Supplemental Materials

Methods

Methods

Study population

FHS conducted 12,272 NP assessments for participants of the Original (Gen I), Offspring (Gen II), Omni 1 and New Offspring Spouse Cohorts. The current FHS NP battery is compromised of 32 NP tests (**eTable 6**). Since some tests were only performed on selected participants and/or were introduced at a later date and lack sufficient longitudinal follow-up, this study focused on the 11 NP tests that were conducted on more than 85% of the participants. The 11 tests were: Wechsler Memory Scale (WMS) Logical Memory – Immediate Recall (LMi), Delayed Recall (LMd), and Recognition (LMr); Visual Reproductions – Immediate Recall (VRi), Delayed Recall (VRd) and Recognition (VRr); Paired Associate Learning – Immediate Recall (PASi), hard scores from Paired Associate Learning – Immediate Recall (PASi_h) and Delayed Recall (PASd_h); Boston Naming Test (BNT30) and the Similarities Test (SIM). 3,029 (24.7%) NP assessments with missing information of any of these 11 tests were excluded from analyzes. In addition, given that sporadic AD is a disease that primarily affects individual of advanced age, NP assessments conducted on participants who were younger than 70 years of age, at the time of testing, were excluded (n=4,731). The final sample size for our study was 4,512 NP assessments. Refer to **eFigure1** for sample selection flowchart.

CHAID decision tree for clinical screen

Decision tree is one of the most widely used methods for inductive inference and concept learning. Chi-square Automatic Interaction Detection (CHAID) decision tree was used to explore the relationship of NP tests with different cognitive conditions [1]. This method determines the optimal combination of NP tests to predict cognitive status in the form of "if-then" rules by portioning each NP test score into mutually exclusive subsets based on heterogeneity of the AD. Nodes that can be separated into sub-nodes are called parent nodes of these sub-nodes (child nodes). Nodes that could not be further branched out are called terminal nodes. CHAID recursively partitions samples into separate and distinct subgroups. First, CHAID chooses the NP test that has the strongest chi-squared automatic interaction with cognitive status. Values of each test are merged if they are not significantly different, which is measured by the Pearson chi-squared test (P value cutoff = 0.05). After the initial segmentation of the population into two or more nodes, the branching-out process is repeated at each of the child nodes until the rules of termination are met. We set the maximum tree depth as five that represent the maximum levels of growth beneath the root node and we set the minimum number of participants at the parent nodes as 50 and 10 for that of the child nodes similar to previous studies [2,3]. If either of the aforementioned rules of termination is satisfied at a sub-node, then there will be no further branching. A detailed stepwise explanation of the decision tree growth is illustrated in the following paragraph as a method to identify cut-off values. The ten-fold cross validation method was applied during the clinical decision tree construction process [4]. The performance of clinical decision trees was evaluated in terms of overall accuracy, AD sensitivity, NAD sensitivity and All-cause dementia sensitivity.

ChiMerge for identifying cut-off values

ChiMerge – a discretization method – was utilized to disperse the numeric score of NP tests and automatically identify cut-off values [5]. This algorithm comprises of two stages: initialization and bottom-up merging. In the initialization stage, all observations are sorted by NP scores and assigned to the respective score intervals, which are randomly selected. This is followed by the bottom-up merging stage, a two-step process: 1) X^2 test statistics are computed for the difference of each pair of adjacent intervals and 2) ranked in order of statistical significance (i.e. p-value). The pair of intervals with the least significant difference merges, forms a combined interval and the bottom-up merging process restarts from step one, where X^2 test statistics are calculated for the each remaining pair of adjacent intervals, including the newly-formed combined interval. The bottom-up merging stage is completed when the p-values of the differences of all remaining pairs of intervals are less than 0.05.

Consider LMd from our study as an example. In the initialization stage, all observations were sorted based on their LMd scores and 24 score intervals for LMd were formed, which corresponded to its score range of 0 to 23. For the first cycle of the bottom-up merging stage, X^2 test statistics were computed for 23 pairs of adjacent intervals. Among them, intervals of $13 < x \leq 14$ and $14 < x \leq 15$, where x is the NP score, were found to be least different in terms of statistical significance ($p= 0.9935$). Therefore, both intervals were merged to form the interval of $13 < x \leq 15$. The bottom-up merging stage restarted from step one, with the

Supplemental Materials

Methods

calculation of X^2 test statistics for the remaining 22 pairs of adjacent intervals, including the newly-formed interval of $13 < x \leq 15$. In the second cycle, the adjacent intervals of $9 < x \leq 10$ and $10 < x \leq 11$ were found to be least different ($p=0.8143$), hence they were merged to form the interval of $9 < x \leq 11$. The difference between each pair of adjacent intervals was considered significant at $X^2 > 5.99$, determined by α-level of 0.05 with 2 degrees of freedom. In the case of LMd, the bottom-up merging stage concluded with six distinct intervals: [0,1], (1,4], (4,6], (6,9], (9,12] and (12,23], where brackets [ ] mean inclusive and parentheses ( ) mean exclusive; and the p-values of the differences between the adjacent intervals were $p<1.0 \text{ X } 10^{-15}$, $p<1.0 \text{ X } 10^{-11}$, $p<1.0 \text{ X } 10^{-4}$, $p<1.0 \text{ X } 10^{-5}$ and $p<1.0 \text{ X } 10^{-4}$ respectively.

Feature selection for identifying optimal NP tests

Considering the interaction between different NP tests, it is possible that information redundancy exists. Feature selection narrows down a subset of relevant features, which can efficiently describe the data and enhance generalizability by reducing model overfitting. In contrast to other dimensionality reduction techniques like those based on projection or compression, feature selection does not transform the values of variables, but instead selects a subset of variables of greater importance to the designated outcomes. Therefore, the original semantics of variables is preserved, which enables easier interpretability by the end-users.

Classification and Regression Trees (CART) are iteratively built by splitting the data based on each feature [6]. The 'splitting' feature is chosen according to its importance for the classification task. It selects the most important NP test according to Gini improvement measure, which then results in binary groups that are most different with respect to the cognitive status. The importance score of a NP test are based on the sum of the improvements in all nodes in which the NP test appears as a splitter. It is weighted by the fraction of the training data in each node split. One important concern for evaluating NP test importance is how to rank them when some tests may obscure the significance of another with slightly higher splitting scores, but another could provide accurate results if used instead. CART can address variable masking and include surrogate variables in the importance calculation [7].

Apart from this embedded method, another representative method was used in our study. Information Theoretic based filters evaluate the importance of a NP test by mutual information maximization [8]. It is robust to overfitting because it introduces bias but has considerably less variance [9]. The relationship between individual NP tests and cognitive status were examined from the view of information theory. We evaluated how much information about cognitive status is involved in individual NP tests according to Shannon entropy [10]. Information gain, a term-goodness criterion in the field of machine learning [11], was employed to rank NP tests in decreasing order.

Correlation-based feature selection adapts greedy search (CBFSGS) and assesses the usefulness of individual NP test in cognitive status prediction, while considering the correlation among them [12]. Based on the training data, it calculates the correlation matrix between the NP test and cognitive outcomes and between each possible pair of NP tests (i.e. 55 combinations) among the 11 NP tests. Using information from the correlation matrix, the algorithm selects a NP test in accordance to its importance for the subset, one at a time, out of the pool of 11 tests. This greedy best first manner search process continues until a subset of five NP tests is chosen. Such selection method minimizes redundancy as correlations – both among features and between outcomes – are computed in a global way.

The top five most optimal NP tests were decided through majority voting of the three feature selection methods (**Table 2**) and new clinical decision trees were constructed based on these reduced feature sets.

Supplemental Materials

Methods

Validation method

In order to validate our approach, we implemented unsupervised machine learning techniques to learn the inherent structure of our data without using explicitly-provided labels (i.e. cognitive outcomes) [13]. Both K-means [14] and Hierarchical Clustering [15], were utilized as validation methods for the importance changes of NP tests in different subpopulations and whether the selected optimal NP profiles could reduce feature redundancy.

K-means method is one of the top ten algorithms used in data mining and can reveal the data structure [14]. We partitioned observations into two clusters (AD & HC) with the nearest mean of the specific NP test serving as a prototype for the cluster. First, 2 observations are randomly selected as the initial 2 cluster centroids. Then each observation is assigned to its closet centroid according to the Euclidean distance of 11 NP tests. Subsequently, the centroids of two clusters are reselected to the center (mean) of the new cluster, which include the newly added observations. The algorithm iterates the aforementioned two steps until convergence. We compared the actual cognitive status – determined by the FHS adjudication panel – with our clustering results to create a confusion matrix [16] and we evaluated the AD sensitivity of specific NP test in different subpopulations.

Hierarchical clustering is a bottom-up method to construct a hierarchy of clusters [15]. Dendrogram is used to demonstrate the degree of discrimination between clusters. Firstly, each observation is treated as an individual cluster. Secondly, the pairwise Euclidean distance matrix is calculated for all observations. Thirdly, according to the distance matrix, the two clusters, which are most similar, are combined to form a new merged cluster. Fourthly, the distances between the newly formed cluster and all other remaining clusters are re-tabulated to update the distance matrix. . Lastly, this clustering process is repeated until all observations are grouped into one cluster. Dissimilarities between clusters of observations are defined as the distance between their two farthest-apart members, which is also known as complete linkage. Feature redundancy will affect the separability of cluster. This approach is used to validate whether the clusters formed by the selected optimal NP profiles are more distinguishable than that formed by all NP tests.

Supplemental Materials

Results

Results

An illustration of decision tree construction

Consider the leftmost pathway (LMd<1 ➔ BNT30<23 ➔ VRd =0 ➔ PASi_h=0) as an example. At the start of the decision tree, CHAID found LMd to have the strongest interaction with cognitive outcomes ($P<1.0x10^{-15}$) among the entire sample, thus LMd was designated as the root node. ChiMerge method was performed to discover the optimal adjacent score intervals for LMd, which resulted in six branched-out sub-nodes. Moving down from the root node to the leftmost sub-node on the first level (LMd<1), CHAID again examined the chi-squared automatic interaction between all the NP tests and the cognitive outcomes, but this time among participants who scored less than 1 for LMd and found BNT30 to have the strongest interaction ($P<1.0x10^{-11}$) at this sub-node. Three optimal adjacent score intervals of BNT30, which were [0,23], (23,28], (28,30], were determined by ChiMerge method. Moving down from the first level to the leftmost sub-node on the second level (BNT30<23), VRd was selected as the next deciding NP test (p=0.0036) with two optimal adjacent scores intervals of [0,0] and (0,14]. Similarly, from the second level to the leftmost sub-node on the third level (VRd=0), CHAID determined PASi_h as the next NP test of interest (p=0.0023), with two optimal adjacent score intervals of [0,0] and (0,12]. Finally at the leftmost sub-node on the fourth level (PASi_h=0), as there was no other significant NP test to examine along this pathway, the splitting terminates with AD diagnostic sensitivity of 88.1%.

Optimal NP profiles and the validation

In order to verify the change of importance for specific NP tests in different populations, K-means method was used to generate clusters based on NP scores only, in an unsupervised manner. Confusion matrixes were constructed by comparing the actual labels (i.e. cognitive outcomes) of the clusters with the predicted labels. As demonstrated in **eTable 3**, AD sensitivity of BNT30 for men was higher than that for women. In contrast, AD sensitivities of PASi and VRd for women were higher than that of men. From an unsupervised learning perspective, the results indicated that PASi, BNT30 and VRd had sex-specific differences. Similar trend could be observed for APOE ε4 allele- and education-stratified analyzes (**eTable 4** and **eTable 5** respectively).

eFigure 14 shows a hierarchical clustering for the total population using all tests and as well as the first five selected tests. Both dendrograms revealed three distinct clusters; however, the distinguishability among them was more pronounced for the selected features dendrogram. This indicated that by using the optimal set of NP tests to construct the decision tree would potentially minimize data redundancy and better represent the inherent patterns within the NP data.

Supplement
References

Reference

[1] Van Diepen M, Franses PH. Evaluating chi-squared automatic interaction detection. *Information Systems*. 2006;31(8):814-31.

[2] Laliberte AS, Fredrickson EL, Rango A. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric engineering & Remote sensing*. 2007;73(2):197-207.

[3] McKee LA, Fabres J, Howard G, Peralta-Carcelen M, Carlo WA, Ambalavanan N. PaCO2 and neurodevelopment in extremely low birth weight infants. *J Pediatrics*. 2009;155(2):217-21.

[4] Seni G, Elder JF. Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis Lectures on Data Mining and Knowledge Discovery*. 2010;2(1):1-26.

[5] Kerber R. Chimerge: Discretization of numeric attributes. *InProceedings of the tenth national conference on Artificial intelligence.* 1992; 123-128.

[6] Breiman, Leo. Classification and regression trees. *Routledge*. 2017.

[7] Tuv E, Borisov A, Runger G, Torkkola K. Feature selection with ensembles, artificial variables, and redundancy elimination. *J Mach Learn Res*. 2009;1341-66.

[8] Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;3:1157-82.

[9] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. *Springer series in statistics*. New York, NY; 2001.

[10] Akaike H. Information theory and an extension of the maximum likelihood principle. *InSelected papers of hirotugu akaike*. Springer, New York, NY. 1998; 199-213.

[11] Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. *InIcml.* 1997; 97: 412-420.

[12] Hall MA. Correlation-based feature selection of discrete and numeric class machine learning. 2000.

[13] Hastie T, Tibshirani R, Friedman J. Unsupervised learning. *InThe elements of statistical learning*. Springer, New York, NY. 2009; 485-585.

[14] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY, Zhou ZH. Top 10 algorithms in data mining. *Knowledge and information systems*. 2008;14(1):1-37.

[15] Xu D, Tian Y. A comprehensive survey of clustering algorithms. *Annals of Data Science*. 2015;2(2):165-93.

[16] Powers, David Martin. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2011.

Supplemental Materials

Tables

eTable 1. Overall performance of clinical decision tree of all NP tests in different populations

	AD Sensitivity	NAD sensitivity	All-cause dementia sensitivity	Accuracy
Total	73.2%	46.3%	85.0%	73.9%
Men	65.9%	53.6%	81.5%	76.1%
Women	71.0%	57.8%	90.8%	72.9%
APOE ε4 (-)	61.3%	61.5%	86.3%	74.6%
APOE ε4 (+)	81.5%	58.1%	92.2%	69.6%
High school and below	70.9%	63.5%	92.8%	65.9%
Beyond high school	59.2%	55.7%	81.6%	80.3%

eTable 2. Overall performance of NP clinical decision tree of selected five NP tests in different populations

	Selected 5 NP tests	AD Sensitivity	NAD sensitivity	All-cause dementia sensitivity	Accuracy
Total	LMd, VRd, LMi, VRi, BNT30	71.9%	46%	84.5%	73.3%
Men	LMd, VRd, LMi, BNT30, VRi	65.9%	49.1%	79.2%	74.6%
Women	LMd, VRd, PASi_h, LMi, PASi	68.6%	56.1%	88.3%	75.3%
APOE ε4 (-)	LMd, VRd, LMi, BNT30, VRi	56.4%	61.8%	85.7%	73.0%
APOE ε4 (+)	LMd, VRd, LMi, VRi, PASi	81.5%	58.1%	92.2%	69.6%
High school and below	LMd, VRd, BNT30, LMi, VRi	73.2%	50.6%	90.2%	66.1%
Beyond high school	LMd, VRd, LMi, VRi, PASi	56.0%	52.2%	78.8%	80.3%

Supplemental Materials

Tables

eTable 3. Confusion Matrix of the BNT30, PASi and VRd in men and women

			Men				Women	
			Prediction				Prediction	
			Normal	AD			Normal	AD
BNT30	Observed	Normal	1222	299	Observed	Normal	1659	334
		AD	43	136		AD	108	268
			Sensitivity =75.98%				**Sensitivity =71.28%**	
			Prediction				Prediction	
			Normal	AD			Normal	AD
PASi	Observed	Normal	761	760	Observed	Normal	1000	993
		AD	25	154		AD	17	359
			Sensitivity =86.03%				**Sensitivity =95.48%**	
			Prediction				Prediction	
			Normal	AD			Normal	AD
VRd	Observed	Normal	713	808	Observed	Normal	1275	718
		AD	12	167		AD	24	352
			Sensitivity =93.3%				**Sensitivity =93.62%**	

Supplemental Materials

Tables

eTable 4. Confusion Matrix of the BNT30, PASi and VRd in APOE ε4 (-) and APOE ε4 (+)

			APOE ε4 (-)					**APOE ε4 (+)**	
			Prediction					Prediction	
			Normal	AD				Normal	AD
BNT30	Observed	Normal	2307	487		Observed	Normal	517	58
		AD	88	258			AD	85	99
			Sensitivity =74.57%					**Sensitivity =53.80%**	
			Prediction					Prediction	
			Normal	AD				Normal	AD
PASi	Observed	Normal	1421	1373		Observed	Normal	399	176
		AD	30	316			AD	22	162
			Sensitivity =91.33%					**Sensitivity =88.04%**	
			Prediction					Prediction	
			Normal	AD				Normal	AD
VRd	Observed	Normal	1578	1216		Observed	Normal	372	203
		AD	19	327			AD	14	170
			Sensitivity =94.51%					**Sensitivity =92.39%**	

Supplemental Materials

Tables

eTable 5. Confusion Matrix of the BNT30, PASi and VRd for education-stratified analyzes

			High school and below				Beyond high school	
			Prediction				Prediction	
			Normal	AD			Normal	AD
BNT30	Observed	Normal	1290	201	Observed	Normal	1443	576
		AD	118	240		AD	53	138
			Sensitivity =67.04%				**Sensitivity =72.25%**	
			Prediction				Prediction	
			Normal	AD			Normal	AD
PASi	Observed	Normal	797	694	Observed	Normal	968	1051
		AD	40	318		AD	13	178
			Sensitivity =88.83%				**Sensitivity =93.19%**	
			Prediction				Prediction	
			Normal	AD			Normal	AD
VRd	Observed	Normal	884	607	Observed	Normal	1036	983
		AD	28	330		AD	11	180
			Sensitivity =92.18%				**Sensitivity =94.24%**	

Supplemental Materials

Tables

eTable 6. Summary statistics of all available 32 NP tests at FHS

NP Tests*	Acronyms	Range	Mean (sd)	Missing (%)
Logical Memory – IR	LMi	0 – 23	10.55(4.53)	244(2%)
Logical Memory – DR	LMd	0 – 23	9.52(4.74)	335(2%)
Logical Memory – Recognition	LMr	0 –11	9.29(1.59)	2241(15%)
Visual Reproductions – IR	VRi	0 – 14	7.57(3.70)	586(4%)
Visual Reproductions – DR	VRd	0 – 14	6.74(3.88)	685(4%)
Visual Reproductions – Recognition	VRr	0 –4	2.72(1.23)	752(5%)
Paired Associate Learning – IR	PASi	0 –21	13.15(4.06)	743(5%)
Paired Associate Learning – IR (Ease Score)	PASi_e	0 – 18	16.29(2.43)	741(5%)
Paired Associate Learning – IR (Hard Score)	PASi_h	0 – 12	4.91(3.37)	743(5%)
Paired Associate Learning – DR	PASd	0 – 10	8.14(1.76)	2452(16%)
Paired Associate Learning – DR (Ease Score)	PASd_e	0 – 6	5.80(0.65)	2452(16%)
Paired Associate Learning – DR (Hard Score)	PASd_h	0 – 4	2.34(1.40)	2451(16%)
Paired Associate Learning - Recognition	PASr	0 – 10	9.65(1.19)	7360(48%)
Digits Forward Span	DSF	0 – 9	6.42(1.39)	3314(22%)
Digits Backward Span	DSB	0 – 8	4.58(1.40)	3480(23%)
Trail A^{+}	trailsA	0 – 7	0.66(0.64)	2642(17%)
Trail B^{+}	trailsB	0 – 10	1.98(2.07)	2995(19%)
Similarities Test	SIM	0 – 26	15.20(5.14)	500(3%)
Hooper Visual Organization Test	HVOT	0 – 30	24.18(4.47)	2768(18%)
Boston Naming test – 10 items	BNT10	0 – 10	9.25(1.46)	661(4%)
Boston Naming test – 10 items (semantic cue)	BNT10_semantic	0 – 4	0.08(0.29)	2446(16%)
Boston Naming test – 10 items (phonemic cue)	BNT10_phonemic	0 – 4	0.21(0.50)	2446(16%)
Boston Naming test – 30 items	BNT30	0 – 30	26.06(4.80)	2447(16%)
Boston Naming test – 30 items (semantic cue)	BNT30_semantic	0 – 6	0.35(0.66)	2447(16%)
Boston Naming test – 30 items (phonemic cue)	BNT30_phonemic	0 – 15	1.23(1.43)	2447(16%)
Finger Tapping – Right hand	FingTapR	0 – 77.6	45.63(10.30)	5835(38%)
Finger Tapping – Left hand	FingTapL	0 – 74	41.49(9.12)	5835(38%)
Wide Range Achievement Test –Reading	WRAT	15 – 57	48.80(5.35)	3424(22%)
Verbal Fluency Test	FAS	0 – 95	35.00(14.44)	3667(24%)
Verbal Fluency Test – Animal	FAS_animal	0 – 50	18.61(6.05)	7415(48%)
Block Design	BD	0 – 26	19.87(6.34)	10716(70%)
WAIS test	WAIS	0 – 29	17.89(6.17)	10616(69%)

* IR: Immediate Recall; DR: Delayed Recall

$^{+}$ Measured in minutes

Supplemental Materials

Tables

eTable 7. Cut-off values of NP test decision tree for different subpopulations

	Total	Men	Women	APOE ε4 (-)	APOE ε4 (+)	High school and below	Beyond high school
LMd	[0,1], (1,4], (4,6], (6,9], (9,12], (12,23], (0,11], (11,23]	[0,1], (1,4], (4,7], (7,12], (12,23]	[0,1], (1,4], (4,8], (8,12], (12,23], [0,9], (9,23]	[0,2], (2,5], (5,7], (7,9], (9,13], (13,23]	[0,0], (0,2], (2,4], (4,9], (9,23]	[0,0], (0,3], (3,6], (6,9], (9,23]	[0,3], (3,23], [0,6], (6,23], [0,3], (3,12], (12,23]
VRd	[0,1], (1,4], (4,6], (6,14], [0,2], (2,5], (5,8], (8,14]	[0,3], (3,14]	[0,1], (1,4], (4,6], (6,14], [0,2], (2,6], (6,14]	[0,3], (3,4], (4,6], (6,14], [0,2], (2,3], (3,14]	[0,4], (4,14], [0,1], (1,5], (5,14]	[0,0], (0,3], (3,14], [0,1], (1,3], (3,14]	[0,1], (1,3], (3,4], (4,5],(5,10], (10,14],
BNT30	[0,23], (23,30], [0,21], (21,30], [0,24], (24,30]	[0,23], (23,30], [0,24], (24,30]	N.A.	[0,22], (22,25], (22,30], (25,30], [0,19], (19,28], (28,30]	[0,22], (22,30]	[0,24], (24,30], [0,25], (25,30], [0,22], (22,30]	[0,24], (24,30]
VRi	[0,4], (4,14], [0,5], (5,14]	[0,3], (3,6], (6,14]	[0,4], (4,14]	[0,3], (3,14]	N.A.	[0,4], (4,6], (6,7], (7,14]	[0,9], (9,14]
SIM	[0,12], (12,26]	[0,11], (11,15], (15,26], (11,26], (15,16], (16,26], [0,17], (17,18], (18,26]	[0,10], (10,26], [0,14], (14,15], (15,26]	[0,8], (8,26], [0,11], (11,26], [0,13], (13,26]	N.A.	[0,5], (5,10], (10,14], (14,26], [0,5], (5,13], (13,26]	[0,11], (11,26], (11,15], (15,26]
PASi	[0,8.5], (8.5,21], [0,9.5], (9.5,21], [0,10.5], (10.5,21]	[0,8.5], (8.5,21], [0,10], (10,21]	[0,9], (9,21], [0,13], (13,21]	N.A.	[0,11.5], (11.5,21]	[0,8], (8,21], [0,9], (9,21]	[0,11.5], (11.5,21]
VRr	[0,1], (1,4]	N.A.	[0,1], (1,4]	[0,0], (0,4]	[0,1], (1,4]	[0,2], (2,4]	[0,0], (0,4]
LMi	[0,10], (10,23]	N.A.	N.A.	[0,10], (10,23]	N.A.	[0,9], (9,23]	[0,7], (7,23]
LMr	N.A.	N.A.	N.A.	[0,7], (7,11], [0,7], (7,10], (10,11]	N.A.	[0,8], (8,11]	N.A.
PASi_h	[0,0], (0,12]	N.A.	[0,2], (2,12], [0,0], (0,12]	[0,4], (4,12], [0,0], (0,12], [0,1], (1,12]	N.A.	N.A.	[0,1], (1,4], (4,12], (1,12]
PASd_h	N.A.	N.A.	[0,0], (0,4]	[0,0], (0,4], [0,1], (1,4]	N.A.	[0,0], (0,2], (2,4]	[0,0], (0,4]

N.A.: Not represented in the decision tree

eTable 8. Cut-off scores of reduced-feature decision tree for different subpopulations

Supplemental Materials

Tables

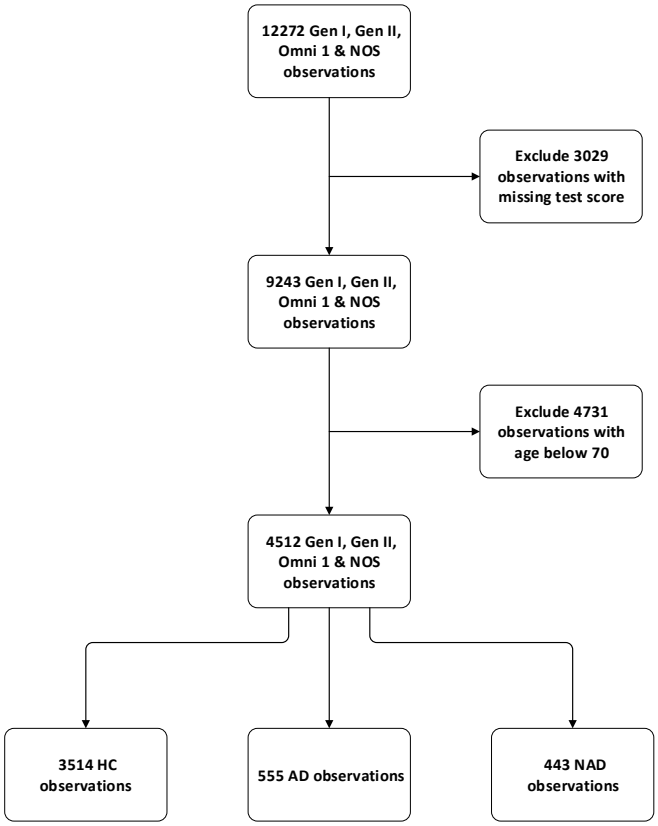
	Total	Men	Women	APOE ε4 (-)	APOE ε4 (+)	High school and below	Beyond high school
LMd	[0,1], (1,4], (4,6], (6,9], (9,12], (12,23], [0,11], (11,23]	[0,1], (1,4], (4,7], (7,12], (12,23]	[0,1], (1,4], (4,8], (8,12], (12,23], [0,9], (9,23]	[0,2], (2,5], (5,7], (7,9], (9,13], (13,23], [0,11], (11,23]	[0,0], (0,2], (2,4], (4,9], (9,23]	[0,0], (0,3], (3,6], (6,9], (9,23], [0,5], (5,23],	[0,3], (3,6], (6,23], (3,23], (3,12], (12,23],
VRd	[0,1], (1,4], (4,6], (6,14], [0,0], (0,2], (2,4], (4,6], (6,14], [0,2], (2,5], (5,8], (8,14]	[0,3], (3,14], [0,4], (4,7], (7,8], (8,14],	[0,0], (0,2], (2,5], (5,14], [0,1], (1,4], (4,6], (6,14], [0,3], (3,6], (6,14], [0,2], (2,6], (6,14]	[0,0], (0,3], (3,14], [0,3], (3,6], (6,14], (3,4], (4,6], [0,2], (2,3]	[0,1], (1,5], (5,14], [0,4], (4,14],	[0,0], (0,3], (3,14], [0,2], (2,3], [0,1], (1,3], (1,6], (6,14]	[0,1], (1,3], (3,4], (4,5], (5,10], (10,14],
BNT30	[0,23], (23,28], (28,30], (23,30], [0,18], (18,30],[0,27], (27,30]	[0,24], (24,27], (27,30], [0,23], (23,30], [0,24], (24,30], [0,20], (20,30],	N.A.	[0,22], (22,30], [0,23], (23,30], [0,25], (25,30], [0,19], (19,22], (22,30], (19,28], (28,30]	N.A.	[0,24], (24,30], [0,22], (22,25], (25,30], (22,30]	N.A.
VRi	[0,4], (4,14]	[0,3], (3,6], (6,14]	N.A.	[0,4], (4,14], [0,3], (3,14]	[0,4], (4,14]	[0,4], (4,14]	[0,9], (9,14]
PASi	N.A.	N.A.	[0,13], (13,21], [0,11], (11,21], [0,10], (10,21]	N.A.	[0,11.5], (11.5,21]	N.A.	[0,8], (8,21], (8,13.5], (13.5,21], [0,11.5], (11.5,21],
LMi	[0,10], (10,23], [0,11], (11,23]	[0,8], (8,23]	N.A.	[0,10], (10,23]	N.A.	[0,10], (10,23], [0,9], (9,23], [0,7], (7,12], (12,23]	[0,5], (5,23]
PASi_h	N.A.	N.A.	[0,0], (0,12], [0,2], (2,12]	N.A.	N.A.	N.A.	N.A.

N.A.: Not represented in the decision tree

Supplement al Materials

Figures

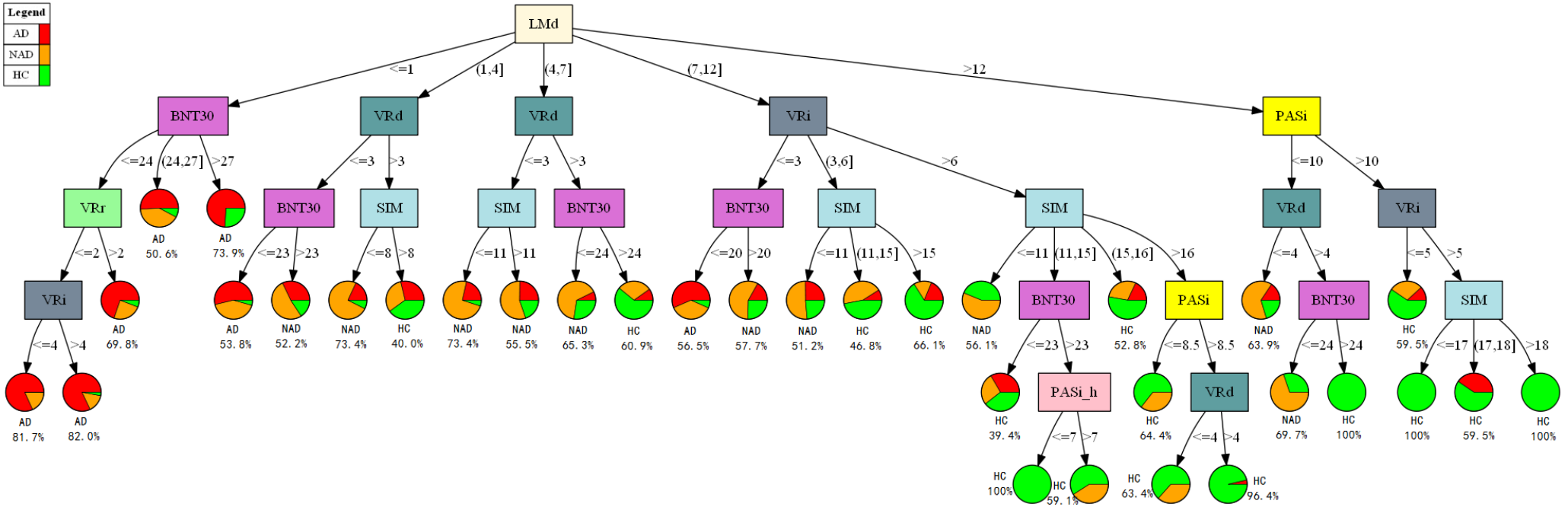
eFigure 1. The process of sample selection



Supplement al Materials

Figures

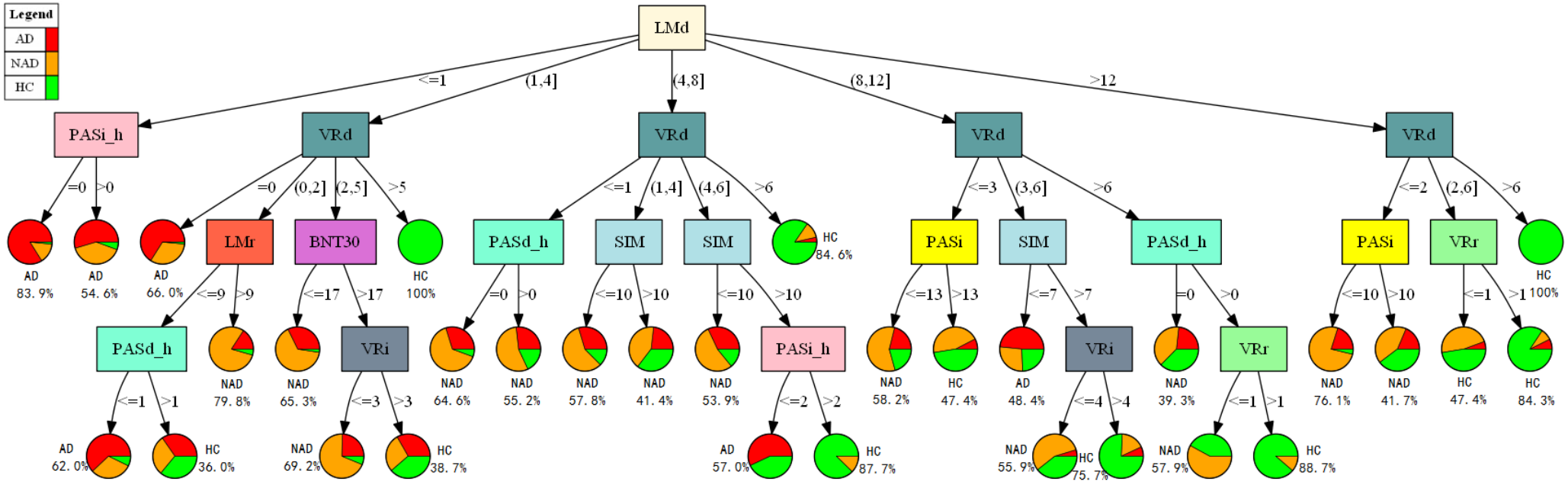
eFigure 2. Clinical cognitive screen decision tree based on all NP tests in men



Supplement al Materials

Figures

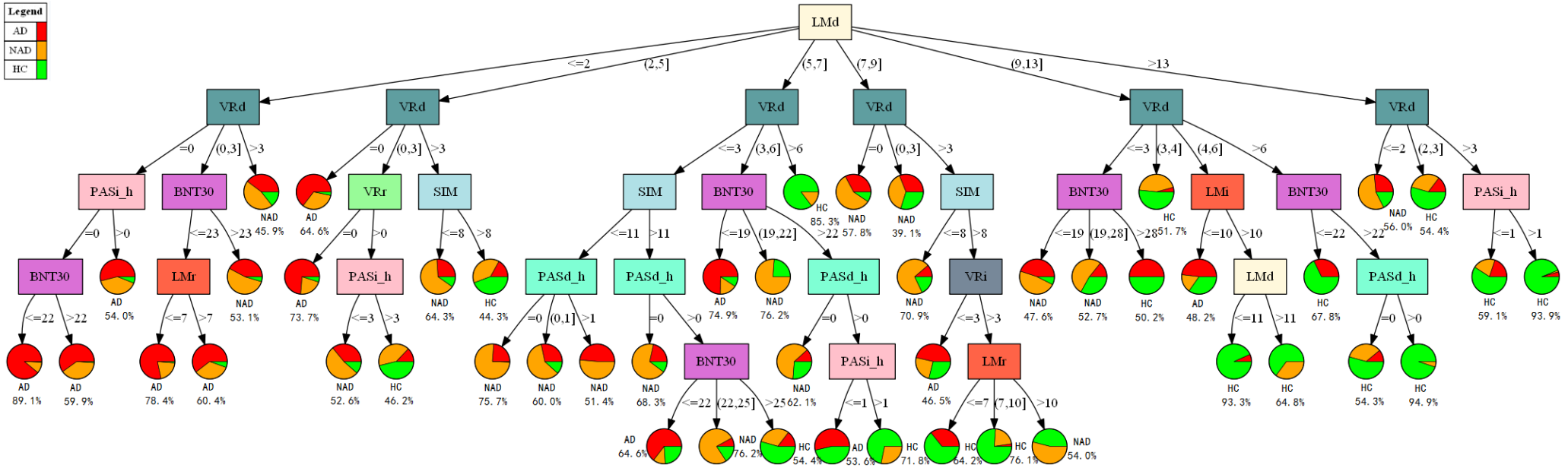
eFigure 3. Clinical cognitive screen decision tree based on all NP tests in women



Supplement al Materials

Figures

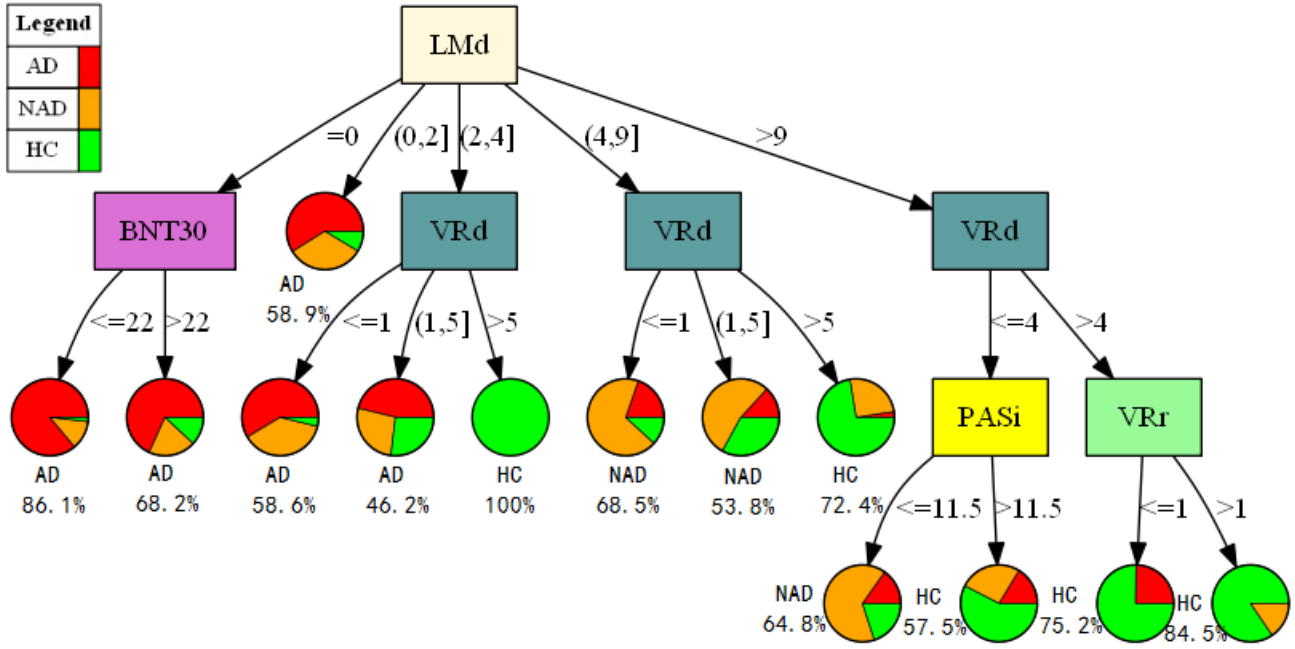
eFigure 4. Clinical cognitive screen decision tree based on all NP tests in APOE ε4 (-) subpopulation



Supplement al Materials

Figures

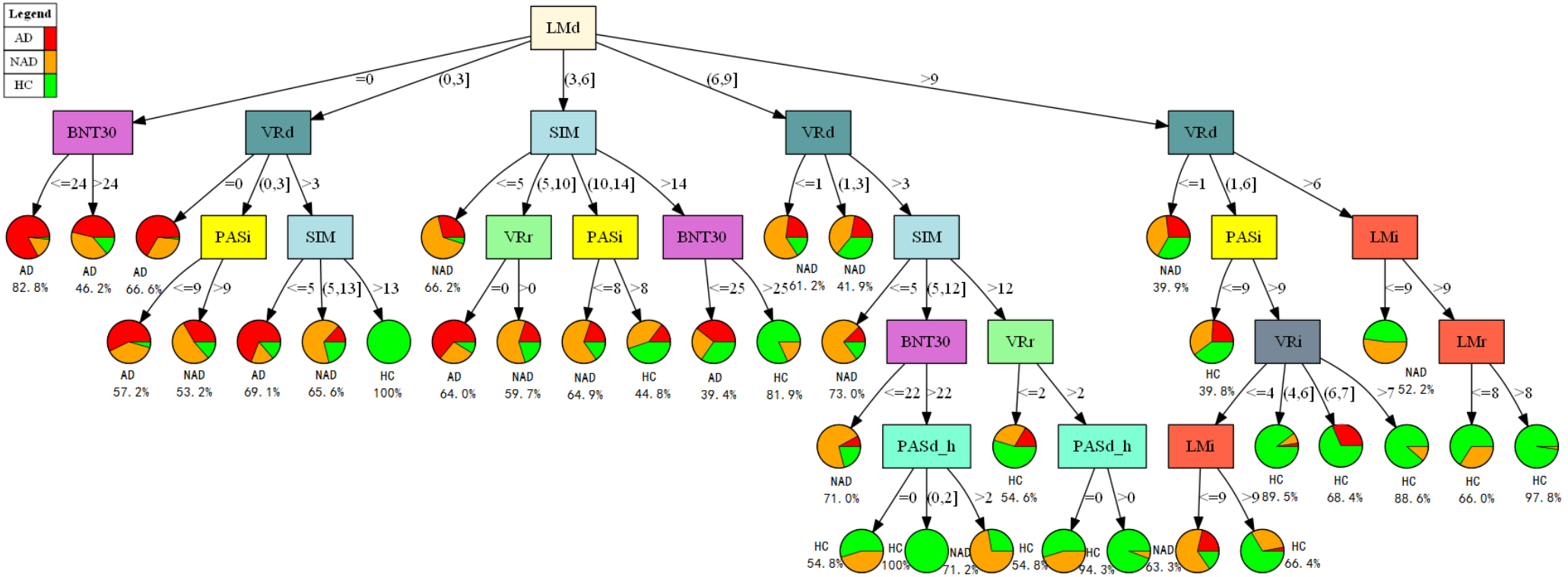
eFigure 5. Clinical cognitive screen decision tree based on all NP tests in APOE ε4 (+) subpopulation



Supplement al Materials

Figures

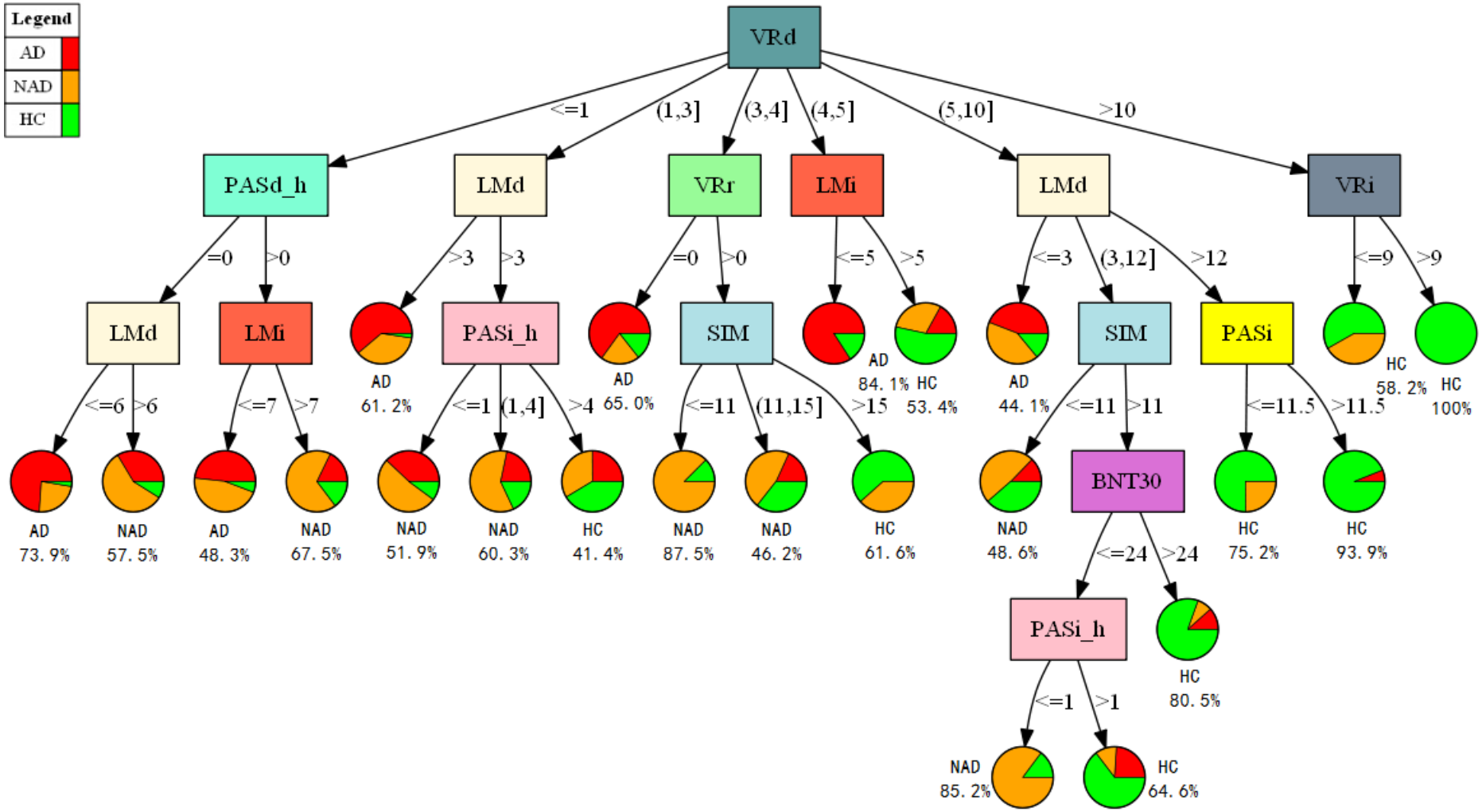
eFigure 6. Clinical cognitive screen decision tree based on all NP tests in subpopulation with high school and below education



Supplement al Materials

Figures

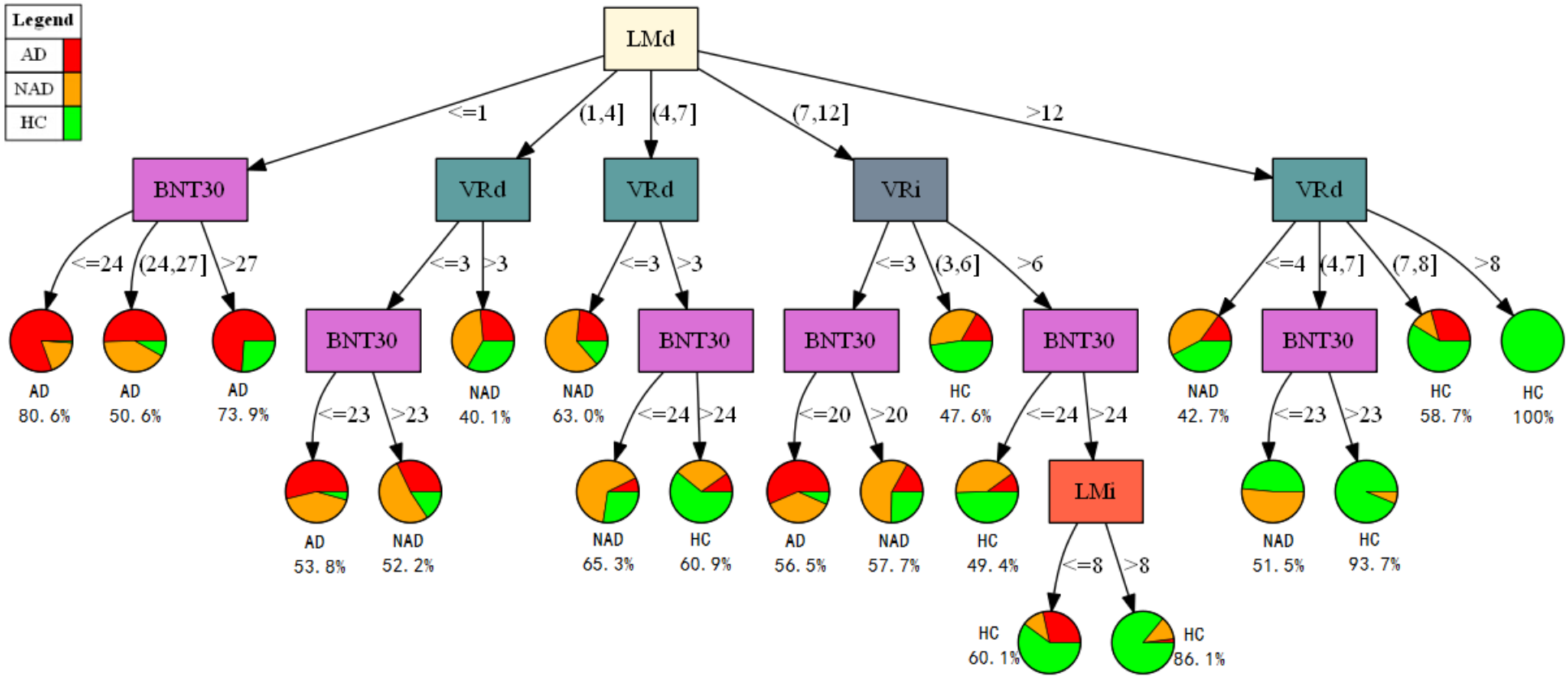
eFigure 7. Clinical cognitive screen decision tree based on all NP tests in subpopulation with beyond high school education



Supplement al Materials

Figures

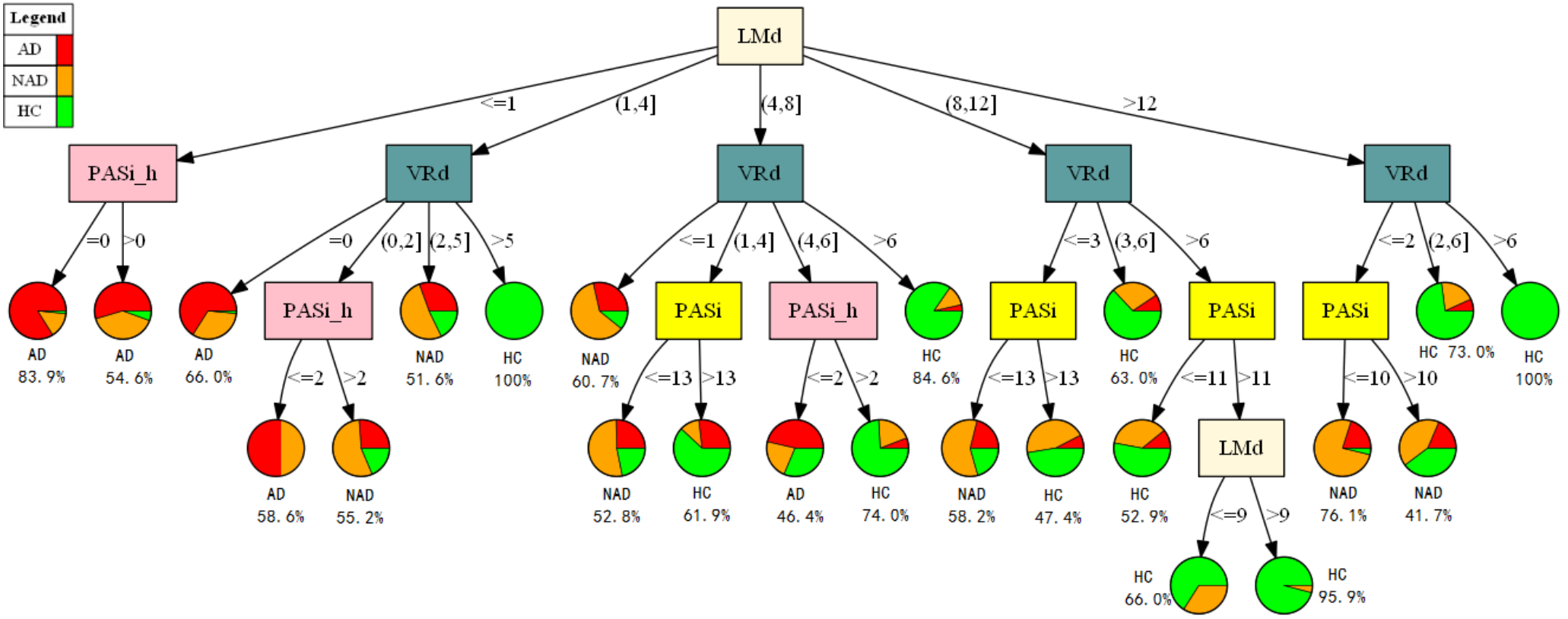
eFigure 8. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in men



Supplement al Materials

Figures

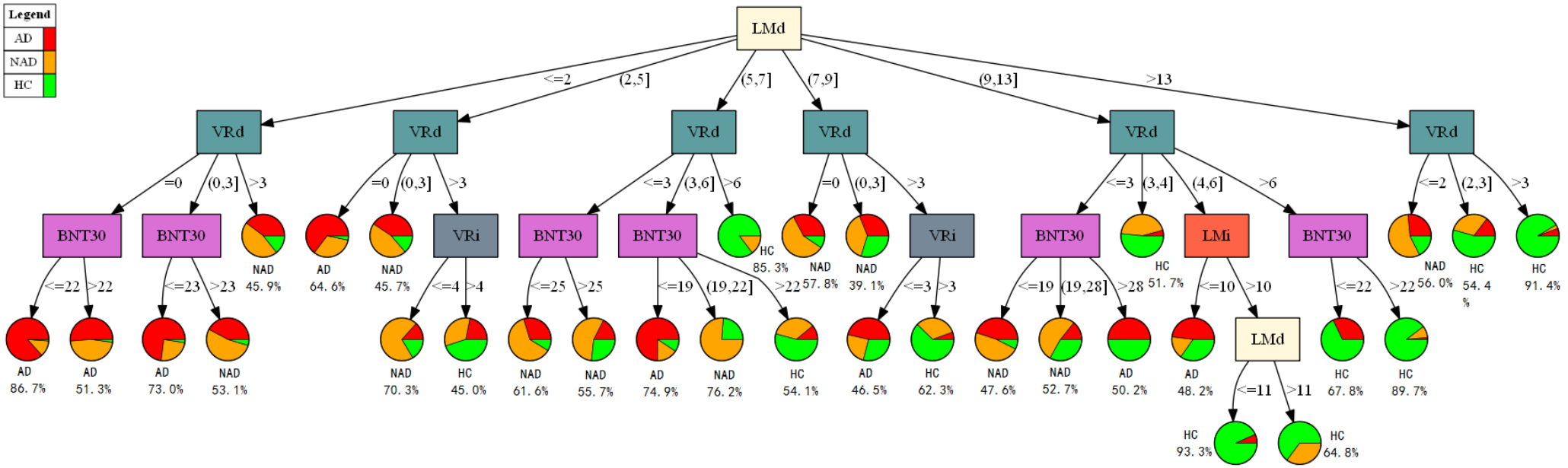
eFigure 9. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in women



Supplement al Materials

Figures

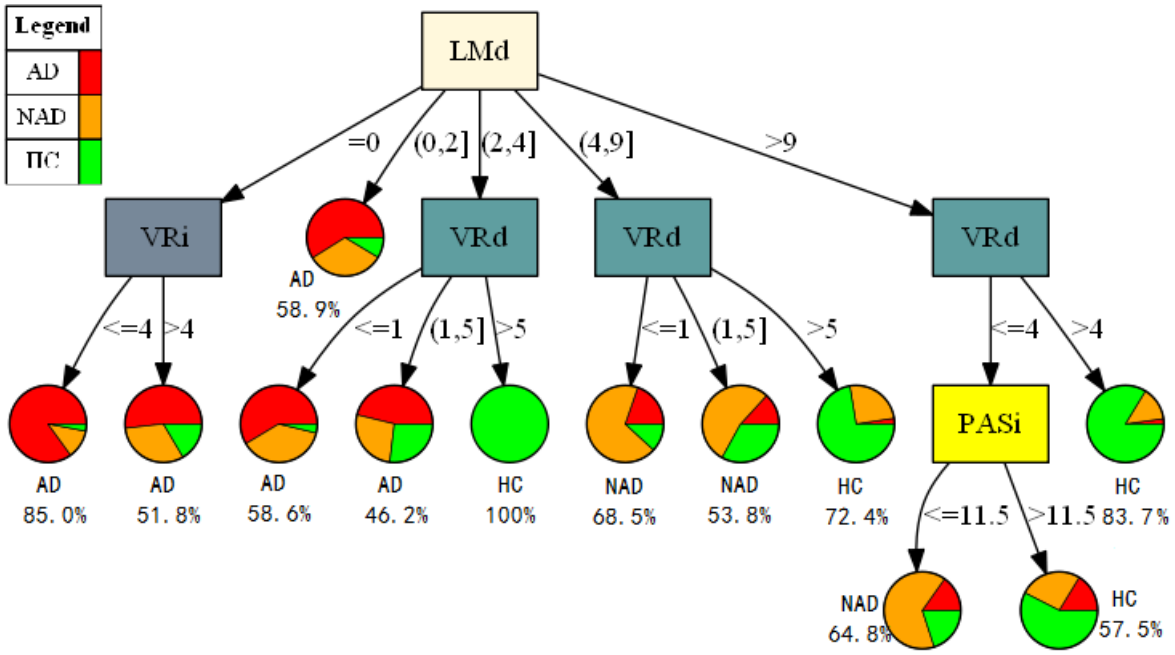
eFigure 10. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in APOE ε4 (-) subpopulation



Supplement al Materials

Figures

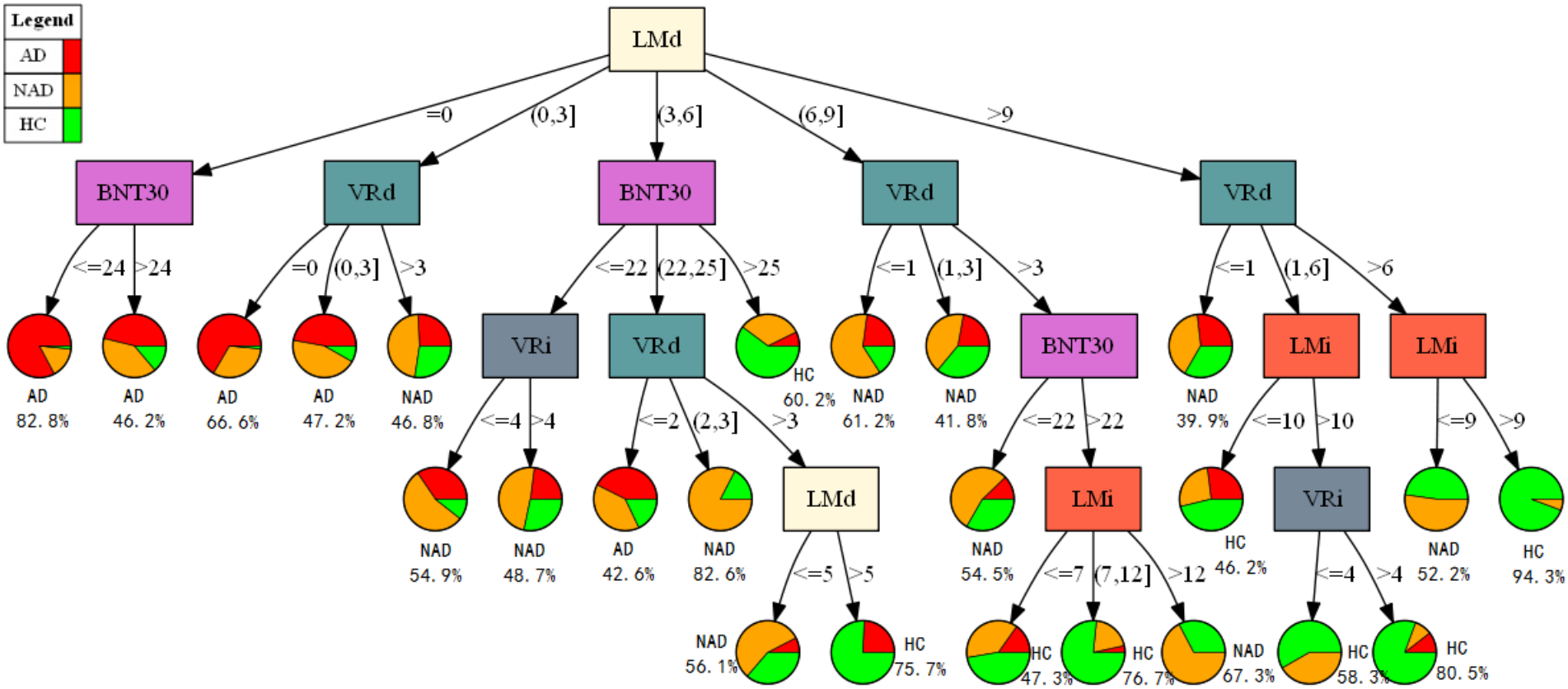
eFigure 11. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in APOE ε4 (+) subpopulation



Supplement al Materials

Figures

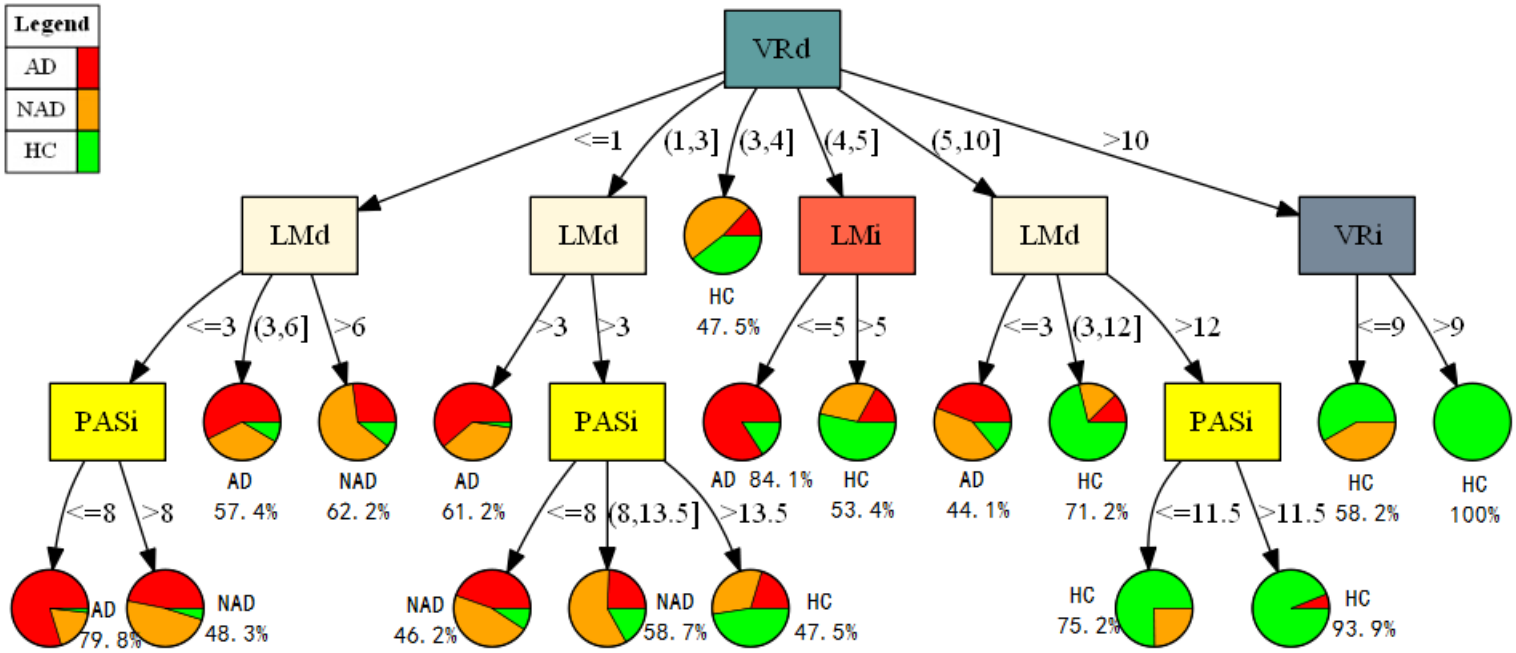
eFigure 12. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in subpopulation with high school and below education



Supplement al Materials

Figures

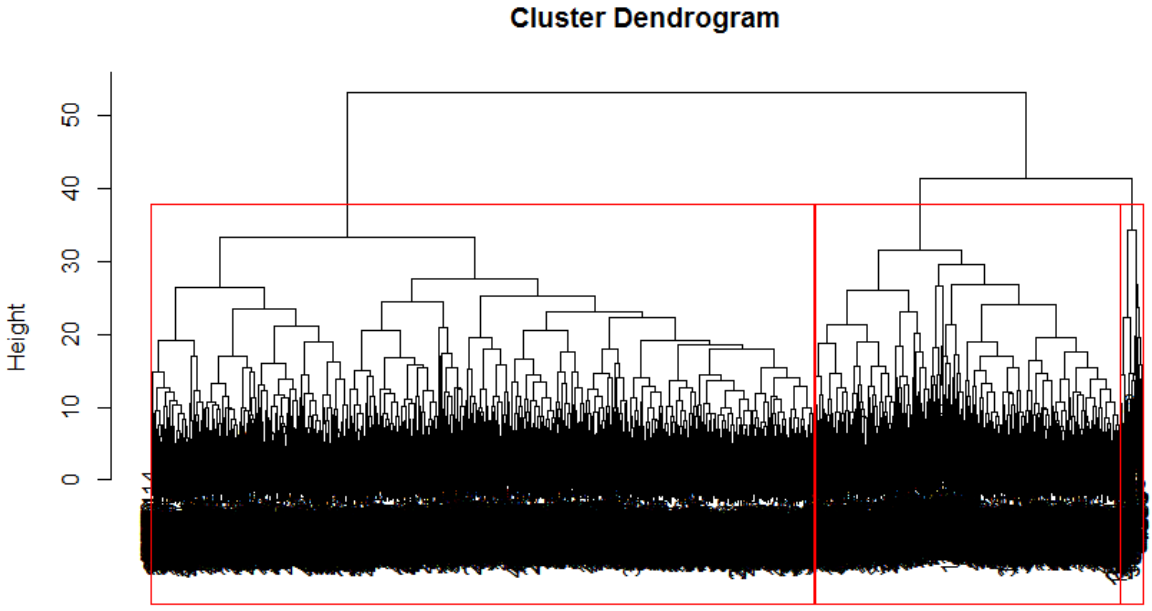
eFigure 13. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in subpopulation with beyond high school education



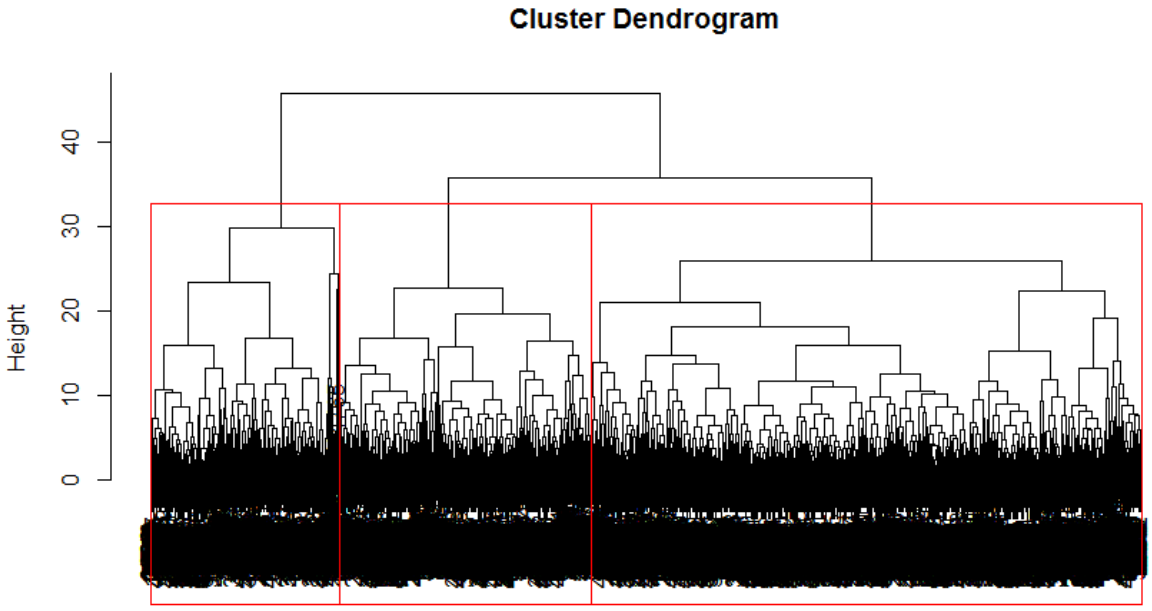
Supplement al Materials

Figures

eFigure 14. Hierarchical clustering of the total population using (1) full NP-test-set versus (2) five selected NP tests. The vertical axis of the dendrogram represents the dissimilarity between clusters. The horizontal axis represents the observations and clusters. The merging of two clusters is represented by the splitting of a vertical line into two vertical lines. Red rectangles indicate observations that are divided into three categories.



(1)



(2)

Supplement al Materials

Figures

eFigure 15. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in total population with re-introduction of 310 observations with valid NP scores for the selected NP tests (LMd, VRd, LMi, VRi and BNT30). The augmented sample comprised of 4,822 sets of NP test scores from 2191 participants. Among them, 652 were AD patients, 485 were NAD and the remaining 1054 were healthy controls. This increased the sample size and enhanced the generalizability of the algorithm.