSA was funded by R01 AG017644 from the National Institute on Aging. The Dementia substudy of the HAALSI cohort was funded by R01 AG054066 from the

National Institute on Aging. The HCAP international network is supported by National Institute on Aging grant R24 AG065182.

Funding Information: This work was supported by the National Institutes of Health [grant numbers T32AG000247 to ELN, R01AG051125 and 2RO1AG030153 to JL, K01AG054693 to JAD, and K01AG050699 to ALG].

Conflicts of Interest: The authors declare no competing financial interests.

References

[1] Prince MJ. World Alzheimer Report 2015: The Global Impact of Dementia 2015. https://www.alz.co.uk/research/world-report-2015 (accessed April 8, 2018).

[2] Nichols E, Szoeke CEI, Vollset SE, Abbasi N, Abd-Allah F, Abdela J, et al. Global, Regional, and National Burden of Alzheimer's Disease and Other Dementias, 1990–2016: a Systematic Analysis for the Global Burden of Disease Study 2016. The Lancet Neurology 2019;18:88–106. https://doi.org/10.1016/S1474-4422(18)30403-4.

[3] Prince M, Ferri CP, Acosta D, Albanese E, Arizaga R, Dewey M, et al. The protocols for the 10/66 dementia research group population-based research programme. BMC Public Health 2007;7. https://doi.org/10.1186/1471-2458-7-165.

[4] Langa KM, Ryan LH, McCammon RJ, Jones RN, Manly JJ, Levine DA, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. NED 2020;54:64–74. https://doi.org/10.1159/000503004.

[5] Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, et al. The Aging, Demographics, and Memory Study: Study Design and Methods. NED 2005;25:181–91. https://doi.org/10.1159/000087448.

[6] Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. J Alzheimers Dis 2018;64:S161–89. https://doi.org/10.3233/JAD-179939.

[7] Guerchet M, Mbelesso P, Ndamba-Bandzouzi B, Pilleron S, Desormais I, Lacroix P, et al. Epidemiology of dementia in Central Africa (EPIDEMCA): protocol for a multicentre population-based study in rural and urban areas of the Central African Republic and the Republic of Congo. SpringerPlus 2014;3:1044. https://doi.org/10.1186/2193-1801-3-338.

[8] César KG, Brucki SMD, Takada LT, Nascimento LFC, Gomes CMS, Almeida MCS, et al. Prevalence of Cognitive Impairment Without Dementia and Dementia in Tremembé, Brazil. Alzheimer Disease & Associated Disorders 2016;30:264–71. https://doi.org/10.1097/WAD.0000000000000122.

[9] Jia J, Wang F, Wei C, Zhou A, Jia X, Li F, et al. The prevalence of dementia in urban and rural areas of China. Alzheimer's & Dementia 2014;10:1–9. https://doi.org/10.1016/j.jalz.2013.01.012.

[10] Jones RN, Gallo JJ. Education and Sex Differences in the Mini-Mental State ExaminationEffects of Differential Item Functioning. J Gerontol B Psychol Sci Soc Sci 2002;57:P548–58. https://doi.org/10.1093/geronb/57.6.P548.

[11] Jones RN. Identification of Measurement Differences Between English and Spanish Language Versions of the Mini-Mental State Examination: Detecting Differential Item Functioning Using MIMIC Modeling. Medical Care 2006;44:S124–33.
[12] Goel A, Gross A. Differential item functioning in the cognitive screener used in the Longitudinal Aging Study in India. International Psychogeriatrics 2019;31:1331–41. https://doi.org/10.1017/S1041610218001746.
[13] Filshtein T, Chan M, Mungas D, Whitmer R, Fletcher E, DeCarli C, et al. Differential Item Functioning of the Everyday Cognition (ECog) Scales in Relation to Racial/Ethnic Groups. Journal of the International Neuropsychological Society 2020;26:515–26. https://doi.org/10.1017/S1355617719001437.
[14] Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological Criteria for Mild Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and Progression Rates. J Alzheimers Dis 2014;42:275–89. https://doi.org/10.3233/JAD-140276.
[15] Perianayagam A, Bloom D, Lee J, Parasuraman S, Sekher TV, Mohanty SK, et al. Cohort Profile: The Longitudinal Ageing Study in India (LASI). International Journal of Epidemiology 2022:dyab266. https://doi.org/10.1093/ije/dyab266.
[16] Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). Int J Epidemiol 2014;43:576–85. https://doi.org/10.1093/ije/dyu067.
[17] Steptoe A, Breeze E, Banks J, Nazroo J. Cohort Profile: The English Longitudinal Study of Ageing. Int J Epidemiol 2013;42:1640–8. https://doi.org/10.1093/ije/dys168.
[18] Wong R, Michaels-Obregon A, Palloni A. Cohort Profile: The Mexican Health and Aging Study (MHAS). International Journal of Epidemiology 2017;46:e2. https://doi.org/10.1093/ije/dyu263.
[19] Gómez-Olivé FX, Montana L, Wagner RG, Kabudula CW, Rohr JK, Kahn K, et al. Cohort Profile: Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI). Int J Epidemiol 2018;47:689–690j. https://doi.org/10.1093/ije/dyx247.
[20] Mejia-Arango S, Nevarez R, Michaels-Obregon A, Trejo-Valdivia B, Mendoza-Alvarado LR, Sosa-Ortiz AL, et al. The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and Methods. Archives of Gerontology and Geriatrics 2020;91:104210. https://doi.org/10.1016/j.archger.2020.104210.
[21] Lee J, Khobragade PY, Banerjee J, Chien S, Angrisani M, Perianayagam A, et al. Design and Methodology of the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia (LASI-DAD). Journal of the American Geriatrics Society 2020;68:S5–10. https://doi.org/10.1111/jgs.16737.
[22] Cadar D, Abell J, Matthews FE, Brayne C, Batty GD, Llewellyn DJ, et al. Cohort Profile Update: The Harmonised Cognitive Assessment Protocol Sub-study of the English Longitudinal Study of Ageing (ELSA-HCAP). International Journal of Epidemiology 2020. https://doi.org/10.1093/ije/dyaa227.
[23] Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: Evaluation of a short form of the CES-D. American Journal of Preventive Medicine 1994;10:77–84.
[24] Rentería MA, Manly JJ, Vonk JMJ, Arango SM, Obregon AM, Samper-Ternent R, et al. Midlife Vascular Factors and Prevalence of Mild Cognitive Impairment in Late-Life in Mexico. Journal of the International Neuropsychological Society undefined/ed:1–11. https://doi.org/10.1017/S1355617721000539.
[25] Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and validation of an innovative approach for estimating general cognitive performance. Neuroepidemiology 2014;42:144–53. https://doi.org/10.1159/000357647.
[26] Gelman A. Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine 2008;27:2865–73. https://doi.org/10.1002/sim.3107.
[27] McCutcheon AL. Latent class analysis. Sage; 1987.

[28] Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The consortium to establish a registry for Alzheimer's disease (CERAD): I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 1989;39:1159–65. https://doi.org/10.1212/WNL.39.9.1159.
[29] Folstein MF, Robins LN, Helzer JE. The Mini-Mental State Examination. Arch Gen Psychiatry 1983;40:812. https://doi.org/10.1001/archpsyc.1983.01790060110016.
[30] Brandt J, Spencer M, Folstein MF. The telephone interview for cognitive status. Cogn Behav Neurol 1988;1:111–7.
[31] Erkinjuntti T, Hokkanen L, Sulkava R, Palo J. The blessed dementia scale as a screening test for dementia. International Journal of Geriatric Psychiatry 1988;3:267–73. https://doi.org/10.1002/gps.930030406.
[32] Ganguli M, Ratcliff G, Chandra V, Sharma S, Gilby J, Pandav R, et al. A Hindi version of the MMSE: the development of a cognitive screening instrument for a largely illiterate rural elderly population in India. International Journal of Geriatric Psychiatry 1995;10:367–77.
[33] Hall KS, Hendrie HC, Brittain HM, Norton JA, Rodgers DD, Prince CS, et al. The development of a dementia screening interview in 2 distinct languages. International Journal of Methods in Psychiatric Research 1993;3:1–28.
[34] Diagnostic and statistical manual of mental disorders : DSM-IV. Fourth edition. Washington, DC : American Psychiatric Association, [1994] ©1994; 1994.
[35] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–8. https://doi.org/10.1001/archneur.56.3.303.
[36] Jak AJ, Urban S, McCauley A, Bangen KJ, Delano-Wood L, Corey-Bloom J, et al. Profile of hippocampal volumes and stroke risk varies by neuropsychological definition of mild cognitive impairment. J Int Neuropsychol Soc 2009;15:890–7. https://doi.org/10.1017/S1355617709090638.
[37] Gemmill A, Weiss J. The relationship between fertility history and incident dementia in the US Health and Retirement Study. The Journals of Gerontology: Series B 2021. https://doi.org/10.1093/geronb/gbab183.
[38] Hong CH, Falvey C, Harris TB, Simonsick EM, Satterfield S, Ferrucci L, et al. Anemia and risk of dementia in older adults: findings from the Health ABC study. Neurology 2013;81:528–33. https://doi.org/10.1212/WNL.0b013e31829e701d.
[39] Kobayashi LC, Gross AL, Gibbons LE, Tommet D, Sanders RE, Choi S-E, et al. You say tomato, I say radish: can brief cognitive assessments in the US Health Retirement Study be harmonized with its International Partner Studies? The Journals of Gerontology: Series B 2020. https://doi.org/10.1093/geronb/gbaa205.

Figure Captions:

Figure 1. Flow chart illustrating the analytic process used throughout the study. We first used Confirmatory Factor Analysis (CFA) to estimate cognitive domain scores in each sample (Step 1). We then used the demographically-corrected norms to define cognitive impairment in each sample using an actuarial neuropsychological approach (Steps 2-4). This process of estimating cognitive domain scores and classifying cognitive impairment was repeated 64 times (the total number of items), leaving out data on each item of interest in turn to prevent circularity in inferences from these analyses (Step 5). The leave one out impairment status for each item was then used in logistic regression analyses to assess associations between cognitive items and cognitive impairment (Step 6). Boxes with rounded edges illustrate data or estimates, whereas boxes with hard edges illustrate analytical steps. Numbers included in the boxes show the order of steps.

Figure 2. Associations between each cognitive test item and cognitive impairment by domain for each Harmonized Cognitive Assessment Protocol Studies (HCAP) conducted in the US (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011) from logistic regression models, controlling for age and sex. Odds ratios are displayed for significant associations. For example, the number 0.14 in the top left hand corner indicates that in the US those who answered the question on the day of the week correctly had an odds of cognitive impairment that was 0.14 times the odds of cognitive impairment for those who did not answer this question correctly. Grey boxes represent instances were an item was not administered or an odds ratio was suppressed due to small cells. Color scale shows differences in associations on the log odds scale.

Figure 3. Distributions of estimated odds ratios describing the association between items on cognition and cognitive impairment across Harmonized Cognitive Assessment Protocol Studies (HCAP) conducted in the US (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011)

from logistic regression models, controlling for age and sex. Odds ratios further from 1 represent stronger associations with cognitive impairment. There is a larger left tail in the distributions for the US, England, and to a smaller extent, Mexico, indicating the presence of some items with stronger associations in these countries as compared to other settings.

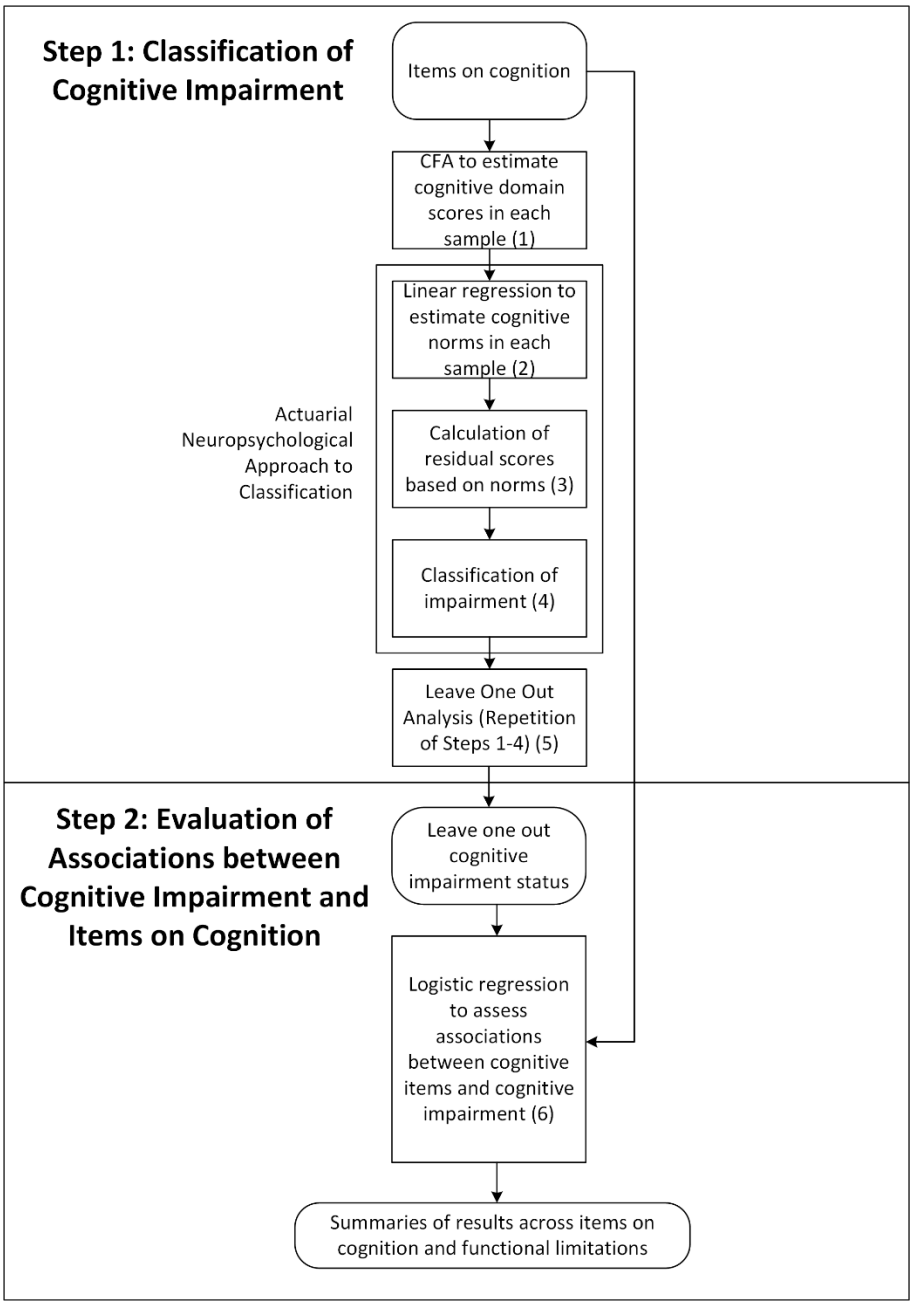

Figure 1. Flow chart illustrating the analytic process used throughout the study. We first used Confirmatory Factor Analysis (CFA) to estimate cognitive domain scores in each sample (Step 1). We then used the demographically-corrected norms to define cognitive impairment in each sample using an actuarial neuropsychological approach (Steps 2-4). This process of estimating cognitive domain scores and classifying cognitive impairment was repeated 64 times (the total number of items), leaving out data on each item of interest in turn to prevent circularity in inferences from these analyses (Step 5). The leave one out impairment status for each item was then used in logistic regression analyses to assess associations between cognitive items and cognitive impairment (Step 6). Boxes with rounded edges illustrate data or estimates, whereas boxes with hard edges illustrate analytical steps. Numbers included in the boxes show the order of steps.

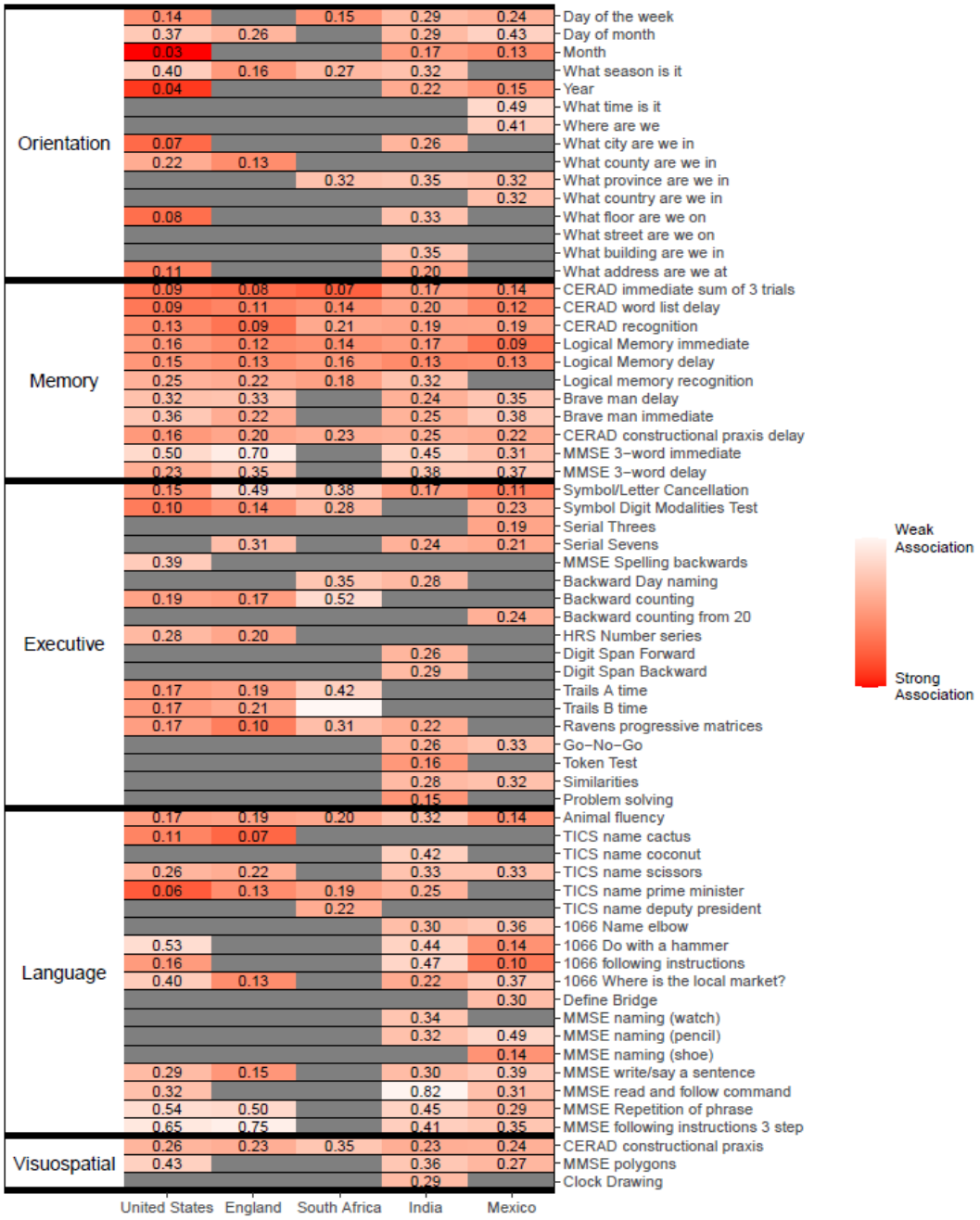

Figure 2. Associations between each cognitive test item and cognitive impairment by domain for each Harmonized Cognitive Assessment Protocol Studies (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011) from logistic regression models, controlling for age and sex. Odds ratios are displayed for significant associations. For example, the number 0.14 in the top left hand corner indicates that in the United States those who answered the question on the day of the week correctly had an odds of cognitive impairment that was 0.14 times the odds of cognitive impairment for those who did not answer this question correctly. Grey boxes represent instances were an item was not administered or an odds ratio was suppressed due to small cells. Color scale shows differences in associations on the log odds scale.

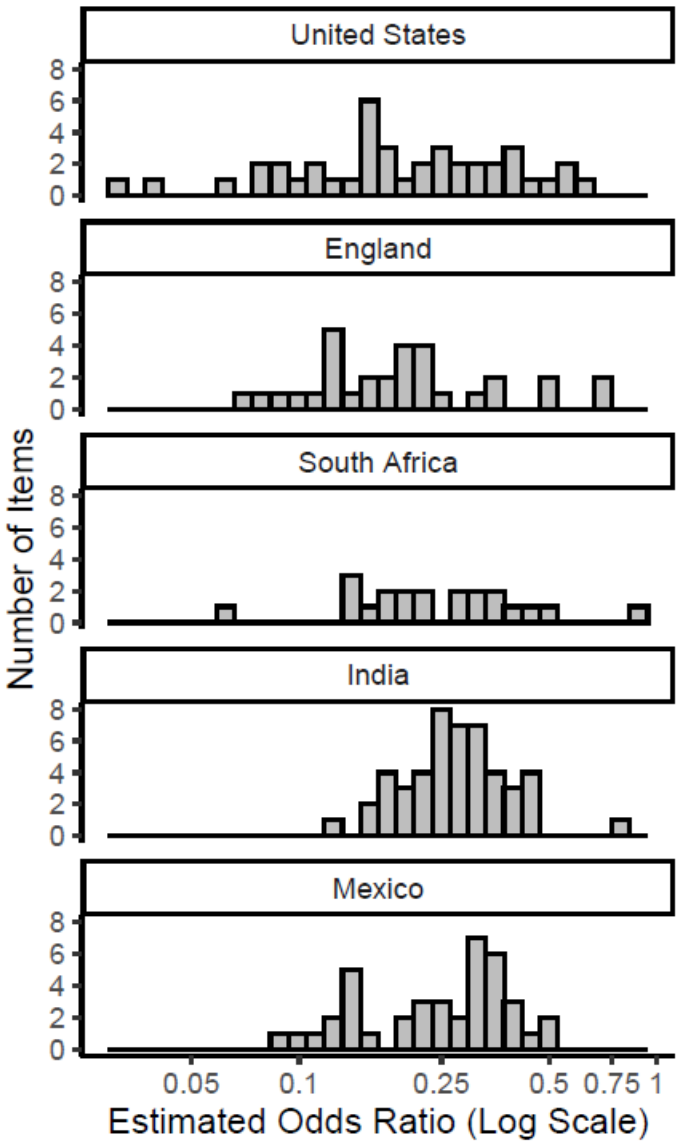

Figure 3. Distributions of estimated odds ratios describing the association between items on cognition and cognitive impairment across Harmonized Cognitive Assessment Protocol Studies (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011) from logistic regression models, controlling for age and sex. Odds ratios further from 1 represent stronger associations with cognitive impairment. There is a larger left tail in the distributions for the US, England, and to a smaller extent, Mexico, indicating the presence of items that have stronger associations with cognitive impairment in these countries as compared to other settings.

Table 1. Cognitive items administered by cognitive domain in each of the US, England, South Africa, India, and Mexico Harmonized Cognitive Assessment Protocol (HCAP) samples

Cognitive Item	US	England	South Africa	India	Mexico
Orientation					
Day Of The Week	X	X	X	X	X
Day Of Month	X	X		X	X
Month	X	X		X	X
Season	X	X	X	X	
Year	X	X		X	X
What Time Is It					X
Where Are We					X
What City Are We In	X	X	X	X	
What County Are We In	X	X			
What Province Are We In	X		X	X	X
What Country Are We In		X			X
What Floor Are We On	X			X	
What Street Are We On		X			
What Building Are We In		X		X	
What Address Are We At	X			X	
Memory					
CERAD Immediate Sum Of 3 Trials	X	X	X	X	X
CERAD Word List Delay	X	X	X	X	X
CERAD Recognition	X	X	X	X	X
Logical Memory Immediate	X	X	X	X	X
Logical Memory Delay	X	X	X	X	X
Logical Memory Recognition	X	X	X	X	
Brave Man Delay	X	X		X	X
Brave Man Immediate	X	X		X	X
CERAD Constructional Praxis Delay	X	X	X	X	X
MMSE 3-Word Immediate	X	X		X	X
MMSE 3-Word Delay	X	X		X	X
Executive Functioning					
Symbol/Letter Cancellation	X	X	X	X	X
Symbol Digit Modalities Test	X	X	X		X
Serial Threes					X
Serial Sevens		X		X	X
MMSE Spelling Backwards	X				
Backward Day Naming			X	X	
Backward Counting	X	X	X		
Backward Counting From 20					X
HRS Number Series	X	X			
Digit Span Forward				X	

Digit Span Backward				X	
Trails A Time	X	X	X		
Trails B Time	X	X	X		
Ravens Progressive Matrices	X	X	X	X	
Go-No-Go				X	X
Token Test				X	
Similarities				X	X
Problem Solving				X	
Language					
Animal Fluency	X	X	X	X	X
TICS Name Cactus	X	X			
TICS Name Coconut				X	
TICS Name Scissors	X	X	X	X	X
TICS Name Prime Minister	X	X	X	X	
TICS Name Deputy President			X		
CSI-D Name Elbow	X	X	X	X	X
CSI-D Do With A Hammer	X	X	X	X	X
CSI-D Following Instructions	X	X	X	X	X
CSI-D Where Is The Local Market?	X	X	X	X	X
Define Bridge					X
MMSE Naming (Watch)	X	X		X	
MMSE Naming (Pencil)	X	X		X	X
MMSE Naming (Shoe)					X
MMSE Write/Say A Sentence	X	X		X	X
MMSE Read And Follow Command	X	X		X	X
MMSE Repetition Of Phrase	X	X		X	X
MMSE Following Instructions 3 Step (Paper)	X	X		X	X
Visuospatial Functioning					
CERAD Constructional Praxis (Copy 4 Figures)	X	X	X	X	X
MMSE Polygons (Copy 1 Figure)	X			X	X
Clock Drawing				X	

* CERAD = Consortium to Establish a Registry for Alzheimer's Disease, MMSE = Mini-Mental State Examination, HRS = Health and Retirement Study, TICS = Telephone Interview for Cognitive Status, CSI-D = Community Screening Instrument for Dementia

Table 2. Characteristics of the US, England, South Africa, India, and Mexico Harmonized Cognitive Assessment Protocol (HCAP) samples

	US	England	South Africa	India	Mexico
Number of Participants (N)	3329	1255	560	4095	2011
Years of Data Collection	2016-2017	2018	2016-2017	2017-2019	2015
Age (Mean [SD])	75.8 (7.5)	75.9 (7.1)	69.2 (11.1)	69.0 (7.6)	68.1 (9.0)
Percent Female (N)	60.5% (2014)	54.9% (689)	56.2% (315)	53.9% (2207)	59.3% (1193)
No education - primary education (% [N])	18.2% (607)	33.1% (416)	92.7% (519)	75.3% (3085)	72.9% (1467)
Some secondary - completed secondary education (% [N])	53.0% (1766)	53.9% (676)	5.4% (30)	20.6% (845)	20.8% (419)
Post-secondary education (% [N])	28.7% (956)	13.0% (163)	2.0% (11)	4.0% (165)	6.2% (125)
White race (% [N])	78.9% (2627)				
Black race (% [N])	16.0% (533)				
Other race (% [N])	5.1% (169)				
Percent Hispanic (N)	10.8% (360)				
Percent Rural (N)				62.0% (2539)	28.3% (569)
Percent Illiterate (N)			58.6% (328)	56.6% (2319)	