Supplemental materials

This document provides additional materials for the paper, “Measurement precision across cognitive domains in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset.”

All content are duplicated from a companion paper, “Cognitive domain harmonization and co-calibration in studies of older adults.”

Table of contents:

Supplement 1: Additional discussion of modern psychometric theory. Page 2.

Supplement 1: Memory domain. Page 6.

Supplement 2: Executive functioning domain. Page 8.

Supplement 3: Language domain. Page 9.

Supplement 4: Visuospatial domain. Page 10.

Supplement 1: Additional discussion of modern psychometric theory.

### Modern Psychometric Approach

Modern psychometric theory has been the dominant paradigm in educational testing settings since at least the 1960s with the publication of Lord and Novick’s highly influential book (F. M. Lord & Novick, 1968). The extension of these approaches from dichotomous correct/incorrect format responses to incorporate ordinal or graded data was outlined in detail in 1969 by Fujiko Samejima (Samejima, 1969). Valuable single volume treatments include a 1999 text from Roderick McDonald (McDonald, 1999). Accessible paperback introductory texts include an introduction to item response theory from 1991 (Hambleton, Swaminathan, & Rogers, 1991) and a second from 2000 focused specifically on implications of these methods for psychology (Embretson & Reise, 2000). These developments and many others have led some to take the field of psychology to task for not using these methods more widely (Borsboom, 2006).

A major fundamental insight from modern psychometric theory is that it is critical to consider the difference between the difficulty of the item and the ability level of the test taker. Immediately from that sentence it follows that item difficulty and person ability must be on the same scale as otherwise it would not make sense to consider a difference between those two values. This fundamental insight and its extensions have led to the development of useful tools to analyze and understand data derived from cognitive tests.

Our primary focus here is on measurement precision. In addition to the works cited above, Lord published a treatise in 1980 that was essentially an extended discussion of measurement precision (Frederick M. Lord, 1980). If we consider a test taker of a particular ability level, a scale that has many items of difficulty levels close to that ability level will provide a lot of measurement precision for that test taker, while a second scale that has few such items will provide much less measurement precision. These intuitive statements are quantified in modern psychometric approaches, producing estimates of *test information*.

One critical aspect of test information is that it may vary across different ability levels, implying that individuals of different ability levels have their ability measured with different levels of precision. When we consider a test to be a group of items of varying difficulty levels, we can imagine that a test that may have many items with difficulty levels appropriate for a very able test taker may have very few items with difficulty levels appropriate for a less able test taker. Such a test would have more information at higher ability levels, and less information at low ability levels. These considerations immediately suggest the somewhat limited value of omnibus single statistics such as Cronbach’s alpha to describe test reliability (McDonald, 1999). Test reliability as represented by Cronbach’s alpha is a complex function of the distribution of ability levels in the population studied together with the information levels of the test at those ability levels. An information curve that plots the information content of the test, along with a summary of the distribution of ability levels in the population of interest, provide the investigator with a much better understanding of the appropriateness of the test for that population of interest than a single number that purports to characterize the performance of the test as a whole. All of this extends from the fundamental insight of considering the difference between person ability and item difficulty.

While practitioners and students of modern psychometric theory are taught to have intuition regarding information content and information curves, others not trained in this way may find it difficult to consider quantifications of measurement precision with completely unfamiliar units. Fortunately, there is a mathematical relationship between information and the standard error of measurement, which is on the same scale as the test score and may be more intuitive. In particular, there is an inverse square root relationship, such that the standard error of measurement is the inverse square root of the information content. The standard error of measurement or SEM = ${1}/{\sqrt{I\left(θ\right)}}$, where theta (θ) is the ability level measured by the test, and I(θ) is the information content at ability level theta. We are thus able to focus on measurement error on a more familiar scale.

Again considering tests administered to a test taker ability level, the test with many items of appropriate difficulty levels will have a lot of information at that ability level, and a correspondingly low SEM, while a test with few items of appropriate difficulty levels will have a small amount of information at that ability level, and a high SEM.

Readers familiar with z scoring approaches will be familiar with the scores produced by modern psychometric approaches, as these are scaled such that the underlying ability level has a mean of 0 and a standard deviation (SD) of 1. It can be useful to consider the range from +3 to -3, or 3 SD above the mean to 3 SD below the mean. The vast majority of the action for the traits or ability levels we tend to measure falls within these 6 units.

A useful number to anchor understanding of SEM units is 0.3. This number is used as the default stopping rule in computerized adaptive testing (CAT) packages (S. W. Choi, Grady, & Dodd, 2010). CATs are designed to pose questions to test takers in an efficient manner, where an algorithm considers all of the items in an item bank and chooses one to administer based on the current estimate of the person’s ability level. The algorithm continues in this fashion until a stopping rule is reached. Items are administered until the individual’s score is known at least as precisely as ± 0.3 units, and then the algorithm stops. We find it useful to note that a SEM of +/- 0.3 units is operationalized as sufficiently precise for individual measurement purposes as the default in CAT software.

Working backwards from an SEM of 0.3, squaring that and taking the inverse leads to an information content of roughly 11.1. Square numbers for information are the easiest to translate into corresponding SEMs. An information content of 9 (32) corresponds to an SEM of 1/3 = 0.33. An information content of 16 (42) corresponds to an SEM of ¼ or 0.25. An information content of 25 corresponds to an SEM of 1/5 or 0.20.

The purpose of this formula and the above discussion is to provide some background information and guidance for interpreting the SEM data we present in this paper. If standard errors of measurement (SEMs) are large this would lead to concern that lack of measurement precision could be influencing results. At the same time, if SEMs are small, this can provide reassurance to investigators finding differences in associations across domains. Additional information about modern psychometrics and SEMs can be found in Supplemental Materials.

Supplement 1. Memory domain.

Final model was a data driven bifactor model with CFI = 0.981, TLI = 0.979, and RMSEA = 0.088 for individuals administered RAVLT version A and CFI = 0.991, TLI = 0.990, and RMSEA = 0.073 for individuals administered RAVLT version B. The following items were included in the CFA analysis:

Supplemental Table 1. Items and secondary structure for memory for the ADNI study

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Variable** | **Description** | **Secondary Structure** |
| ADNI | limmtotal | WMS-R: Logical Memory—Immediate Recall(AT) | F1 |
| ADNI | ldeltotal | WMS-R: Logical Memory—Delayed Recall (AT) | F1 |
| ADNI | avtot1\* | Rey: AVLT Trial 1 Total |  |
| ADNI | avtot2\* | Rey: AVLT Trial 2 Total | F2 |
| ADNI | avtot3\* | Rey: AVLT Trial 3 Total | F2 |
| ADNI | avtot4\* | Rey: AVLT Trial 4 Total | F2 |
| ADNI | avtot5\* | Rey: AVLT Trial 5 Total | F2 |
| ADNI | avtot6\* | Rey: AVLT Trial 6 Total | F2 |
| ADNI | avtotb\* | Rey: AVLT List B Total |  |
| ADNI | avdel30min\* | Rey: AVLT 30 Minute Delay Total | F2 |
| ADNI | avdeltot\* | Rey: AVLT Recognition Score |  |
| ADNI | q1score | ADAS-Cog: Word Recall—score | F3 |
| ADNI | q4score  | ADAS-Cog: Delayed Word Recall | F3 |
| ADNI | q7score | ADAS-Cog: Orientation—score |  |
| ADNI | q8score | ADAS-Cog: Word Recognition—score |  |
| ADNI | mmdate | MMSE: What is today's date? |  |
| ADNI | mmyear | MMSE: What is the year? |  |
| ADNI | mmmonth | MMSE: What is the month? |  |
| ADNI | mmday | MMSE: What day of the week is today? |  |
| ADNI | mmseason | MMSE: What season is it? |  |
| ADNI | mmhospit | MMSE: What is the name of this hospital (clinic, place)? |  |
| ADNI | mmfloor | MMSE: What floor are we on? |  |
| ADNI | mmcity | MMSE: What town or city are we in? |  |
| ADNI | mmarea | MMSE: What county (district, borough, area) are we in? |  |
| ADNI | mmstate | MMSE: What state are we in? |  |
| ADNI | bft1 | MMSE: Ball, flag, tree—immediate recall (collapsed) |  |
| ADNI | bft2 | MMSE: Ball, flag, tree—delayed recall (collapsed) |  |
| ADNI | mocaregi | MoCA: registration, sum of two trials |  |
| ADNI | delsum | MoCA: delayed recall of word list |  |

\* MoCA (blue) items were only administered in ADNI GO/2/3 while orange items were in all ADNI waves (1/GO/2).

Supplement 2. Executive Functioning.

Final models were data driven bifactor models in ADNI 1 (CFI = 0.993, TLI = 0.990, and RMSEA = 0.050) and ADNI GO/2/3 (CFI = 0.972, TLI = 0.967, and RMSEA = 0.045). The following items were included in the CFA analysis:

Supplemental Table 2. Items and secondary structure for executive functioning for the ADNI study

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Variable** | **Description** | **Secondary Structure** |
| ADNI | clockcirc | Approximately circular face |  |
| ADNI | clocksym | Symmetry of number placement |  |
| ADNI | clocknum | Correctness of numbers |  |
| ADNI | clockhand | Presence of the two hands |  |
| ADNI | clocktime | Presence of the two hands, set to ten after eleven |  |
| ADNI | dspanbac | WAIS-R: Digit Span Backward Total Correct | F2 |
| ADNI | traascor | Trails A Time to Complete | F1 |
| ADNI | trabscor | Trails B Time to complete | F1 |
| ADNI | digitscor | WMS-R: Digit Symbol Total Correct | F1 |
| ADNI | dspanfor | WMS-R: Digit Span Forward Total Correct | F2 |
| ADNI | q13score | ADAS-Cog: Number cancellation task | F1 |
| ADNI | rworld | MMSE: Spell WORLD backwards |  |
| ADNI | absmeas | MoCA: Abstraction: watch-ruler |  |
| ADNI | abstran | MoCA: Abstraction: train-bicycle |  |
| ADNI | trails | MoCA: Trails |  |
| ADNI | digback | MoCA: Digits Backward |  |
| ADNI | serial | MoCA: Serial 7 total |  |
| ADNI | digfor | MoCA: Digits Forward |  |
| ADNI | letters | MoCA: List of Letters/Tapping: # Errors |  |

MoCA (blue) items were only administered in ADNI GO/2/3 while gray items were in all ADNI waves (1/GO/2).

Supplement 3. Language domain.

Final models was a single factor model in ADNI 1 (CFI = 0.979, TLI = 0.973, and RMSEA = 0.080);
ADNI GO/2 (CFI = 0.977, TLI = 0.973, and RMSEA = 0.048), and ADNI 3 (CFI = 0.953, TLI = 0.943, and RMSEA = 0.037). The following items were included in the CFA analysis (Supplemental Table 10):

Supplemental Table 3. Items and secondary structure for language for the ADNI study

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Variable** | **Description** | **Secondary Structure** |
| ADNI | catanimsc | Category Fluency (Animals) —Total Correct |  |
| ADNI | catvegesc\* | Category Fluency (Vegetables) —Total Correct |  |
| ADNI | bnttotal\*\* | BNT: Boston Naming Test: Total Number Correct (1+3) |  |
| ADNI | q2score | ADAS-Cog: Commands |  |
| ADNI | q5score | ADAS-Cog: Naming |  |
| ADNI | q6score | ADAS-Cog: Ideational Praxis—score |  |
| ADNI | mmrepeat | MMSE: Repeat after me: no ifs, ands, or buts. |  |
| ADNI | mmhand | MMSE: Takes paper in right hand |  |
| ADNI | mmfold | MMSE: Folds paper in half |  |
| ADNI | mmonflr | MMSE: Puts paper on floor |  |
| ADNI | mmread | MMSE: Present the piece of paper which reads |  |
| ADNI | mmwrite | MMSE: Write a sentence. |  |
| ADNI | camel | MoCA: Camel naming |  |
| ADNI | lion | MoCA: Lion naming |  |
| ADNI | rhino | MoCA: Rhinoceros naming |  |
| ADNI | repeat1 | MoCA: Repeat Sentence |  |
| ADNI | repeat2 | MoCA: Repeat Sentence |  |
| ADNI | ffluency | MoCA: Letter Fluency—F (total number of correct words) |  |

### \* Only in ADNI 1; \*\* Boston Naming Test excluded in ADNI 3; Blue items are MoCA items introduced in ADNI GO/2/3.

\* MMSE items watch and pencil naming were dropped from the model because of sparseness in cells. They are extremely easy items and <1% gets it wrong.

Supplement 4. Visuospatial Domain.

Final model was a single factor model in ADNI 1/GO/2/3 with CFI = 0.988, TLI = 0.981, and RMSEA = 0.043. The following items were included in the CFA analysis (Supplemental Table 14):

Supplemental Table 4. Items and secondary structure for visuospatial functioning for the ADNI study

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Variable** | **Description** | **Secondary Structure** |
| ADNI | copycirc | Clock copy: Approximately circular face |  |
| ADNI | copysym | Clock copy: Symmetry of number placement |  |
| ADNI | copynum | Clock copy: Correctness of numbers |  |
| ADNI | copytime | Clock copy: Presence of the two hands, set to ten after eleven |  |
| ADNI | q3score | ADAS-Cog: Constructional Praxis—score |  |
| ADNI | mmdraw | MMSE: Copy interlocking pentagons |  |

### \* Clock copy (copyhand) item was dropped from the model because of sparseness in a cell. Almost all individuals got it correct.