Is The University Of Pennsylvania Smell Identification Test (UPSIT) Valid for the UK Population?

Abstract

Aims: *UK publications base evidence on University of Pennsylvania Smell Identification Test (UPSIT) results, yet UPSIT normative values are not transferrable outside the USA. This study was designed to (a) Estimate normative scores for the UK population and (b) Identify smells causing cultural bias.*

Methods: *The UPSIT test was applied to 44 healthy individuals in the UK.*

Results: *Results showed a mean score of 32.36(males) and 33.27(females), equivalent to an UPSIT olfactory diagnosis of 'mild microsmia' ($p<0.05$, $p<0.0005$ respectively). Six smells were identified as culturally biased and removed resulting in diagnosis of 'normosmia'($p=0.1$).*

Conclusions: *The present UPSIT test needs to be adapted to produce a reliable, validated and culturally appropriate olfactory test for use on the UK population.*

Keywords

Smell, anosmia, UPSIT-UK, cultural, olfactory test

Introduction

Smell testing is a useful objective clinical tool for assessing changes to the olfactory mucosa, for example during disease processes such as sinusitis and polyps.[1] Furthermore, reduced olfaction has been related to alcoholism, early stages of Alzheimer's disease, cigarette smoking, cystic fibrosis, Down's syndrome, industrial chemical exposure, Korsakoff's psychosis, Parkinson's disease, sinusitis and many others.[2]

The University of Pennsylvania Smell Identification Test (UPSIT) has become a 'gold standard' in olfactory testing. It is now the most widely used smell test in the world.[2] UPSIT has been administered to an estimated 35,000 persons in North America where it has gained wide acceptance within the general medical community.[3] The UPSIT comprises forty different smells released by scratching a panel of microencapsulated odorants using a pencil lead. For each of the forty smells, participants must choose an answer from four possible options; only one is correct. Booklets containing the smells, four, each with ten smells, are in sealed envelopes and a pencil provided (Figure 1).

Data collected on the USA population showed that participants' score out of forty was dependent on gender, age and smoking status.[1] Normative data was collected on 1819 men and 2109 women. A diagnostic ladder was drawn up and six categories devised for olfactory diagnosis depending on score (Figure 3). The term 'microsmia' describes lessened smell function as can be determined by UPSIT. A diagnosis of hyposmia is not used as UPSIT does not test the function of the nerve. The major advantages of UPSIT are that there is already a wealth of data on the results and due to its design it has strong internal consistency reliability.[3]

One of the major drawbacks of using UPSIT in the UK is that it is widely recognized the normative values may not be easily transferable from the USA to other countries, where several of the odours used in the test are not familiar.[4] It has already been shown that cultural and language differences result in mean scores for Chinese and Indian subjects falling below the normosmic range.[5] In the last two decades the sniff magnitude test, alcohol sniff test, anosmia 3 item sniff test, sniffin' sticks test, Barcelona Smell Test

N. Muirhead[1]

E. Benjamin[1]

H. Saleh[1]

[1] Department of ENT, Charing Cross Hospital

Correspondence:
N. Muirhead
Department of ENT,
Charing Cross Hospital,
Fulham Palace Road,
London, W6 8RF
Email: nina.muirhead@btinternet.com

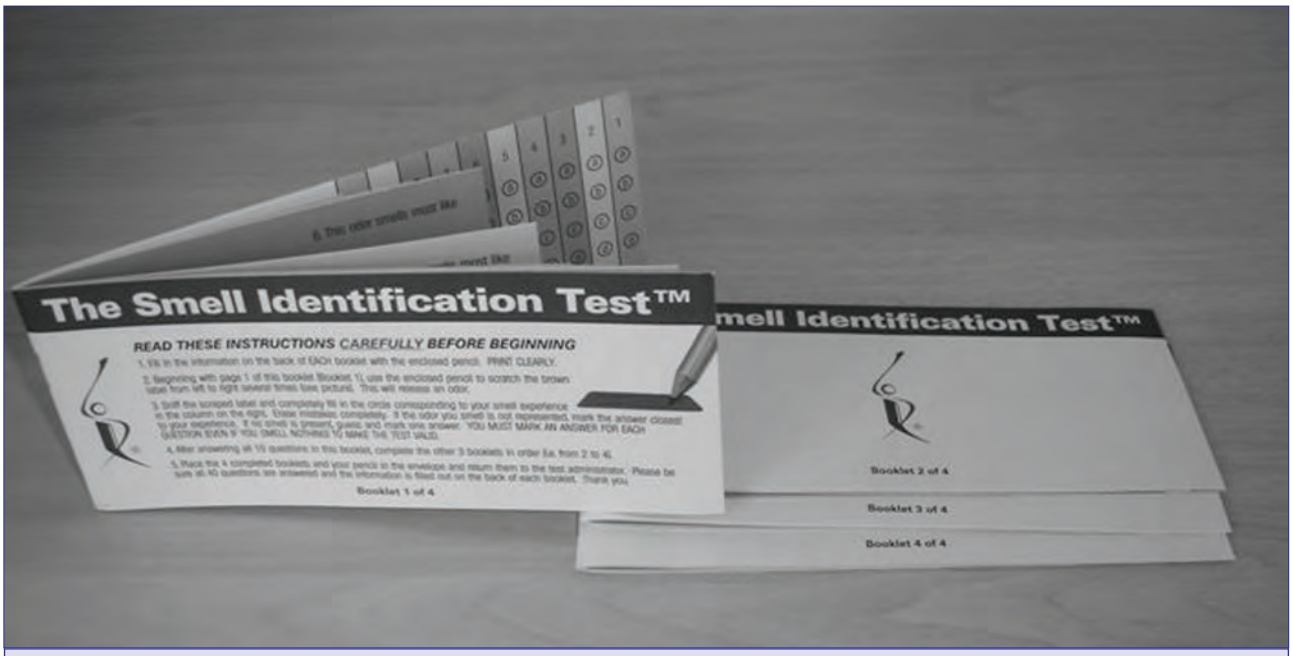

Figure 1: University of Pennsylvania Smell Identification Test (UPSIT)

- 24, Le Nez du Vin, olfactory evoked potentials and forced choice butanol threshold test have all been devised.[6, 7, 8, 9] Doty, Marcus and Lee also worked on the development of a cross-cultural test (CCSIT) based upon twelve of the UPSIT smells.[10] However each time a new test is developed, time and money are required to collect data that will give reliable predictions both for normal members of the population as well as for those with pathology affecting olfactory function.

Culturally modified versions of the UPSIT have now been developed in 12 different languages. Recent additions are a Chinese version being trialed in Taiwan UPSIT-CT and a Brazilian-Portuguese version UPSIT-Br2.[11] Despite the evidence from large control studies based on the UK population that results are below the normal levels[12] data has been published on UK patients using the US normative data.[13]

For these reasons a pilot study was designed to use UPSIT on the UK population to collect data on 'normal' individuals and identify which smells or names of smells were causing confusion. We suspected that language and cultural differences between the USA and the UK would lead to a reduction in test score and worked on the assumption that normal olfactory function of the UK population is equal to normal olfactory function in the USA. Results of this pilot study will be used to identify culturally unfamiliar smells to be replaced with more 'British' smells to create a new 40-item smell test.

Materials & Methods

UPSIT booklets were sent directly from the USA in sealed envelopes to the ENT department at Charing Cross hospital.

Participants for the pilot study were chosen from a random population of healthy individuals in and around London. Participants were selected in equal numbers of males (22) and females (22), between the ages of 18 and 80, mean 44.47 years (16.61) (Figure 2).

The proportion of non-smokers to smokers (18% smokers) roughly reflects that of the USA participants in Doty's original work (19% smokers).[3] Ethnic groups included two Asian, one Afro-Carribean, one Oriental and forty white Caucasian participants. All participants were born, and had lived for the majority of their life in the UK. Other criteria included; no previous surgery on their nose, no known nasal pathology and no other significant medical history correlated with decreased olfactory function. The same operator was used to ask questions and demonstrate the 'scratch and sniff' technique to participants to reduce bias. Questions asked to each participant included: (1) Of the forty smells were there any that were unfamiliar, or difficult to detect? (2) Were any of the names of smells unfamiliar?

Results

Scores out of forty were obtained for all 44 participants and had mean 32.82, median 33.5 and mode 37. Females scored higher than males with mean female score 33.27 (95% CI= 33.22 to

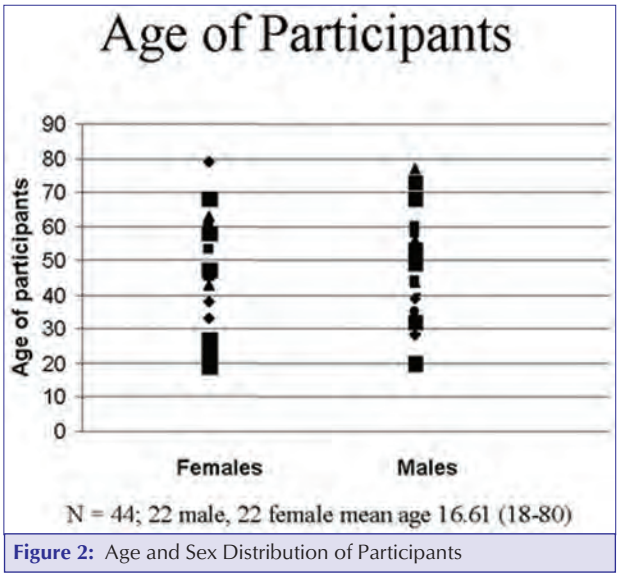

Figure 2: Age and Sex Distribution of Participants

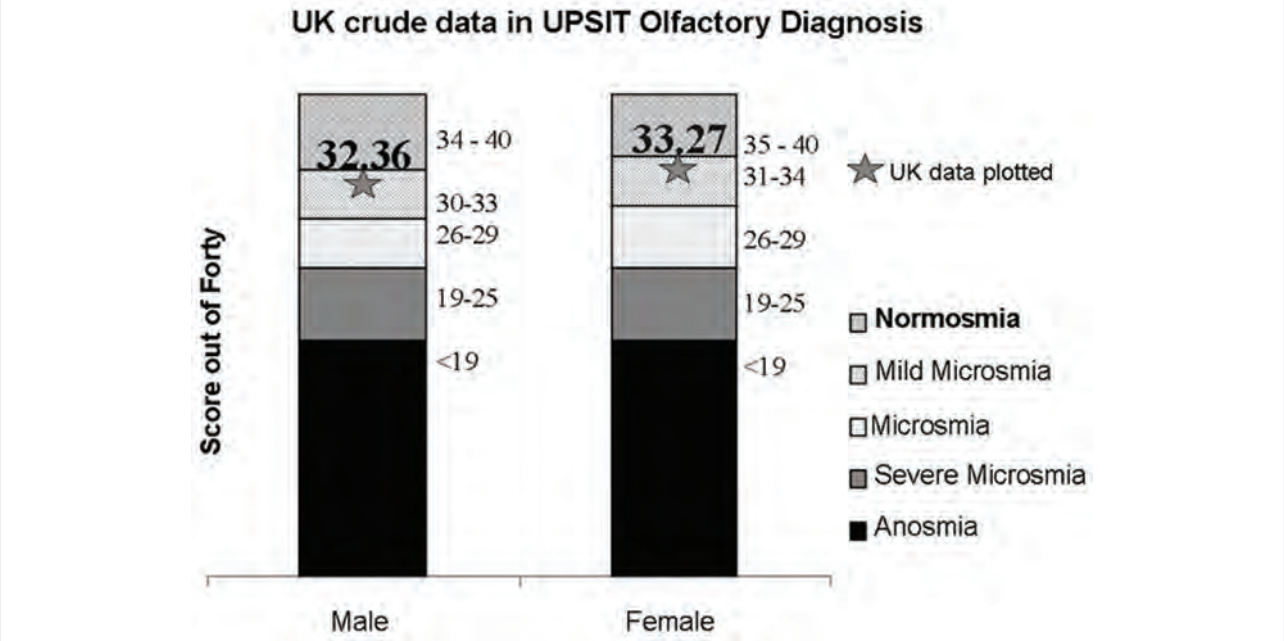

Figure 3: The UK mean UPSIT crude scores plotted on the UPSIT 'diagnostic ladder'

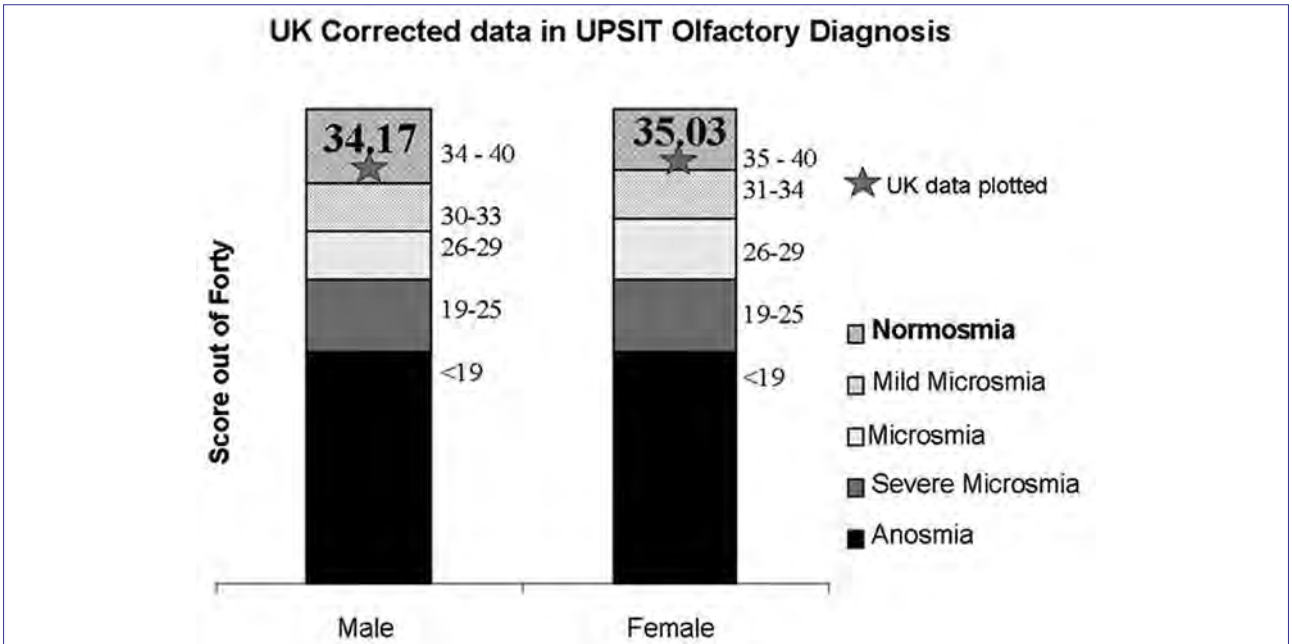

Figure 4: The UK mean UPSIT corrected scores plotted on the UPSIT 'diagnostic ladder'

33.33) compared to mean male score 32.36 (95% CI=32.30 to 32.43). Standardised UPSIT scores from normative data from an American population show that a diagnosis of normosmia results from males scoring greater than 34, whilst women need a score above 35 (Figure 3). UK results for males and females fall below the UPSIT normosmic category therefore giving an olfactory diagnosis of mild microsmia. (Figure 3).

Chi squared analysis, showed that the proportion of expected subjects who would be normosmic, based on the American normative data compared to those observed in our UK population was highly significantly different, (Female; Chi Squared = 11.51, $p<0.05$, Male ; Chi Squared = 12.21 $p<0.005$)

Results for individual smells showed that fifteen out of forty were correctly identified by more than 90% of participants with motor oil, mint and paint thinner being correctly sensed by all 44 participants. However six smells were identified correctly by less than 70% of the participants: cheddar cheese (52.3%), root beer (52.3%), lemon (56.8%), lime (56.8%), dill pickle (61.4%) and turpentine (65.9%).

If all questions were equally difficult, the statistical prediction (binomial distribution) determines that no single question should be answered incorrectly by more than 13 subjects, as this was not the case, then a bias must be present for these questions, which we presume to be a cultural bias.

Results were then re-calculated by removing these six smells and finding the sum of remaining scores out of 34. For the purpose of direct comparison with original UPSIT data, scores were then converted back to a predicted score out of forty. Results had mean 34.6. Females still scored higher than males with a mean of 35.03 (95%CI = 34.97 to 35.08) compared to 34.17 (95%CI = 34.11 to 34.22) respectively. These converted scores put the UK population in the UPSIT normosmic category (Figure 4).

Re-calculation of the Chi squared analysis, once the culturally biased smells had been removed showed no difference between the expected proportion of normosmics, based on the American normative data compared to those observed in our UK population. (Female; Chi Squared = 1.73, p=0.1, Male; Chi Squared = 2.29 p=0.1)

15 of the 44 (34.1%) of participants reported that smell number 14, cheddar cheese was difficult to recognize. Other names of smells that were reported as unfamiliar were wintergreen (27%), root beer (18%), pumpkin pie (9%), dill pickle, turpentine (7%) skunk, musk, gasoline (4.5%). Other general comments included that fruit smelled sweet, and therefore reflected the smell of sweets rather than fruit, and that the smell of soap or pizza may vary depending on flavor.

Discussion

UPSIT is a useful clinical test for sense of smell. This pilot study on the UK population showed results that fell below the mean expected for a normosmic population. This was attributed to two factors. Firstly some of the *smells* were culturally unfamiliar, such as root beer and dill pickle and American-style cheddar cheese. Secondly unfamiliar *names* of smells including turpentine, skunk, root beer and musk served to put participants off when selecting the answer. This is similar to the findings of the Brazilian study where at least 10 items were not highly identifiable by subjects including dill pickle (mistakenly translated as "pepino", which means cucumber) and soap.[11]

It was initially surprising to find cheddar cheese amongst the list of smells that participants did not recognize. It became apparent that the smell in the UPSIT booklet had a much sweeter, synthetic quality than the sharp English cheddar smell that UK participants would have been expecting. Although 'wintergreen' was a name that was identified as unfamiliar by a quarter of participants, only 15.9% chose the wrong option. This is likely to be due to the fact that it is a distinctive smell and the other three options were easily eliminated. To simply re-name this, and change any other unfamiliar names, such as exchanging 'petrol' for 'gasoline' should be a sufficient modification in the development of a UK test.

This study was a simple, quick and effective way of highlighting problems associated with applying the UPSIT to the UK population. Design ideas aimed at reducing bias included recruiting equal numbers of males and females and the proportion of smokers to non-smokers also reflected Doty's original cohort. Participants were exposed to the same level of supervision, and given the same briefing before attempting the test.

We have identified some minor weaknesses with the study. The tests used were outside their recommended 'use by' date (by a few months), however Doty and Agrawal showed that even four years after manufacture there was no significant overall decline of test results.[14] Number 36 (lemon) was identified as one smell that did significantly decline over time, falling from an accuracy of 100% with recent manufacture to 97% at 2.5 years and only 67% at four years.[14] This was one of the six smells removed when pilot results were re-calculated in our study. Another of the six smells removed was lime. Although less abundant in the UK than the USA, lime should still be a familiar smell to the UK population. On closer inspection there were two other options for this scent that could have confounded results; musk and turpentine. Both odors occurred in the list of names that participants reported they had difficulty recognizing. Additionally, turpentine is one of the microencapsulated odors tested before 'lime', it appears that those who did not select the answer 'turpentine' the first time, often selected it instead of lime. The incorrect option frequently chosen for turpentine was skunk; accounting for nine of the fifteen incorrect answers. Although skunk was not one of the microencapsulated odors, it appears that the unfamiliar name of the option was enough to distract participants.

Average age of participants in the pilot study was 44.47(+/-16.61) years whilst the average age of Doty's UPSIT participants was 35.24(+/-19.21) years. It is widely accepted that sense of smell diminishes with increasing age and therefore our UK results may be an underestimate when compared to UPSIT normative data. A large control group of 145 UK patients were recruited for comparing olfactory function in parkinsons disease, pure autonomic failure and autonomic atrophy. The mean age was also higher 63.8 (+/-9.5) years as were smoking rates (only 51% had never smoked). Results for these 'normal' UK controls fell into the UPSIT microsmic range with an average score of 29.7 (SD5.4),[15] despite the increased age this is a large cohort scoring well below the expected normal and provides further evidence that UPSIT is culturally inappropriate.

Our findings of a cultural bias are brought with comparison to other studies. The 12-Item Cross Cultural Smell Identification Test (CC-SIT) was designed in 1996 by Doty *et al.*[9] CC-SIT is a short, easily used, empiric assessment of olfactory function. It was also designed to reduce the probability of misdiagnosis due to cultural differences. After consulting representatives from countries in Europe, South America and China, it was decided that only 12 of the original 40 smells would be used for the CCSIT test.

We suspect that cultural differences are the main reason why UPSIT is not appropriate for the UK population. We therefore removed all scores apart from scores on the 12 CC-SIT smells, to find out if the theory of removing cultural bias would put our 'normal' UK population above the level of mild microsmia.

Using results from our pilot study for just the 12 'cross cultural' smells, mean score out of 12 was 9.91 (95%CI 9.89 to 9.93). This fell well within CC-SIT normosmic category (Figure 5). Other low scoring smells highlighted by the research carried out in the development of CC-SIT included root beer and dill pickle. By removing these from UPSIT to produce a UK test it may transpire that a new UK test is more applicable worldwide.

Cardesin *et al* have also tried to produce a test that would gain general acceptance.[16] They segregated volunteers by gender, age and smoking habit and employed a new olfactory test, the Barcelona Smell Test 24 (BAST 24). The test consists of 24 odours scoring three variables; smell detection, identification and forced choice. Although they concluded that this was a 'good and reliable method' there were some

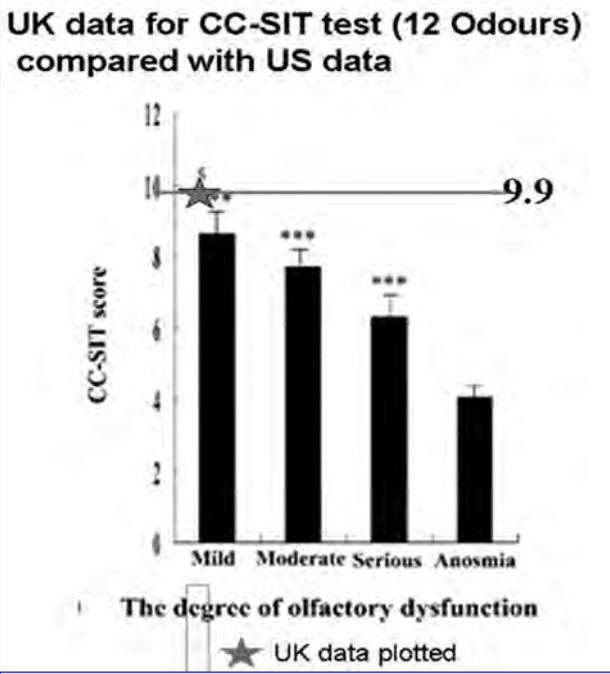

Figure 5: UK data (for 12 CC-SIT odours) plotted on original CC-SIT data 'diagnostic ladder'

shortcomings. Although it is a 'shorter' test in terms of numbers of odours, the time taken to brief the patient on the three variable tests may be longer.

There are two factors that would make development of a UK version of the 40-item UPSIT preferable to either of these tests. Firstly the size is important. It has already been shown that the longer the test the more repeatable, reliable and sensitive the test. Using the Spearman Brown Formula that relates test length to reliability a 12-item test only has 73% reliability compared to 92% reliability of the UPSIT.[10] Due to its size UPSIT can detect malingering based on a statistically improbable score (≤5 out of 40), shorter tests produce no score distinction between malingering and anosmia. Secondly there is a wealth of background data available, not only for normal population members but much more relevantly for pathological conditions related to olfactory function based on a 40-item UPSIT score.

We are now in the process of developing a 40-item smell test that would be suitable for the UK population. The UPSIT-UK test will be designed to avoid both smells, and names of smells, that are culturally unfamiliar. Once this has been produced then normative data for this new 40-item test can be collected for the UK. There was a substantial overlap between smells highlighted as culturally unfamiliar in this study and during research for development of CC-SIT.[10] By removing these smells that are seemingly specific to North America, a new 40-item UK test may be more applicable worldwide.

Conclusions

It has been shown that the present UPSIT test when applied to the UK population is inappropriate due to cultural differences and olfactory experiences between the UK and American populations. Any publications or legal evidence based on UPSIT scores for UK patients are not currently valid without direct comparison with suitable UK controls. The test needs to be adapted to produce a reliable, validated and culturally appropriate olfactory test for use on the UK population.

Conflict of Interest

All authors have no conflict of interest to declare. No extraneous funding was obtained.

References

1. KernCandidate's Thesis: Chronic Sinusitis and Anosmia: Pathologic Changes in the Olfactory Mucosa. *Laryngoscope.* 2003; **110**(7):1071-1077
2. Jiang RS, Su MC, Liang KL, Shiao JY, Wu SH, Hsin CH. A pilot study of a traditional Chinese version of the University of Pennsylvania Smell Identification Test for application in Taiwan. *Am J Rhinol Allergy.* 2010 Jan-Feb;**24(1)**:45-50.
3. Doty RL, Frye RE and Agrawal U. 1989 Internal consistency reliability of the fractionated and whole University of Pennsylvania Smell Identification Test.
4. Katzenschlager R, Zijlmans J, Evans A, Watt H, Lees AJ. Olfactory function distinguishes vascular parkinsonism from Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 2004 Dec;**75(12)**:1749-52.
5. Frank *et al.* A comparison of the sniff magnitude test and the University of Pennsylvania Smell Identification Test in Children and non-native English speakers. *Physiology and Behavior.* 2004;475-480
6. Asherdorf *et al.* Performance of community dwelling adults ages 55-75 University of Pennsylvania Smell Identification Test; an item analysis *Applied Neuropsychology.* 2005;**12(1)**:24-9
7. Jackman, Alexis H, Doty Richard L. Utility of a Three-Item Smell Identification Test in Detecting Olfactory Dysfunction. *Laryngoscope.* 2005 Dec;**115(12)**:2209-12
8. Forster G, Damm M, Gudziol H, Hummel T, Huttenbrink KB, Just T, Seeber H, Welge-Luessen A Testing the sense of smell using validated procedures *Z Arztl Fortbild Qualitatssich.* 2004; Jun;**98(4)**:279-81
9. McMahon C, Scadding GK. Le Nez du Vin –a quick test of olfaction *Clin Otolaryngol Allied Sci.* 1996; Jun;**21(3)**:278-80
10. Doty RL Marcus A, Lee WW. Development of the 12-Item Cross-Cultural Smell Identification Test. *Laryngoscope.* 1996; **106(3)**:353-356
11. Silveira-Moriyama L, Azevedo AM, Ranvaud R, Barbosa ER, Doty RL, Lees AJ Applying a new version of the Brazilian-Portuguese UPSIT smell test in Brazil. *Arq Neuropsiquiatr.* 2010 Oct;**68(5)**:700-5.
12. Silveira-Moriyama L, Mathias C, Mason L, Best C, Quinn NP, Lees AJ. Hyposmia in pure autonomic failure. *Neurology.* 2009 May 12;**72(19)**:1677-81.
13. Gibberd FB, Feher MD, Sidey MC, Wierzbicki AS Smell testing: an additional tool for identification of adult Refsum's disease *J Neurol Neurosurg Psychiatry* 2004;**75**:1334-1336
14. Doty, Agrawal On the Shelf Life of University of Pennsylvania Smell Identification Test (UPSIT) *Laryngoscope.* 1989;**99**:402-405
15. Silveira-Moriyama L, Mathias C, Mason L, Best C, Quinn NP, Lees AJ. Hyposmia in pure autonomic failure. *Neurology.* 2009 May 12;**72(19)**:1677-81
16. Cardesin A *et al.* Barcelona Smell Test - 24 (BAST - 24): 2006 validation and smell characteristics in the healthy Spanish population.

