HRS | HEALTH AND RETIREMENT STUDY
A Longitudinal Study of Health, Retirement, and Aging
Sponsored by the National Institute on Aging

Cognitive Test Selection for the Harmonized Cognitive Assessment Protocol (HCAP)

Documentation Report

David Weir, Ryan McCammon, Lindsay Ryan, and Kenneth Langa

Survey Research Center
Institute for Social Research
University of Michigan
Ann Arbor, Michigan

October 2014

The goal of this effort, set by the National Institute on Aging (NIA), is to devise a cognitive assessment that meets the following criteria: 1) Can be administered in the home by a survey interviewer in about one hour, 2) Has sufficient overlap with the 2002/03 HRS-ADAMS study that it can be used to establish trends in prevalence of dementia and cognitive impairment without dementia, and 3) Can be administered comparably and lead to comparably valid diagnoses in other developed and developing countries where HRS-type surveys are conducted. This goal is achievable because of substantial investment by the NIA in data collection in population studies focused on dementia. There is thus now an empirical basis for making test selection. Informant reports are a key element of most population-based approaches to assessment to establish the presence or absence of change in cognitive ability and of limitation in activity resulting from change. We intend to include informant reports in our data collection protocol and in the diagnostic algorithm. This document is focused on cognitive testing of the subject.

We began by surveying the tests that were available in ADAMS and some other large population studies. Table 1 lists these tests, organized by domain. Domain assignments come from the studies based on face validity. Rigorous factor analysis to determine the validity of these domains and these assignments has in general not been done. Brief descriptions and key citations for tests are given in the appendix.

We see from Table 1 that there is considerable overlap between ADAMS and the Rush studies--the Religious Orders Study (ROS) and the Memory and Aging Project (MAP) (Bennett, et al 2012a and Bennett et al 2012b)—as well as between ADAMS and the CSI-D used by Hendrie and more recently by Martin Prince in the 10/66 studies of dementia in developing countries.

Generally speaking, most of these studies have at least two tests in most domains. For episodic memory, word recall and the Wechsler logical memory test are the most common. ADAMS also included the Fuld object memory test which no other study has done, while the Rush studies are the only ones to include the East Boston memory test (a story recall test similar to Wechsler). For semantic fluency, animal naming and the Boston naming test are the most common. ADAMS had a third test using controlled oral word association, in which subjects are asked to name words starting with specific letters, whereas Rush used fruit naming as its third fluency test.

Beyond these two domains there is somewhat less congruence, and somewhat less agreement on the nature of the domains themselves. We have labeled as attention/executive function what the Rush studies call perceptual speed. The only test in common is symbol digit modalities (substitution), a kind of switching task. ADAMS used Trail-making A and B, whereas MAP adopted the Stroop test and a shorter number comparison test. Rush uses a domain labeled perceptual orientation and includes two tests, the Benton line orientation test and Raven's progressive matrices (16 items). ADAMS had only one test that could plausibly be included in this domain and that is the CERAD constructional praxis. Finally, for working memory both studies used the digit span test, to which Rush added a number ordering task.

Both ADAMS and Rush included vocabulary-based measures intended to capture crystallized intelligence, or premorbid ability—the WRAT in ADAMS and the NART in Rush. These tests are intended to measure cognitive abilities that change very little with age or the progression of dementia. Exactly how they ought to be used in an algorithm for dementia assessment is not clear.

Two other domains—processing speed and fluid intelligence—are not part of ADAMS, the Rush studies or most other population studies we could find. There is reason to believe these may underlie important age-related changes in function and so we include them here even though we have no data available to appraise them against diagnosed dementia or impairment.

Table 1. Cognitive tests used in US population studies, by domain

	Time	HRS core	ADAMS	Hendrie	ROS/MAP	MIDUS BTACT	Mayo MCI
Informant scales							
Blessed			X				
DSRS			X				
CSI-D				X			
AD8		(x)					
SBT							
CDR-clinical judgment			X	X			X
Short multi-domain screeners							
MMSE	12		X	X	X		
3MSE	20						
MOCA							
Episodic Memory							
CERAD word recall	10	(x)	X	X	X		
Wechsler Logical Memory	8.5		X		X		X
Fuld Object Memory	13		X				
Benton Visual Retention	7		X				
Wechsler Visual Reproduction	?						X
Rey Auditory Verbal Learning	?					X	X
East Boston word recall					X		
Language/Semantic Memory/Retrieval Fluency							
Animal Naming	2	X	X	X	X	X	X
Boston Naming test	3		X		X		X
Controlled Oral Word Assoc	4		X				
Fruit Naming	2				X		
Attention / Executive							

Function							
Trail Making	7.5		X				X
Symbol Digit Substitution	3.5		X		X		X
Serial 7s	2	X	X				
Number comparison	?				X		
Stroop	?				X		
IU token test				X			
Stop and Go	2					X	
Visuospatial skills / perceptual orientation							
Benton Line Orientation	7				X		
Raven's Progressive Matrices	8				X		
CERAD Constructional Praxis	3		X	X			
WAIS-R Block Design	?						X
WAIS-R Picture Completion	?						X
Working memory							
Digit Span	4		X		X	X	
Digit ordering	?				X		
Processing speed							
Timed backward count	1.5					X	
Deary-Liewald choice RT	4						
Crystallized (pre-morbid) intelligence							
NART					X		
WRAT			X				
Fluid intelligence							
Number Series	6	X				X	
Verbal Analogies	6	X					
Raven's Progressive Matrices	8				X		

Based on the inventory of tests it is clear that there is considerable overlap between the Rush studies and ADAMS. Indeed, we were able to complete both assessments in a pilot study with only modest increase in time over the ADAMS itself. Diagnoses made independently by ADAMS and Rush investigators showed good agreement (89% for demented vs not demented, and 69% for a three-way classification including cognitive impairment without dementia). With the assistance of David Bennett, PI of the Rush studies and a member of the HRS Data Monitoring Committee, we have been able to analyze data from ROS/MAP to examine the relationship between diagnosis and cognitive testing in their data as well as in ADAMS.

The first thing we examined was item missing data rates (see Table 2). People with dementia are unable to do some tests. They may express this, or the test administrator may exercise their own judgment and skip a test, or end the testing altogether before all tests are complete. We would prefer in our own test administration to minimize the role of interviewer judgment, but in analyzing the available data we must deal with what was done in the past. Missing data rates serve as an indication of which tests are feasible for demented subjects, and which tests can meaningfully be used to calibrate a diagnostic algorithm.

In both studies the MMSE was the first test administered, and it had very low rates of missing data.

Table 2 shows missing data rates for persons diagnosed with dementia from ADAMS and MAP. The Rush MAP study is of a general population, whereas the ROS is of a very specialized and highly-educated population. They received similar tests and generally speaking the relationships between test performance and diagnosis are very similar. However, for analyses about population distributions we will focus on MAP.

The first two columns focus on persons diagnosed with dementia at that testing. Missing data rates among the demented are fairly high for all these tests, in both studies. In a few cases, such as symbol digit and digit span, ADAMS had significantly higher missing data rates than MAP. This serves as a caution against thinking that cognitive status diagnosis is based on complete scoring across the entire range of tests.

A large fraction of the missing tests come from persons with severe dementia. To examine missingness further, we exclude persons with MMSE of 15 or lower and calculate missing data rates for the remaining set of persons who received a diagnosis of dementia or mild impairment. For most of the tests administered by both studies, the missing data rates are in the single digits for this group. The highest rates of missing data are for the executive function tests (Trail-making, Stroop, and to a lesser extent symbol digits). The Rush tests of perceptual orientation have high missingness rates, as well.

Based on the analysis of missing data, it would seem that a strong common core across the two studies consists of MMSE, CERAD word recall and recognition, Wechsler logical memory, animal naming, Boston naming, digit span, and symbol digit modalities. A second set of candidates unique to one study or the other would include Fuld, Controlled Oral Word association, and CERAD constructional praxis from ADAMS, and East Boston, fruit naming, number comparison, Benton line orientation, Raven's matrices, and digit ordering from MAP. The trail-making and Stroop tests had high missing rates that limit our ability to evaluate them.

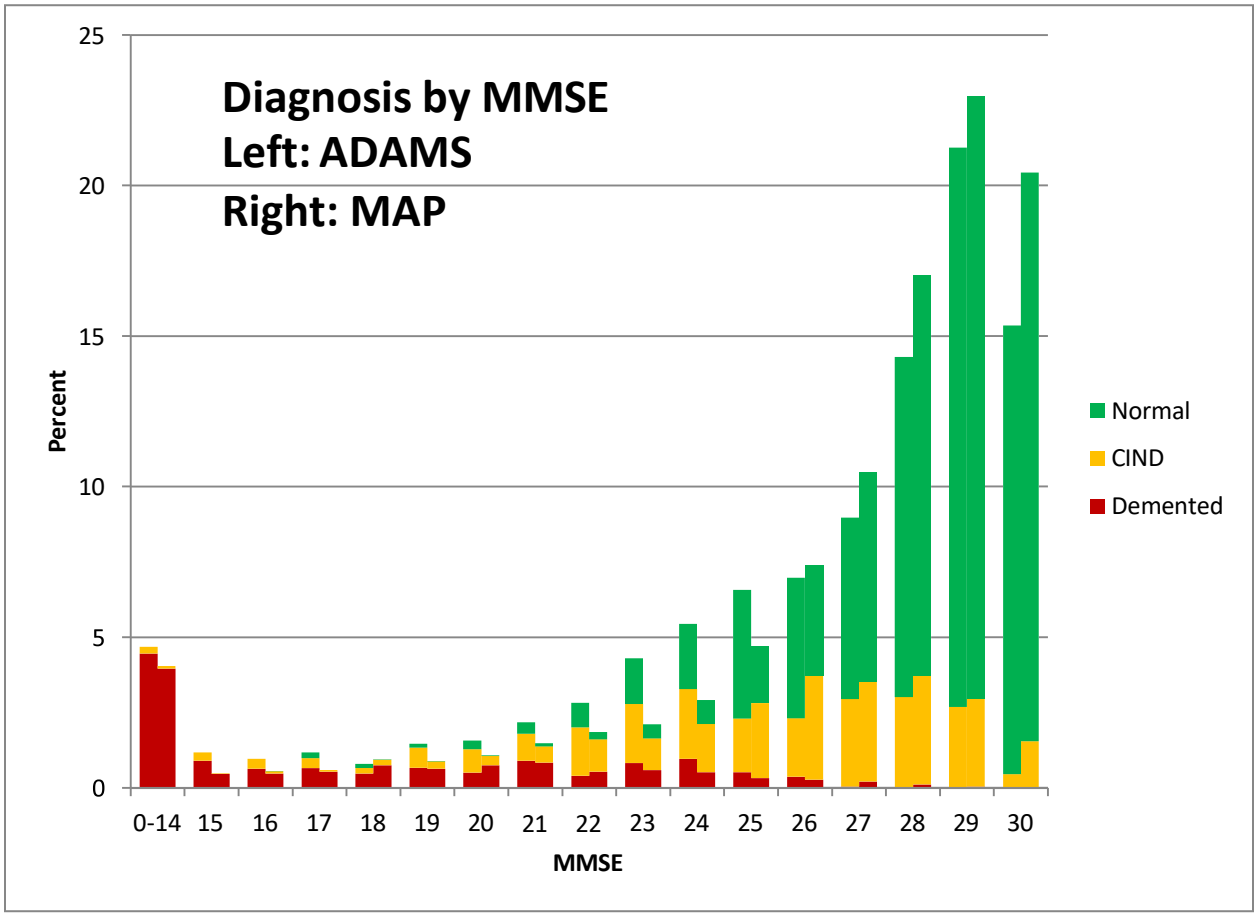
The next step is to study the association of these tests with diagnoses to determine their prognostic value and, in particular, to determine whether any are redundant to the others. We begin with an evaluation of the diagnoses in the two studies.

Table 2. Percent missing score on indicated test, by final diagnosed cognitive status, in ADAMS and MAP.

	Demented		Impaired with MMSE>15	
	ADAMS	MAP	ADAMS	MAP
Episodic Memory				
CERAD word recall	21.4%	23.0%	2.0%	6.6%
Wechsler Logical Memory	34.1%	24.8%	8.3%	7.2%
Fuld Object Memory	41.9%		10.9%	
Benton Visual Retention	52.9%		18.4%	
East Boston word recall		15.2%		5.2%
Language/Semantic Memory/Retrieval Fluency				
Animal Naming	20.8%	13.4%	0.3%	4.4%
Boston Naming test	16.9%	23.7%	1.4%	8.6%
Controlled Oral Word Assoc	49.7%		10.9%	
Fruit Naming		13.7%		4.5%
Attention / Executive Function				
Symbol Digit Substitution	74.0%	31.6%	33.6%	12.0%
Serial 7s	15.6%		0.3%	
Trail Making	83.4%		43.7%	
Number comparison		30.9%		9.9%
Stroop		40.9%		18.8%
Visuospatial skills / perceptual orientation				
Benton Line Orientation		35.5%		11.2%
Raven's Progressive Matrices		40.1%		17.8%
CERAD Constructional Praxis	36.0%		7.8%	
Working memory				
Digit Span	35.1%	12.1%	7.5%	4.4%
Digit ordering		26.6%		8.0%
Crystallized Intelligence				
NART		27.4%		10.7%
WRAT	60.7%		39.0%	

Diagnosis in ADAMS and MAP.
Although the two studies differed operationally in how they made diagnoses, they were similar in that cognitive testing was an input to an evaluation based on clinical judgment. It appears that this produced striking similarities in the relationship of diagnosis to cognitive assessment. Figure 1 shows the frequency distributions of diagnoses by MMSE score in ADAMS and MAP, side by side. Several points are very clear from this figure. First, ADAMS (weighted) and MAP are broadly similar in the overall distribution of MMSE scores, with MAP showing somewhat higher frequencies at the very top of the range. Secondly, ADAMS and MAP are also very similar in the mix of diagnoses at each level of MMSE. In both studies, MMSE under 15 is almost exclusively dementia, and nearly so through a score of 17. Then, from 18 to about 24 diagnoses are a mix of dementia and milder impairment with very few normal. From 25 to 27 it is a mix of mild impairment and normal, with a small number of dementia diagnoses. At 28 to 29 there are still a few 'mild impairment', while normal predominates. Only at 30 is impairment nearly absent.

Figure 1. Study diagnosis by MMSE score, ADAMS and MAP



It is clear from Figure 1 that ADAMS and MAP are very comparable studies and that in both studies diagnosis is very imperfectly tied to the single multi-domain MMSE screener.

What is perhaps more surprising is that diagnosis is also imperfectly tied to all the cognition measures taken together. Ordered probits of diagnosis on all the cognition measures have

pseudo-R-squares in the range of .55 to .61. Classification based on cognitive measures alone agrees with diagnosis about 85% of the time, in both studies. That is, the very best one can do using the full battery of cognitive data in each study is to match the diagnosis about 85% of the time. Clinical judgment, not reproducible by an analyst, accounts for the difference. In ADAMS, one potential source for the lack of one-to-one correspondence between cognitive testing and dementia diagnosis is the DSM-IV and new NIA-AA requirement for disability in daily activities for a dementia diagnosis to be made. The assessment for whether a disability was present was made by the consensus diagnosis panel mainly based on informant reports (the DSRS and structured clinical history).

Imputation.

To cope with the item missing data rates shown earlier while working on complete samples, we imputed missing test scores. All imputations were done by nearest-neighbor methods. For each test, a prediction equation was estimated using tests with complete data (including prior imputations). Predicted values were generated for all cases, including those missing on the item. Cases were sorted by the predicted value, and missing values were given the observed value of the case nearest to them in predicted value. Predicted values were used only to establish matches. This preserves the variability in the data as well as the correlations across tests.

Modeling Diagnosis

To assess the predictive power of the individual tests, we use a multivariate ordered probit over the three diagnostic categories. Using un-ordered multinomials, or doing binary models for each diagnosis separately did not suggest that the ordered model was inappropriate. A bigger concern is multi-collinearity. Correlation coefficients across tests average .6 in ADAMS and .51 in Rush, and these are only slightly lower across domains than within them. This makes it difficult to be certain that statistical analysis can point to the strongest tests. At the same time, it suggests the tests are reasonably good substitutes for one another so overall precision does not depend greatly on the choice of specific tests. As a final stage, we can test the loss from elimination of a specific test by likelihood-ratio tests of the loss of explanatory power comparing models with and without a specific test.

Bearing these caveats in mind, Table 3 shows the ordered probit results for ADAMS and the combined Rush studies. Looking first at the middle columns showing results for the tests shared in both studies, we see that the overall rate of classification based on this subset is similar in the two studies at about 81% correctly classified. All the tests were strongly significant in the Rush models. Several were not in the much smaller ADAMS sample. Coefficients were smaller in ADAMS for most tests.

Table 3. Ordered Probit Results for ADAMS and the Combined Rush Studies.

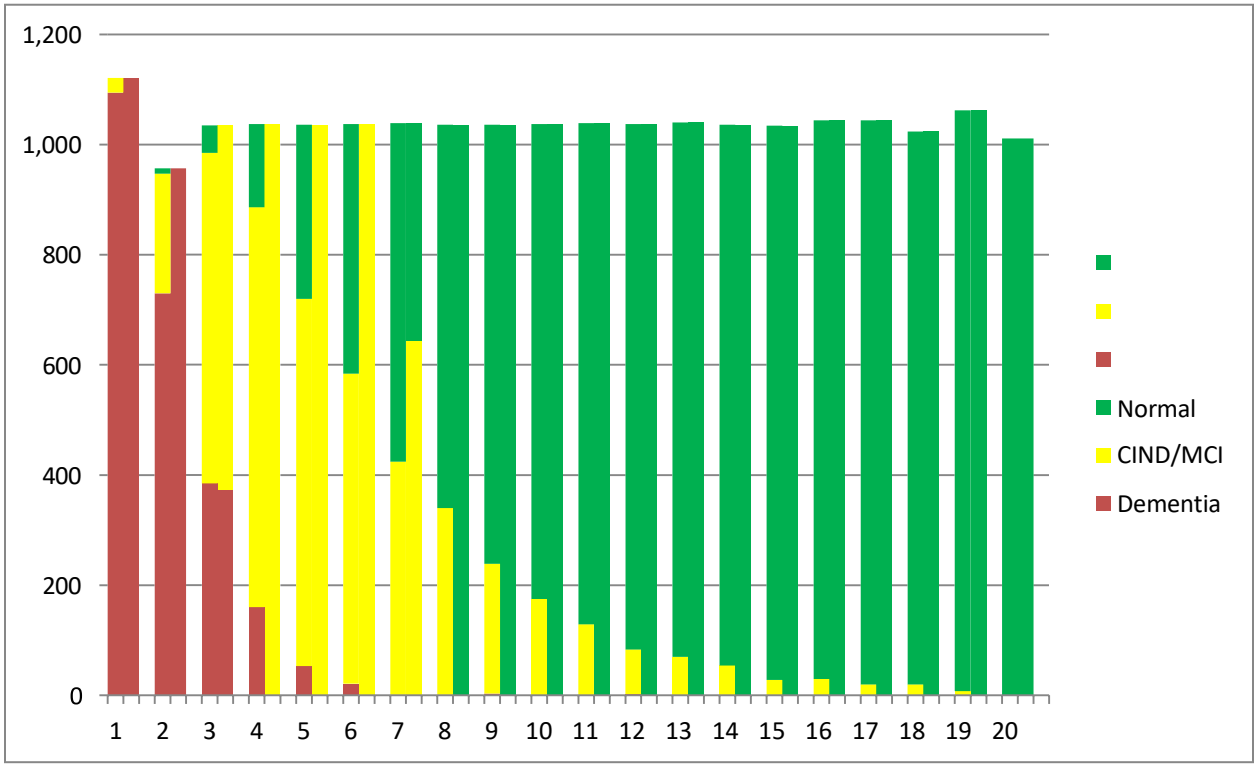
	ADAMS			ROS/MAP	
	All	Shared		Shared	All
MMSE	-0.0590	-0.0951***		-0.122***	-0.112***
	(0.0317)	(0.0241)		(0.00501)	(0.00527)
Immed rec	-0.125*	-0.101		0.0104**	0.0153***
	(0.0635)	(0.0580)		(0.00361)	(0.00378)
Delayed rec	-0.140*	-0.179**		-0.146***	-0.164***
	(0.0638)	(0.0562)		(0.00720)	(0.00752)
Word list recog	-0.0806	-0.0794		-0.135***	-0.141***
	(0.0497)	(0.0433)		(0.00849)	(0.00880)
Wechsler del	-0.0310*	-0.0357**		-0.0623***	-0.0604***
	(0.0146)	(0.0129)		(0.00296)	(0.00325)
Fuld	-0.0463*				
	(0.0193)				
Benton visual	-0.0183				
	(0.0168)				
East Boston del					-0.0185**
					(0.00614)
animals	-0.0346	-0.0666***		-0.0246***	-0.00801**
	(0.0205)	(0.0180)		(0.00266)	(0.00311)
Boston naming	-0.0338	0.0113		-0.108***	-0.0597***
	(0.0546)	(0.0418)		(0.00785)	(0.00813)
controlled owa	0.0178*				
	(0.00819)				
fruit naming					-0.0249***

					(0.00313)
Symbol digit	-0.00697	-0.0298***		-0.0152***	-0.00713***
	(0.00985)	(0.00846)		(0.00117)	(0.00154)
Constr praxis	-0.0381				
	(0.0293)				
Number comparison					0.00775***
					(0.00209)
Line orientation					-0.0920***
					(0.00385)
Raven's					-0.0774***
					(0.00410)
Digit span back	0.0753	0.0588		-0.0759***	-0.0405***
	(0.0514)	(0.0393)		(0.00611)	(0.00661)
Digit ordering					-0.00213
					(0.00698)
Trails B	0.00310*				
	(0.00126)				
DSRS	0.141***				
	(0.0174)				
cut1					
_cons	-4.217***	-6.093***		-8.158***	-8.681***
	(0.999)	(0.587)		(0.151)	(0.158)
cut2					
_cons	-1.363	-4.097***		-6.188***	-6.568***
	(0.959)	(0.545)		(0.140)	(0.147)
N	856	856		20742	20742
pseudo R-sq	0.631	0.521		0.503	0.541
ll	-7934414.3	-10299823.5		-8775.1	-8100.4
ll_0	-21512273.4	-21512273.4		-17639.9	-17639.9
classification	0.85	0.79		0.82	0.83

An important evaluative statistic is the agreement or correct classification rate. We determine this by creating a score based on a model, establishing study-specific cutpoints for that score to match the overall distribution of diagnoses, and comparing. The smaller shared model in ADAMS had the least agreement, whereas the full model in ADAMS had the best agreement. Agreement was not much higher in Rush for the full model compared with the smaller model.

The agreement can be seen in Figure 2 below. Here we compare the actual Rush diagnoses to the statistical assignments from the full model, by ventile of the cognition score implied by the full model (rather than just MMSE as in Figure 1). In the statistical assignments (right-hand bars), all the bars are of a single color except for the two ventiles in which the cutpoint between categories lies (third for dementia and seventh for normal). In contrast, the actual diagnoses are more dispersed even across this cognition score based on all testing. Mild impairment is fairly common at ventiles 8-11 and present all the way to the 19th --the 90-94th percentile group. Most of the disagreement between diagnoses is this spread of MCI up through the distribution of cognitive performance.

Figure 2. Study diagnosis and statistical assignment in ROS/MAP, by ventile of cognition score



We also compared agreement between the full model and the smaller shared model . It was .92 in Rush and .87 in ADAMS. This is likely the result of the informant report DSRS and clinical history having substantial influence in ADAMS. Finally, we used the scoring system from the Rush estimates to classify ADAMS subjects. The agreement was essentially identical to that using ADAMS estimates from the shared model at 79%.

Other validation tests

We compared the assigned diagnoses based on the models with the actual diagnoses in both study sets. First, we compared the percent of each diagnostic category with at least one apoE e4 allele. Neither study used genetic information in making its diagnosis. Because the e4 allele is associated with higher risk of dementia, it occurs at greater frequency among the diagnosed. Table 4 shows these results. ADAMS and Rush have very similar patterns using the study diagnoses. In Rush the genetic makeup of the demented category does not change much when using statistical estimates for diagnosis. The genetic distinction between normal and MCI narrows very slightly. In ADAMS, the genetic makeup of the normal group does not change much when using statistical estimates for diagnosis. The genetic distinction between CIND and demented narrows somewhat. Overall, the statistical estimates show good validity.

Table 4. Percent with any e4 allele, by study, diagnosis, and source of diagnosis

		Normal	CIND/MCI	Demented
Rush	Study	20.8	27.7	39.2
	Full model	21.2	26.7	39.2
	Shared model	21.3	26.2	39.7
ADAMS	Study	23.5	27.2	39.1
	Full model	23.3	29.0	37.2
	Shared model	22.9	30.9	35.9

For Rush only, we are able to compare the consistency of diagnosis over time by source of diagnosis. Using the study diagnoses, there was a reversion from MCI to normal in 32.9% of the cases where an MCI diagnosis was followed by another assessment. There was a reversion from demented to MCI or normal in 17.9% of the cases where a dementia diagnosis was followed by another assessment. Using statistical diagnoses reduces the amount of reversion. In the full model it was 20.8% and 11.2%, and in the shared model it was 19.8% and 8.8%. Thus, the statistical models produce a more stable time series of diagnosis.

Determining the best final set of tests.

Selection priority.
Tests in both studies that contribute significantly to diagnosis in both (definitely in)
MMSE
CERAD word recall and recognition
Animal naming
Symbol Digit substitution

Tests in both studies that contribute significantly to diagnosis in one (Rush) (probably in)
Digit span backward
Boston naming

Tests in one study that contribute significantly to diagnosis (need to choose)
CERAD constructional praxis
Raven's matrices
Line orientation
Number comparison
Fuld object memory
Trails
Stroop

Tests in neither study that experts believe capture important other dimensions
BTACT timed backward count
Deary-Liewald 4-choice reaction time

Tentative choices pending full committee review:

By domain
Episodic Memory word recall, recognition, Wechsler (and possibly visual retention)
Orientation first ten items of MMSE
Language animals and either Boston or another naming
Attention/exec function Serial-7s (in MMSE), SDMT, Trails B
Visuo-spatial CERAD constructional praxis, Raven's matrices (and possibly line orientation)
Working memory digit span backward
Processing speed BTACT timed backward count, Deary-Liewald RT
Fluid intelligence Number series (and Raven's)
Crystallized intelligence NART

APPENDIX Test Descriptions and Key Citations (Incomplete on Rush tests)

Mini Mental State Exam (MMSE)
Measures Cognitive Orientation/ General Cognitive Status

The MMSE assesses general cognitive status with measures of cognitive orientation, language, and memory (Folstein, Folstein, & McHugh, 1975). This test has a maximum score of 30 and is often used in clinical and research settings to identify individuals with likely cognitive impairment or dementia.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini Mental State': A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatriatric Research, 12,* 189 – 198.

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and education level. *JAMA, 269*, 2368 – 2391.

TICS (Reduced)
Measures of Cognitive Orientation and Vocabulary

This set of measures is based on the Telephone Interview for Cognitive Status (Brandt, Spencer, & Folstein, 1988), and overlaps to some extent with MMSE. The items not in MMSE are two object naming questions and naming the president and vice president of the United States.

Brandt, J., Spencer, M. and Folstein, M. (1988). The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuopsychology, and Behavioral Neurology, 1,* 111-117.

Neurological Praxis
Identifies the presence of Ideomotor Apraxia (IMA)

Ideomotor apraxia is characterized by the inability to perform tool-use pantomiming tasks either to command or to imitation (Wheaton and Hallet, 2007). The underlying cause of IMA is thought to be a disconnection of areas of the cerebral cortex caused by focal brain lesions, often due to Alzheimer's disease (Parakh et al., 2004), Parkinson's disease (Leiguarda et al., 1997), or stroke. Ideomotor apraxia is tested by asking the subject to follow the following verbal pantomiming commands: "comb your hair," "hammer a nail," and "brush your teeth."

Wheaton LA and Hallett M. Ideomotor apraxia: a review. *J Neurol Sci.* 2007; 260 (1-2): 1–10

Parakh R, Roy E, Koo E, Black S. Pantomime and imitation of limb gestures in relation to the severity of Alzheimer's disease. *Brain Cogn*. 2004; 55(2): 272-274.

Leiguarda RC, Pramstaller PP, Merello M, Starkstein S, Lees AJ, Marsden CD. Apraxia in Parkinson's disease, progressive supranuclear palsy, multiple system atrophy and neuroleptic-induced Parkinsonism. *Brain*. 1997; 120(Pt 1): 75-90.

EPISODIC MEMORY

CERAD Word List Memory and Recall

This task includes 10 high imagery words that are visually presented for 2 seconds each (Morris et al., 1989). The subject reads each word aloud as it was presented and is then tested on an immediate recall procedure and later on a delayed recall procedure and recognition. In the recognition task, a word is read to the subject and he or she indicates whether or not they recognize the word as having been read earlier as part of the word list. There are 3 trials with immediate recall, one trial of delayed recall, and one trial of recognition.

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum,G., Mellits, E.D., Clark,C., and the CERAD investigators (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*, 1159-1165.

HRS variant of CERAD Immediate and Delayed Word Recall

Poor performance of episodic memory is a key indicator of Alzheimer’s Disease. The Immediate Word Recall measure asks the respondent to listen to a list of 10 words, spoken with roughly 1 second between words, and to repeat as many words as can be remembered. After roughly five minutes, the respondent is asked again to recall as many words from the initial list to get a score for Delayed Recall.

Ofstedal, M.B., Fisher, G.G., & Herzog, A. R. (2005). Documentation of cognitive functioning measures in the health and retirement study. HRS/AHEAD Documentation Report DR-006. Available through the Survey Research Center at the Institute for Social Research, University of Michigan. http://hrsonline.isr.umich.edu/sitedocs/userg/dr-006.pdf

McArdle, J. J., Fisher, G. G., & Kadlec, K. M. (2007). Latent variable analyses of age trends of cognition in the Health and Retirement Study, 1992 – 2004. *Psychology and Aging, 22,* 525 – 545.

Wechsler Logical Memory

Respondents are read two brief stories and they are asked to recall as much of the story as possible. Subsequently they are asked to recall as much of the story as possible for a delayed recall score.

Wechsler, D. (1987). Wechsler Memory Scale - Revised Manual. San Antonio, TX: The Psychological Corporation.

Fuld Object Memory

This test is based on recall of ten common household objects, which the Subject first identifies by touch. The modified 3-trial version was used. As per standard instructions, the subject was selectively reminded of items he or she did not recall. Each recall trial was preceded by a distractor task requiring word generation.

Fuld, P. A. (1981). The Fuld Object Memory Evaluation. Stoelting Instrument Co, Chicago.

Lowenstein, D. A., Arguelles, T., Acevedo, A., Freeman, R.Q., Mendelssohn, E., Ownby, R.L., White, G., Mogosky, B. J., Schram, L., Barker, W., Rodriguez, I., & Duara, Ranjan. (2001). The utility of a modified object memory test in distinguishing between different age groups of Alzheimer's disease patients and normal controls. *Journal of Mental Health and Aging, 7*, 317-324.

SEMANTIC FLUENCY

Animal Naming

Animal naming asks respondents to name as many animals as possible in a minute. This measure assesses verbal retrieval and processing speed. The first verbal fluency measure was included in Thurstone's Primary Mental Abilities (Thurstone, 1938), a phonemic task asking participants to list words based on a common first letter. Animal Naming is the most frequently used semantic retrieval fluency measure (e.g. Goodglass & Kaplan, 1983; Kertesz, 1982). The CogVal Animal Naming instrument was adapted by McArdle and Woodcock from the WJ- III Tests of Achievement: Retrieval Fluency (© Riverside Publishing).

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders (2nd ed.). Philadelphia: Lea & Febiger.

Kertesz, A. (1982). Western Aphasia Battery. San Antonio, TX. The Psychological Corporation.

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum,G., Mellits, E.D., Clark,C., and the CERAD investigators (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*, 1159-1165.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

CERAD Boston Naming Test

This test requires that respondents provide the names of a series of line drawings (common man-made and naturally occurring objects)

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum,G., Mellits, E.D., Clark,C., and the CERAD investigators (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*, 1159-1165.

Controlled Oral Word Association Test (COWA)

Description: The COWA test was developed as part of the Multilingual Aphasia Examination and requires retrieval and oral production of spoken words beginning with a designated letter (Benton et al., 1983). The letters used in the ADAMS neuropsychological assessment were CFL. The letters used in Spanish-language administration were PSV (Jacobs et al., 1997).

Benton, A.L., Hamsher, K., Varney, N., and Spreen, O. (1983). *Contributions to Neuropsychological Assessment.* New York: Oxford University Press.

ATTENTION / EXECUTIVE FUNCTION

Symbol Digit Modalities (substitution) Test

The test requires the subject to substitute a number for randomized presentations of geometric figures. A printed key is provided which pairs the Arabic numbers 1-9 with a specific symbol so that each number has its own unique symbol.

Smith, A. (1968). The symbol-digit modalities test: a neuropsychologic test of learning and other cerebral disorders. In J. Helmuth (Ed.), Learning disorders (pp. 83–91). Seattle: Special Child Publications.

Smith, A. (1982). Symbol Digits Modalities Test. Los Angeles: Western Psychological Services.

Trail Making Test

Description: The test is administered in two parts, A and B. The Subject is asked to draw lines connecting consecutively numbered circles on a worksheet (Part A) and connect consecutively numbered and lettered circles on another worksheet (Part B) by alternating between the numbers and letters.

Reitan, R.M. (1992). *Trail Making Test: Manual for Administration and Scoring.* Tuscon, AZ: Reitan Neuropsychological Laboratory.

Ricker, J.H. and Axelrod, B.N. (1994). Analysis of an Oral Paradigm for the Trail Making Test. *Assessment, 1,* 47-51.

VISUOSPATIAL SKILLS / PERCEPTUAL ORIENTATION

Raven's Standard Matrices

Raven's Standard Progressive Matrices was designed to be a language-free assessment of general intelligence that could be useful in assessing ability in childhood through late life (Raven, 1981). The instrument shows respondents a series of illustrations, typically geometric shapes and patterns, and asks the respondent to select the missing picture from six to eight possible answers. The Rush studies administered a sub-set of 17 items from the full set of 60.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41,* 1 – 48.

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement, 26,* 1–16.

Raven, J. (1981). Manual for Raven's Progressive Matrices and Vocabulary Scales. Research supplement no. 1: The 1979 British standardization of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, together with comparative data from earlier studies in the UK, US, Canada, Germany, and Ireland. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

CERAD Constructional Praxis

The constructional praxis test in CERAD is adapted from Rosen's assessment of constructional praxis (Rosen et al., 1984), and tests the ability of the subject to copy four geometric forms of varying difficulty (circle, overlapping rectangles, diamond and cube).

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum,G., Mellits, E.D., Clark,C., and the CERAD investigators (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*, 1159-1165.

Rosen, W.G., Mohs, R.C., & Davis, K.L. (1984). A new rating scale for Alzheimer's disease. American Journal of Psychiatry, 141, 1356-1364.

WORKING MEMORY

Digit Span

Digit Span (Wechsler Adult Intelligence Scales, 3rd ed.; WAIS-III) includes separate tasks of both forward and backward repetition. On both tasks, the examiner read a series of number sequences to the examinee. For each Digits Forward item, the subject is required to repeat the

number sequence in the same order as presented. For Digits Backward, the subject is required to repeat the number sequence in the reverse order.

Wechsler, D. (1997). Wechsler Adult Intelligence Scales-Third Edition. San Antonio, TX: The Psychological Corporation.

PROCESSING SPEED

Backward Count (MIDUS)

Measures Processing Speed

The CogVal Backward Count measure asks respondents to begin with the number 100 and to count back as fast as possible for 30 seconds. This instrument measures the speed at which respondents process a simple cognitive task. A similar task is included in the MIDUS cognition protocol (Agrigoroaei & Lachman, 2011).

Agrigoroaei, S., & Lachman, M. E. (2011). Cognitive functioning in midlife and old age: Combined effects of psychosocial behavioral factors.

FLUID INTELLIGENCE

Number Series

Measures Quantitative Reasoning

The Number Series test was to be a quick six item measure of quantitative reasoning ability (Fisher, McArdle, McCammon, Sonnega, & Weir, 2013). Respondents are asked to look at a series of numbers and correctly identify the missing number in the series. All participants answer the same first set of three problems, and depending on their score on the first set they are given a second set of three items ranging from very easy to very difficult. This is described as a block-adaptive test.

Gwenith G. Fisher, John J. McArdle, Ryan J. McCammon, Amanda Sonnega, and David R. Weir. New Measures of Fluid Intelligence in the HRS (2013).