# Supplementary Materials

## Table of Contents

# Supplementary Materials 1: List of cognitive tests

Orientation:

- Orientation to time: day of month, month, year, day of week, season
- Orientation to place: state, city, floor of building, area of town/street name, hospital name

Memory:

- Immediate & delayed word recall, word recognition (10 words)
- Immediate & delayed word recall (3 words)
- Immediate & delayed logical memory, logical memory recognition (Robbery story)
- Immediate & delayed Brave man story recall
- Constructional praxis delayed recall

Executive functioning:

- Problem solving
- Raven's progressive matrices
- Similarities & differences (judgement)
- Token test
- Digit span forward and backward
- Go-no-go test
- Symbol cancellation test
- Serial 7s
- Backward day naming

Language/fluency:

- Animal fluency
- Name coconut
- Name scissors
- Name watch
- Name pencil
- Name elbow
- What does one do with a hammer
- Point at window/door
- Write/say a sentence
- Read and follow a command/follow example
- Phrase repetition
- Where is the local market?
- Follow 3-stage instruction
- Name prime minister

Visuospatial:

- Interlocking pentagons
- Constructional praxis

Underlined items are items that we have timing data for. Items that are not underlined could did not have accompanying timing data due to the design of the CAPI instrument.

# Supplementary Materials 2: Outlier identification

The items varied in the skewness of the observed timing data with a bimodal distribution of variable skewness across tests. Therefore, we defined extreme as the 99th percentile of the data distribution for items with low skew, and the 98th percentile of the data distribution for items with high skew.

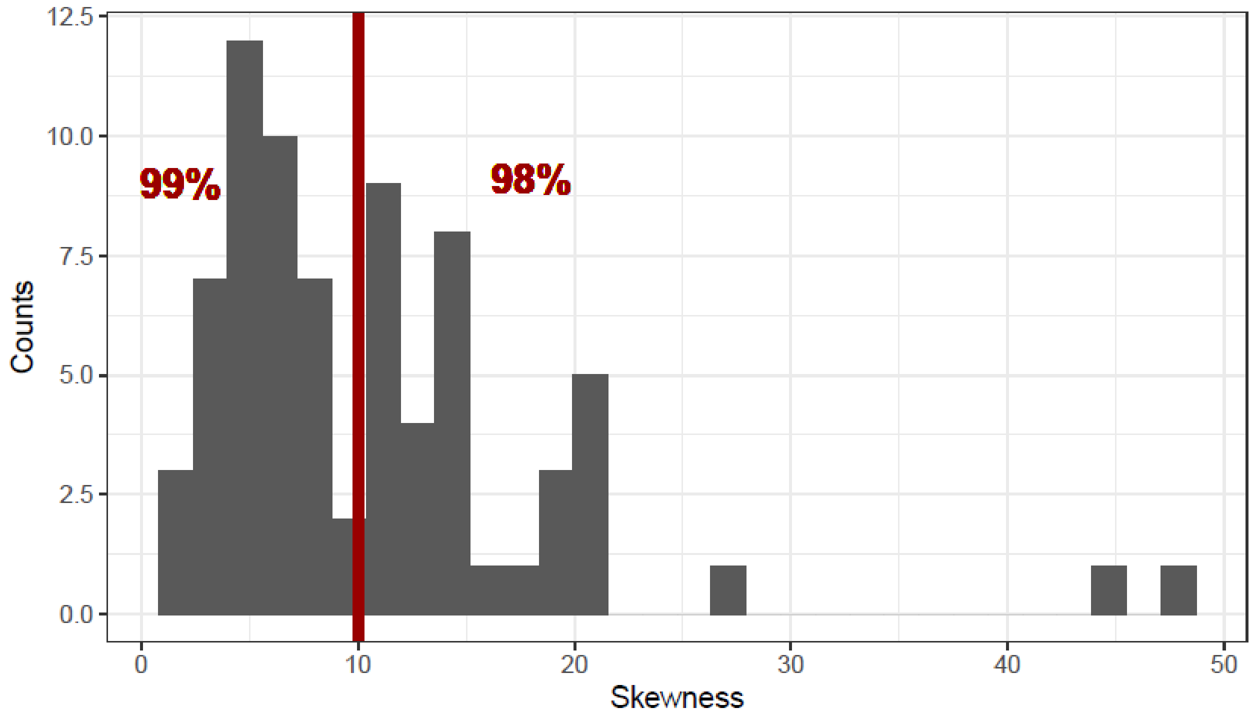
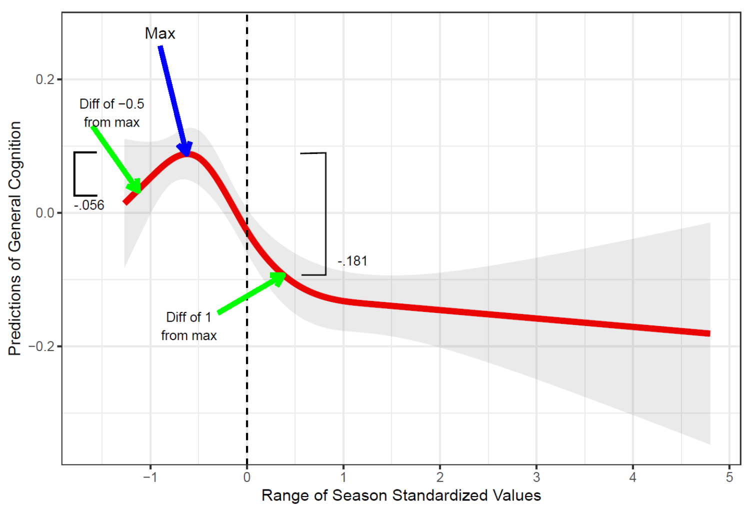



**Figure S1.** Distribution of the skewness of timing information across items and the threshold (skewness=10) used to determine whether data beyond the 99$^{th}$ or 98$^{th}$ percentiles were identified as outliers.

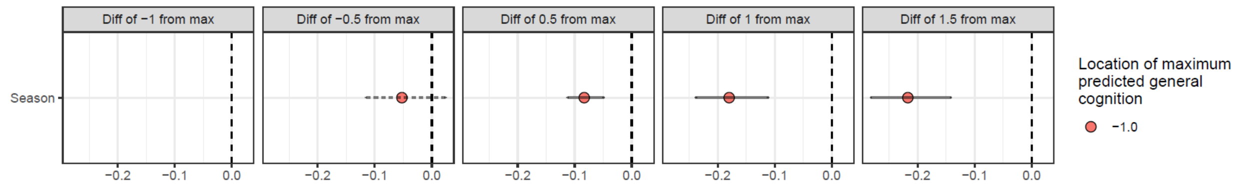
# Supplementary Materials 3: Comparison across different points on continuous splines

Primary analytic models estimate associations using quintiles of time spent on cognitive tests, which simplifies analysis and interpretation of data, but potentially obscures important information contained in the continuous data on timing. To assess the sensitivity of primary findings to the choice to using quintiles of timing as the primary exposure of interest, we estimated linear regression models using cubic splines. We used restricted cubic splines with 3 degrees of freedom and boundary knots at the 5th and 90th percentiles, which constrain the tail segments to be linear to prevent potential outliers from having undue influence. We standardized all continuous item-specific timing data prior to estimating spline models to facilitate comparisons across items. To evaluate the statistical significance of comparisons between different points on the estimated spline terms (analogous to the comparison of quintiles in primary analyses), we used a non-parametric bootstrap. Figure 3 in the main text shows the estimated splines; however, the visual assessment of spline fits does not yield conclusions about the statistical significance of observed patterns. To assess statistical significance of differences and generate estimates that are comparable to those from primary analyses, it is necessary to compare estimates from various points on the estimated curves.

For example, for the item on naming the month, the below graph shows the continuous spline function, with several characteristics of the spline fit illustrated.




The maximum prediction for general cognitive functioning is at about 0.5 standard deviation less than the average time taken to answer the item. Predicted general cognitive functioning is .108 less at 1 standard deviations from the mean time or 0.5 standard deviations from the maximum when compared to the maximum of the curve, which is not significant. However, predicted general cognition is .225 less at 1 standard deviation after the maximum when compared to the maximum of the curve, which is statistically significant. These differences, as shown visually in the plot, can be depicted as estimates with corresponding 95% confidence intervals, calculated using a non-parametric bootstrap. In addition to the two differences, we can estimate a range of differences based on different standard deviation differences from the curve maximum. The plot below shows the two differences included in the example above, as well as differences for 1 and 1.5 standard deviations above the maximum. The panel for 1 standard deviation below the maximum is blank because 1 standard deviation below the maximum is outside of the estimated range for the example curve.




We estimated these differences using estimated spline fits for each item included in analyses; results are shown below in Figure S6. Generally, we observe that differences between different points on the curve are largely statistically significant.

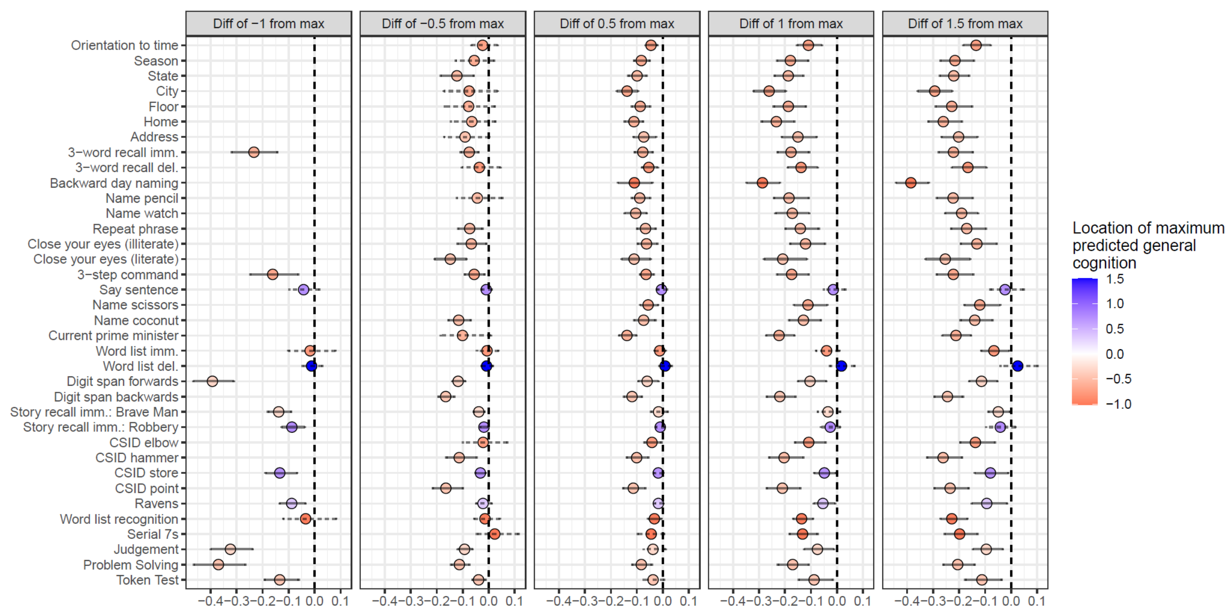



**Figure S2.** Statistical significance of differences in general cognitive functioning for differences of -1, -0.5, 0.5, 1, and 1.5 standard deviations of time spent on cognitive items. All differences are calculated in reference to the amount of time taken and mean estimated cognitive level at the maximum point on estimated spline curve, which represents the highest observed prediction of mean cognitive level. All spline models adjusted for age, gender, and item-specific scores.

# Supplementary Materials 4: Quantification of item difficulty using item response theory methods

*Text has been adapted from Nichols, E., Ng, D. K., James, B. D., Deal, J. A., & Gross, A. L. (2023). The application of cross-sectionally derived dementia algorithms to longitudinal data in risk factor analyses. Annals of Epidemiology, 77, 78-84.*

We used item response theory methods to estimate item difficulty for each item. The item difficulty parameter in item response theory models represents the latent ability at which the probability of answering an item correctly (or scoring one point higher on an ordinal scale) is 50%. The difficulty parameters from within a single model provide a relative metric to anchor comparisons. Binary items have a single threshold, whereas ordinal items with *k* categories have *k-1* thresholds. All continuous items were discretized to 10 categories or fewer using equal interval discretization to enable the estimation of item difficulty [1]. To collapse information across thresholds and quantify a single item difficulty for ordinal or discretized items, we calculated the weighted average of thresholds, weighted by the number of people with scores determined by each threshold (those who scored in either the category directly above or below a given threshold). Item response theory models were estimated in Mplus Version 8. We used the 2-parameter logistic model for binary items and the graded response model for ordinal items [2].


1. Rucker DD, McShane BB, Preacher KJ. A researcher's guide to regression, discretization, and median splits of continuous variables. J. Consum. Psychol. 2015;25(4):666–678.
2. Samejima F. Graded response model. In: Handbook of modern item response theory. Springer; 1997:85–100

# Supplementary Materials 5: Estimated item difficulties

**Table S1.** Item difficulty scores. Test difficulty was estimated using item response theory methods.

| Test item | Difficulty |
|---|---|
| Name pencil | -2.422 |
| City | -1.760 |
| CSID elbow | -1.754 |
| Floor | -1.283 |
| CSID store | -1.267 |
| CSID point | -1.249 |
| Repeat phrase | -1.120 |
| Close your eyes (illiterate) | -1.100 |
| Say sentence | -1.088 |
| Address | -1.040 |
| 3-word recall immediate | -0.951 |
| Name watch | -0.877 |
| Name scissors | -0.848 |
| Season | -0.829 |
| Orientation to time | -0.573 |
| 3-step command | -0.569 |
| Home | -0.500 |
| CSID hammer | -0.187 |
| Word list recognition | -0.130 |
| Backward day naming | 0.040 |
| Current prime minister | 0.112 |
| 3-word recall delayed | 0.178 |
| State | 0.192 |
| Problem Solving | 0.235 |
| Token Test | 0.295 |
| Name coconut | 0.296 |
| Ravens | 0.617 |
| Judgement | 0.650 |
| Story recall imm.: Brave Man | 0.687 |
| Word list total | 0.687 |
| Serial 7s | 0.918 |
| Word list delayed | 0.950 |
| Close your eyes (literate) | 0.982 |
| Story recall imm.: Robbery | 1.231 |
| Digit span backwards | 1.342 |
| Digit span forwards | 1.392 |

## Supplementary Materials 6: Distributions of general cognitive functioning by quintiles of total time taken

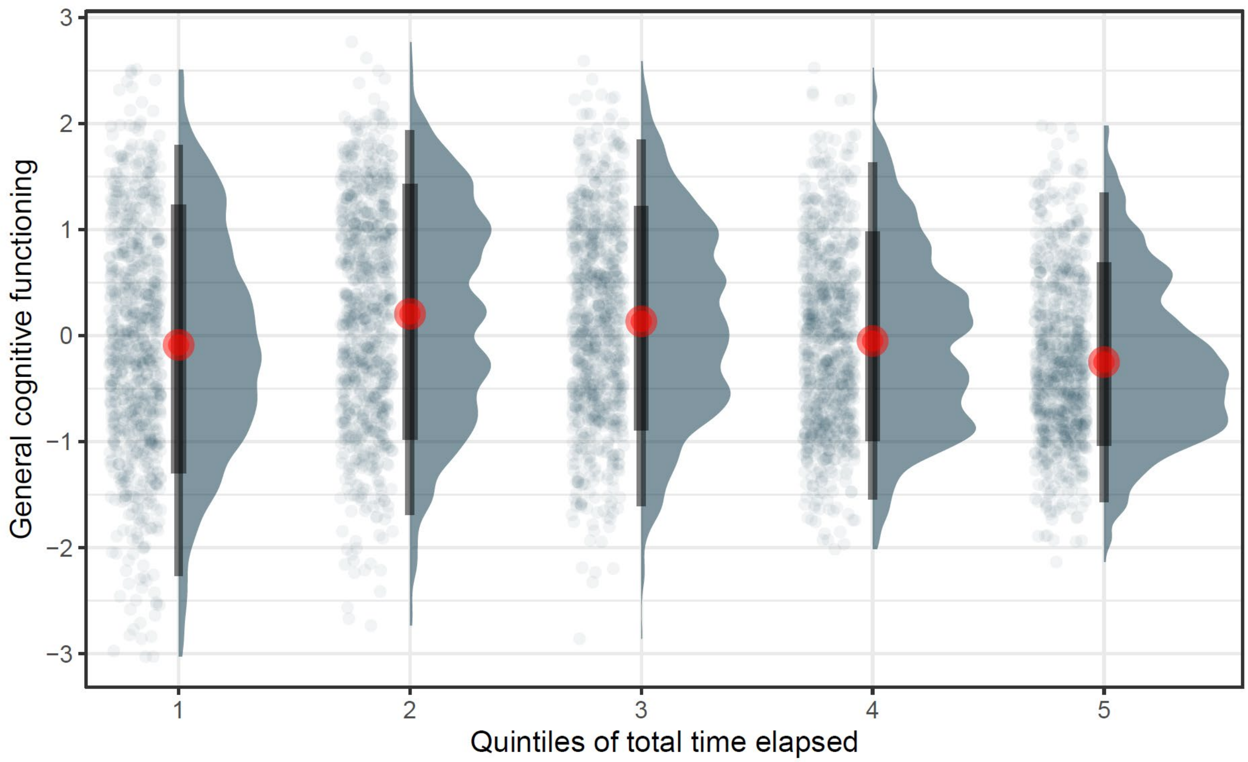



**Figure S3.** Distributions of general cognitive functioning by quintiles of total time taken to complete cognitive tests. Points represent distribution means, and thick and thin lines show the 75% and 95% percentiles of the data distributions.

# Supplementary Materials 7: Item-specific regression models for executive functioning

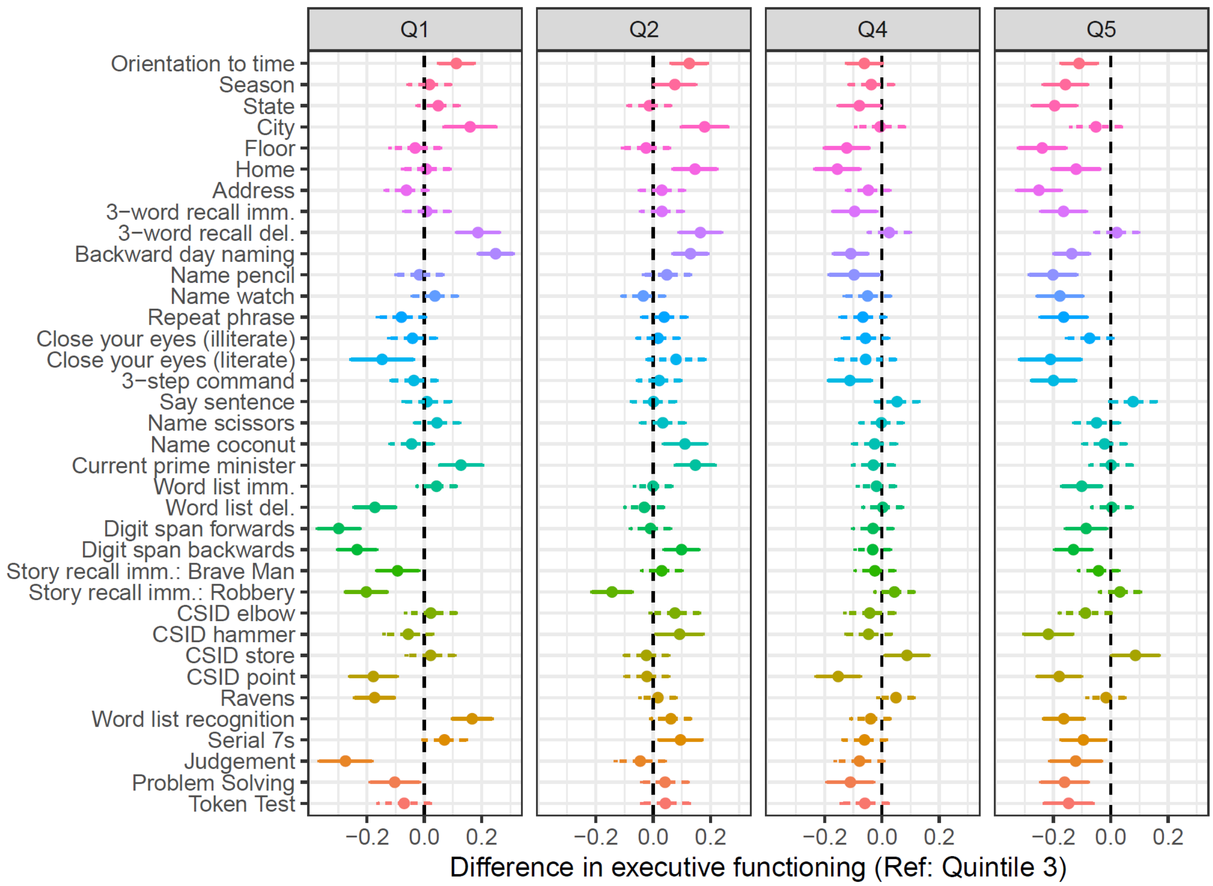



**Figure S4.** Differences in mean executive functioning for each quintile of time taken to complete individual cognitive tests compared to Quintile 3. Estimates were derived from item-specific regression models for the association between executive functioning and quintile of time taken on each specific test controlling for age, gender, interviewer, and score of the test. Uncertainty intervals show 95% confidence intervals; lines are solid if the 95% confidence interval does not include 0 and dotted if it does.

# Supplementary Materials 8: Item-specific regression models adjusting for educational attainment

Adjustments for education, urbanicity, and hearing impairment were considered sensitivity analyses because these factors may have large effects on cognition in addition to their potential effects on timing, independent of cognitive functioning. Because there is no way to separate out the impacts of these constructs on cognition and on timing independent of cognition this creates challenges analytically. While we want to adjust for factors that affect timing data, models are designed to capture the utility of test timing as a marker of underlying cognition, so we also want to avoid adjustment for variables strongly associated with underlying cognition, because to the extent that adjustment removes important variability in cognitive functioning, the signal of timing data as marker for cognitive functioning will be diminished.

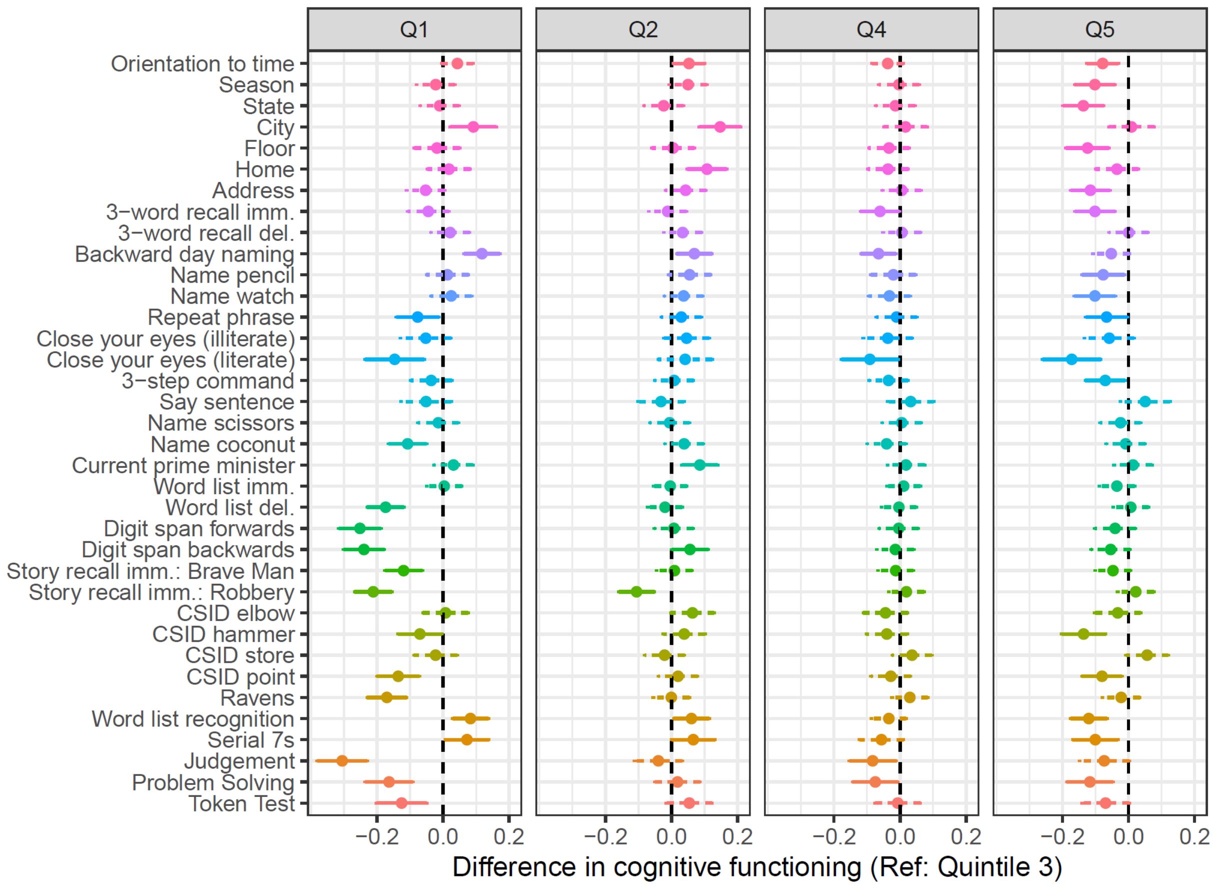



**Figure S5.** Differences in mean cognitive functioning for each quintile of time taken to complete individual cognitive tests compared to Quintile 3. Estimates were derived from item-specific regression models for the association between general cognitive functioning and quintile of time taken on each specific test controlling for age, gender, interviewer, education, and score of the test. Uncertainty intervals show 95% confidence intervals; lines are solid if the 95% confidence interval does not include 0 and dotted if it does. Education was divided into three categories, none, less than secondary education, and secondary or higher.

# Supplementary Materials 9: Item-specific regression models adjusting for urbanicity

Adjustments for education, urbanicity, and hearing impairment were considered sensitivity analyses because these factors may have large effects on cognition in addition to their potential effects on timing, independent of cognitive functioning. Because there is no way to separate out the impacts of these constructs on cognition and on timing independent of cognition this creates challenges analytically. While we want to adjust for factors that affect timing data, models are designed to capture the utility of test timing as a marker of underlying cognition, so we also want to avoid adjustment for variables strongly associated with underlying cognition, because to the extent that adjustment removes important variability in cognitive functioning, the signal of timing data as marker for cognitive functioning will be diminished.

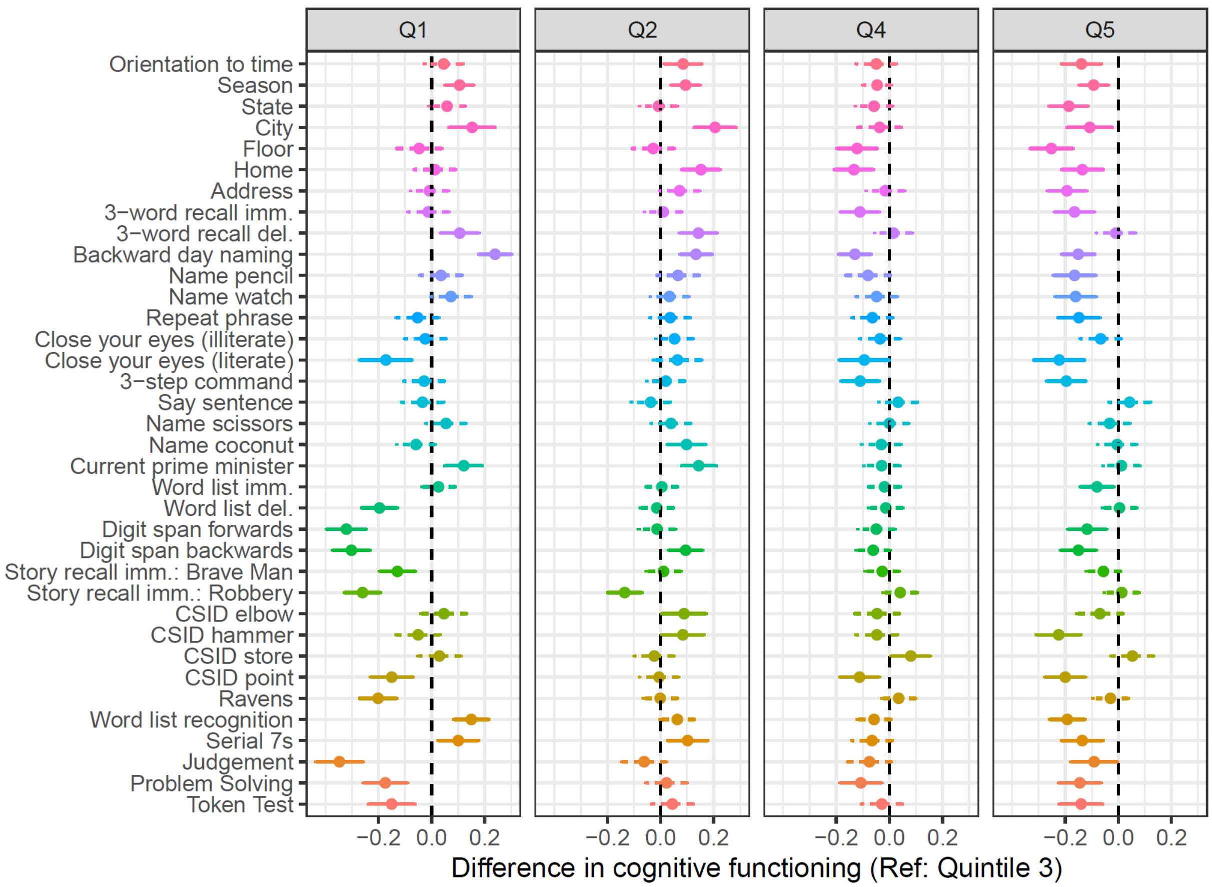



**Figure S6.** Differences in mean cognitive functioning for each quintile of time taken to complete individual cognitive tests compared to Quintile 3. Estimates were derived from item-specific regression models for the association between general cognitive functioning and quintile of time taken on each specific test controlling for age, gender, interviewer, urbanicity, and score of the test. Uncertainty intervals show 95% confidence intervals; lines are solid if the 95% confidence interval does not include 0 and dotted if it does.

## Supplementary Materials 10: Item-specific regression models adjusting for language of administration

Models adjusting for language of administration were considered as a sensitivity analysis because language was somewhat collinear with interviewer and added a significant number of terms to models, reducing precision.

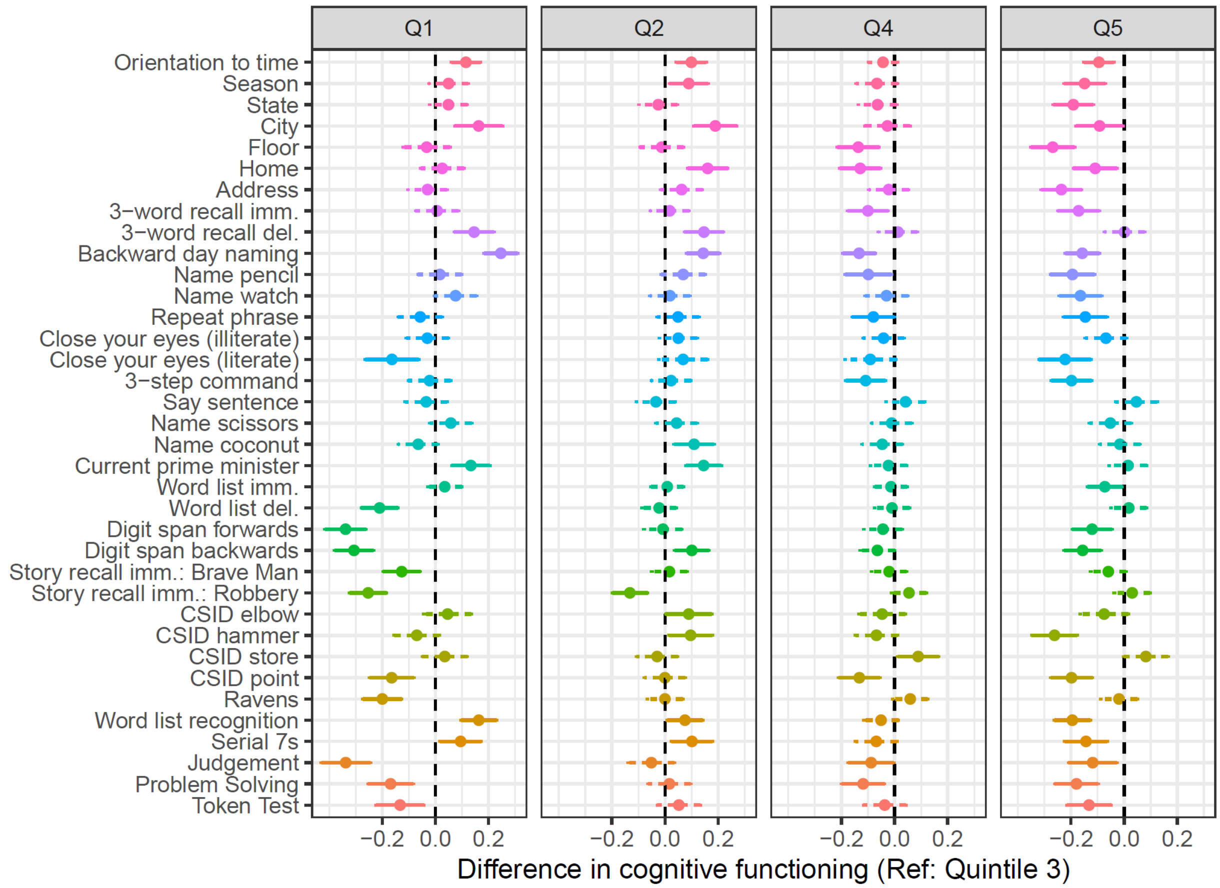



**Figure S7.** Differences in mean cognitive functioning for each quintile of time taken to complete individual cognitive tests compared to Quintile 3. Estimates were derived from item-specific regression models for the association between general cognitive functioning and quintile of time taken on each specific test controlling for age, gender, interviewer, language in which the test was given, and score of the test. Uncertainty intervals show 95% confidence intervals; lines are solid if the 95% confidence interval does not include 0 and dotted if it does.

# Supplementary Materials 11: Item-specific regression models adjusting for hearing impairment

Adjustments for education, urbanicity, and hearing impairment were considered sensitivity analyses because these factors may have large effects on cognition in addition to their potential effects on timing, independent of cognitive functioning. Because there is no way to separate out the impacts of these constructs on cognition and on timing independent of cognition this creates challenges analytically. While we want to adjust for factors that affect timing data, models are designed to capture the utility of test timing as a marker of underlying cognition, so we also want to avoid adjustment for variables strongly associated with underlying cognition, because to the extent that adjustment removes important variability in cognitive functioning, the signal of timing data as marker for cognitive functioning will be diminished.

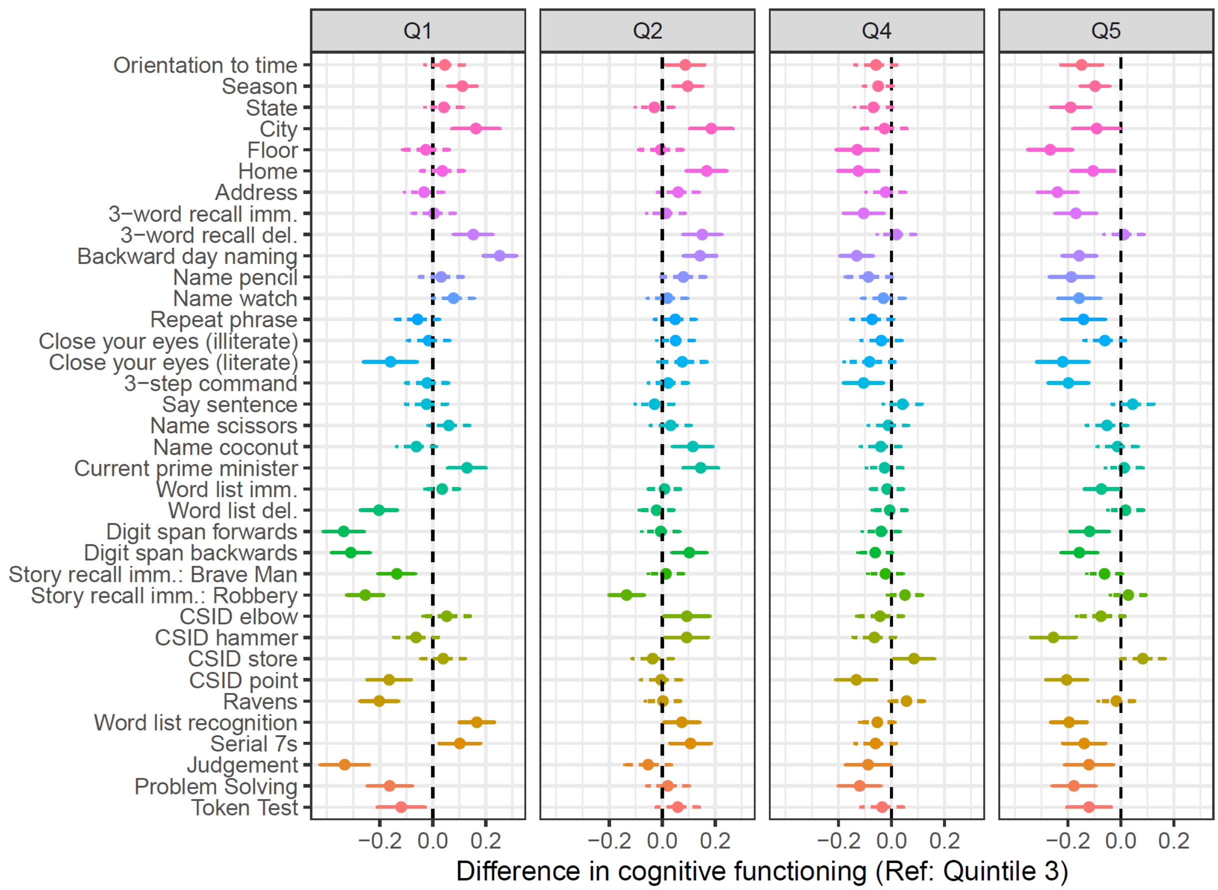



**Figure S8.** Differences in mean cognitive functioning for each quintile of time taken to complete individual cognitive tests compared to Quintile 3. Estimates were derived from item-specific regression models for the association between general cognitive functioning and quintile of time taken on each specific test controlling for age, gender, interviewer, hearing impairment, and score of the test. Uncertainty intervals show 95% confidence intervals; lines are solid if the 95% confidence interval does not cross 0 and dotted if it does.

# Supplementary Materials 12: Correlations with item test difficulty

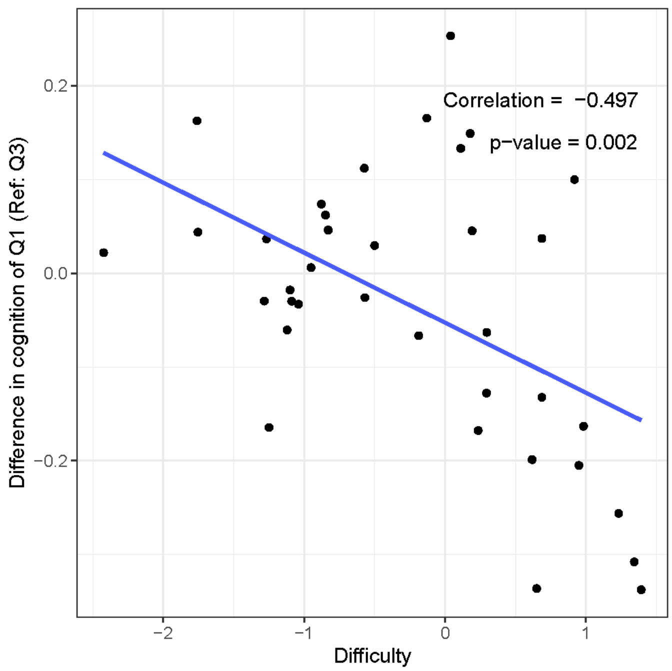



**Figure S9.** Correlations between test difficulty and the difference in general cognitive functioning between the first and third quintile of the of time taken to complete individual cognitive tests. Test difficulty was quantified using either item response theory methods.

# Supplementary Materials 13: Proportion of respondents with above average cognition, but timing data suggestive of lower cognitive functioning

**Table S2.** Proportion of respondents with above average cognitive functioning, but timing data suggestive of lower cognitive functioning. More specifically, the proportion indicates the number of respondents with predicted cognition based on timing data below the mean but cognitive scores above the mean out of all respondents with cognitive scores above the mean. Right tails indicate the percent who took longer to answer that are predicted to have below average cognitive functioning and left tails indicate the percent who took less time that are predicted to have below average cognitive functioning.

| Test Item | Left Tail% | Right Tail% | Total% |
|---|---|---|---|
| Token Test | 0.131 | 0.303 | 0.434 |
| Problem Solving | 0.147 | 0.255 | 0.402 |
| CSID point | 0.139 | 0.202 | 0.341 |
| Name coconut | 0.141 | 0.176 | 0.316 |
| Floor | 0.315 | 0.000 | 0.315 |
| Name scissors | 0.000 | 0.309 | 0.309 |
| 3-word recall del. | 0.305 | 0.000 | 0.305 |
| City | 0.304 | 0.000 | 0.304 |
| Digit span backwards | 0.119 | 0.164 | 0.283 |
| 3-word recall imm. | 0.058 | 0.222 | 0.280 |
| 3-step command | 0.276 | 0.000 | 0.276 |
| Home | 0.270 | 0.000 | 0.270 |
| Orientation to time | 0.270 | 0.000 | 0.270 |
| Current prime minister | 0.000 | 0.269 | 0.269 |
| Name watch | 0.249 | 0.000 | 0.249 |
| Season | 0.242 | 0.000 | 0.242 |
| Word list recognition | 0.234 | 0.000 | 0.234 |
| Repeat phrase | 0.047 | 0.181 | 0.228 |
| Backward day naming | 0.223 | 0.000 | 0.223 |
| CSID elbow | 0.222 | 0.000 | 0.222 |
| Name pencil | 0.218 | 0.000 | 0.218 |
| CSID hammer | 0.031 | 0.186 | 0.217 |
| Ravens | 0.000 | 0.210 | 0.210 |
| State | 0.026 | 0.182 | 0.208 |
| Judgement | 0.000 | 0.207 | 0.207 |
| Word list imm. | 0.187 | 0.000 | 0.187 |
| Story recall imm.: Robbery | 0.000 | 0.165 | 0.165 |
| Word list del. | 0.000 | 0.158 | 0.158 |
| Address | 0.154 | 0.000 | 0.154 |
| Digit span forwards | 0.000 | 0.152 | 0.152 |
| Story recall imm.: Brave Man | 0.000 | 0.078 | 0.078 |
| Close your eyes (illiterate) | 0.000 | 0.000 | 0.000 |
| Close your eyes (literate) | 0.000 | 0.000 | 0.000 |
| Say sentence | 0.000 | 0.000 | 0.000 |
| CSID store | 0.000 | 0.000 | 0.000 |
| Serial 7s | 0.000 | 0.000 | 0.000 |