Development of algorithmic dementia ascertainment for racial/ethnic disparities research in the U.S. Health and Retirement Study

Supplemental Digital Content

Contents

eTABLE 1. List of predictors used in existing algorithms, Expert Model, and machine learning models

Variable	Included in one or more existing algorithm	Included in Expert Model	Available for Machine Learning Models
Sociodemographic/socioeconomic variables			
Age	X	X	X
Gender	X	X	X
Race	X	X	X
Ethnicity		X	X
Education	X	X	X
Respondent status (proxy- vs. self-response)	X	X	X
Self-respondent cognition			
Immediate & delayed word recall	X	X	X
Dates recall	X	X	X
Backwards count	X		X
Serial 7s	X	X	X
Object recall (cactus)	X	X	X
Object recall (scissors)	X	X	X
President recall	X	X	X
VP recall	X	X	X
Proxy-respondent cognition			
Proxy-rated memory	X	X	X
Proxy-assessment of Jorm symptoms	X	X	X
IQCODE average score	X	X	X
Interview-assessment of respondent cognition [a]	X		
Physical functioning			
ADLs	X	X	X
IADLs	X	X	X
Physical health			
Self-reported health status		X	X
History of diabetes		X	X
History of high blood pressure			X
BMI			X
Health behaviors			
Smoking			X
Drinking			X
Social engagement			
Partnered/Married		X	X
Time spent volunteering in past year		X	X
Time spent helping others in past year		X	X
Retirement status			X
Change variables			
Change in cognition (self and proxy)	X		X
Change in ADLs and IADLs	X		X
Change in self-reported health status			X
Change in social engagement			X

eTABLE 1 (cont.). List of predictors used in existing algorithms, Expert Model, and machine learning models

Variable	Included in one or more existing algorithm	Included in Expert Model	Available for Machine Learning Models
Interactions with sex			
Delayed word recall	X		X
IQCODE average score	X		X
Interactions with race and ethnicity			
Education		X	X
Serial 7s		X	X
Object, president, and VP naming		X	X
Interactions with respondent status			
Age		X	X
Education		X	X
Sex		X	X
Physical functioning (ADLs and IADLs)		X	X
Self-reported health status		X	X
Other interactions			
Date recall with ADLs		X	X
Object/president/VP naming with ADLs		X	X
Age with IADLs		X	X
Serial 7s with self-reported health		X	X
Serial 7s with education		X	X
Word recalls with self-reported health		X	X

Abbreviations: VP = vice-president; IQCODE = Informant Questionnaire on Cognitive Decline; BMI = Body mass index; ADLs = Basic activities of daily living; IADLs = Instrumental activities of daily living

[a] We excluded interviewer assessment of cognition in the covariate set for the Expert Model and Machine Learning Models despite it being used in the Crimmins and Langa-Kabeto-Weir algorithms given that it is not available prior to 2000

eTABLE 2. Coefficients from re-estimated Hurd ordered probit regression model (dementia = 1, CIND = 2, normal = 3), estimated on training dataset used in our previous work[1]

Predictor variable	Self-respondents (N=595)		Proxy-respondents (N=165)	
	Coefficient	SE	Coefficient	SE
Age (reference: < 75)				
75 - 79	-0.281	0.156	0.114	0.549
80 - 84	-0.570	0.152	-1.502	0.530
85 - 89	-0.726	0.182	-1.184	0.559
90 +	-0.892	0.227	-0.898	0.598
Education (reference: LTHS)				
HSGED	-0.332	0.135	0.081	0.432
GTHS	-0.325	0.200	-0.018	1.074
Female	-0.041	0.121	-0.389	0.366
ADL limitations (0-5)	0.046	0.074	-0.230	0.176
IADL limitations (0-5)	-0.195	0.088	-0.231	0.179
Change in ADL limitations	-0.061	0.073	0.286	0.168
Change in IADL limitations	0.077	0.086	0.034	0.173
Date recall (0-4)	0.314	0.096	-	-
TICS backwards count	-0.321	0.202	-	-
Serial 7	0.167	0.046	-	-
Scissors recall	-0.783	0.505	-	-
Cactus recall	-0.055	0.169	-	-
President recall	0.488	0.241	-	-
Immediate word recall	0.185	0.062	-	-
Delayed word recall	0.262	0.053	-	-
Change in date recall [a]	-0.050	0.089	-	-
Change in backwards counting [a]	0.293	0.197	-	-
Change in serial 7 [a]	-0.033	0.040	-	-
Change in scissors recall [a]	0.398	0.428	-	-
Change in cactus recall [a]	0.393	0.181	-	-
Change in president recall [a]	-0.123	0.204	-	-
Change in immediate word recall [a]	-0.043	0.049	-	-
Change in delayed word recall [a]	-0.086	0.039	-	-
IQCODE	-	-	-2.501	0.603
Proxy respondent status two waves prior	-	-	2.098	1.196
Change in IQCODE [b]	-	-	1.643	0.689
Date recall two waves prior [c]	-	-	0.571	0.362
Serial 7 two waves prior [c]	-	-	0.425	0.239
President recall two waves prior [c]	-	-	-0.847	1.167
Immediate word recall two waves prior [c]	-	-	0.167	0.268
Delayed word recall two waves prior [c]	-	-	0.150	0.325
cut point 1	0.274	0.552	-7.634	2.328
cut point 2	1.780	0.557	-5.861	2.269

Abbreviations: CIND = cognitive impairment, no dementia; LTHS = less than high school education; HSGED = completed high school or general educational development deploma; GTHS = greater than high school education; ADL = basic activities of daily living; IADL = instrumental activities of daily living; IQCODE = Informant Questionnaire on Cognitive Decline

[a] Change in score between immediate prior HRS wave and two waves prior for those who were self-repsondents in both waves

[b] Change in score between immediate prior HRS wave and two waves prior for those who had proxy-repsondents in both waves

[c] Self-cognition item scores from two waves prior for those who had a proxy-respondent in immediate prior HRS wave, but were self-respondents two waves prior

eTABLE 3. Descriptive statistics of HRS-ADAMS linked datasets used in training the expert logistic model and in training the machine learning models

Predictors	**Expert logistic model sample (N=1917)**		**Machine Learning sample (N=1688)**		**Common Sample (N=1571)**	
	Unweighted	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
Dementia status (%)	40%	**17%**	38%	**15%**	37%	**15%**
Proxy respondent (%)	21%	**9%**	20%	**9%**	22%	**9%**
Demographics						
Age (mean)	81.9	**79.2**	81.8	**79.1**	81.7	**79.0**
Male (%)	42%	**41%**	42%	**41%**	41%	**40%**
LTHS (%)	48%	**31%**	47%	**29%**	46%	**28%**
Race/ethnicity						
Non-Hispanic White (%)	70%	**84%**	71%	**84%**	72%	**85%**
Non-Hispanic Black (%)	21%	**9%**	20%	**9%**	20%	**9%**
Hispanic (%)	10%	**7%**	9%	**7%**	8%	**6%**
Physical Health						
At least 1 ADL (%)	43%	**29%**	42%	**28%**	42%	**28%**
At leas 1 IADL (%)	44%	**24%**	43%	**24%**	43%	**23%**
Excellent, very good, or good self-rated health (%)	55%	**68%**	57%	**69%**	57%	**70%**
Social engagement						
Married with spouse present, or partnered (%)	39%	**47%**	40%	**48%**	39%	**47%**
Spent any time volunteering in past year (%)	18%	**29%**	19%	**29%**	19%	**30%**
Spent any time helping friends, neighbors, or relatives in past year (%)	31%	**43%**	32%	**43%**	32%	**44%**
Cognition (self-respondents)	(N=1520)		(N=1349)		(N=1233)	
Immediate word recall, 0-10 (mean)	3.8	**4.9**	3.9	**4.9**	4.0	**5.0**
Delayed word recall, 0-10 (mean)	2.5	**3.7**	2.6	**3.8**	2.7	**3.8**
Serial 7's, 0-5 (mean)	2.5	**3.5**	2.6	**3.5**	2.6	**3.5**
At least one date recall item incorrect (%)	43%	**25%**	42%	**24%**	40%	**23%**
At least one of scissors, cactus, president, or VP naming incorrect (%)	53%	**28%**	51%	**27%**	50%	**26%**
Cognition (proxy-respondents)	(N=397)		(N=339)		(N=338)	
Proxy-rated memory, 1 (excellent) - 5 (poor) (mean)	4.3	**4.0**	4.3	**4.1**	4.3	**4.1**
16-item IQCODE, 1 (much improved) - 5 (much worse) (mean)	4.1	**3.8**	4.1	**3.9**	4.1	**3.9**
Jorm symptoms, 0-5 (mean) [a]	1.8	**1.5**	1.9	**1.7**	1.9	**1.7**

Abbreviations: LTHS = Less than high school education; ADLs = Basic activities of daily living; IADLs = Instrumental activities of daily living; VP = vice-president; IQCODE = the Jorm Informant Questionnaire for Cognitive Decline

[a] Jorm symptoms include: hallucinations, getting lost, memory, wandering, ability to be left alone

eTABLE 4. Expert logistic model regression coefficients [a], estimated on full dataset containing all HRS-ADAMS linked observations with complete data on model predictors (N=1917)

Predictor variable	Coefficient estimate	SE	p-value
Intercept	-1.023	0.832	0.219
Age (centered at 70) [b]	0.197	0.050	<.001
Age squared	-0.005	0.002	0.008
Male	-0.191	0.229	0.404
Non-Hispanic Black	1.598	0.646	0.014
Hispanic	-0.754	0.739	0.308
LTHS	-0.183	0.430	0.671
LTHS X non-Hispanic black	-1.137	0.566	0.045
LTHS X Hispanic	-0.127	0.627	0.839
Self-reported Health status [c]	-1.292	0.687	0.060
Has at least 1 ADL	-0.309	0.514	0.548
Has at least 1 IADL	0.830	0.504	0.100
Age X IADLs	-0.005	0.034	0.878
Past diagnosis of diabetes	0.407	0.250	0.103
Social engagement [d]	0.773	0.288	0.007
Proxy respondent status	0.517	1.354	0.703
Serial 7 score	-0.164	0.149	0.269
Serial 7 score X non-Hispanic black	-0.283	0.221	0.200
Serial 7 score X Hispanic	-0.184	0.203	0.365
Serial 7 score X LTHS	-0.065	0.141	0.645
Serial 7 score X self-reported health status	0.240	0.146	0.101
Immediate word recall score	-0.774	0.160	<.001
Delayed word recall score	-0.269	0.121	0.026
Immediate word recall X self-reported health status	0.434	0.191	0.024
Delayed word recall X self-reported health status	-0.426	0.150	0.005
Date recall [e]	0.813	0.297	0.006
Date recall X has at least 1 ADL	0.842	0.481	0.080
Name recall [f]	1.367	0.330	<.001
Name recall X non-Hispanic black	-0.981	0.500	0.050
Name recall X Hispanic	0.230	0.747	0.758
Name recall X has at least 1 ADL	-0.460	0.505	0.363
IQCODE (centered at 5)	0.541	0.372	0.146
Proxy memory score (centered at 5)	0.341	0.334	0.308
Jorm symptoms [g]	0.652	0.528	0.217
Proxy status X age	0.031	0.038	0.414
Proxy status X LTHS	-0.241	0.756	0.751
Proxy status X self-reported health	1.575	0.813	0.053
Proxy status X male	-1.711	0.597	0.004
Proxy status X has at least 1 ADL	0.295	0.756	0.697
Proxy status X has at least 1 IADL	0.160	0.679	0.814

Abbreviations: LTHS = Less than high school education; ADLs = Basic activities of daily living; IADLs = Instrumental activities of daily living; VP = vice-president; IQCODE = the Jorm Informant Questionnaire for Cognitive Decline

[a] Coefficients are unexponentiated

[b] All interaction terms with age use age centered at 70

[c] Coded as binary: 1 = excellent, very good, or good; 0 = fair or poor

[d] Social engagement coded as binary: 1 = 0 or 1 in summary score; 0 = 2-7 in summary score. Summary score computed by summing spouse/partner present (1 pt), spent 1-99 hours (1 pt) or 100-100 hrs (2 pts) or 200+ hrs (3pts) volunteering in past 12 months, and spent 1-99 hours (1 pt) or 100-100 hrs (2 pts) or 200+ hrs (3pts) helping friends, neighbors, or relatives in past 12 months

[e] Date recall coded as 1 if respondent recalled at least one of the following incorrectly: date day of the week, month, year

[f] Name recall coded as 1 if respondent named at least one of the following incorrectly: scissors, cactus, president, vice-president

[g] Jorm symptoms include: hallucinations, getting lost, memory, wandering, ability to be left alone

eTABLE 5. LASSO model regression coefficients,[a] estimated on full dataset containing all HRS-ADAMS linked observations with complete data on model predictors (N=1688)

Predictor variable	**Coefficient estimate**
Intercept	2.758
Male	-0.327
Hispanic	-0.291
LTHS	-0.134
LTHS X non-Hispanic black	-0.576
Age (centered at 70)	0.048
Immediate word recall score	-0.134
Delayed word recall score	-0.453
Immediate word recall score squared	-0.016
Serial 7 score	0.399
Backwards counting score	0.075
Wu algorithm TICS score [b]	-0.346
Wu algorithm TICS score squared	-0.007
Number of IADLs	0.250
Has at least one ADL	-0.130
Difficulty eating	0.498
Difficulty bathing	0.077
Difficulty dressing	-0.505
Difficulty using the phone	-0.051
Self-reported health (excellent or very good)	0.214
Change in self-reported health status [c]	0.001
Change in number of ADLs	-0.003
Has history of high blood pressure	0.023
BMI (centered at 25)	-0.006
BMI category [d]	-0.100
Drinking status [e]	-0.151
Retirement status [f]	0.038
Hours spent volunteering in past 12 months [g]	-0.035
Spent any time helping others in past 12 months	-0.439
Social engagement score (out of 7) [h]	-0.027
Change in number of IADLs	-0.113
IQCODE (centered at 5)	1.239
Proxy memory score (centered at 5)	0.112
Jorm symptoms [i]	0.616
Current wave self-respondent, last wave proxy-respondent indicator	0.287
Change in delayed word recall score [j]	0.167
Change in TICS score [j]	-0.020

eTABLE 5 (cont.). LASSO model regression coefficients [a],estimated on full dataset containing all HRS-ADAMS linked observations with complete data on model predictors (N=1688)

Predictor variable	**Coefficient estimate**
Change in date recall [j, k]	-0.034
Change in name recall [j, l]	0.281
Change in backwards counting score [j]	-0.161
Change in IQCODE [m]	-1.224
Change in proxy memory score [m]	-0.229
Lagged squared immediate word recall score [n]	-0.020
Lagged delayed word recall score [n]	-0.135
Lagged squared Wu algorithm TICS score [n]	-0.003
Lagged date recall [n, k]	0.053
Delayed wore recall X Male	-0.040
IQCODE X Male	0.100
Serial 7 score X self-reported health (excellent or very good)	0.114
Serial 7 score X self-reported health (excellent, very good, or good)	0.017
Delayed word recall score X self-reported health (excellent or very good)	-0.138
Date recall X at least 1 ADL	0.387
Name recall X at least 1 ADL	-0.037
Name recall X non-Hispanic black	-0.231
Proxy status X self-reported health (excellent or very good)	0.435
Proxy status X Male	-0.430
Proxy status X Bathing difficulty	0.164

Abbreviations: LTHS = Less than high school education; TICS = Telephone interview for cognitive status;ADLs = Basic activities of daily living; IADLs = Instrumental activities of daily living; VP = vice-president; IQCODE = the Jorm Informant Questionnaire for Cognitive Decline

[a] Coefficients are unexponentiated

[b] Wu algorithm TICS score (0-13) is the sum of date recall (date, day of week, month, year), name recall (cactus, president, vice-president), serial 7 score, and backwards counting (first attempt only)

[c] Change in continuous self-reported health status ranges from -4 to 4 (continuous self-reported health status ranges from 1 (excellent) to 5 (poor)

[d] 1 = Underweight, 2 = Normal, 3 = Overweight, 4 = Obese (https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html)

[e] 0 = Non-drinker, 1 = Moderate drinker (1 drink/day for women, 2 drinks/day for men), 2 = More than moderate drinker

[f] -1 = Proxy/Irrelevnt, 0 = Not retired, 1 = 0-2 years since retirement, 2 = 3-5 years since retirement, 3 = 6-19 years since retirement, 4 = 11-15 years since retirement, 5 = 15+ years since retirement

[g] 0 = 0 hrs, 1 = 1-99 hours, 2 = 100-199 hours, 3 = 200+ hours

[h] Sum of spouse/partner present (1 pt), spent 1-99 hours (1 pt) or 100-100 hrs (2 pts) or 200+ hrs (3pts) volunteering in past 12 months, and spent 1-99 hours (1 pt) or 100-100 hrs (2 pts) or 200+ hrs (3pts) helping friends, neighbors, or relatives in past 12 months

[i] Jorm symptoms include: hallucinations, getting lost, memory, wandering, ability to be left alone

[j] Change in self-respondent cognition scores between current and prior waves is set to 0 for current wave proxy-respondents

[k] Date recall coded as 1 if respondent recalled at least one of the following incorrectly: date day of the week, month, year

[l] Name recall coded as 1 if respondent named at least one of the following incorrectly: scissors, cactus, president, vice-president

[m] Change in proxy-respondent cognition scores between current and prior waves is set to 0 for current wave self-respondents, and current wave proxy-respondents who were self-respondents in prior wave

[n] Lagged self-respondent cognition scores are available for current wave proxy-respondents who were self-respondents in prior wave; set to 0 for current wave self-respondents, and current wave proxy respondents also with proxy respondents in prior wave.

eTABLE 6. Comparison of race/ethnicity-specific dementia prevalences, and prevalence ratios between race/ethnicity groups based on ADAMS gold-standard diagnoses vs. algorithmic diagnoses in waves A & B, weighted and unweighted

	ADAMS (true) diagnoses	**Algorithmic diagnoses** Estimated prevalence (difference between estimated and true prevalence) Estimated prevalence ratio (ratio of estimated to true prevalence ratio)					
		Hurd probit model	**Expert Model**	**Gradient boosting model**	**Conditional random forests**	**LASSO**	**Super Learner**
Analyses weighted to represent the US age >70 population in 2002							
True and estimated dementia prevalence							
Non-Hispanic white	**13.7%**	16% (2.3%)	18.2% (4.5%)	18.2% (4.5%)	18.2% (4.5%)	19.4% (5.7%)	19.3% (5.6%)
Non-Hispanic black	**23.3%**	28.1% (4.8%)	27.4% (4.1%)	29.2% (5.9%)	27.7% (4.4%)	31.4% (8.1%)	30.4% (7.1%)
Hispanic	**17.9%**	16.9% (-1%)	22.7% (4.8%)	26.5% (8.6%)	28% (10.1%)	31.5% (13.6%)	30.5% (12.6%)
True and estimated prevalence ratios							
Non-Hispanic black vs. white	**1.70**	1.75 (1.03)	1.5 (0.89)	1.6 (0.94)	1.52 (0.9)	1.62 (0.95)	1.57 (0.93)
Hispanic vs. Non-Hispanic white	**1.31**	1.06 (0.81)	1.25 (0.95)	1.45 (1.11)	1.54 (1.18)	1.62 (1.24)	1.58 (1.21)
Non-Hispanic black vs. Hispanic	**1.30**	1.66 (1.27)	1.21 (0.93)	1.1 (0.85)	0.99 (0.76)	1 (0.77)	1 (0.77)
Unweighted analyses							
True and estimated dementia prevalence							
Non-Hispanic white	**40.6%**	48.1% (7.5%)	49.4% (8.7%)	50.4% (9.8%)	50.5% (9.9%)	52% (11.3%)	51.7% (11.1%)
Non-Hispanic black	**46.9%**	57.2% (10.3%)	56.5% (9.6%)	61.2% (14.3%)	59.4% (12.5%)	63.1% (16.2%)	60.1% (13.2%)
Hispanic	**39.7%**	52.3% (12.6%)	42.3% (2.6%)	48.5% (8.8%)	50.6% (10.9%)	61% (21.3%)	56.9% (17.2%)
True and estimated prevalence ratios							
Non-Hispanic black vs. white	**1.15**	1.19 (1.03)	1.15 (0.99)	1.21 (1.05)	1.18 (1.02)	1.21 (1.05)	1.16 (1.01)
Hispanic vs. Non-Hispanic white	**0.98**	1.09 (1.11)	0.86 (0.88)	0.96 (0.99)	1 (1.03)	1.17 (1.2)	1.1 (1.13)
Non-Hispanic black vs. Hispanic	**1.18**	1.09 (0.92)	1.34 (1.13)	1.26 (1.07)	1.17 (0.99)	1.03 (0.87)	1.06 (0.89)

eTABLE 7. Comparison of race/ethnicity-specific dementia prevalences, and prevalence ratios between race/ethnicity groups based on ADAMS gold-standard diagnoses vs. algorithmic diagnoses in waves C & D, weighted and unweighted

	ADAMS (true) diagnoses	**Algorithmic diagnoses** Estimated prevalence (difference between estimated and true prevalence) Estimated prevalence ratio (ratio of estimated to true prevalence ratio)					
		Hurd probit model	**Expert Model**	**Gradient boosting model**	**Conditional random forests**	**LASSO**	**Super Learner**
Analyses weighted to represent the US age >70 population in 2006							
True and estimated dementia prevalence							
Non-Hispanic white	**17.9%**	18.2% (0.4%)	17.2% (-0.6%)	18.9% (1%)	17.1% (-0.7%)	18.2% (0.4%)	18% (0.2%)
Non-Hispanic black	**28.1%**	27.3% (-0.8%)	29.3% (1.2%)	29.4% (1.3%)	29.2% (1.1%)	29.5% (1.4%)	28.1% (0%)
Hispanic	**21.4%**	28.3% (7%)	22.4% (1%)	22% (0.7%)	22.3% (0.9%)	20.7% (-0.7%)	21.9% (0.6%)
True and estimated prevalence ratios							
Non-Hispanic black vs. white	**1.57**	1.5 (0.95)	1.7 (1.08)	1.56 (0.99)	1.7 (1.08)	1.62 (1.03)	1.56 (0.99)
Hispanic vs. Non-Hispanic white	**1.20**	1.55 (1.3)	1.3 (1.08)	1.17 (0.98)	1.3 (1.09)	1.13 (0.95)	1.22 (1.02)
Non-Hispanic black vs. Hispanic	**1.31**	0.96 (0.73)	1.31 (1)	1.33 (1.01)	1.31 (1)	1.43 (1.09)	1.28 (0.98)
Unweighted analyses							
True and estimated dementia prevalence							
Non-Hispanic white	**31.6%**	34.6% (3%)	32.7% (1.1%)	34.7% (3.1%)	33.2% (1.6%)	33.9% (2.3%)	34.1% (2.5%)
Non-Hispanic black	**47.5%**	54.4% (6.9%)	54.5% (7%)	57.1% (9.6%)	55% (7.5%)	57.1% (9.6%)	54.8% (7.2%)
Hispanic	**51.6%**	58.7% (7.1%)	46.5% (-5.1%)	48.3% (-3.2%)	48.3% (-3.2%)	46.7% (-4.9%)	48.3% (-3.2%)
True and estimated prevalence ratios							
Non-Hispanic black vs. white	**1.50**	1.57 (1.05)	1.67 (1.11)	1.65 (1.1)	1.65 (1.1)	1.69 (1.12)	1.6 (1.07)
Hispanic vs. Non-Hispanic white	**1.63**	1.7 (1.04)	1.42 (0.87)	1.39 (0.85)	1.45 (0.89)	1.38 (0.84)	1.42 (0.87)
Non-Hispanic black vs. Hispanic	**0.92**	0.93 (1.01)	1.17 (1.27)	1.18 (1.28)	1.14 (1.23)	1.22 (1.33)	1.13 (1.23)

eTABLE 8. Comparison of race/ethnicity-specific dementia prevalences, and prevalence ratios between race/ethnicity groups based on ADAMS gold-standard diagnoses vs. algorithmic diagnoses in unweighted analyses

	ADAMS (true) diagnoses	**Algorithmic diagnoses** Estimated prevalence (difference between estimated and true prevalence) Estimated prevalence ratio (ratio of estimated to true prevalence ratio)					
		Hurd probit model	**Expert Model**	**Gradient boosting model**	**Conditional random forests**	**LASSO**	**Super Learner**
Full training sample (waves A, B, C and D)							
True and estimated dementia prevalence							
Non-Hispanic white	**37.3%**	43.1% (5.8%)	43.2% (5.9%)	44.4% (7.1%)	43.9% (6.6%)	45.1% (7.8%)	45% (7.7%)
Non-Hispanic black	**47.1%**	56.2% (9.1%)	55.8% (8.7%)	59.7% (12.6%)	57.8% (10.6%)	60.9% (13.8%)	58.2% (11.1%)
Hispanic	**43.8%**	54.4% (10.7%)	43.8% (0%)	48.1% (4.3%)	49.4% (5.6%)	55.4% (11.6%)	53.4% (9.6%)
True and estimated prevalence ratios							
Non-Hispanic black vs. white	**1.26**	1.31 (1.03)	1.29 (1.02)	1.34 (1.06)	1.31 (1.04)	1.35 (1.07)	1.29 (1.02)
Hispanic vs. Non-Hispanic white	**1.17**	1.26 (1.08)	1.01 (0.86)	1.08 (0.92)	1.12 (0.96)	1.23 (1.05)	1.19 (1.01)
Non-Hispanic black vs. Hispanic	**1.08**	1.03 (0.96)	1.28 (1.19)	1.24 (1.15)	1.17 (1.09)	1.1 (1.02)	1.09 (1.01)
Wave A only							
True and estimated dementia prevalence							
Non-Hispanic white	**35.1%**	42.5% (7.3%)	43.8% (8.7%)	45.9% (10.7%)	45.3% (10.2%)	46.4% (11.3%)	46.8% (11.7%)
Non-Hispanic black	**42.1%**	56.7% (14.5%)	53.5% (11.3%)	59.7% (17.5%)	55.5% (13.3%)	61.1% (18.9%)	59% (16.9%)
Hispanic	**27.4%**	45.7% (18.3%)	38.4% (11%)	46.1% (18.7%)	47.7% (20.3%)	58.4% (31%)	55.3% (27.9%)
True and estimated prevalence ratios							
Non-Hispanic black vs. white	**1.20**	1.33 (1.11)	1.22 (1.02)	1.3 (1.08)	1.22 (1.02)	1.32 (1.1)	1.26 (1.05)
Hispanic vs. Non-Hispanic white	**0.78**	1.08 (1.38)	0.87 (1.12)	1.01 (1.29)	1.05 (1.35)	1.26 (1.61)	1.18 (1.51)
Non-Hispanic black vs. Hispanic	**1.54**	1.24 (0.81)	1.39 (0.91)	1.29 (0.84)	1.16 (0.76)	1.05 (0.68)	1.07 (0.69)

eTABLE 9. Bootstrapped out-of-sample predictive performance of recommended models at chosen race/ethnicity-specific probability thresholds [a]

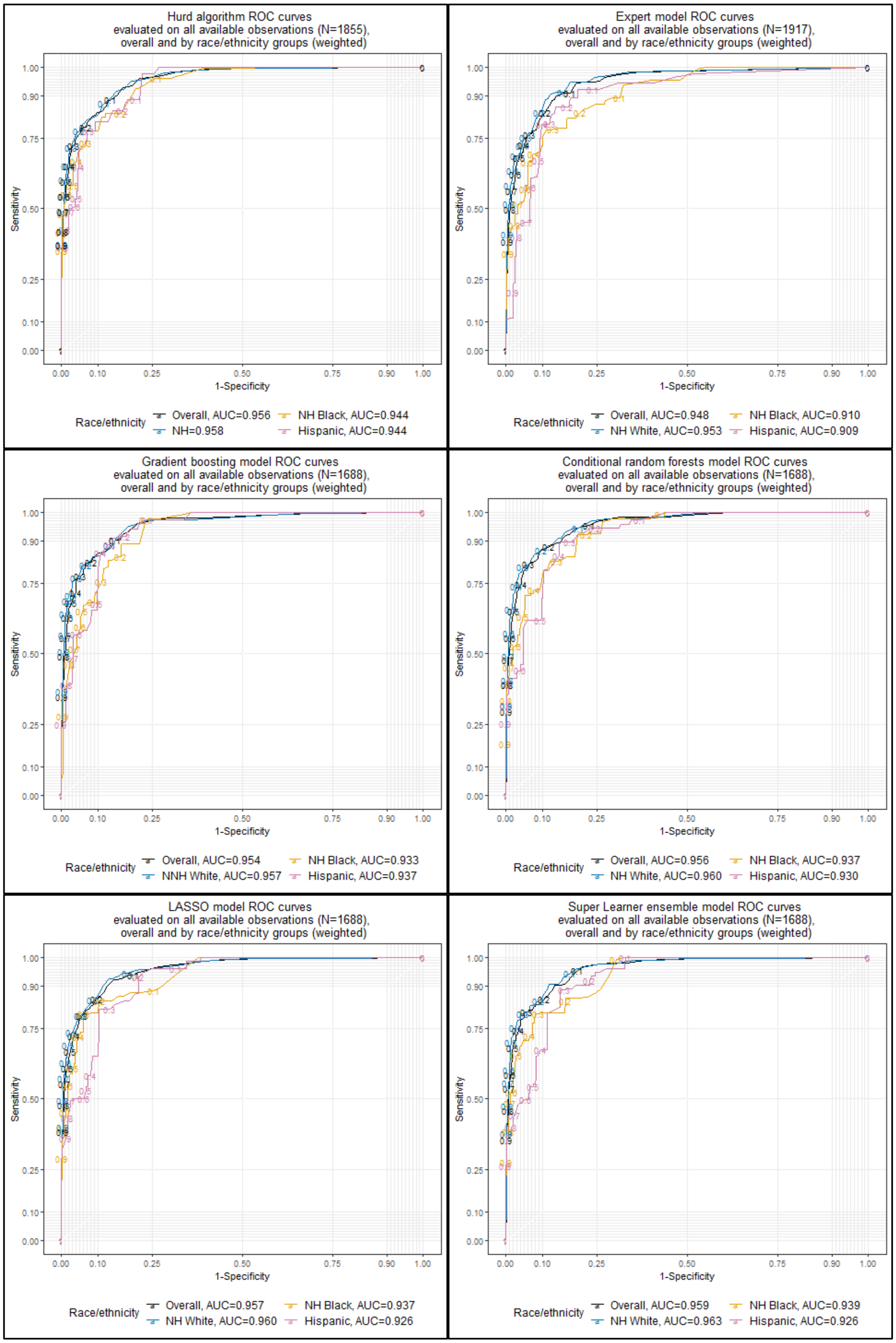
	Sensitivity % (95% CI)	**Specificity** % (95% CI)	**Overall accuracy** % (95% CI)
Hurd model			
non-Hispanic white	79% (73%-86%)	95% (93%-96%)	92% (91%-94%)
non-Hispanic black	78% (66%-89%)	90% (85%-94%)	87% (82%-91%)
Hispanic	82% (63%-95%)	91% (82%-96%)	89% (81%-95%)
Overall	79% (73%-85%)	94% (92%-95%)	92% (90%-93%)
LASSO model			
non-Hispanic white	83% (76%-89%)	93% (91%-95%)	91% (90%-93%)
non-Hispanic black	85% (72%-96%)	89% (83%-93%)	88% (83%-92%)
Hispanic	82% (68%-93%)	90% (80%-96%)	88% (81%-94%)
Overall	83% (77%-88%)	92% (91%-94%)	91% (89%-92%)
Expert Model			
non-Hispanic white	77% (70%-83%)	93% (91%-95%)	91% (89%-92%)
non-Hispanic black	79% (65%-90%)	89% (83%-93%)	86% (80%-91%)
Hispanic	75% (60%-88%)	91% (83%-96%)	88% (80%-93%)
Overall	77% (71%-82%)	93% (91%-94%)	90% (88%-92%)

[a] Point estimates provided here are obtained through the bootstrap percentile method (i.e. the 50th percentile of the statistic distribution across bootstrap samples), and thus may differ slightly from those presented in Table 3.

eTABLE 10. Bootstrapped, weighted in-sample performance of Expert model and LASSO model

	Sensitivity % (95% CI)	**Specificity** % (95% CI)	**Accuracy** % (95% CI)
LASSO Model			
Non-Hispanic White	84% (78%-90%)	93% (92%-95%)	92% (90%-94%)
Non-Hispanic Black	85% (72%-96%)	90% (85%-93%)	89% (84%-92%)
Hispanic	82% (68%-93%)	91% (82%-96%)	89% (82%-94%)
Overall	84% (79%-89%)	93% (91%-94%)	92% (90%-93%)
Expert Model			
Non-Hispanic White	80% (73%-85%)	94% (93%-95%)	92% (90%-93%)
Non-Hispanic Black	81% (71%-90%)	90% (85%-94%)	88% (83%-91%)
Hispanic	78% (63%-90%)	95% (90%-98%)	91% (86%-95%)
Overall	80% (75%-84%)	94% (92%-95%)	91% (90%-93%)

eFIGURE 1. ROC curves of models that meet pre-specified criteria using racial/ethnic-specific threshold



eAPPENDIX 1. Computation of Weights Used in Analyses

Not all participants in ADAMS were assessed at each wave.

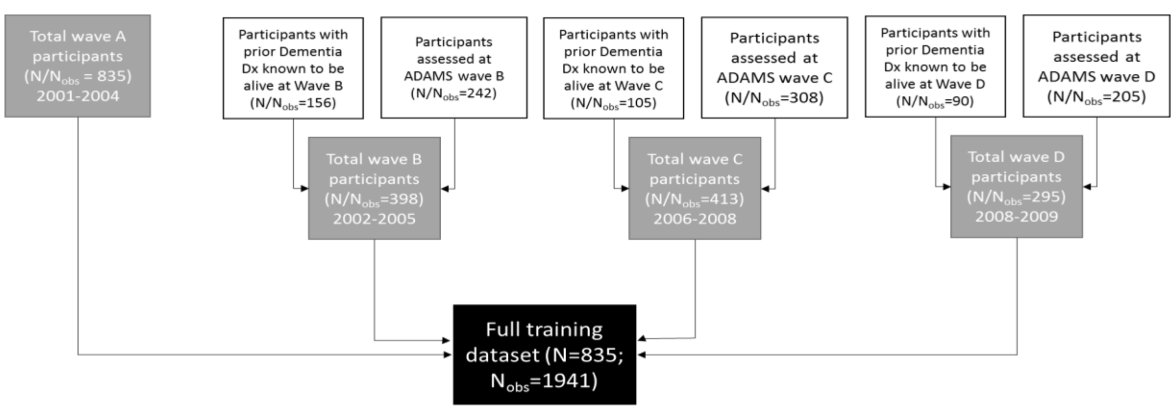
Wave A contains data on all ADAMS participants. ADAMS investigators invited a non-random subset of ADAMS participants to Wave B, with selection based on prior cognitive status. Thus, Wave B is not representative of all ADAMS participants.

All living ADAMS participants without a prior dementia diagnosis were invited to participate in Wave C. Wave D, like Wave B, included a non-random and non-representative subset of Wave C participants with participation based on prior cognitive status.

For the purpose of our study, we created a dataset that included observations for ADAMS participants after their initial dementia diagnosis (**Figure 1**). For participants who were not re-examined at Waves B, C, or D due to a prior dementia diagnosis but were known to be alive, we assigned a diagnosis of dementia at the median ADAMS wave date and linked this to HRS data from the corresponding prior HRS interview wave.

Given this change, we did not think use of the derived longitudinal weights provided by ADAMS to recover the US >70 population to be appropriate in our analyses. Instead, we constructed a new set of weights using the procedures described in this technical appendix. Our larger goal was to be able to create an algorithm that could assign dementia status at a given time point based on HRS data from that time point. Thus, we created a set of weights allowing us to weight the Wave A and B data to be representative of the 2002 US population and to weight the Wave C and D data to be representative of the 2006 US population. Taken together, we view these weights as enabling estimation of our algorithm using the combined data of two, nationally-representative, cross-sectional surveys. These weights derived as we describe here are not appropriate for persons interested in estimating transitions across ADAMS waves – the derived longitudinal weights provided by ADAMS are available for this purpose.

Figure 1. Derivation of dataset



Construction of wave A and B weights

We began with ADAMS wave A cross-sectional analyses weights provided by ADAMS.[1] Because follow-up assessments in wave B were only attempted for a non-random subset of respondents who were diagnosed with dementia, cognitive impairment no dementia (CIND), mild dementia, or borderline normal cognition in wave A,[2] we treated wave A and wave B observations as one unit in our weighting procedure. For participants who only appeared in wave A, we retained the original ADAMS wave A weights; for participants who appeared in both waves A and B, we halved the original ADAMS wave A weights, and assigned the new weight across both waves. When properly weighted, the sample of pooled observations from waves A and B is representative of the July 2002 US >70 population. (Note, however, that this is based on a minimum age of 70 at the corresponding 2000 or 2002 HRS study visit, as the minimum age at the ADAMS assessment was closer to 71 given the lag between the HRS and ADAMS study visits).[1]

Construction of wave C and D weights

At wave C, follow-up visits were attempted for all wave A participants who were still alive in 2006 and who had not previously been diagnosed with dementia. In the dataset we constructed for training our algorithm, wave C observations include all those who were successfully assessed at wave C, as well as those who were not assessed due to a prior dementia diagnosis but were known to be alive on April 1, 2007 (the median date of wave C assessments). Thus, we treated wave C observations as a new cohort, and constructed a new set of weights for wave C observations to be representative of the US >70 population (described below). As with wave B and wave A, wave D assessments were attempted for a non-random subset of those identified for follow-up during wave C.[2] Thus, we similarly treated observations from waves C and D as one unit. Participants who only appeared in wave C retained their newly constructed wave C weights. Participants who appeared in both waves C and D received half of the newly constructed wave C weight in each wave. When properly weighted, the sample of pooled observations from waves C and D is representative of the July 2006 US 72+ population; we were unable to recover the >70 population (as in wave A) due to the aging of the cohort: the youngest wave C observation was 72 at the time of completing the closest prior HRS interview.

We followed the ADAMS procedure for deriving wave A cross-sectional analyses weights,[1] with some necessary modification, to derive our new wave C weights. We chose to weight the wave C sample to the July 2006 population given that the closest prior HRS interview occurred in 2006 for 84% of all persons eligible for Wave C participation and all those who were not eligible due to a previous dementia diagnosis but known to be alive at wave C. Parallel to the process used to derive ADAMS wave A cross-sectional analysis weights, we began with wave C participants' HRS base weights and made a series of adjustments (described below) to arrive at our final analyses weights assigned to our wave C and D observations.

STEP 1 – adjustment for ADAMS sample sub-selection:

To adjust for selection into the wave C training dataset sample, we first assigned all Wave C observations (participants selected for Wave C participation and Wave C non-participants with dementia for whom we created a new observation) to one of the original 18 ADAMS strata, which are based on combinations of respondent status (self- vs. proxy-response), gender, age, and cognition. We used participants' age at HRS interview for this step and all subsequent stratification steps given that age at ADAMS assessment is not available for participants who were lost to follow-up or died prior to ADAMS assessment in the original ADAMS wave C sample, or for participants included in our expanded sample who were not selected for ADAMS wave C assessment but who were previously diagnosed with dementia and known to be alive at the time wave C assessments took place.

While the original ADAMS procedure used 18 strata, we had to collapse two of the original strata (proxy-response men aged 70-79 (stratum 15) and 80+ (stratum 17) with "normal: medium" cognition) into a single stratum due to having no remaining survivors in stratum 15 survivors by wave C, resulting in a total of 17 strata. We then computed a sub-selection weight adjustment factor $W_{ADAMSsub,}$ by taking the inverse of the sampling rate for ADAMS observations from HRS observations in each stratum, *h*, as follows:

$$W_{h,ADAMSSub} = \left(\frac{ADAMS_{h,n}}{HRS_{h,n}}\right)^{-1}$$

where:

$ADAMS_{h,n}$ = total unweighted ADAMS wave C observations (including those sampled for wave C assessment but who died prior to contact or who did not respond for other reasons) in stratums h=1 to h=17

$HRS_{h,n}$ = total unweighted 2006 HRS observations in stratums h=1 to h=17.

The total ADAMS cases, HRS cases, and sub-selection weights for each stratum are provided in **Table 1**. Due to sample attrition, the ADAMS sub-selection weight adjustment factors computed for wave C are approximately three to eight times greater than correspondent sub-selection adjustment factors computed for ADAMS wave A (Heeringa et al. 2009, p. 18).

Note that 3 observations from the full wave C sample (N=1 assessed in wave C, and N=2 not assessed at wave C, but previously diagnosed with dementia and known to be alive at wave C) were dropped at this stage due to missing cognition data in the immediate prior HRS interview. Similarly, 26 observations from the 2006 HRS survey and 12 observations from the 2004 HRS survey that met the age criterion were dropped due to missing cognition data. Note that we used information from the 2004 HRS interview to assign persons to strata for the 16% of observations whose closest prior HRS interview occurred in 2004. We believe that stratum frequencies and mean base weights in the 2004 and 2006 HRS waves are sufficiently similar (**Table 2**) to ensure that the use of 2004 data on age, gender, cognition, and response status for the 16% subset of observations does not significantly affect our computed weights to recover the 2006 U.S. 70+ population.

Table 1. ADAMS sub-sample weight adjustment factors by strata							
Stratum	Cognition	Cognitive score	Age	Gender	$ADAMS_{h,n}$	$HRS_{h,n}$	$W_{h,ADAMSSub}$
Self-response (summary self-response cognition score, 0-35)							
1	Low function	0-9	70+	M, F	54	227	4.204
2	Borderline	9-22	70+	M, F	30	229	7.633
3	Normal: Low	12-16	70+	M, F	111	995	8.964
4	Normal: Med	17-19	70-79	M	20	442	22.100
5	Normal: Med	17-19	70-79	F	22	487	22.136
6	Normal: Med	17-19	80+	M	32	261	8.156
7	Normal: Med	17-19	80+	F	38	431	11.342
8	Normal: High	20-35	70-79	M	48	1267	26.396
9	Normal: High	20-35	70-79	F	41	1704	41.561
10	Normal: High	20-35	80+	M	38	432	11.368
11	Normal: High	20-35	80+	F	57	682	11.965
Proxy-response (cognitive score based on Jorm IQCODE, 1-5)							
12	Low function	3.90-5.00	70+	M, F	38	279	7.342
13	Borderline	3.35-3.89	70+	M, F	12	102	8.500
14	Normal: Low	3.10-3.34	70+	M, F	9	102	11.333
15	Normal: Med	1.00-3.10	70+	M	9	168	18.667
16	Normal: Med	1.00-3.10	70-79	F	4	48	12.000
17	Normal: Med	1.00-3.10	80+	F	8	60	7.500

Table 2. Comparison of stratum frequencies and mean HRS base weight in the 2004 and 2006 HRS and ADAMS wave C samples								
	HRS Sample				ADAMS Sample			
	2004 (N=7701)		2006 (N=7916)		2004 (N=92)		2006 (N=479)	
Stratum	%	Mean weight	%	Mean weight	%	Mean weight	%	Mean weight
1	2.4	2,982	2.9	3,142	6.5	2,404	10.0	3,105
2	2.9	3,053	2.9	3,129	4.4	4,528	5.4	2,453
3	12.9	3,182	12.6	3,155	23.9	3,196	18.6	2,951
4	5.2	3,408	5.6	3,237	4.4	3,563	3.3	3,031
5	6.0	3,543	6.2	3,284	2.2	4,031	4.2	3,543
6	3.4	3,206	3.3	3,434	5.4	4,259	5.6	3,122
7	5.1	3,407	5.4	3,725	6.5	3,001	6.7	3,396
8	15.4	3,373	16.0	3,255	9.8	2,916	8.1	3,249
9	21.2	3,556	21.5	3,432	6.5	5,435	7.3	4,246
10	5.0	3,423	5.5	3,607	8.7	2,992	6.3	3,328
11	8.7	3,577	8.6	3,845	12.0	4,506	9.6	3,696
12	4.5	3,089	3.5	3,310	1.1	4,718	7.7	3,235
13	1.5	3,211	1.3	3,283	1.1	3,451	2.3	3,842
14	1.3	3,433	1.3	3,176	0.0	-	1.9	2,966
15	2.9	3,164	2.1	3,053	4.4	2,189	1.0	2,191
16	0.7	3,396	0.6	3,004	1.1	6,905	0.6	4,037
17	1.2	2,845	0.8	2,964	2.2	3,537	1.3	2,664

The distributions of weights after adjustment for ADAMS sample sub-selection, $Wgt_{ADAMSSel}$, for the total sample and for the survivor sample are provided in **table 3**.

Table 3. Distribution of weights after adjustment for ADAMS sample sub-selection, *Wgt$_{ADAMSSel}$*		
Weight distribution descriptive statistic	Total Wave C Sample	Survivors only
N	*571*	*491*
Mean	*48,520*	*50,129*
Std. Dev.	*52,445*	*55,103*
100%-tile	486,430	486,430
99%-tile	253,896	281,432
95%-tile	145,463	151,781
90%-tile	99,207	102,310
75%-tile	52,107	57,717
50%-tile	32,686	32,504
25%-tile	20,491	20,527
10%-tile	12,847	12,862
5%-tile	9,075	9,804
1%-tile	4,754	4,754
0%-tile	3,830	3,897
Sum	**27,705,205**	**24,613,377**

STEP 2 – adjustment for ADAMS non-response:

Next, we adjusted for non-response by the inverse of the weighted response rate among survivors, by stratum. The adjustment factor was computed as follows:

$$W_{h,ADAMSnr} = \left(\frac{Sum\ of\ \boldsymbol{Wgt_{ADAMSSel}}\ among\ respondents\ in\ stratum\ h}{Sum\ of\ \boldsymbol{Wgt_{ADAMSSel}}\ among\ respondents + nonrespondents\ in\ stratum\ h}\right)^{-1}$$

The stratum-specific non-response adjustment factors, and distribution of resulting weights, *Wgt$_{ADAMSSel,nr}$* across wave C observations (excluding deaths and non-responses) are respectively provided in **Tables 4** and **5.** The sum of weights after adjustment for non-response was 24,613,377, which is comparable to the sum of weights of respondents aged 72+ in the full HRS of 23,112, 557.

Table 4. ADAMS non-response weight adjustment factors by strata							
Stratum	Cognition	Cognitive score	Age	Gender	*Sum of $Wgt_{ADAMSSel}$, Resp + Non-resp.*	*Sum of $Wgt_{ADAMSSel}$, Resp. only*	$W_{h,ADAMSnr}$
Self-response (summary self-response cognition score, 0-35)							
1	Low function	0-9	70+	M, F	615,788	591,150	1.0417
2	Borderline	9-22	70+	M, F	460,450	394,590	1.1669
3	Normal: Low	12-16	70+	M, F	2,351,400	2,014,507	1.1672
4	Normal: Med	17-19	70-79	M	1,055,452	996,423	1.0592
5	Normal: Med	17-19	70-79	F	1,561,167	1,457,193	1.0714
6	Normal: Med	17-19	80+	M	721,102	584,069	1.2346
7	Normal: Med	17-19	80+	F	1,232,592	955,913	1.2894
8	Normal: High	20-35	70-79	M	3,799,786	3,380,409	1.1241
9	Normal: High	20-35	70-79	F	7,211,785	5,671,868	1.2715
10	Normal: High	20-35	80+	M	1,223,106	994,316	1.2301
11	Normal: High	20-35	80+	F	2,384,344	2,028,304	1.1755
Proxy-response (cognitive score based on Jorm IQCODE, 1-5)							
12	Low function	3.90-5.00	70+	M, F	869,702	843,116	1.0315
13	Borderline	3.35-3.89	70+	M, F	359,176	280,517	1.2804
14	Normal: Low	3.10-3.34	70+	M, F	282,087	148,331	1.9017
15	Normal: Med	1.00-3.10	70+	M	220,248	220,248	1.0000
16	Normal: Med	1.00-3.10	70-79	F	145,320	145,320	1.0000
17	Normal: Med	1.00-3.10	80+	F	119,873	119,873	1.0000

Table 5. Distribution of weights after adjustment for ADAMS sample sub-selection and non-response, $Wgt_{ADAMSSel,nr}$	
Weight distribution descriptive statistic	Wave C sample, excluding deaths and non-response
N	*423*
Mean	*58,188*
Std. Dev.	*67,519*
100%-tile	618,496
99%-tile	297,358
95%-tile	167,412
90%-tile	115,003
75%-tile	63,167
50%-tile	39,172
25%-tile	22,935
10%-tile	14,210
5%-tile	10,086
1%-tile	4,953
0%-tile	4,059
Sum	**24,613,377**

STEP 3 – Trimming extreme weight values

Following the original ADAMS procedure, we truncated non-response adjusted weights to the 5th and 95th percentiles of the full sample distribution in order to minimize the influence of extreme values. Weights falling below the 5th percentile (10,086) were replaced by 10,086, and weights above the 95th percentile (167,412) where replaced by 167,412. We then restored the total sum of weights by rescaling the trimmed weights linearly *within* each stratum by the ratio of the sum of weights prior to trimming to the sum of weights post trimming, as follows:

$$W_{h,trimrescale} = \left(\frac{Sum\ of\ Wgt_{ADAMSSel,nr}\ in\ stratum\ h}{Sum\ of\ trimmed\ Wgt_{ADAMSSel,nr}\ in\ stratum\ h}\right)$$

In total, 21 observations had weight values adjusted upwards, and 21 observations had weight values adjusted downwards; the number of trimmed weights and post-trim adjustment factors in each stratum are provided in **Table 6**. Notably, 19 out of the 31 cases in stratum 9 (self-responding females aged 70-79 with “normal: high” cognition) had weights higher than the 95th percentile of the full sample distribution and were trimmed down. This is consequent of (1) the disproportionately high mean weight values of stratum 9 ADAMs participants relative to the weight values of stratum 9 participants in the full HRS cohort (**Table 2**), and (2) the large ADAMS sub-selection adjustment factor (41.6, **Table 1**) applied to stratum 9 participants. The sum of pre-trimmed weights of these 19 cases is 5,703,118, constituting 79% of the total stratum sum of pre-trimmed weights (7,211,785); subsequently, the post-trimming rescaling adjustment factor for stratum 9 is much higher than that of other strata to restore the pre-trimming total weight sum. In contrast, even though the weights of 16 of the 45 cases in stratum 1 (self-responding males and females aged 70+ with low cognitive function) were trimmed up, the sum of the pre-trimmed weights of these 16 cases is 106,376, constituting just 17% of the total stratum sum of pre-trimmed weights (615,788). Subsequently, the post-trimmed weights in stratum 1 required only minimal adjustment to restore the original stratum non-response adjusted weight sum total. The distribution of weights after trimming for extreme weights and rescaling, $Wgt_{ADAMSSel,nr,trim}$, are provided in **Table 7**, demonstrating a narrower distribution of weights compared to the distribution of pre-trimmed weights provided in **Table 5** (while retaining the sum total).

Table 6. Number of observations with trimmed weights, and post-trimming rescaling adjustment factors								
Stratum	Cognition	Cognitive score	Age	Gender	N	N with weights adj. up	N with weights adj. down	Post-trimming adj. factor to restore stratum weight total
Self-response (summary self-response cognition score, 0-35)								
1	Low function	0-9	70+	M, F	45	16	0	0.918
2	Borderline	9-22	70+	M, F	21	2	0	0.992
3	Normal: Low	12-16	70+	M, F	74	1	0	1.000
4	Normal: Med	17-19	70-79	M	15	0	0	1.000
5	Normal: Med	17-19	70-79	F	18	0	0	1.000
6	Normal: Med	17-19	80+	M	23	1	0	0.999
7	Normal: Med	17-19	80+	F	25	0	0	1.000
8	Normal: High	20-35	70-79	M	39	0	2	1.047
9	Normal: High	20-35	70-79	F	30	0	19	1.538
10	Normal: High	20-35	80+	M	27	0	0	1.000
11	Normal: High	20-35	80+	F	43	0	0	1.000
Proxy-response (cognitive score based on Jorm IQCODE, 1-5)								
12	Low function	3.90-5.00	70+	M, F	35	1	0	0.999
13	Borderline	3.35-3.89	70+	M, F	9	0	0	1.000
14	Normal: Low	3.10-3.34	70+	M, F	5	0	0	1.000
15	Normal: Med	1.00-3.10	70+	M	5	0	0	1.000
16	Normal: Med	1.00-3.10	70-79	F	3	0	0	1.000
17	Normal: Med	1.00-3.10	80+	F	6	0	0	1.000
TOTAL					**423**	**21**	**21**	

Table 7. Distribution of weights after trimming for extreme weight values and rescaling to restore pre-trim stratum totals, ***Wgt***$_{ADAMSSel,nr,trim}$	
Weight distribution descriptive statistic	Wave C sample excluding deaths and non-response
N	*423*
Mean	*58,188*
Std. Dev.	*59,871*
100%-tile	257,456
99%-tile	257,456
95%-tile	248,354
90%-tile	120,316
75%-tile	64,649
50%-tile	39,163
25%-tile	22,650
10%-tile	13,493
5%-tile	10,078
1%-tile	9,259
0%-tile	9,259
Sum	**24,613,377**

STEP 4 – Adjusting weights for nursing home residents

The ADAMS wave A cross-sectional weight derivation protocol noted the importance of maintaining an accurate representation of the aged nursing home population in ADAMS, using data from the 2000 Census and the Centers for Medicare and Medicaid Services (CMS) Minimum Data Set (MDS) to determine post-stratification weight adjustments for this population.[1] Because we lacked access to the MDS, we instead used the 2006 HRS nursing home resident indicator and base weights to derive post-stratification nursing home adjustment factors. We believe that this alternate approach is comparable to the original ADAMS approach given that the HRS base weights for nursing home residents were also developed using the census and MDS data. The adjustment factor for each nursing home stratum, *g*, was computed as follows:

$$W_{g,nh} = \left(\frac{Sum\ of\ weights\ in\ 2006\ HRS\ in\ stratum\ g}{Sum\ of\ Wgt_{ADAMSSel,nr,trim}\ in\ stratum\ g}\right)$$

Additionally, while the original ADAMS procedure used nursing home status at time of ADAMS assessment, we used nursing home status at time of HRS interview, given that our expanded training sample included participants who were not assessed, but who were alive and included in our expanded sample due to a dementia diagnosis from ADAMS waves A or B. Finally, the original ADAMS procedure stratified nursing home residents into two age groups (70-79 and 80+) for each gender. Because there were no surviving nursing home males aged 70-79 by ADAMS wave C, we had to collapse across the two male age groups. The size and post-stratification factor for each stratum are provided in **Table 8**. As Heeringa et al. note in the documentation of the weighting procedure for wave A, cell sizes for post-stratification adjustment should ideally be greater than 20; however, given the importance of achieving representation of nursing home residents within gender-age groups, Heeringa et al. chose not to collapse cells across gender or age. Following this logic, we also chose not to collapse further across cells despite small cell sizes in strata 1 and 2.

Of note, observations in nursing home stratum 2 (females aged 70-79) are adjusted down by over 50% in this step. This large adjustment factor is consequent of the disproportionately high weight values of ADAMS stratum 9 cases relative to average weight values of stratum 9 cases from the full HRS cohort (**Table 2**), coupled with this stratum's high sub-selection adjustment factor (**Table 1**), as described previously. This resulted in substantial over-representation of 70-79 year old females in the ADAMS sample compared to the true U.S. population after weighting by the ADAMS sample sub-selection adjusted weights. Trimming of extreme values in step 3 did not remedy this over-representation because weights were subsequently rescaled linearly to restore the pre-trim weight total within each stratum (as noted previously, this resulted in a high adjustment factor of 1.5 in stratum 9).

Table 8. Number of observations and post-stratification adjustment factors, by nursing home stratum

Post-stratification stratum	Age	Gender	N	Nursing home residents		
				Sum of weights in 2006 HRS	Sum of Weights in ADAMS, $Wgt_{ADAMSSel,nr,trim}$	Post-stratification factor
1	70+	M	10	346,814	285,230	1.216
2	70-79	F	5	186,981	413,613	0.452
3	80+	F	35	792,898	947,361	0.837
TOTAL			**50**	**1,326,693**	**1,646,204**	-

STEP 5 – Post-stratification to census 2006 population estimates

In this final step, we post-stratified weights to July 2006 U.S. Census population estimates by age and gender groups (**Table 9**). We believe that the original ADAMS procedure used age at ADAMS assessment to stratify their participants in this final post-stratification step due to use of age 71-74 as the youngest age group, a decision attributed to the one year lag between ADAMS sample recruitment (at age 70 eligibility) and the time at which ADAMS assessment actually took place. While we continued using age at HRS interview for this step (given that our sample includes those who were not assessed at wave C), we followed the same logic and we restricted our youngest age group in this step to 72-74, reflecting the youngest age at which wave C participants drawn into our analytical dataset completed their immediate prior HRS interview (selected for wave A participation at age 70 in 2002, with HRS interview prior to wave C completed in 2004, at age 72).

Additionally, while the original ADAMS weighting procedure separated age groups 85-89 and 90+, we were unable to do so due to data limitations. We used census population estimates from the public-use July 2006 census population survey data (https://www.nber.org/data/cps_basic.html), which censored ages above 85, requiring us to collapse all ages above 85 in our stratification procedure.

The final post-stratification adjustment factors are provided in **Table 9**. Compared to the original ADAMS procedure for deriving wave A weights (Heeringa et al., p. 26), our post-stratification adjustment factors are substantially greater for the two strata in the youngest age group (72-74), and substantially smaller in the oldest age groups (85+) due to the aging of the cohort over time. Additionally, for reasons described in the previous section, our female 75-79 sample is disproportionately over-represented compared to the true US population after weighting by ADAMS sample sub-selection adjusted weights; thus, our census post-stratification adjustment factor for stratum 4 is substantial, reducing weight values by almost 50% to recover the true US population.

Again, following the logic of the original ADAMs weighting procedure, we did not collapse cells further across age or gender despite small cell sizes in strata 1 and 2 due to the importance of representing each gender-age group as accurately as possible.

Table 9. Final post-stratification to 2006 Census adjustment factors, by age-gender stratum

Stratum	Age	Gender	N	July 2006 census population	Sum of weights in ADAMS after nursing home post-stratification adjustment, $\boldsymbol{Wgt}_{ADAMSSel,nr,trim,nh}$	Post-stratification factor
1	72-74	M	16	2,159,760	1,251,006	1.726
2	72-74	F	19	2,624,954	1,837,005	1.429
3	75-79	M	63	3,145,144	4,326,047	0.727
4	75-79	F	58	4,288,615	7,648,412	0.561
5	80-84	M	44	2,058,590	1,501,207	1.371
6	80-84	F	64	3,325,347	2,619,164	1.270
7	85+	M	57	1,376,312	1,776,905	0.775
8	85+	F	102	2,606,731	3,334,119	0.782

The distribution of weights after the final post-stratification adjustment step is provided in **table 9**, with total sum of 21,585,452 corresponding exactly to the total census population estimate for adults aged 72+ in July 2006. **Table 10** demonstrates that the sum of our final weights by age and gender to correspond exactly to the July 2006 US population estimates:

Table 9. Distribution of weights after post-stratification to 2006 census, $Wgt_{ADAMSSel,nr,trim,nh,census}$

Weight distribution descriptive statistic	Wave C sample excluding deaths and non-response
N	*423*
Mean	*51,029*
Std. Dev.	*52,880*
100%-tile	367,887
99%-tile	316,560
95%-tile	144,361
90%-tile	115,060
75%-tile	64,681
50%-tile	35,363
25%-tile	18,986
10%-tile	11,139
5%-tile	7,975
1%-tile	6,830
0%-tile	3,663
Sum	**21,585,452**

Table 10. Comparison of CPS population estimates and sum of final weights by gender and age groups

Gender	Age	July 2006 CPS estimates	Sum of final weights
Male	72-74	2,159,760	2,159,760
Female	72-74	2,624,954	2,624,954
Male	75-79	3,145,144	3,145,144
Female	75-79	4,288,615	4,288,615
Male	80-84	2,058,590	2,058,590
Female	80-84	3,325,347	3,325,347
Male	85+	1,376,312	1,376,312
Female	85+	2,606,731	2,606,731
TOTAL		21,585,452	21,585,452

Finally, to assess reliability of the estimated performance metrics of each algorithm when applied to the larger HRS cohort, we demonstrate that our newly derived weights are adequate to recover annual, nationally-representative HRS samples when applied to our full training dataset based on comparisons of summary statistics of Expert Model predictors (**Table 11**). Similar to the cross-sectional weights provided by ADAMS, our newly derived weights adequately recovered the nationally-representative statistics, as estimated using the full weighted HRS cohorts, with few exceptions: compared to the weighted 2002 HRS sample, the percent of individuals that respond incorrectly to at least one of the four naming recall items (cactus, scissors, president, vice-president) were under-represented in wave A observations (weighted by the cross-sectional weights provided by ADAMS) and wave A and B observations combined (weighted by our newly derived weights) overall, and among non-Hispanic whites and Hispanics. Compared to the weighted 2006 HRS sample, proxy cognition (as measured by proxy memory score and Jorm symptoms) was substantially worse in weighted ADAMS wave C and D observations among non-Hispanic whites and Hispanics. Additionally, percent of non-Hispanic blacks with a less than a high school education were over-represented in the weighted ADAMS samples when compared to the corresponding HRS waves; percent of non-Hispanic blacks with diabetes were over-represented in ADAMS wave A and wave A/B observations compared to the weighted 2002 HRS sample. Due to the small number of Hispanics selected into ADAMS and the lack of stratification by race/ethnicity in the weighting procedure, differences in summary statistics across predictors between the weighted training sample and the weighted HRS cohorts were generally larger among Hispanics than among non-Hispanics.

Table 11. Comparison of summary statistics of expert model predictors in the weighted training dataset vs. the full HRS cohort of individuals aged 70+

Expert model predictors	Training sample wave A only (representative of 2002 population)[a]	Training sample waves A & B (representative of 2002 population)[b]	**HRS wave 2002**	Training sample waves C & D (representative of 2006 population)[b]	**HRS wave 2006**
ALL RACE/ETHNICITY GROUPS					
Proxy respondent (%)	11%	11%	**14%**	8%	**9%**
Male (%)	40%	40%	**39%**	41%	**40%**
Non-Hispanic black (%)	8%	8%	**8%**	10%	**8%**
Hispanic (%)	5%	5%	**5%**	9%	**6%**
Age (mean)	77.9	78.2	**78.9**	80.5	**79.2**
LTHS (%)	32%	32%	**31%**	29%	**27%**
At least 1 ADL limitation (%)	26%	27%	**26%**	31%	**27%**
AT least 1 IADL limitation (%)	22%	23%	**25%**	26%	**26%**
History of diabetes diagnosis (%)	20%	20%	**18%**	18%	**21%**
Excellent/very good/good self-reported health (%)	66%	66%	**65%**	70%	**65%**
Social engagement (%)	55%	56%	**58%**	62%	**60%**
Self-cognition					
Immediate word recall, 0-10 (mean)	4.8	4.8	**4.8**	5.0	**4.7**
Delayed word recall, 0-10 (mean)	3.8	3.7	**3.7**	3.7	**3.5**
Serial 7s, 0-5 (mean)	3.4	3.4	**3.4**	3.6	**3.4**
Date recall[c] (%)	22%	23%	**23%**	26%	**27%**
Object/president/VP recall[d] (%)	29%	30%	**37%**	27%	**29%**

Proxy cognition					
IQCODE, -4(best)-0 (worst) (mean)	-1.3	-1.3	**-1.3**	-1.0	**-1.2**
Proxy memory score, -4(best)-0 (worst) (mean)	-1.4	-1.3	**-1.3**	-0.4	**-1.1**
Jorm symptoms, 0-5 (mean)	1.0	1.1	**1.1**	2.2	**1.3**
NON-HISPANIC WHITES					
Proxy respondent (%)	9%	10%	**13%**	7%	**9%**
Male (%)	40%	40%	**39%**	41%	**40%**
Age (mean)	77.9	78.2	**78.9**	80.5	**79.3**
LTHS (%)	26%	26%	**26%**	21%	**22%**
At least 1 ADL limitation (%)	25%	26%	**25%**	28%	**25%**
AT least 1 IADL limitation (%)	21%	22%	**24%**	24%	**25%**
History of diabetes diagnosis (%)	17%	18%	**17%**	16%	**19%**
Excellent/very good/good self-reported health (%)	68%	68%	**67%**	75%	**68%**
Social engagement (%)	53%	54%	**56%**	58%	**58%**
Self-cognition					
Immediate word recall, 0-10 (mean)	4.9	4.9	**4.9**	5.2	**4.8**
Delayed word recall, 0-10 (mean)	3.8	3.8	**3.8**	3.9	**3.6**
Serial 7s, 0-5 (mean)	3.6	3.6	**3.5**	3.8	**3.6**
Date recall[c] (%)	21%	22%	**22%**	23%	**25%**
Object/president/VP recall[d] (%)	24%	25%	**32%**	20%	**24%**
Proxy-cognition					
IQCODE, -4(best)-0 (worst) (mean)	-1.3	-1.3	**-1.3**	-0.9	**-1.2**
Proxy memory score, -4(best)-0 (worst) (mean)	-1.3	-1.3	**-1.3**	-0.2	**-1.1**
Jorm symptoms, 0-5 (mean)	1.0	1.1	**1.1**	2.4	**1.4**
NON-HISPANIC BLACKS					
Proxy respondent (%)	14%	14%	**20%**	12%	**14%**
Male (%)	44%	44%	**37%**	44%	**36%**
Age (mean)	78.5	78.8	**78.5**	80.9	**78.4**
LTHS (%)	75%	75%	**60%**	68%	**53%**
At least 1 ADL limitation (%)	32%	33%	**34%**	44%	**38%**
AT least 1 IADL limitation (%)	31%	30%	**34%**	35%	**36%**
History of diabetes diagnosis (%)	38%	39%	**26%**	33%	**31%**
Excellent/very good/good self-reported health (%)	50%	51%	**49%**	52%	**50%**
Social engagement (%)	67%	66%	**67%**	68%	**69%**
Self-cognition					
Immediate word recall, 0-10 (mean)	4.3	4.4	**4.2**	3.9	**4.1**
Delayed word recall, 0-10 (mean)	3.1	3.2	**2.8**	2.6	**2.6**

Serial 7s, 0-5 (mean)	1.5	1.5	**1.9**	1.8	**1.9**
Date recall[c] (%)	34%	32%	**34%**	37%	**35%**
Object/president/VP recall[d] (%)	65%	65%	**68%**	68%	**61%**
Proxy-cognition					
IQCODE, -4(best)-0 (worst) (mean)	-1.1	-1.2	**-1.3**	-1.5	**-1.3**
Proxy memory score, -4(best)-0 (worst) (mean)	-1.1	-1.1	**-1.3**	-1.3	**-1.1**
Jorm symptoms, 0-5 (mean)	1.1	1.1	**1.2**	1.0	**1.3**
HISPANICS					
Proxy respondent (%)	27%	14%	**25%**	11%	**14%**
Male (%)	35%	44%	**41%**	36%	**40%**
Age (mean)	77.0	78.8	**77.9**	79.3	**78.5**
LTHS (%)	66%	75%	**73%**	63%	**70%**
At least 1 ADL limitation (%)	40%	33%	**31%**	41%	**34%**
AT least 1 IADL limitation (%)	27%	30%	**28%**	36%	**29%**
History of diabetes diagnosis (%)	30%	39%	**28%**	25%	**35%**
Excellent/very good/good self-reported health (%)	49%	51%	**46%**	50%	**40%**
Social engagement (%)	78%	66%	**79%**	92%	**81%**
Self-cognition					
Immediate word recall, 0-10 (mean)	4.5	4.5	**4.1**	4.1	**4.0**
Delayed word recall, 0-10 (mean)	3.5	3.5	**3.0**	2.9	**2.9**
Serial 7s, 0-5 (mean)	2.1	2.2	**2.2**	3.2	**2.5**
Date recall[c] (%)	23%	25%	**31%**	42%	**35%**
Object/president/VP recall[d] (%)	62%	62%	**68%**	49%	**65%**
Proxy-cognition-cognition					
IQCODE, -4(best)-0 (worst) (mean)	-1.7	-1.6	**-1.4**	-1.3	**-1.4**
Proxy memory score, -4(best)-0 (worst) (mean)	-1.9	-1.8	**-1.3**	-0.5	**-1.2**
Jorm symptoms, 0-5 (mean)	1.2	1.1	**0.9**	1.9	**1.1**

Abbreviations: LTHS = Less than high school education; ADLs = Basic activities of daily living; IADLs = Instrumental activities of daily living; VP = vice-president; IQCODE = the Jorm Informant Questionnaire for Cognitive Decline

[a] Weighted by original wave A cross-sectional weights provided by ADAMS
[b] Weighted by newly derived weights described in eAppendix 1
[c] Responded incorrected to at least one of four date recall items (date, day of week, month, year)
[d] Responded incorrected to at least one four naming items (cactus, scissors, president, vice president)

eAPPENDIX 2. Description of LKW and HW algorithms

Langa-Kabeto-Weir (L-K-W) algorithm[3]

The LKW algorithm for self-respondents is the sum of immediate word recall, delayed word recall, serial 7s, and backward counting for a score range of 0-27; those with scores 0-6 were classified demented. The algorithm sums number of limitations in proxy-rated memory score (scaled 0-4) instrumental activities of daily living, interviewer assessment of cognitive status (0-2) for a score range of 0-11; those with scores 6-11 were classified demented.

Herzog-Wallace (H-W) algorithm[4]

The H-W algorithm for self-respondents is the sum of immediate word recall, delayed word recall, serial 7s, backward counting, recall of date, day of the week, month and year, cactus and scissors naming, as well as president and vice-president recall for a score range of 0-35; those scoring 0-8 were classified demented. The original H-W algorithm for proxy-respondents summed the presence of 7 total Jorm symptoms (memory, judgment organization, hallucinations, getting lost, ability to be left alone, and wandering), classifying those exhibiting 2 or more as demented. However, because the judgement and organization items were dropped from the HRS proxy survey after 2002, we modified the H-W proxy algorithm to a scale of 0-5 that was applied uniformly across waves.

REFERENCES

1. Heeringa SG, Fisher GG, Hurd M, et al. Aging, Demographics and Memory Study (ADAMS): Sample Design, Weighting and Analysis for ADAMS. http://hrsonline.isr.umich.edu/sitedocs/userg/ADAMSSampleWeights_Jun2009.pdf. Published 2009. Accessed January 2, 2018.

2. Health and Retirement Study. Health and Retirement Study - Aging, Demographics, and Memory Study (ADAMS) Supplement: Early Data Release (v. 7.0). http://hrsonline.isr.umich.edu/modules/meta/adams/adamstrk/desc/adams1dd.pdf. Published 2010.

3. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of Cognition Using Surveys and Neuropsychological Assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study. *Journals Gerontol Ser B Psychol Sci Soc Sci*. 2011;66B(Supplement 1):i162-i171.

4. Herzog AR, Wallace RB. Measures of Cognitive Functioning in the AHEAD Study. *Journals Gerontol Ser B,*. 1997;52B:37-48.