Appendix A

Table of Contents

Survey weights

(Reprinted from "Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research" by Nichols et al. *Alzheimer's and dementia* 2022)

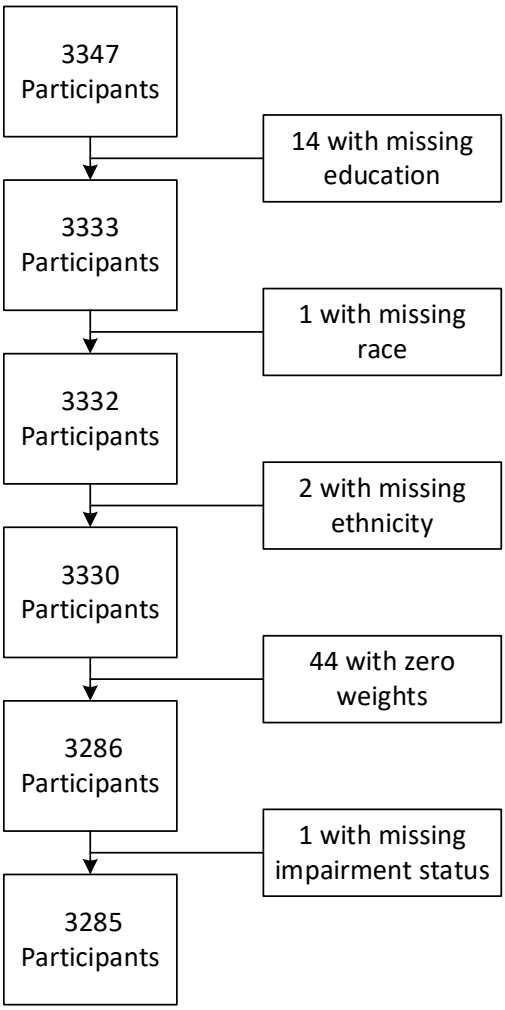
Survey weights were used in analyses to ensure that results are generalizable to broader populations, and therefore can better inform future survey design. In the United States and Mexico, study participants were randomly sampled from the prior HRS IPS studies. Therefore, in this analysis we used the survey weights provided in these broader HRS IPS studies. In the United States, additional data was provided on individuals who were selected to be in the HCAP sample but declined participation. We accounted for potential selection bias by calculating stabilized inverse probability of selection weights and multiplying these weights by the provided survey weights. We used a logistic regression models with predictor variables for gender and 5-year age group to predict selection. The England, India, and South Africa HCAP samples over-selected individuals with low cognition, and provided survey weights to allow for the generalization of results to the broader samples. For England, and India, survey results enable generalization to nationally-representative samples. Survey weights for South Africa enable generalization to the population of the Agincourt sub-district of the Mpumalanga Province in South Africa.

Sample exclusions

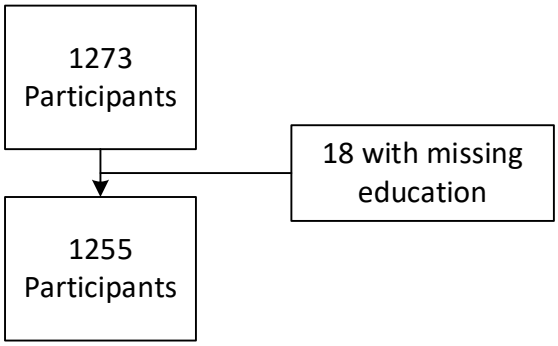
(Reprinted from "Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research" by Nichols et al. *Alzheimer's and dementia* 2022)

For each of the HCAP samples, we included individuals with valid responses on the variables required to ascertain cognitive impairment status. For each of the samples, we required non-missing values for age, sex, and education. Additionally, we required non-missing values for race and ethnicity in the United States, rurality in Mexico and India, and illiteracy status in India. In the United States and Mexico, a few participants had estimated survey weights of 0 for the HRS IPS survey wave from which HCAP participants were selected. As we used these survey weights to ensure results would be applicable to the general population, these individuals were excluded from our analyses. Finally, we excluded all individuals with missingness on cognitive impairment status due to high missingness in cognitive items and low reliability of scores across all cognitive domains. The flow charts for sample exclusions for each one of the HCAP samples are shown below.

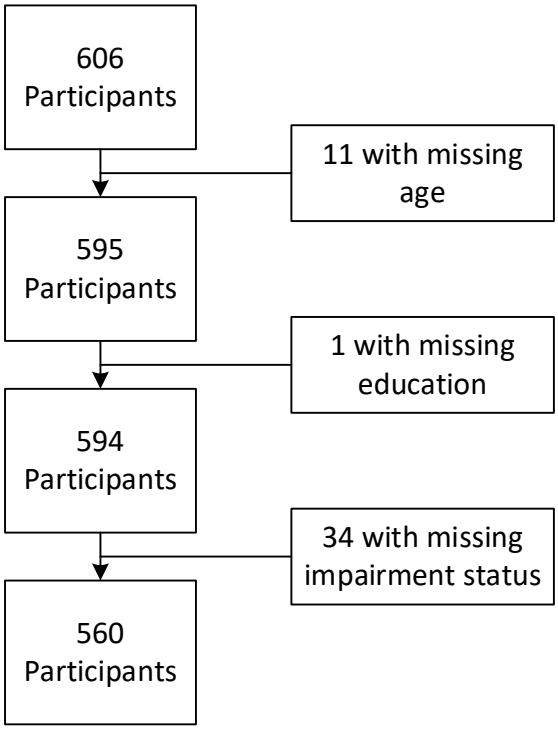
United States



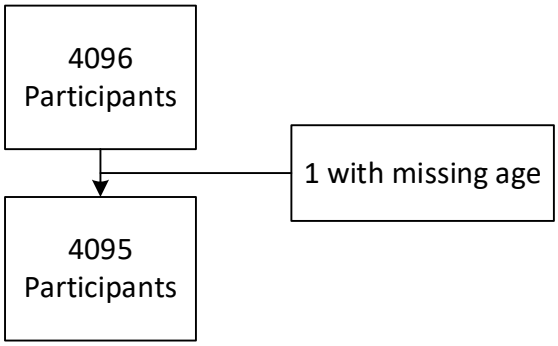
England



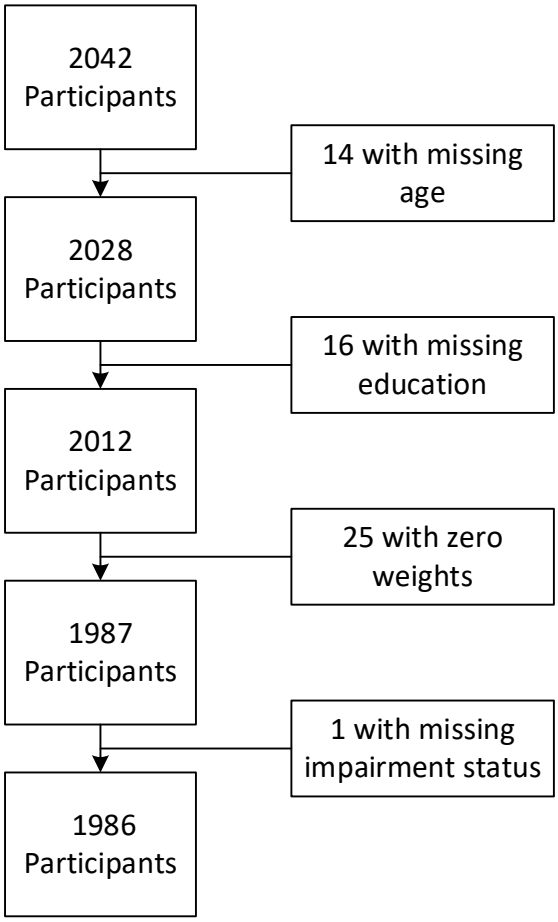
South Africa



India



Mexico



Cognitive test items included in classification of cognitive impairment

(Adapted from "Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research" by Nichols et al. *Alzheimer's and dementia* 2022)

Table S1. 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

* 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

The estimation of cognitive domains using confirmatory factor analysis

(Reprinted from "Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research" by Nichols et al. *Alzheimer's and dementia* 2022)

All models were scaled such that the scores on the latent cognitive domain estimated would have a mean of 0 and standard deviation of 1 within each specific HCAP study. To ascertain fit of estimated models to the data, we initially fit CFA models using a Weighted Least Squares estimator and evaluated model fit using the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI) and the Standardized Root Mean Residual (SRMR) [2]. Model fit was considered excellent if the RMSEA was <0.05, CFI >0.95, TLI >0.95, and SRMR <0.05. Model fit was considered poor if the RMSEA >0.1, CFI <0.9, TLI <0.9, and SRMR >0.1. Where model fit was poor, we added additional methods factors to the model structure to explain higher correlations between similar items or items from the same cognitive test (i.e. immediate and delayed recall) [3]. These methods factors were selected by examining evidence of model misfit in conjunction with a priori knowledge of the content of items.

The following methods factors were used:

United States

Memory:

- Methods factor 1: Logical memory immediate, logical memory delayed, logical memory recognition
- Methods factor 2: Brave man immediate, brave man delayed
- Methods factor 3: CERAD word list immediate, CERAD word list delayed, CERAD word list recognition

Language:

- Methods factor 1: Naming a watch, naming a pencil

England

Memory:

- Methods factor 1: Logical memory immediate, logical memory delayed
- Methods factor 2: Brave man immediate, brave man delayed

Executive functioning:

- Methods factor 1: Trail-making test part A, trail-making test part B

Language:

- Methods factor 1: Naming a cactus, naming scissors

South Africa

Memory:

- Methods factor 1: CERAD word list immediate, CERAD word list delayed
- Methods factor 2: Logical memory immediate, logical memory delayed

India

Memory:

- Methods factor 1: Logical memory immediate, logical memory delayed, logical memory recognition
- Methods factor 2: Brave man immediate, brave man delayed
- Methods factor 3: CERAD word list immediate, CERAD word list delayed, CERAD word list recognition

Language:

- Methods factor 1: Naming scissors, naming a coconut

Mexico

Memory:

- Methods factor 1: Brave man immediate, brave man delayed
- Methods factor 2: MMSE 3-word immediate, MMSE 3-word delayed
- Methods factor 3: Logical memory immediate, logical memory delayed

Executive functioning:

- Methods factor 1: Serial 7's, serial 3's

Language:

- Methods factor 1: Naming a pencil, naming a shoe

To estimate factor scores used in the classification of cognitive impairment using the actuarial neuropsychological approach, a second set of CFA models was fit using the Maximum Likelihood (MLR) estimator to generate estimates of cognitive impairment for all study participants. CFA models estimated with an MLR estimator make an assumption that data are missing at random (MAR) and allow for the estimation of cognitive ability even with large amounts of missing data by relying on information from each of the other non-missing cognitive items included in the model. However, the reliability of the estimation of cognition in individuals with large amounts of missing data may be poor [4]. To prevent the scores estimated with low reliability from having an outsized influence on the estimation of cognitive impairment, we estimated reliability using the formula: $Reliability = 1 - standard\ error^2$. Reliability as measured using this formula is akin to a measure of internal consistency and reflects the precision of estimated cognitive scores [5]. We set all scores with a reliability of under 0.6 and greater than 50% missingness on cognitive items to be missing [6].

Final CFA models by study and cognitive domain generally showed adequate to good fit (Table S2).

Table S2. Fit statistics for confirmatory factor analysis models of cognitive domains in the United States, England, South Africa, India, and Mexico Harmonized Cognitive Assessment Protocol (HCAP) samples

Sample	Domain	Parameters	CFI	TLI	RMSEA	SRMR
United States	Orientation	20	0.971	0.963	0.028 (0.023-0.034)	0.064
England	Orientation	20	0.997	0.997	0.015 (0.000-0.027)	0.064
South Africa	Orientation	8	0.989	0.967	0.042 (0.000-0.100)	0.089
India	Orientation	20	0.954	0.941	0.039 (0.034-0.043)	0.089
Mexico	Orientation	16	0.924	0.894	0.062 (0.053-0.070)	0.066
United States	Memory	53	0.986	0.977	0.040 (0.035-0.046)	0.019
England	Memory	46	0.954	0.938	0.075 (0.067-0.082)	0.044
South Africa	Memory	24	0.978	0.959	0.084 (0.063-0.106)	0.036
India	Memory	53	0.982	0.971	0.044 (0.040-0.049)	0.023
Mexico	Memory	46	0.986	0.979	0.045 (0.038-0.052)	0.028
United States	Executive Functioning	27	0.927	0.898	0.110 (0.104-0.117)	0.036
England	Executive Functioning	28	0.886	0.832	0.140 (0.130-0.151)	0.050
South Africa	Executive Functioning	23	0.914	0.857	0.086 (0.063-0.110)	0.062
India	Executive Functioning	41	0.989	0.986	0.033 (0.028-0.037)	0.023
Mexico	Executive Functioning	31	0.971	0.953	0.086 (0.076-0.097)	0.034
United States	Language	32	0.973	0.967	0.020 (0.016-0.024)	0.067
England	Language	28	0.997	0.997	0.007 (0.000-0.019)	0.070
South Africa	Language	17	0.975	0.965	0.030 (0.000-0.050)	0.120
India	Language	32	0.922	0.906	0.032 (0.029-0.035)	0.060
Mexico	Language	30	0.975	0.970	0.021 (0.016-0.027)	0.076

* CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Residual

There was some evidence of poor fit in models for executive functioning in the United States and England, likely due to the lower inter-correlations between the more diverse set of items measuring executive functioning. The reliability, or precision of the measurement, of the cognitive scores was generally fair to good, although lower reliabilities were observed for the orientation domain, as this domain was composed solely of binary cognitive items which provide less information than continuous items, particularly in the less impaired range of cognitive performance (Figure 2). Outside of the orientation domain, only language scores in England (41%) and India (47%) had greater than 25% of scores with reliabilities under 0.6. Many CFA models had scores with high reliabilities, and models of executive functioning in the United States, England, India, and Mexico, memory in the United States, England, Mexico, and South Africa all had scores with mean reliabilities of over 0.8.

Definition of the cognitively robust group for the neuropsychological norms classification approach

(Reprinted from "Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research" by Nichols et al. *Alzheimer's and dementia* 2022)

We used questions from the CSID and the 10/66 assessment of functional limitations to select individuals who had no reports of: (1) changes in daily activities, (2) general decline, (3) difficulty remembering, (4) changes in the ability to handle money, (5) forgetting friends or family's names, (6) using the wrong words, (7) forgetting where they are, when they last saw the informant, or what happened yesterday, (8) getting lost in the community or at home, (9) changes in ability to think or reason things through. We also excluded individuals with the highest 10% of depressive symptom burden in each sample, and individuals who either used a proxy informant or had a self-reported stroke or heart attack in the prior HRS IPS wave. Additionally, we excluded individuals with informant-reported stroke, Alzheimer's disease or memory problems in all samples except Mexico, as these data were not available in the Mexican HCAP survey. Due to low endorsement of limitations in the South African sample as compared to other HCAP samples, we further excluded individuals with fair or poor self-reported health in the South African sample. While individuals with missing data on individual items used to create the normative sample were considered to be "not impaired" on these individual items, we additionally excluded individuals with greater than 50% missing data on the items used for the selection of the normative group from the normative group.

Definition of cognitive impairment using the neuropsychological norms approach

(Adapted from "Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research" by Nichols et al. *Alzheimer's and dementia* 2022)

We estimated multiple linear regression models within the normative sample of each HCAP study to quantify the relationship between cognitive functioning on each domain, adjusting for basic demographic factors. For all studies we included age, gender, and educational attainment (dichotomized). We included this crude marker of education to ensure that we accounted for differences in educational attainment that would confer advantages or disadvantages on the cognitive testing used to measure cognitive functioning (bias), but that we did not control for variation in education that is expected to be associated with true variability in cognitive functioning [28]. We additionally included race and ethnicity in regression model for the US HCAP study, rurality in the regression model for the India and Mexico studies, and literacy status in the regression models for the India and South Africa studies.

Using the coefficients from these models we calculated expected cognitive performance on each domain for each respondent in the full sample by predicting scores based on participants' demographic characteristics. We then used these predictions to calculate residual scores, scaled by the standard error of the regression equation, using the formula:

$$Residual\ Score = \frac{True\ Cognitive\ Score - Predicted\ Cognitive\ Score}{SE\ of\ the\ Regression\ Equation}.$$

This ensures that the variability in the residual scores (which is used to determine cutoffs of cognitive impairment) is proportional to the variability that remained unexplained by the regression equation, and is therefore not attributable to demographic factors.

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. Individuals with missing scores on all cognitive domains were excluded (N=36 across all studies).

Percent of data missing for items on functional impairment

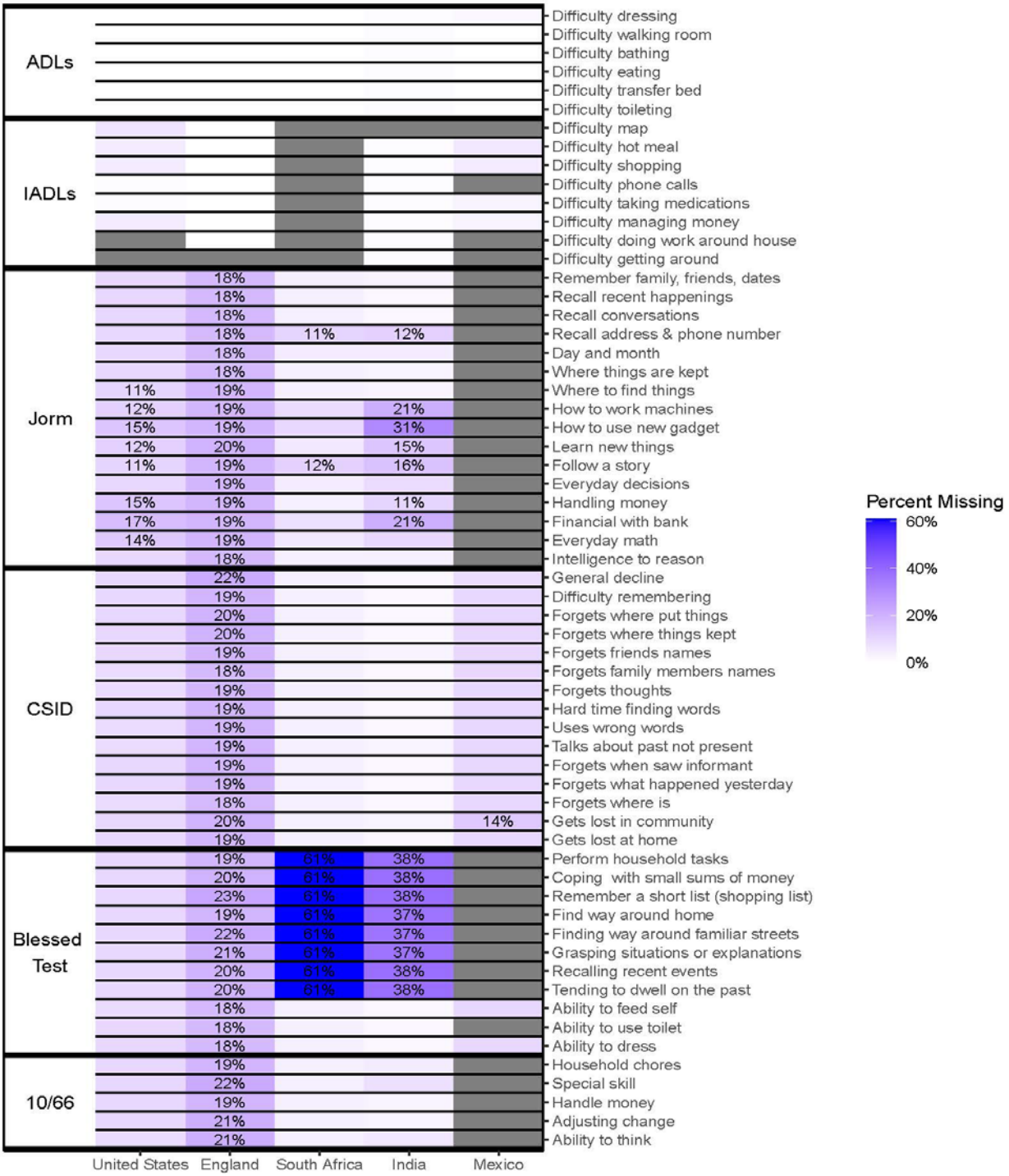


Figure S1. Percentage of data missing on each item on functional limitations across the Harmonized Cognitive Assessment Protocol Surveys (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011). Numbers are shown when the percent missingness is higher than 10%, grey boxes are shown when the item was not administered in a study. ADL = Activity of Daily Living, IADL = Instrumental Activity of Daily Living, CSID = Community Screening Instrument for Dementia.

There was patterned missingness for some of the items of functional limitations. For the England HCAP study, only 82.5% of participants who completed the cognitive testing had informant interviews. For the India HCAP study, some items on the Blessed test were only administered to participants below a cutoff score based on the Jorm IQCODE test during Phase 1 of data collection efforts. The same was true for data collection in the South African HCAP study. A number of items on the Jorm IQCODE test also had higher levels of missingness due to informants who were unaware or did not know about respondent decline or improvement on specific activities. For example, the item on learning how to use a new gadget had 31% missingness in India, 15% missingness in the United States, and 10% missingness in South Africa. The item on dealing with financial matters at the bank also had 21% missingness in India, 17% missingness in the United States, and 8% missingness in South Africa; missingness fell along expected gendered patterns.

Items on functional limitations with suppressed odds ratios

Figure S2. Items on functional limitations with suppressed odds ratios due to low variability in responses across the Harmonized Cognitive Assessment Protocol Surveys (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011). Grey boxes indicate items that either weren't assessed. ADL = Activity of Daily Living, IADL = Instrumental Activity of Daily Living, CSID = Community Screening Instrument for Dementia.

Variability (proportion endorsed limitations) for binary items on functional limitations

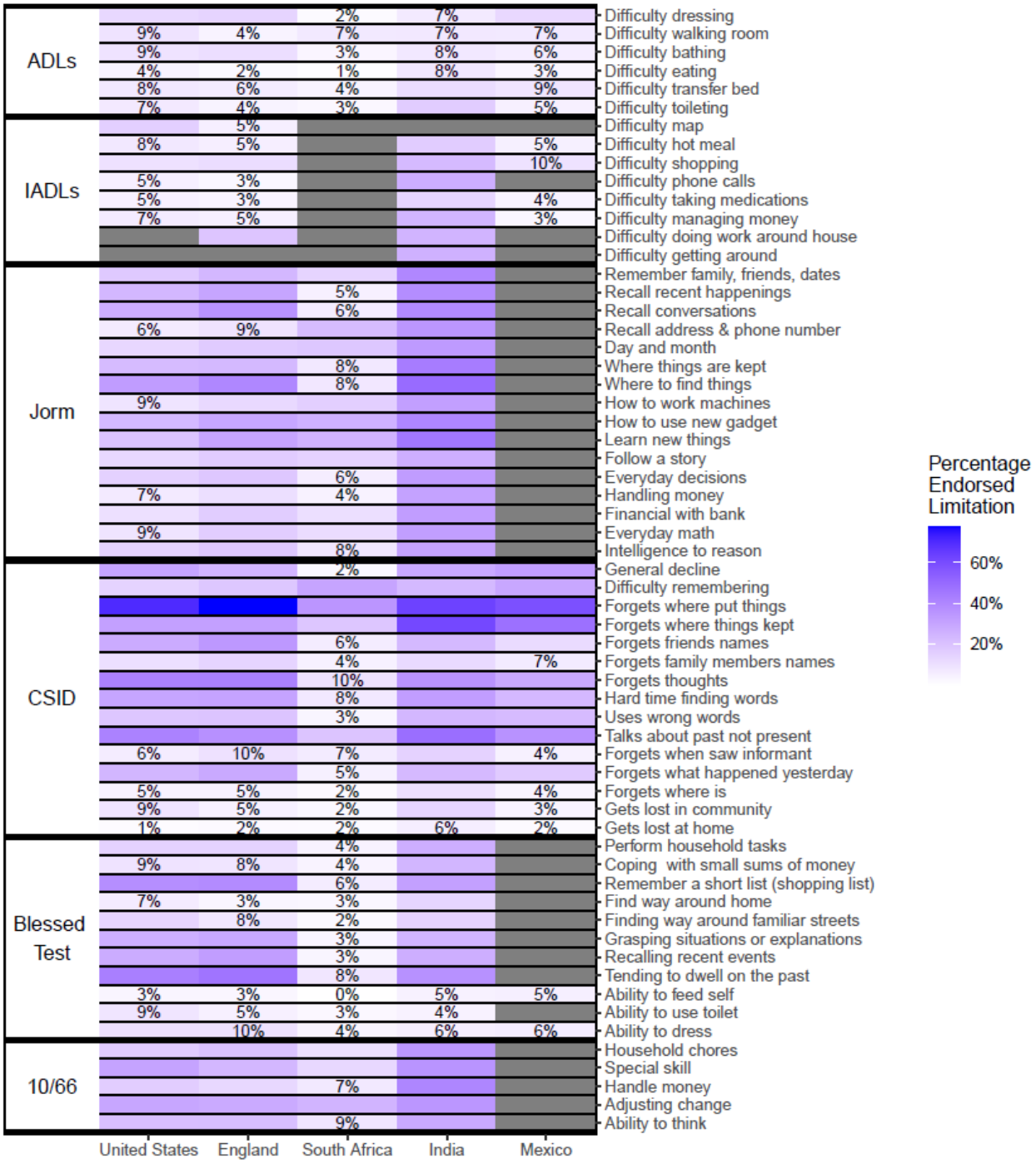


Figure S3. Variability (proportion correct) for binary items on functional limitations across the Harmonized Cognitive Assessment Protocol Surveys (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011). Grey boxes indicate items that either weren't assessed. ADL = Activity of Daily Living, IADL = Instrumental Activity of Daily Living, CSID = Community Screening Instrument for Dementia.

Sensitivity analysis for items on Blessed test

We conducted a sensitivity analysis for items on the Blessed test that included the reason for limitations. In primary analyses, responses that specified that limitations were due to physical restrictions were treated as having no limitations. In this sensitivity analysis, we treated such responses as missing and found that results were very similar.

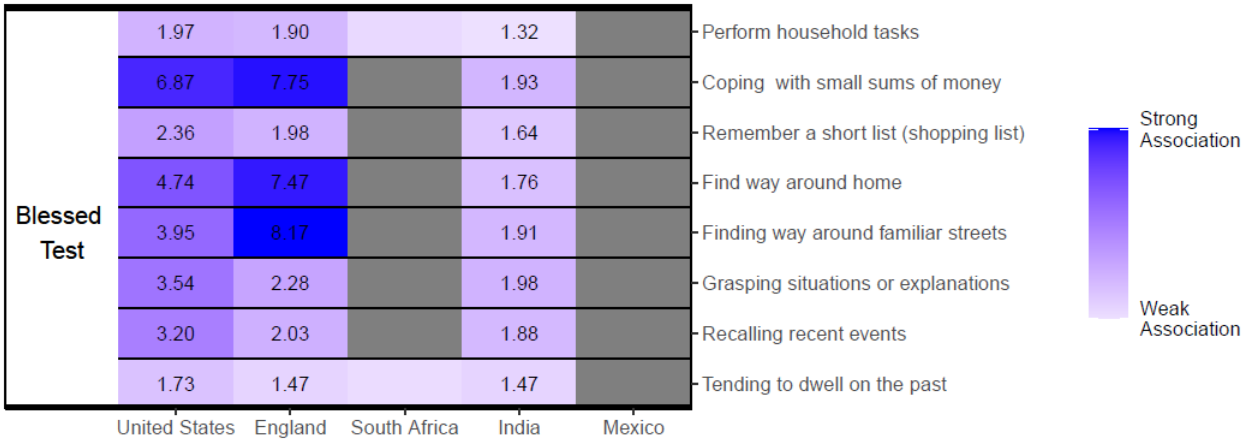


Figure S4. Associations between each item from the Blessed test that included reasons for reported difficulties and cognitive impairment by domain for each Harmonized Cognitive Assessment Protocol Studies (HCAP) conducted in the United States (N = 3,329), England (N = 1,255), South Africa (N = 560), India (N = 4,095), and Mexico (N = 2,011) from logistic regression models, controlling for age and gender. Responses that specified that difficulties were due to physical reasons were set to missing. Odds ratios are displayed for significant associations. 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.

Sensitivity analysis using LCA (items on functional limitations)

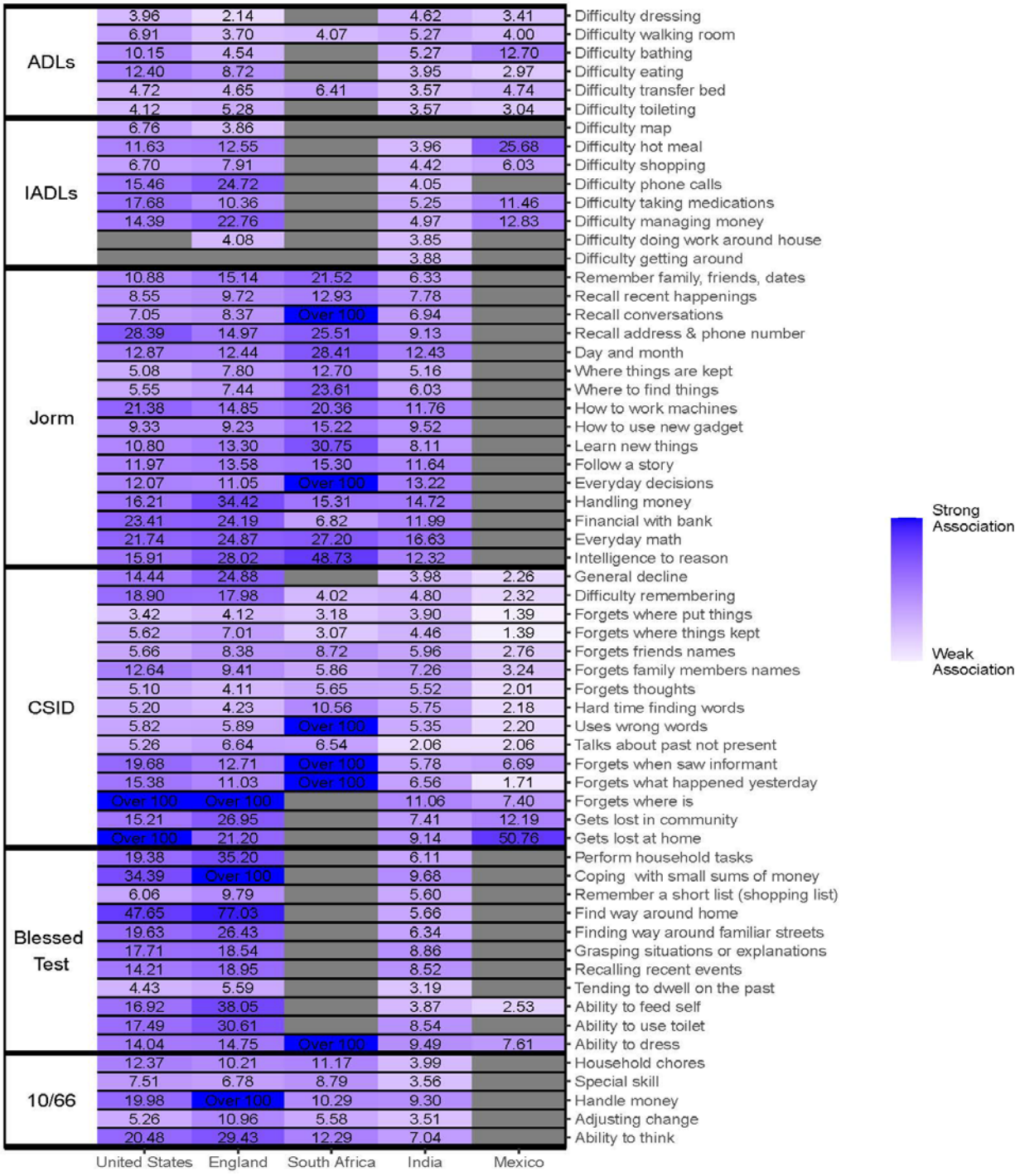


Figure S5. Associations between each item on cognitive functioning and cognitive impairment by domain for participants age 65+ in each Harmonized Cognitive Assessment Protocol Surveys (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011) from latent class analysis models. Numbers indicate odds ratios compared individuals in the impaired class to individuals in the unimpaired class. Grey boxes represent instances were an item was not administered. Color scale represents log odds ratios.

Sensitivity analysis for functional limitations items (restricted to age 65+)

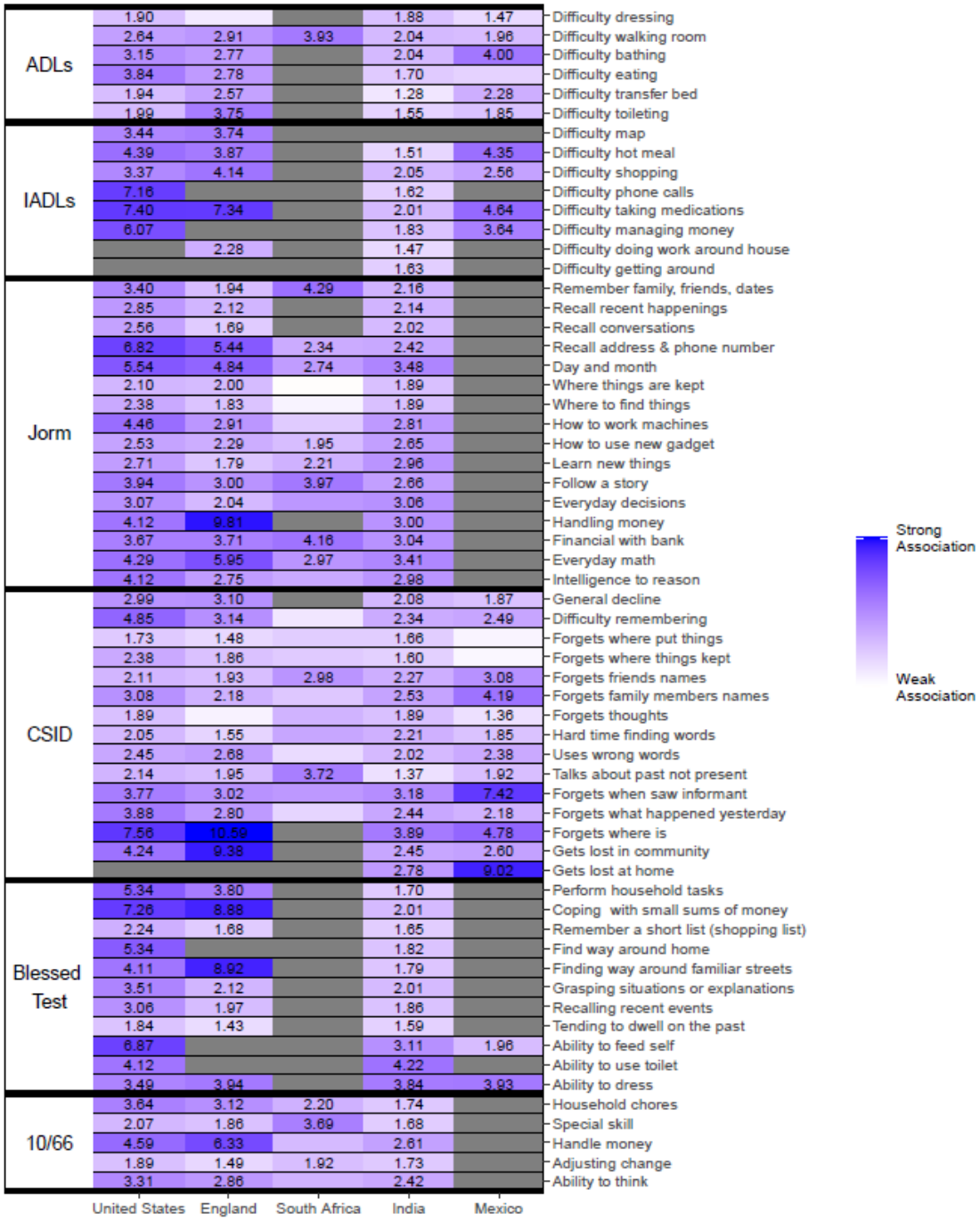


Figure S6. Associations between each item on cognitive functioning and cognitive impairment by domain for participants age 65+ in each Harmonized Cognitive Assessment Protocol Surveys (HCAP) conducted in the United States (N = 3329), England (N = 1255), South Africa (N = 560), India (N = 4095), and Mexico (N = 2011) from logistic regression models, controlling for age and gender. Odds ratios are displayed for significant associations. Grey boxes represent instances were an item was not administered or an odds ratio was suppressed due to small cells. Color scale represents log odds ratios.

Results for the US and England are identical or almost identical to primary analyses, given that all (England) or almost all (US) of participants were over 65 years old. Items with the strongest associations in the US included difficulty with phone calls, taking medications, and managing money (IADLs), as well as difficulty coping with small sums of money and feeding oneself (Blessed). Items with the strongest associations in England included difficulty taking medications (IADLs), and forgetting where one is (CSID), and coping with small sums of money (Blessed). Many of the items with the strongest associations were similar in the US and England.

While there were some changes to results in other settings, the broader patterns and comparisons across countries remained similar. Therefore, study conclusions remained consistent. However, some changes worth highlighting include lower associations across a number of items in South Africa for those over 65 years old, including items on where things are kept and where to find things. There were similar but slightly stronger associations for many of the proxy-reported items in India, but similar and slightly weaker associations for many of the self-reported items in India for those over 65 years old. Patterns of differences were less apparent in Mexico, but some items had differences; for example, the item on forgets when saw informant had an odds ratio of 5.46 in the full sample but 7.42 in those over 65 years old.