

PATTERN OF SMELL IDENTIFICATION IMPAIRMENT IN ALZHEIMER'S DISEASE

Latha Velayudhan¹✉, Petroula Proitsi², Amy Gasper¹, Megan Pritchard², Sarah Baillon¹, Charlotte Messer³, Simon Lovestone^{2,4}

¹ Psychiatry for the Elderly, Department of Health Sciences, University of Leicester, United Kingdom

² King's College London, Institute of Psychiatry, London and NIHR Biomedical Research Centre for Mental Health, London, United Kingdom

³ Mental Health Services for Older People, Leicestershire Partnership NHS Trust. Leicester, United Kingdom

⁴ Department of Psychiatry, University of Oxford, United Kingdom

✉ Corresponding author

Dr Latha Velayudhan

Senior Clinical Research Fellow, Psychiatry for the Elderly

Academic Department, Leicester General Hospital, Leicester, UK, LE5 4PW

Tel: 0116 258 4158

Fax: 0116 258 4078

E-mail: lv24@le.ac.uk

Competing financial interests: None

Abstract: 248 words

Text: 2602 words

2 tables, 2 figures and 30 references

1 supplementary table

Running title: Pattern of smell impairment in AD

Abstract

Olfactory dysfunction in general and impaired odor identification in particular, have been reported in Alzheimer's disease (AD). Olfactory testing may be a useful diagnostic aid for AD, but the types of odor most commonly affected need to be identified. This study aimed to determine pattern and types of odor affected in AD with the goal of improving clinical applicability. 54 outpatients with mild to moderate AD and 40 age and sex matched non-demented controls (NDC) were tested using British version of University of Pennsylvania Smell Identification Test (UPSIT; Sensonics, Inc, Haddon Heights, NJ) and data analysed to identify an optimal subset of UPSIT to best differentiate AD patients from controls. AD subjects had significantly lower UPSIT total scores than NDC. Random Forest with backward elimination identified 12 UPSIT items which accurately differentiated AD patients compared to controls (sensitivity, 0.89 and specificity, 0.83, positive predictive value of 0.889 and negative predictive value of 0.833). With MMSE, the accuracy improved to 96.7% with sensitivity of 100% and specificity 91.7%. The 12 smell items found to be most affected in AD subjects, falls into the categories of Chemicals, Fruits and Nuts, that reflects important attributes such as danger and food, known to be affected in people with AD and that has the potential to impair activities of daily living. The 12 items of British UPSIT most affected in AD subjects, provides a potential brief scale for early detection of AD in clinical settings. Independent replication is needed to validate these findings.

Key words: Alzheimer's disease, smell identification, olfaction, odor, pattern of deficits

INTRODUCTION

Olfactory dysfunction in general and impaired odor identification in particular have been reported in AD and are found to occur at early stages of the disease [1]. It has been indicated that involvement of the olfactory bulb and tract is one of the earliest events in the degenerative process on the central nervous system in AD [2] and also that tau pathology in the olfactory bulb increases with severity of AD [3]. All published reports of smell identification in AD have demonstrated deficits relative to healthy elderly [4]. Olfactory dysfunction has been suggested to be included in the diagnostic criteria of AD [5]. A recent meta-analysis (4) suggest that smell identification and then smell detection, is more strongly affected in AD than in Parkinson's disease (PD) - suggesting that AD patients are more strongly impaired on higher-order olfactory tasks involving specific cognitive processes.

The University of Pennsylvania Smell Identification Test (UPSIT) is widely used in AD research to assess smell identification deficits. Its applicability in many clinical practices, particularly in those in which geriatric or demented populations are tested, is limited by administration time. The AD subjects needed approximately 35 minutes to complete the test, whereas the controls needed about 20 minutes, when an examiner assisted all subjects with the test and allowed them unlimited time [6]. In a previous study we have found that AD patients required 27 minutes and NDC 18 minutes [7].

Tests of olfaction are commonly based on total olfactory performance after exposure to multiple odors. Although it is well established that patients with AD perform significantly

more poorly than matched controls, there is little information on whether their ability to identify specific odors is more impaired than the control subjects. We used British version of UPSIT to identify the selective pattern of olfactory deficits and compile an optimal subset of UPSIT items discriminating AD patients from control subjects, with the goal of improving clinical applicability.

SUBJECTS AND METHODS

Subjects:

AD Participants: Mild to moderate AD participants (Mini Mental State Examination score, MMSE; 15-25) (n=54) were recruited from Mental Health for Older Adults (MHOA) services of the South London and Maudsley (SLaM) NHS Foundation Trust, UK and Mental Health Services for Older People (MHSOP), Leicestershire Partnership NHS Trust, UK. Diagnosis of probable Alzheimer's disease was made according to the National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria [8]. The *exclusion criteria* were: dementia other than AD; history of psychiatric disorder, including substance abuse; medical conditions that may alter cerebral functioning or other conditions known to affect olfactory functioning (e.g., common cold, blocked nasal passages, polyps etc.). Subjects had either no history at all of cigarette smoking or had stopped smoking for 20 years or more. Informed consent or assent, as appropriate, was taken from all the patients. The study had ethics approval from both the Joint Institute of Psychiatry and the South London and Maudsley ethics committee and the local ethics committee for Leicestershire Partnership NHS Trust.

Non demented control subjects (NDC):

Eligible and interested non-demented control subjects (n=40) were recruited from the ongoing AD biomarker studies at the centre for this olfaction study. As part of these biomarker studies, subjects with AD and non-demented controls were recruited and assessed as per standard protocol described previously [9, 10]. *Inclusion criteria:* MMSE score range between 24 and 30 (inclusive), medication stable and good general health. *Exclusion criteria:* Meeting DSM-IV criteria for dementia, significant neurological or psychiatric illness, significant unstable systematic illness or organ failure.

Assessments

Assessment included cognitive testing with MMSE [11] ; non-cognitive symptoms were assessed using Neuropsychiatric Inventory (NPI) [12], daily activities with Bristol Activities of Daily Living Scale (BADL) [13] and smell identification function using University of Pennsylvania smell identification test (UPSIT) [14, 15]. UPSIT is a standardised test of smell identification with good test-retest reliability ($r=0.95$) and strong correlation with detailed olfactory threshold tests ($r=0.80$) [15].

Statistics

Comparisons were made on demographic information, clinical characteristics, and cognitive and behavioural test results using the Chi-square Test for categorical outcomes and using parametric (Student's *t*-test, Pearson's ρ correlations) and nonparametric (Mann-Whitney tests, Spearman rank correlations) statistics, as appropriate, for continuous outcomes., using SPSS 20.0, STATA 12 (Stata Statistical Software: Release 10. College Station, TX:

StataCorp LP). Alpha level was set at 0.05; corrections for multiple comparisons were not made given the exploratory nature of the study. A Random Forest approach (using 'rf' and 'rfe' in the R package 'CARET') was used to select a subset of UPSIT items which can accurately classify AD patients compared to controls. AD cases and controls were divided into a training (2/3) and an independent test dataset (1/3) such that the training set comprised equal ratios of each diagnostic group. We additionally tried to maintain equal female/male ratios as well as similar age, education and MMSE means for each diagnostic group between the two datasets.

Random Forest with RFE (i.e. backwards elimination) was used on the Training dataset using 100 Bootstraps on the top 40 UPSIT items in order to rank the UPSIT items according to their predictive power. 100 Bootstraps were used such that the training dataset was randomly split into a bootstrap training set (75%) which was used to build the Random Forest model and a bootstrap test set (25%) which was used to test the model. The bootstrapping was repeated keeping the top 35 to 10 analytes in steps of 5 and down to two analytes in steps of one. In each model the default setting for ntree=500 were used. For each subset of predictors the mean bootstrap testing performance was calculated using the "AUC" and based on this the optimal number of variables was identified. This was then used to build a final model in the complete training data, which was then tested with the independent test set. Finally, we used the "sizeTolerance" function in 'CARET', which takes into account the whole profile during 'rfe' and tries to pick a subset size that is small without sacrificing too much performance. This model was compared to models including a) Final model plus covariates (age, sex, and education) b) Final model plus MMSE.

RESULTS

None of the participants reported subjective impairment in olfaction. The socio-demographic and clinical comparisons between the AD subjects and the NDC are as described in Table 1. The AD group scored higher on the baseline MMSE and UPSIT measures than NDC, and took longer to complete the UPSIT. There were no gender differences in the UPSIT scores in both groups ($\chi^2(1) = 0.92, P = 0.34$). The NDC group had higher education years ($p = 0.001$) than the AD group.

Multivariate analysis

Random Forest with recursive feature elimination was fitted on the 40 UPSIT items down to 2 UPSIT items of the training dataset and showed that the highest mean training performance was for a model with 30 UPSIT items (Training dataset Sensitivity= 86.2%, Specificity=82.8%, ROC=0.90 average after 100 bootstraps). However, in order to identify a model with high accuracy whilst keeping the number of UPSIT items as low as possible, a model with 12 items was chosen (Training dataset Sensitivity= 84.5%, Specificity=82.0%, ROC=0.915 average after 100 bootstraps) (Supplementary Figure 1). As observed in Supplementary Figure 1, this was the point where the average ROC on the Training dataset started to level off and did only slightly worse compared to a model including all 40 UPSIT items and compared to a model using 30 UPSIT items which showed the highest mean performance in the Training datasets.

We then fitted a Random Forest model with the 12 selected UPSIT variables on the whole Training dataset which predicted the Training dataset with 84.5% sensitivity, 82.0% specificity and an Area Under the Curve of 0.919 and could classify our test dataset with 86.7% accuracy, 88.9% sensitivity, 83.3% specificity, a positive predictive value of 0.889, a negative predictive value of 0.833 and an area under the curve of 0.861 (Table 2).

Results for the Random Forest models, including a model with covariates only, a model which includes MMSE and a model including the 40 UPSIT items are displayed in Table 2 and ROC curves in Figure 1.

The 12 UPSIT items most affected in AD, in the order of their significance, were: Motor oil, Banana, Coconut, Apple, Orange, Cherry, Watermelon, Leather, Cedar, Pine, Paint thinner and Peanut. Logistic regression analyses on each of the 12 UPSIT items against disease status for the whole dataset, after adjusting for age, gender and education are presented in Supplementary Table 1. When looking at the Variable Importance measures during "rfe" after 100 Bootstraps it is seen that the greatest contribution to the model was made by motor oil, banana and coconut which produced a ROC curve of >80% for the rfe training set.

DISCUSSION

A subset of 12 of the 40 UPSIT items (British version) differentiated AD subjects from NDC accurately (86.7%) with a sensitivity of 88.9% and specificity of 83.3% on its own. With MMSE, the accuracy improved to 96.7% with sensitivity of 100% and specificity 91.7%. The

accuracy with 12 items were as good as for 40 items, indicating that these 12 items are sufficient to differentiate the groups and there is no added advantage of doing the complete 40 items as the effect seem to level off.

The AD participants were deliberately chosen to be in the early stages (mild-moderate) of the disease, so that there is little question of their ability to understand and perform the smell test. None of the AD participants had speech or language difficulties. All participants were White Europeans except for 4 patients and 1 NDC.

The total UPSIT scores for AD and NDC subjects were similar to those reported in previous studies [16-20]. We did not find any significant effects of gender on olfactory identification deficits which is in keeping with previous studies [21]. A possible explanation could be that, as has been shown previously, gender differences decrease with increasing age [22]. The mean age of both our groups were 72 years (with the range between 45 and 90), which could be a possible reason why women in our normal elderly control group did not show an expected superiority on the odour identification task. Secondly, it is possible that the effects of pathological processes on olfactory functions are so large that they blur gender effects existing in the AD patients [21].

There have been several studies that have looked at the specific UPSIT items (US version) which differentiated AD patients from controls. Of the 40 UPSIT items, the 5 items that maximally differentiated ($p < 0.005$) the Alzheimer's and control groups were bubble gum (No. 2), root beer (No. 24), watermelon (No. 30), grass (No. 32), and rose (No. 39), in a

sample of 17 patients and 17 controls [6]. In another study, UPSIT data from control subjects, patients with mild cognitive impairment and AD patients were analysed to derive an optimal subset of items related to risk for Alzheimer's disease (i.e., healthy through mild cognitive impairment to early and moderate disease stages) [23]. The derived 10-item scale performed comparably with the UPSIT in classifying subjects, and it strongly predicted conversion to Alzheimer's disease on 2-year follow-up evaluation in patients with mild cognitive impairment. The items were Menthol, Clove, Leather, Strawberry, Lilac, Pineapple, Smoke, Soap, Natural gas and Lemon.

Our study, using the British version of UPSIT, identified the 12 items most affected in AD patients, which in the order of their significance were: Motor oil (No.5), Banana (No.7), Coconut (No.10), Apple (No.20), Orange (No.28), Cherry (No.4), Watermelon (No.30), Leather (No.9), Cedar (No.18), Pine (No.34), Paint thinner (No.31) and Peanut (No.40). It's interesting that the first 3 items, motor oil, banana and coconut itself made the greatest contribution to produce a ROC curve of >80%.

According to a recent classification of odours, into 10 categories using non-negative matrix factorization (NMF) to odor profiling data for human odor percepts, the 12 items most affected in our AD patients fall into the categories of Chemicals, Fruits (citrus and non-citrus) and Popcorn (Nuts) [24]. Interestingly these categories reflect important attributes such as danger and food. Deficits in identification of odors in these categories could explain certain symptoms of AD, such as decreased sense of danger such as cooking related

incidents, leaving gas cooker on due to inability to detect leaking gas, storage or ingestion of spoiled food and loss of weight.

There have also been reports looking into one specific odor identification deficit in AD patients. In one study participants with probable AD (N=18), mild cognitive impairment (MCI, N=24), other causes of dementia (OD, N=26) and matched controls (OC, N=26) were tested, with closed eyes, for their ability to detect an odor, one nostril at a time, using peanut butter [25]. A left nostril impairment of odor detection was present in all the patients with probable AD compared to controls (100% sensitive and 92% specific).

However, a recent study sought to replicate and expand this work by conducting 2 studies; one using the same methodology as Stamps et al and another using UPSIT 20 items on each nostril, and found that no meaningful asymmetry in smell perception was apparent [26].

The limitations of the study are: firstly, although differences between the odors identified here and those identified in previous reports, can be accounted for to some degree by methodological differences (e.g., sample size, US vs British versions, statistical approaches), the lack of consistency in the odors identified as optimal across studies emphasizes the need for independent replication. Another explanation for variability across studies regarding specificity of odorants is that such odorants are not unitary entities and made up of number of chemicals [27, 28]. Even a single chemical can stimulate a range of receptors that overlap the receptive fields of other single compound stimuli [29]. Importantly, the number of receptors involved with different stimuli is unknown, even at the receptor sites between components of a single odor and much less the interactions that occur within the olfactory

bulb [29, 30]. Hence, there can be considerable variability in stimuli across studies. The greater risk of impairment, associated with certain UPSIT items may be explained by AD-related pathology in olfactory related brain areas which may differentially impact the neural circuitry, mediating odor processing [23] and this needs further investigation.

We did not test and compare UPSIT performance with other causes of dementia such as Lewy body dementia. However this is beyond the remit of our study and further work is needed to look at this.

For olfactory tests to be clinically useful for early detection of AD, they must be brief, sensitive to early pathological and functional changes and significantly add to the predictive accuracy of known demographic and clinical risk factors [23]. The proposed 12-item scale meets these criteria and can be self-administered by patients or carers with minimal supervision. Scoring is objective and needs minimal training, which can be reviewed and interpreted in the context of a clinical workup. Pending independent replication, this 12-item subset's potential diagnostic and predictive utility, make a strong case for including it as part of a standard clinical workup for patients with suspected AD. Specificity against other disorders such as PD also needs to be established, with neuropathological confirmation if possible. Finally, future cross-cultural validation studies will be needed to ascertain the discriminative and predictive potential of these derived items in a multi-ethnic British population.

In conclusion, if the proposed UPSIT subset's predictive accuracy for AD diagnosis is replicated and is comparable or superior to other early markers, it may, in combination with other diagnostic tests, help to improve early detection of AD in clinical settings.

Funding: This work was supported by the National Institute for Health Research Specialist Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, King's College London and Alzheimer's Research UK (ARUK) East Midlands Network funding.

Author responsibilities: LV designed the study, carried out the data collection, performed statistical analysis and interpretation. AG, SB, MP and CM contributed to the data collection. PP contributed to the statistical analysis and interpretation. LV wrote the paper with contributions from all authors. LV had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

- [1] Mesholam RI, Moberg PJ, Mahr RN, Doty RL (1998) Olfaction in neurodegenerative disease: a meta-analysis of olfactory functioning in Alzheimer's and Parkinson's diseases. *Arch Neurol* **55**, 84-90.
- [2] Christen-Zaech S, Kraftsik R, Pilleuit O, Kiraly M, Martins R, Khalili K, Miklossy J (2003) Early olfactory involvement in Alzheimer's disease. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques* **30**, 20-25.
- [3] Attems J, Konig C, Huber M, Lintner F, Jellinger KA (2005) Cause of death in demented and non-demented elderly inpatients; an autopsy study of 308 cases. *Journal of Alzheimer's disease : JAD* **8**, 57-62.

- [4] Rahayel S, Frasnelli J, Joubert S (2012) The effect of Alzheimer's disease and Parkinson's disease on olfaction: a meta-analysis. *Behav Brain Res* **231**, 60-74.
- [5] Foster J, Sohrabi H, Verdile G, Martins R (2008) Research criteria for the diagnosis of Alzheimer's disease: genetic risk factors, blood biomarkers and olfactory dysfunction. *International psychogeriatrics / IPA* **20**, 853-855.
- [6] Warner MD, Peabody CA, Flattery JJ, Tinklenberg JR (1986) Olfactory deficits and Alzheimer's disease. *Biological psychiatry* **21**, 116-118.
- [7] Velayudhan L, Pritchard M, Powell JF, Proitsi P, Lovestone S (2013) Smell identification function as a severity and progression marker in Alzheimer's disease. *International psychogeriatrics / IPA* **25**, 1157-1166.
- [8] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944.
- [9] Thambisetty M, Simmons A, Velayudhan L, Hye A, Campbell J, Zhang Y, Wahlund LO, Westman E, Kinsey A, Guntert A, Proitsi P, Powell J, Causevic M, Killick R, Lunnon K, Lynham S, Broadstock M, Choudhry F, Howlett DR, Williams RJ, Sharp SI, Mitchelmore C, Tunnard C, Leung R, Foy C, O'Brien D, Breen G, Furney SJ, Ward M, Kloszewska I, Mecocci P, Soininen H, Tsolaki M, Vellas B, Hodges A, Murphy DG, Parkins S, Richardson JC, Resnick SM, Ferrucci L, Wong DF, Zhou Y, Muehlboeck S, Evans A, Francis PT, Spenger C, Lovestone S (2010) Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Archives of general psychiatry* **67**, 739-748.
- [10] Velayudhan L, Killick R, Hye A, Kinsey A, Guntert A, Lynham S, Ward M, Leung R, Lourdasamy A, To AW, Powell J, Lovestone S (2012) Plasma transthyretin as a candidate marker for Alzheimer's disease. *Journal of Alzheimer's disease : JAD* **28**, 369-375.
- [11] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* **12**, 189-198.
- [12] Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J (1994) The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology* **44**, 2308-2314.
- [13] Bucks RS, Ashworth DL, Wilcock GK, Siegfried K (1996) Assessment of activities of daily living in dementia: development of the Bristol Activities of Daily Living Scale. *Age and ageing* **25**, 113-120.
- [14] Doty RL, McKeown DA, Lee WW, Shaman P (1995) A study of the test-retest reliability of ten olfactory tests. *Chemical senses* **20**, 645-656.
- [15] Doty RL, Shaman P, Dann M (1984) Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiology & behavior* **32**, 489-502.
- [16] Warner MD, Peabody CA, Flattery JJ, Tinklenberg JR (1986) Olfactory deficits and Alzheimer's disease. *Biol Psychiatry* **21**, 116-118.
- [17] Serby M, Larson P, Kalkstein D (1991) The nature and course of olfactory deficits in Alzheimer's disease. *Am J Psychiatry* **148**, 357-360.

- [18] Kareken DA, Doty RL, Moberg PJ, Mosnik D, Chen SH, Farlow MR, Hutchins GD (2001) Olfactory-evoked regional cerebral blood flow in Alzheimer's disease. *Neuropsychology* **15**, 18-29.
- [19] Djordjevic J, Jones-Gotman M, De Sousa K, Chertkow H (2008) Olfaction in patients with mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging* **29**, 693-706.
- [20] Li W, Howard JD, Gottfried JA (2010) Disruption of odour quality coding in piriform cortex mediates olfactory deficits in Alzheimer's disease. *Brain* **133**, 2714-2726.
- [21] Djordjevic J, Jones-Gotman M, De Sousa K, Chertkow H (2008) Olfaction in patients with mild cognitive impairment and Alzheimer's disease. *Neurobiology of aging* **29**, 693-706.
- [22] Larsson M, Nilsson LG, Olofsson JK, Nordin S (2004) Demographic and cognitive predictors of cued odor identification: evidence from a population-based study. *Chemical senses* **29**, 547-554.
- [23] Tabert MH, Liu X, Doty RL, Serby M, Zamora D, Pelton GH, Marder K, Albers MW, Stern Y, Devanand DP (2005) A 10-item smell identification scale related to risk for Alzheimer's disease. *Ann Neurol* **58**, 155-160.
- [24] Castro JB, Ramanathan A, Chennubhotla CS (2013) Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. *PLoS one* **8**, e73289.
- [25] Stamps JJ, Bartoshuk LM, Heilman KM (2013) A brief olfactory test for Alzheimer's disease. *J Neurol Sci* **333**, 19-24.
- [26] Doty RL, Bayona EA, Leon-Ariza DS, Cuadros J, Chung I, Vazquez B, Leon-Sarmiento FE (2014) The lateralized smell test for detecting Alzheimer's disease: Failure to replicate. *J Neurol Sci* **340**, 170-173.
- [27] Amoore JE (1974) Evidence for the chemical olfactory code in man. *Annals of the New York Academy of Sciences* **237**, 137-143.
- [28] Johnson BA, Leon M (2007) Chemotopic odorant coding in a mammalian olfactory system. *The Journal of comparative neurology* **503**, 1-34.
- [29] Raman B, Stopfer M, Semancik S (2011) Mimicking biological design and computing principles in artificial olfaction. *ACS chemical neuroscience* **2**, 487-499.
- [30] Doty RL, Kamath V (2014) The influences of age on olfaction: a review. *Frontiers in psychology* **5**, 20.

Table 1: Socio demographic clinical parameters comparison between AD subjects and NDC

Variables Mean (SD) or %	AD (n=54)	NDC (n=40)	p-value
Female	44%	55%	NS ^a
Age (years)	73.5 (11)	70.5 (9)	NS [*]
Education (years)	10.9 (2)	12.9 (3)	0.001 [*]
Disease Duration (months)	38.5 (33)	n/a	n/a
MMSE	22.5(5)	29.2 (1)	<0.001 [*]
40 UPSIT items	19.7 (7)	30.9 (5)	<0.001 [*]
40 UPSIT, time taken (minutes)	24.3 (10)	18.7 (5)	0.004 [*]
12 UPSIT	5.4 (3)	10.4 (2)	<0.001 [*]
BADL	8.2 (8)	n/a	n/a
NPI	8.6 (14)	n/a	n/a

Values are mean (SD) or n (%); ^a Calculated using the χ^2 test, ^{*} Calculated using the t test, AD, Alzheimer's disease; NDC, non-demented controls; MMSE, Mini Mental State Examination; UPSIT, University of Pennsylvania Smell Identification Test; BADL, Bristol Activities of daily living, NPI, Neuropsychiatric Inventory; n/a, not applicable

Table 2: Multivariate analysis results

AD subjects vs NDC									
a. Training Set (n=64)				b. Data Set (n=30)					
Random Forest Models	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC	PPV	NPV
12 UPSIT items	84.5%	82.0%	0.915	86.7%	88.9%	83.3%	0.861	0.889	0.833
12 UPSIT items plus covariates	88.2%	84.0%	0.929	93.3%	100%	83.3%	0.917	0.900	1.000
12 UPSIT +MMSE	88.3%	85.1%	0.948	96.7%	100%	91.7%	0.958	0.947	1.000
40 UPSIT items	79.1%	84.9%	0.893	86.7%	88.9%	83.3%	0.861	0.889	0.833
40 UPSIT items plus covariates	81.0%	84.4%	0.895	90%	88.9%	91.67%	0.903	0.941	0.846
40 UPSIT +MMSE	88.8%	83.3%	0.931	96.7%	100%	91.7%	0.958	0.947	1.000

AD, Alzheimer's disease; NDC, non-demented controls; MMSE, Mini Mental State

Examination; UPSIT, University of Pennsylvania Smell Identification Test; Covariates= Age, gender and education

Figure 1: ROC values for the recursive feature elimination (rfe) on the 40 UPSIT items using the Random Forest function and a decreasing number of variables in each step (40 down to 2)

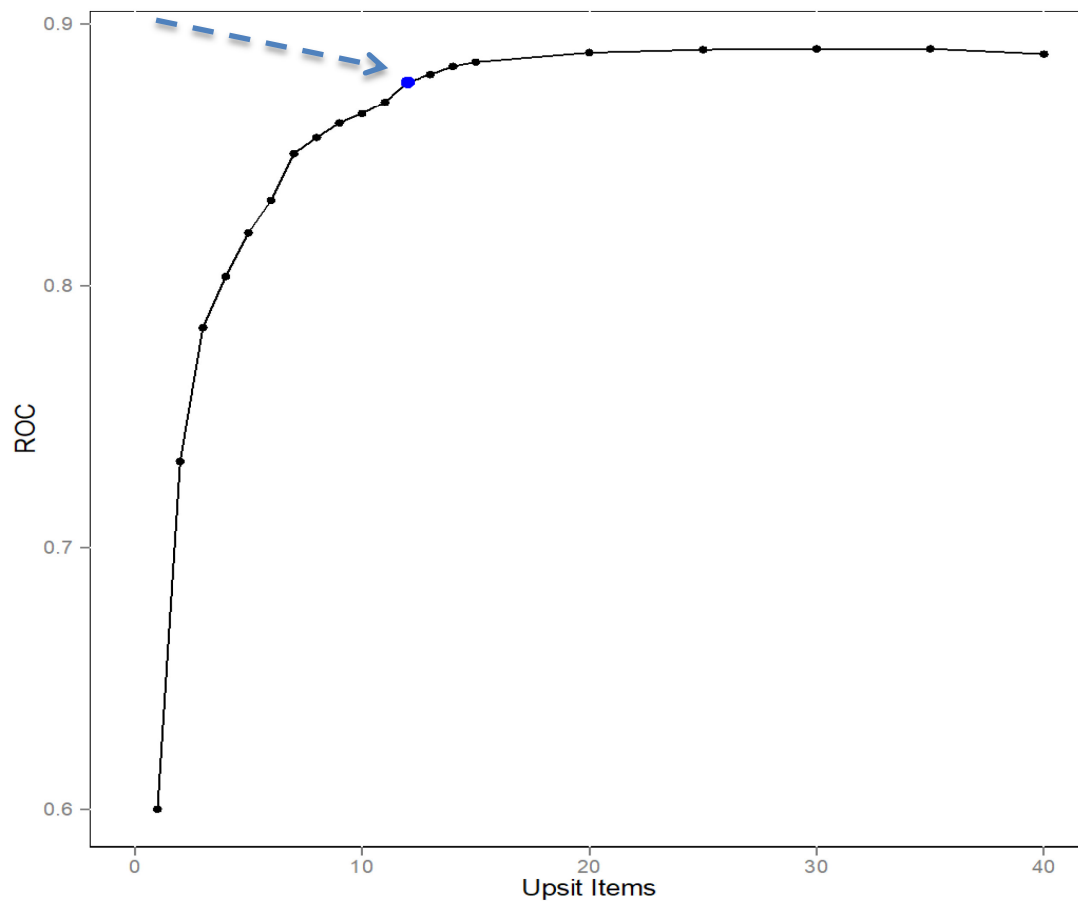
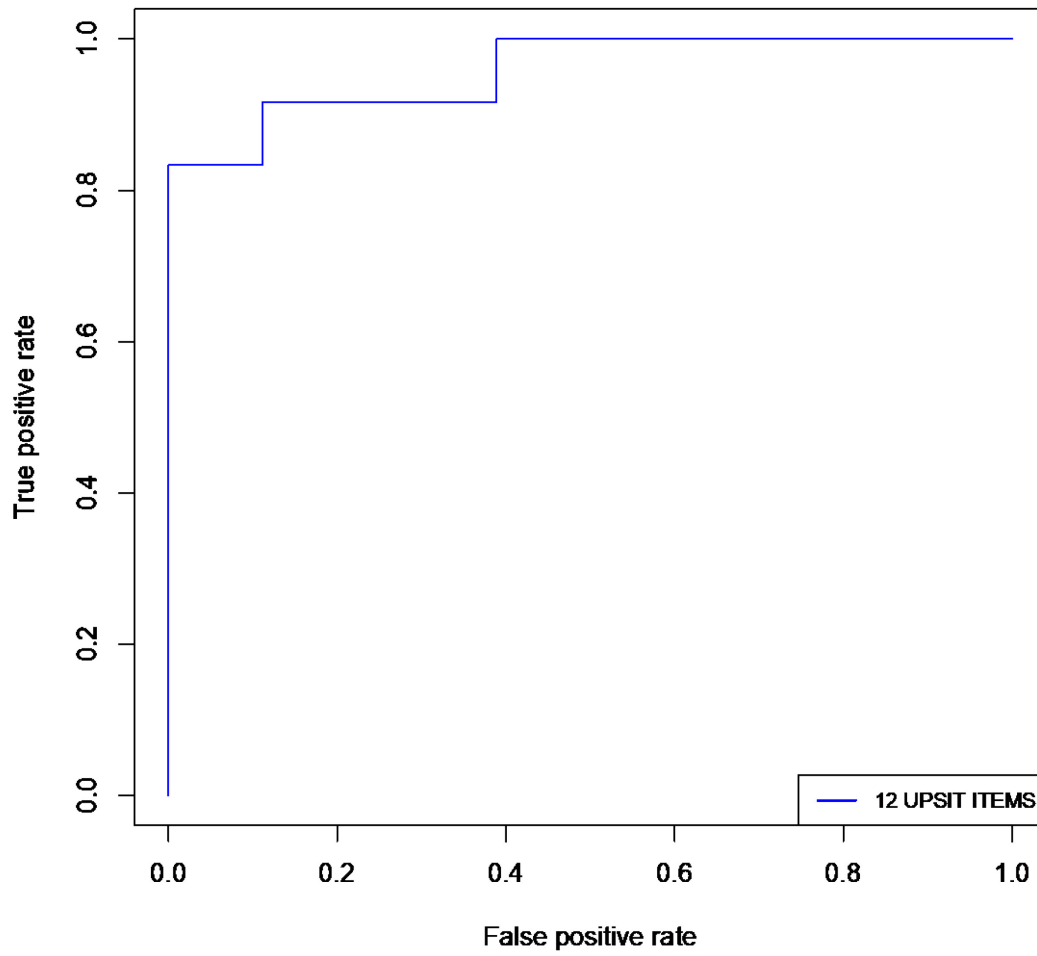


Figure 2. . Receiver operator curve (ROC) for the independent test dataset for the selected 12 UPSIT items after RFE for AD versus controls.

ROC plot: Random Forest Results for the Test dataset



Supplementary Table 1: Results of univariate logistic regression on the 10 items selected after adjusting for age, gender and education. The UPSIT items are ordered according to their importance during recursive feature elimination.

UPSIT item	OR	L95	U95	P value
UPSIT motor oil	6.427	2.44	18.88	3.17E-04
UPSIT banana	12.088	3.67	55.39	2.08E-04
UPSIT coconut	5.879	2.18	17.79	8.17E-04
UPSIT apple	17.620	5.38	76.75	1.56E-05
UPSIT orange	8.460	2.79	32.17	4.72E-04
UPSIT cherry	8.390	2.96	27.32	1.48E-04
UPSIT watermelon	10.920	3.43	43.55	1.79E-04
UPSIT leather	15.503	3.87	107.76	7.40E-04
UPSIT cedar	4.376	1.54	14.09	8.00E-03
UPSIT pine	9.808	3.12	38.64	2.86E-04
UPSIT paint thinner	3.762	1.41	11.35	1.182-02
UPSIT peanut	8.256	2.64	32.45	7.642E-04

