Diagnostic accuracy, item analysis and age effects of the UPSIT Spanish version in Parkinson's disease

A. Campabadal[1]; B. Segura[1]; H.C. Baggio[1]; A. Abos[1]; C. Uribe[1]; A.I. Garcia-Diaz[1]; M.J. Marti [2,3]; F. Valldeoriola[2,3]; Y. Compta[2,3]; N. Bargallo,[4]; C. Junque[1,2,5*].

1 Medical Psychology Unit, Department of Medicine. Institute of Neuroscience, University of Barcelona. Barcelona, Catalonia, Spain.

2 Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Hospital Clínic de Barcelona. Barcelona, Spain.

3 Movement Disorders Unit, Neurology Service, Hospital Clínic de Barcelona. Institute of Neuroscience, University of Barcelona, Barcelona, Catalonia, Spain.

4 Centre de Diagnòstic per la Imatge, Hospital Clínic, Barcelona, Catalonia, Spain.

5 Institute of Biomedical Research August Pi i Sunyer (IDIBAPS). Barcelona, Catalonia, Spain.

Running title: Effects of ageing on olfaction in PD

Abstract word count: 217

Manuscript word count: 3561

Supplementary data:

Supplementary Table 1. Demographic and clinical characteristics of PD patients and HC

Supplementary table 2. Olfactory measures in PD and HC.

Corresponding author: Prof. Carme Junque

Medical Psychology Unit, Department of Medicine. University of Barcelona

Casanova 143 (08036) Barcelona, Spain

Phone: (+34) 93 402 45 70 // Fax: (+34) 93 403 52 94 // E-mail: cjunque@ub.edu

Authors' contact information:

A Campabadal: anna.campabadal@ub.edu

B Segura: bsegura@ub.edu

HC Baggio: hbaggio@ub.edu

A Abos: alexandraabos@ub.edu

C Uribe: carme.uribe@ub.edu

AI Garcia-Diaz: ai.garciadiaz@ub.edu

MJ Marti: mjmarti@clinic.cat

F Valldeoriola: fvallde@clinic.cat

Y Compta: YCOMPTA@clinic.cat

N Bargallo: bargallo@clinic.cat

A Campabadal: anna.campabadal@ub.edu

B Segura: bsegura@ub.edu

HC Baggio: hbaggio@ub.edu

Diagnostic accuracy, item analysis and age effects of the UPSIT Spanish version in Parkinson's disease

Running title: Effects of ageing and sex on olfaction in PD

Abstract word count: 217

Manuscript word count: 3561

Supplementary data:

Supplementary Table 1. Demographic and clinical characteristics of PD patients and HC

Supplementary table 2. Olfactory measures in PD and HC.

Abstract

Objective. The University of Pennsylvania Smell Identification Test (UPSIT) is the most commonly used test to detect olfactory impairment in Parkinson's disease (PD), but the cut-off score for clinical purposes is often difficult to establish because of age and sex effects. The current work aims to study the sensitivity and specificity of the UPSIT Spanish version and its accuracy in discriminating PD patients at different age groups from healthy controls (HC), and to perform an item analysis.

Method. Ninety-seven non-demented PD patients and 65 HC were assessed with the UPSIT Spanish version. Sensitivity, specificity, and diagnostic accuracy for PD were calculated. Multiple regression analysis was used to define predictors of UPSIT scores.

Results. Using the normative cut-off score for anosmia (≤18), the UPSIT showed a sensitivity of 54.6% with a specificity of 100.0% for PD. We found that, using the UPSIT cut-off score of ≤25, sensitivity was 81.4% and specificity 84.6% (area under the receiver operating characteristic curve=.908). Diagnosis and age were good predictors of UPSIT scores (B=-10.948; P<.001; B=-.203; P<.001). When optimal cut-off scores were considered according to age ranges (≤60; 61-70; ≥71), sensitivity and specificity values were >80.0% for all age groups.

Conclusions. In the Spanish UPSIT version, sensitivity and specificity are improved when specific cut-off scores for different age groups are computed.

KEYWORDS: Parkinson's disease; Elderly/Geriatrics/Aging; Olfaction; Assessment; Movement disorders.

Introduction

Olfactory impairment has been found to be a prodromal manifestation in Parkinson's disease (PD) (Berendse, Roos, Raijmakers, & Doty, 2011; Doty, 2012) and other neurodegenerative disorders (Doty, 2012). The prevalence of olfactory deficits in PD is estimated to be about 90% (Fullard et al., 2016; Haehner et al., 2009). Olfactory dysfunction substantially affects wellbeing and quality of life, reducing enjoyment from food, beverages, personal care products, and the natural environment. Moreover, decreased smell function increases the risk of danger from fire, environmental toxins, leaking natural gas, and spoiled food (Doty, 2017).

Odour identification tests, such as the University of Pennsylvania Smell Identification Test (UPSIT) (Doty, 1995), are used in both research and clinical settings to assess olfactory dysfunction in PD patients. The UPSIT has demonstrated high capability in differentiating PD patients from healthy subjects (Bohnen, Studenski, Constantine, & Moore, 2008), and has been used as a tool for the diagnosis of PD (Doty, Bromley, & Stern, 1995; Silveira-Moriyama et al., 2008).

It is well known that aging per se is associated with olfactory decline; McKinnon et al. (2010), using data collected from 732 subjects, estimated that the decrease in UPSIT scores is around 3.2 points for every 10 years of age (McKinnon et al., 2010). Loss of olfactory identification ability is seen from the 5th decade onwards but is most remarkable between the 7th and 8th decades of life. In PD patients, it has also been found that UPSIT scores decreased as a function of age (Doty et al., 1995).

Olfactory functions in normal subjects are clearly influenced by gender (Doty, Ford, Preti, & Huggins, 1975; Doty, Orndorff, Leyden, & Kligman, 1978; Doty, Green, Ram, & Yankell, 1982). Sex-related differences in the UPSIT have also been described in PD, with lower performance for men (Doty et al., 1995). These results suggest that age and sex should be taken into consideration when using olfactory measures as a tool for PD diagnosis. In this

setting, Doty et al. (Doty et al., 1995) proposed optimal cut-off scores that best discriminate between PD and healthy subjects, according to different age groups and according to gender. As far as we know, these results have not been validated in the UPSIT Spanish version.

Considering that performance in smell identification tests are subject to cultural variation, the UPSIT has been adapted to several countries (Ogihara, Kobayashi, Nishida, Kitano, & Takeuchi, 2011; Picillo et al., 2014; Rodriguez-Violante, Lees, Cervantes-Arriaga, Corona, & Silveira-Moriyama, 2011; Silveira-Moriyama et al., 2008, 2010; Yücepur et al., 2012). Even though changes in some specific items have improved sensitivity and specificity, it is hard to generalise the results to other populations (Picillo et al., 2014; Rodriguez-Violante, Gonzalez-Latapi, Camacho-Ordoñez, et al., 2014a; Silveira-Moriyama et al., 2008).

Furthermore, previous works found a decreased identification of some UPSIT items in healthy populations (Picillo et al., 2014; Silveira-Moriyama et al., 2008; Rodriguez-Violante, Gonzalez-Latapi, Camacho-Ordonez, et al., 2014b; Kondo, Matsuda, Hashiba, 1998). In this sense, the analysis of individual smell scores of UPSIT Spanish version may allow identifying items with greater accuracy in PD diagnosis, and items that are culturally unsuitable.

In the current work, we aim to investigate the sensitivity and specificity of the UPSIT Spanish version, its accuracy in discriminating PD patients from healthy controls (HC), age and sex related effects on test performance, and individual item reliability. It was hypothesised that: (1) PD patients will have poorer performance than HC; (2) the UPSIT will discriminate PD patients from healthy controls; (3) Some UPSIT items will show low percentage of success in control subjects; (4) Age will have an effect in UPSIT scores; (5) Sex will have an effect in UPSIT scores; (6) Diagnostic accuracy will improve when the effects of age and sex are considered.

Methods

Participants

The study sample included 97 non-demented PD patients and 65 HC who underwent olfactory evaluation.

PD patients were recruited from an outpatient movement disorders clinic (Parkinson's disease and Movement Disorders Unit, Department of Neurology, Hospital Clinic, Barcelona, Catalonia, Spain). HC consisted of individuals who had previously volunteered who take part in studies of aging at the "*Fundació Salut i Envelliment, Universitat Autònoma de Barcelona*" (Barcelona, Spain).

Exclusion criteria for both groups were: (1) significant psychiatric, neurological, or systemic comorbidity; (2) low average premorbid intelligence quotient estimated by the Vocabulary subtest of the Wechsler Adult Intelligence Scale, 3rd edition (scaled score ≤ 7); and (3) Mini Mental State Examination (MMSE) score < 25. Specific exclusion criteria for the olfaction test were: (1) history of nasal bone fracture, (2) diagnosis of rhinitis or nasal polyps, and (3) upper respiratory tract infections in the two weeks prior to testing.

The inclusion criteria for PD were: (1) the fulfilment of UK PD Society Brain Bank diagnostic criteria for PD (Hughes, Daniel, Kilford, & Lees, 1992); (2) no surgical treatment with deep brain stimulation. Exclusion criteria for PD were: (1) Hoehn and Yahr (H&Y) (Hoehn & Yahr, 1967) score > 3 (defined as mild to moderate bilateral disease with some postural instability but physically independent); (2) dementia according to Movement Disorder Society criteria (Dubois et al., 2007): (a) PD developed prior to the onset of dementia, (b) PD associated with a decreased Global Cognitive Efficiency, (c) cognitive impairment in more than one cognitive domain, (d) impact on daily living resulting from cognitive deficits, over and above those imposed by motor and autonomic problems (Dubois et al., 2007).

All PD patients were taking antiparkinsonian medication consisting of different combinations of L-DOPA, COMT inhibitors, MAO inhibitors, dopamine agonists, and amantadine. In order to standardise doses, levodopa equivalent daily dose (LEDD) was calculated as suggested by Tomlinson et al. (Tomlinson et al., 2010). All assessments were done in the on state. Motor disease severity was evaluated using H&Y staging and the Unified Parkinson's Disease Rating Scale motor section (UPDRS-III).

The study was approved by the Ethics Committee of the University of Barcelona (IRB00003099). All subjects provided written informed consent to participate after full explanation of the procedures involved.

Olfactory and clinical assessment

Odour identification was assessed using the Spanish version of the University of Pennsylvania Smell Identification Test (UPSIT-40) (Doty, 1995).

The UPSIT is a standardised multiple-choice scratch-and-sniff test consisting of four test booklets with 10 items each. In accordance with normative instructions, subjects scratch the impregnated area and are asked to select one of four possible answers for each item. Following normative data presented in the UPSIT manual (Doty et al., 1995), which includes adjustment for age and sex, scores greater than 33 in males and 34 in females are considered to reflect normosmia, and scores lower or equal to 18 reflect anosmia. Microsmia ranges from 33 or 34 (depending on the subject's gender) to 19.

In order to achieve higher cultural familiarity, the Smell Identification Test™ (UPSIT) Spanish version commercialised by Sensonics (https://sensonics.com) substitutes some items with respect to the original UPSIT. Specifically, in item 13, *talcum powder* replaces *liquorice*; in item 20, *apple* replaces *gingerbread*; in item 24, *rubber tire* replaces *root beer*; in item 27, *raspberry* replaces *lime*; and in item 29, *walnut* replaces *wintergreen*.

6

The Beck Depression Inventory II (Beck, Steer, 1996), Starkstein's Apathy Scale (Starkstein et al., 1992), and the Neuropsychiatric Inventory (Cummings et al., 1994) were administered to all subjects to explore the presence of psychiatric symptoms.

Demographic, clinical, and olfactory data analysis

Statistical analysis of demographic, clinical, and olfactory data were performed using IBM SPSS Statistics 24.0.0 (2016; Armonk, NY: IBM Corp). Group differences in demographic and clinical variables between HC and PD patients were analysed with independent Student's t-tests for quantitative measures or with Pearson's chi-squared test for categorical measures. Bonferroni correction was used to correct for multiple testing. A multiple regression analysis with stepwise method was used to define possible UPSIT score predictors in the whole sample and in the PD group. Predictors were chosen based on previous literature (Deeb et al., 2010; Doty et al., 1978; Doty, 2012; Frye, Schwartz, & Doty, 1990; Hubert, Fabsitz, Feinleib, & Brown, 1980; Joyner, 1964; McKinnon et al., 2010) and correlation analysis. Age, PD diagnosis, sex, years of education, and smoking status were introduced as possible predictors for the whole sample regression. For the PD subsample, PD diagnosis was excluded, and clinical measures such as H&Y, UPDRS-III, and LEDD were added.

UPSIT item analysis

Sensitivity, specificity, diagnostic accuracy, positive predictive value, and negative predictive value for PD were calculated for the UPSIT-40 total score, and for each item. Pearson's chi-squared test was used to compare the percentage of HC and PD patients who correctly identified each item of the 40-items UPSIT version.

Incorrect answers were considered as positive findings (*i.e.,* altered performance), whereas correct answers defined negatives. True positives (TP) were defined as the number of PD patients who got a wrong response in a certain item; true negatives (TN) as the number of HC who got a correct response; false positives (FP) were calculated as the number of HC who got a wrong response and, finally, false negatives (FN) were described as the number

7

of PD patients who had a correct answer. The following diagnostic measures were calculated: (1) Sensitivity, defined as the probability of testing positive when PD is present: TP/(TP+FN); (2) Specificity, defined as the probability of testing negative when PD is absent: TN/(FP+TN); (3) Positive predictive value (PPV), calculated as the percentage of subjects with a positive test who actually have PD: TP/(TP+FP); (4) negative predictive value (NPV), defined as the percentage of subjects with a negative test who do not have PD: TN/(TN+FN); finally, (5) overall diagnostic accuracy was defined as the proportion of correctly classified subjects among all subjects: (TP+TN)/(TP+TN+FP+FN).

Results

Demographic and clinical characteristics.

There were no differences between groups in sex (X2=3.447; P=.063), age (t=-1.2798; P=.437), years of education (t=.639; P=.429), or estimated IQ scores (t=.041; P=.977). Regarding psychiatric symptoms, there were significant differences between groups in Beck Depression Inventory II (t=-4.1329; P=.001), Starkstein's Apathy Scale (t=-4.5463; P=<.001), and in Cummings' Neuropsychiatric Inventory scores (t=-6.6537; P<.001). **Supplementary Table 1.**

Olfactory performance

PD patients correctly identified 18.6±6.8 UPSIT items and HC 29.8±4.2 (t=11.961; p<.001). Considering the normative cut-off score of ≤18, 54.6% of PD patients and no HC were identified as anosmic. **Supplementary Table 2.**

UPSIT diagnostic measures

For the normative cut-off score of ≤18 to define anosmia, UPSIT total score had a sensitivity of 54.6% with a specificity of 100.0%, a PPV of 100.0%, a NPV of 59.6%, and a diagnostic accuracy of 72.8% for PD. The receiver operating characteristic (ROC) curve showed that the optimal cut-off score for UPSIT-40 was ≤25 with a sensitivity of 81.4%, a specificity of

84.6 %, a PPV of 88.8%, a NPV of 75.3% and a diagnostic accuracy of 82.7% (area under the ROC curve = .908). **Table 1.**

When considering the whole sample, UPSIT total score correlated negatively with age (r=-.300; p<.001). In the HC group, significant correlations were found between UPSIT total score and age (r=-.406; p=.001) and years of education (r=.296; p=.017). In the PD patient group, a negative correlation was found between UPSIT total score and age (r=-.366; p<.001). Sex did not achieve significant correlation with UPSIT total score, either in HC (r=.221; p=.076) or in PD (r=-.016; P=.875). Diagnosis (B=-10.948; t=-11.333; P<.001) and age (B=-.203; t=-4.370; P<.001) were found to be the best predictors of UPSIT scores when the whole sample was considered. For the PD patient sample, clinical measures were included in the regression analysis, and only age was identified as a significant UPSIT score predictor (B=-.250; t= -3.545; P<.001). According to these data, we plotted ROC curves for three age groups (≤60, 61-70, and ≥71 years). **Table 2** shows cut-off scores with higher sensitivity and specificity for each age group. Sensitivity and specificity for PD diagnosis considering age groups for each UPSIT cut-off score are represented in **Figure 1**. In addition, we plotted ROC curves considering the same three groups divided by gender. **Table 3**.

UPSIT item analysis

Diagnostic measures and group differences for each item are described in **Table 4** and **Figure 2.**

Items that did not show significant differences between groups in the chi-squared test after Bonferroni correction were: *Bubble gum* (X2=9.727; P=.002), *cherry* (X2=.063; P=.802), *fruit juice* (X2=8.645; P=.003), *cheese* (X2=3.580; P=.058), *cedar* (X2=1.849; P=.174), *peach* (X2=7.183; P=.007), *pickle* (X2=.404; P=.507), *orange* (X2=3.836; P=.050), *walnut* (X2=.059; P=.808), *lemon* (X2=1.652; P=.199), and *soap* (X2=9.556; P=.002).

Items with less than 50.0% correct answers in HC were: *cedar* (44.6%), *lemon* (40.0%), *cherry* (33.8%), *pickle* (26.2%), and *walnut* (23.1%). In particular, according to age

grouping, items with an error rate above 50.0% for HC under 60 years old were *dill pickle* (33.3%), *walnut* (20.8%), and *lemon* (37.5%); for the middle-age group, *pizza* (38.1%), *dill pickle* (23.8%), *walnut* (9.5%), and *cherry* (28.6%); lastly, for HC above 71, *pizza* (40.0%), *cedar* (30.0%), *dill pickle* (20.0%), *orange* (40.0%), *walnut* (40.0%), *lemon* (35.0%), and *cherry* (10.0%).

Items with sensitivity above 75.0% in the PD patient group were *walnut* (75.3%), *pickle* (78.4%), and *pizza* (78.4%). Sensitivity below 25.0% was seen for *onion* (23.7%). Items with a specificity above 90.0% were *chocolate* (93.8%), *coconut* (93.8%), *lilac* (95.4%), *onion* (96.9%), *talcum powder* (98.5%), and *smoke* (98.5%). *Walnut* (23.1%) had a specificity below 25.0%. Items with a diagnostic accuracy above 70.0% were *paint thinner* (71.6%), *clove* (72.3%), and *mint* (73.3%), and those below 55.0% were *onion* (53.1%), *cherry* (54.3%), and *walnut* (54.3%).

Discussion

Since olfactory deficits have been found to be a relevant clinical feature in PD (Berendse, Roos, Raijmakers, & Doty, 2011; Doty, 2012), there has been growing interest in incorporating odour tests in routine protocols both in patient assessment and in data collection. Our results show age and sex should be taken into consideration when using UPSIT Spanish version as a tool for identifying olfactory deficits associated with PD.

In the current study, the normative cut-off score of anosmia proposed by the UPSIT manual (≤18) demonstrated high specificity (100.0%) and low sensitivity (54.6%). In our sample, the optimal cut-off score was identified as ≤25, with a diagnostic accuracy of 82.7%, a sensitivity of 81.4%, and a specificity of 84.6%. This observed sensitivity is similar to that found in a Mexican population using the same cut-off score (Rodriguez-Violante, Gonzalez-Latapi, Camacho-Ordonez, et al., 2014a, 2014b). In an Italian study, the cut-off score of ≤21 differentiated PD and HC with similar sensitivity (82.0%) and specificity (88.2%) (Picillo et al., 2014). Given that cross-cultural aspects may influence these results (Rodriguez-

Violante, Gonzalez-Latapi, Camacho-Ordoñez, et al., 2014a), it could be of interest to perform a more detailed analysis of the UPSIT-40 items to identify those with less capability to detect olfactory deficits in a given population due to cultural biases.

We performed an item analysis similar to previous works (Kondo, Matsuda, Hashiba, 1998; Picillo et al., 2014; Rodriguez-Violante, Gonzalez-Latapi, Camacho-Ordonez, et al., 2014b; Silveira-Moriyama et al., 2008). Our results showed that some items had a rate of correct answers of less than 50.0% in HC (*cedar, lemon, cherry, dill pickle and walnut*), specifically, the three age groups coincided on *dill pickle* and *walnut*. Such items showed no significant differences in the percentage of success between HC and PD patients. As has been proposed in the literature (Picillo et al., 2014), a low percentage of success in control subjects may be an indicator of culturally unsuitable items. A decreased ability to identify of some odours has been reported as well in Italian (Picillo et al., 2014), Brazilian (Silveira-Moriyama et al., 2008), Mexican (Rodriguez-Violante, Gonzalez-Latapi, Camacho-Ordonez, et al., 2014b), and Japanese (Kondo, Matsuda, Hashiba, 1998) control subjects. Out of the five items with low identification rates by HC in our study, three items (*cherry, cedar, lemon*) also showed a low percentage of correct identification in a previous study obtained from an Italian population (Picillo et al., 2014). Rodríguez-Violante et al. also found that *cherry*, *walnut*, and *dill pickle* were identified by less than 50.0% of subjects in a Mexican population (Rodriguez-Violante, Gonzalez-Latapi, Camacho-Ordonez, et al., 2014b). In Brazil, Silveira-Moriyama et al. found an identification rate of less than 50.0% in *cedar*, *dill pickle*, and *lemon* as well (Silveira-Moriyama et al., 2008). In this sense, a large percentage (86.0%) of HC fell in the microsmic range of performance, with only a small number (13.8%) being classified as normosmic. This increased rate of microsmia may reflect, among other reasons, the impact of cultural differences on UPSIT performance. These findings suggest a need to provide and to promote the use of country-specific normative data.

Our results show that UPSIT performance was explained mainly by group, but also by age. Contrary to previous studies (Campabadal et al., 2017; Sharma & Turton, 2012; Silveira-

Moriyama et al., 2008), clinical measures such as motor scales and disease duration did not significantly predict smell identification performance in our PD subsample. One explanation could be that olfactory loss is an early symptom that progresses mainly in prodromal stages and may reach maximum severity levels even earlier than motor symptoms, which progress through advanced stages (Doty, Deems, & Stellar, 1988). Further studies in de novo samples could help identify a more accurate cut-off score for PD patients in initial stages, which could eventually have a positive impact on their quality of life.

Age contribution to olfactory deficits is well known both in normal subjects (McKinnon et al., 2010; Sorokowska et al., 2015) and in PD patients (Doty et al., 1995). This is demonstrated by the fact that, in the original English version, discriminatory UPSIT scores are presented for different age groups (Doty et al., 1995). Even though changes in some specific items have been made, these age-related discrimination criteria have not been reviewed for the Spanish version.

Age-group analysis showed that the UPSIT-40 cut-off score of ≤25 demonstrated higher sensitivity for all age ranges than the normative cut-off score for anosmia. When cut-off scores for each age group were considered, sensitivity and specificity values were above 80.0% for all age groups. These values were similar when groups were classified by age and gender. In particular, our results show cut-off scores for the youngest group (≤60 years) were 29 for males and 30 for females; for the intermediate age group (61-70 years), they were 22 for males and 26 for females; and for the oldest group (≥71 years) they were 24 and 27, respectively. Cut-off scores for all groups were similar but below those proposed by Doty et al. (Doty et al., 1995). With a sample of 180 PD patients and 612 HC, Doty et al. (1995) found that the sensitivity and specificity of the UPSIT in distinguishing between male PD patients and HC under the age of 61 years was 91.0% and 88.0%, respectively (cut-off score 31). For females of the same age, the corresponding values were 79.0% and 85.0% (cut-off score 33). Sensitivity and specificity of the UPSIT-40 for subjects between 61 and 70 years of age were 81.0% and 82.0% for males (cut-off score 25) and 80.0% and 88.0%

for females, respectively (cut-off score 30). Finally, for subjects older than 71, sensitivity and specificity for males were 76.0% and 78.0% (cut-off score 22), and 78.0% and 82.0% for females (cut-off score 25) (Doty et al., 1995). Considering that the onset of motor features in PD takes place around the age of 60, there is a need for developing more accurate procedures for detecting olfactory impairment in young PD patients and in the prodromal stages.

Screening for potential early indicators of PD such as olfaction, constipation or sleep disturbances facilitates the identification of individuals at high risk for PD who could participate in medical trials addressed to prevent or slow disease progression (Ross et al., 2008). Therefore, although the UPSIT has high capability in differentiating PD patients from healthy controls, it should be used within the context of a comprehensive clinical assessment.

In conclusion, our results demonstrate that the sensitivity and specificity of the UPSIT Spanish version were improved when specific cut-off scores for different age and sex groups were considered.

Limitations of this study include the fact that most of our PD patients are in mild to moderate stages of the disease. Since the PD patients of our study were already identified as having PD, our findings cannot address the issue of early diagnosis of PD. Further studies examining early clinical or prodromal stages of PD are needed to address this question. In this study we have excluded patients with low IQ and dementia; thus, our findings are not generalizable to the whole PD population. Finally, our data is preliminary, and the proposed cut-offs need to be replicated before being used clinically.

Funding

Disclosure

References

Beck A, Steer R, B. G. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, Texas: Psychological Corporation.

Berendse, H. W., Roos, D. S., Raijmakers, P., & Doty, R. L. (2011). Motor and non-motor correlates of olfactory dysfunction in Parkinson's disease. *Journal of the Neurological Sciences, 310*(1–2), 21–24. https://doi.org/10.1016/j.jns.2011.06.020

Bohnen, N. I., Studenski, S. A., Constantine, G. M., & Moore, R. Y. (2008). Diagnostic performance of clinical motor and non-motor tests of Parkinson disease : a matched

case –control study. *European Journal of Neurology*, 685–691.

https://doi.org/10.1111/j.1468-1331.2008.02148.x

Campabadal, A., Uribe, C., Segura, B., Baggio, H. C., Abos, A., Garcia-Diaz, A. I., … Junque, C.

(2017). Brain correlates of progressive olfactory loss in Parkinson's disease.

Parkinsonism and Related Disorders, 1–7.

https://doi.org/10.1016/j.parkreldis.2017.05.005

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J.

(1994). The Neuropsychiatric Inventory: comprehensive assessment of

psychopathology in dementia. *Neurology*, *44*(12), 2308–2314.

Deeb, J., Shah, M., Muhammed, N., Gunasekera, R., Gannon, K., Findley, L. J., & Hawkes, C. H.

(2010). A basic smell test is as sensitive as a dopamine transporter scan: Comparison of

olfaction, taste and DaTSCAN in the diagnosis of Parkinson's disease. *Qjm*, *103*(12),

941–952. https://doi.org/10.1093/qjmed/hcq142

Doty, R. L. (1995). *The smell identification test administration manual* (3rd ed.). Haddon

Heights: Sensonics.

Doty, R. L. (2012). Olfaction in Parkinson's disease and related disorders. *Neurobiol Dis.,*

46(3), 527–552. https://doi.org/10.1016/j.nbd.2011.10.026.Olfaction

Doty, R. L. (2012). Olfactory dysfunction in Parkinson disease. *Nature Publishing Group*,

8(6), 329–339. https://doi.org/10.1038/nrneurol.2012.80

Doty, R. L. (2017). Olfactory dysfunction in neurodegenerative diseases: is there a common

pathological substrate? *The Lancet Neurology*, *16*(6), 478–488.

https://doi.org/10.1016/S1474-4422(17)30123-0

Doty, R. L., Bromley, S. M., & Stern, M. B. (1995). Olfactory testing as an aid in the diagnosis

of Parkinson's disease: development of optimal discrimination criteria.

Neurodegeneration : A Journal for Neurodegenerative Disorders, Neuroprotection, and

Neuroregeneration, 4(1), 93–7. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/7600189

Doty, R. L., Deems, D. A., & Stellar, S. (1988). Olfactory dysfunction in Parkinsonism: a

general deficit unrelated to neurologic signs, disease stage, or disease duration.

Neurology, *38*(8), 1237–44. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/3399075

Doty, R. L., Ford, M., Preti, G., & Huggins, G. R. (1975). Changes in the intensity and

pleasantness of human vaginal odors during the menstrual cycle. *Science (New York,*

N.Y.), 190(4221), 1316–8. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/1239080

Doty, R. L., Green, P. A., Ram, C., & Yankell, S. L. (1982). Communication of gender from

human breath odors: Relationship to perceived intensity and pleasantness. *Hormones*

and Behavior, 16(1), 13–22. https://doi.org/10.1016/0018-506X(82)90002-2

Doty, R. L., Orndorff, M. M., Leyden, J., & Kligman, A. (1978). Communication of gender

from human axillary odors: relationship to perceived intensity and hedonicity.

Behavioral Biology, 23(3), 373–80. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/697690

Dubois, B., Burn, D., Goetz, C., Aarsland, D., Brown, R. G., Broe, G. A., … Poewe, W. (2007).

Diagnostic Procedures for Parkinson' s Disease Dementia : Recommendations from the

Movement Disorder Society Task Force*, 22*(16), 2314–2324.

https://doi.org/10.1002/mds.21844

Frye, R. E., Schwartz, B. S., & Doty, R. L. (1990). Dose-related effects of cigarette smoking

on olfactory function*. JAMA, 263*(9), 1233–6. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/2304239

Fullard, M. E., Tran, B., Xie, S. X., Toledo, J. B., Scordia, C., Linder, C., … Morley, J. F. (2016). Olfactory impairment predicts cognitive decline in early Parkinson's disease. *Parkinsonism and Related Disorders, 25*, 45–51. https://doi.org/10.1016/j.parkreldis.2016.02.013

Haehner, A., Boesveldt, S., Berendse, H. W., Mackay-Sim, A., Fleischmann, J., Silburn, P. A., … Hummel, T. (2009). Prevalence of smell loss in Parkinson's disease - A multicenter study. *Parkinsonism and Related Disorders, 15*(7), 490–494. https://doi.org/10.1016/j.parkreldis.2008.12.005

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: onset, progression and mortality. *Neurology, 17*(5), 427–442.

Hubert, H. B., Fabsitz, R. R., Feinleib, M., & Brown, K. S. (1980). Olfactory sensitivity in humans: genetic versus environmental control. *Science (New York, N.Y.), 208*(4444), 607–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7189296

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry, 55*(3), 181–4. https://doi.org/10.1136/jnnp.55.3.181

Joyner, R. E. (1964). Effect of cigarette smoking on olfactory acuity. *Archives of Otolaryngology (Chicago, Ill. : 1960)*, 80, 576–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14216019

Kondo H, Matsuda T, Hashiba M, B. S. (1998). A study of the relationship between the T&T olfactometer and the University of Pennsylvania Smell Identification Test in a Japanese population. *American Journal of Rhinology & Allergy, 2*(5), 353–8.

McKinnon, J., Evidente, V., Driver-Dunckley, E., Premkumar, A., Hentz, J., Shill, H., … Adler, C. (2010). Olfaction in the elderly: a cross-sectional analysis comparing Parkinson's

disease with controls and other disorders. *The International Journal of Neuroscience,*
120(1), 36–9. https://doi.org/10.3109/00207450903428954

Ogihara, H., Kobayashi, M., Nishida, K., Kitano, M., & Takeuchi, K. (2011). Applicability of
the cross-culturally modified University of Pennsylvania Smell Identification Test in a
Japanese population. *American Journal of Rhinology & Allergy, 25*(6), 404–10.
https://doi.org/10.2500/ajra.2011.25.3658

Picillo, M., Pellecchia, M. T., Erro, R., Amboni, M., Vitale, C., Iavarone, A., … Barone, P.
(2014). The use of University of Pennsylvania Smell Identification Test in the diagnosis
of Parkinson's disease in Italy. *Neurological Sciences : Official Journal of the Italian*
Neurological Society and of the Italian Society of Clinical Neurophysiology, 35(3), 379–
83. https://doi.org/10.1007/s10072-013-1522-6

Rodriguez-Violante, M., Gonzalez-Latapi, P., Camacho-Ordoñez, A., Martínez-Ramírez, D.,
Morales-Briceño, H., & Cervantes-Arriaga, A. (2014a). Comparing the accuracy of
different smell identification tests in Parkinson's disease: Relevance of cultural aspects.
Clinical Neurology and Neurosurgery, 123, 9–14.
https://doi.org/10.1016/j.clineuro.2014.04.030

Rodriguez-Violante, M., Gonzalez-Latapi, P., Camacho-Ordonez, A., Martinez-Ramirez, D.,
Morales-Briceno, H., & Cervantes-Arriaga, A. (2014b). Low specificity and sensitivity of
smell identification testing for the diagnosis of Parkinson's disease. *Arquivos de Neuro-*
Psiquiatria, 72(1), 33–37. https://doi.org/10.1590/0004-282X20130190

Rodriguez-Violante, M., Lees, A. J., Cervantes-Arriaga, A., Corona, T., & Silveira-Moriyama,
L. (2011). Use of smell test identification in Parkinson's disease in Mexico: A matched
case-control study. *Movement Disorders, 26*(1), 173–176.
https://doi.org/10.1002/mds.23354

Ross, G. W., Petrovitch, H., Abbott, R. D., Tanner, C. M., Popper, J., Masaki, K., … White, L. R. (2008). Association of olfactory dysfunction with risk for future Parkinson's disease. *Annals of Neurology, 63*(2), 167–173. https://doi.org/10.1002/ana.21291

Sharma, J. C., & Turton, J. (2012). Olfaction, dyskinesia and profile of weight change in Parkinson's disease: Identifying neurodegenerative phenotypes. *Parkinsonism and Related Disorders, 18*(8), 964–970. https://doi.org/10.1016/j.parkreldis.2012.05.004

Silveira-Moriyama, L., Azevedo, A. M. ., Ranvaud, R., Barbosa, E. R., Doty, R. L., & Lees, A. J. (2010). Applying a new version of the Brazilian-Portuguese UPSIT smell test in Brazil. *Arquivos de Neuro-Psiquiatria, 68*(5), 700–705. https://doi.org/10.1590/S0004-282X2010000500005

Silveira-Moriyama, L., de Jesus Carvalho, M., Katzenschlager, R., Petrie, A., Ranvaud, R., Barbosa, E. R., & Lees, A. J. (2008). The use of smell identification tests in the diagnosis of Parkinson's disease in Brazil. *Movement Disorders, 23*(16), 2328–2334. https://doi.org/10.1002/mds.22241

Sorokowska, A., Schriever, V. A., Gudziol, V., Hummel, C., Hähner, A., Iannilli, E., … Hummel, T. (2015). Changes of olfactory abilities in relation to age: odor identification in more than 1400 people aged 4 to 80 years. *European Archives of Oto-Rhino-Laryngology, 272*(8), 1937–1944. https://doi.org/10.1007/s00405-014-3263-4

Starkstein, S. E., Mayberg, H. S., Preziosi, T. J., Andrezejewski, P., Leiguarda, R., & Robinson, R. G. (1992). Reliability, validity, and clinical correlates of apathy in Parkinson's disease. *The Journal of Neuropsychiatry and Clinical Neurosciences, 4*(2), 134–139. https://doi.org/10.1176/jnp.4.2.134

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders, 25*(15), 2649–2653. https://doi.org/10.1002/mds.23429

Yücepur, C., Ozücer, B., Değirmenci, N., Yıldırım, Y., Veyseller, B., & Ozturan, O. (2012). University of Pennsylvania smell identification test: application to Turkish population. *Kulak Burun Bogaz Ihtisas Dergisi : KBB = Journal of Ear, Nose, and Throat, 22*(2), 77–80. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22548263

Table 1. Diagnostic accuracy measures for UPSIT-40

UPSIT version	TP	TN	FP	FN	Diagnostic accuracy	Sensitivity	Specificity	PPV	NPV
UPSIT-40 (cut-off ≤18)	53	65	0	44	72.8	54.6	100.0	100.0	59.6
UPSIT-40 (cut-off ≤25)	79	55	10	18	82.7	81.4	84.6	88.8	75.3

FN: false negative; FP: false positive; HC: healthy controls; NPV: negative predictive value; PD: Parkinson's disease; PPV: positive predictive value; TN: true negative; TP: true positive; UPSIT: *University of Pennsylvania Smell Identification Test;* UPSIT-40 (cut-off ≤18): normative cut-off for anosmia; UPSIT-40 (cut-off ≤25): optimal cut-off for this study sample.

Table 2. Optimal cut-off scores and associated sensitivity and specificity for PD diagnosis for three age groups.

Age group	Cut-off score	Sensitivity (%)	Specificity (%)	AUROC
≤60 years	29	83.3	87.5	0.905
61-70 years	25	82.1	81.0	0.903
≥71 years	22	85.7	85.0	0.954

AUROC: area under the receiver operating characteristic curve; HC: healthy controls; PD: Parkinson's disease. Optimal cut-off scores and associated sensitivity and specificity were calculated through ROC curve analysis for three age groups: ≤60 years (PD=30; HC=24), 61-70 years (PD=39; HC=21), and ≥71 years (PD=28; HC=20).

Table 3. Optimal cut-off scores and associated sensitivity and specificity for PD diagnosis considering age and gender measures.

Age group	Sample size		Cut-off score		Sensitivity (%)		Specificity (%)		AUROC	
	M	F	M	F	M	F	M	F	M	F
Whole sample	32/62	33/35	24	27	80.6	88.6	81.2	87.9	.891	.922
≤60 years	9/25	15/5	29	30	84.0	80.0	77.8	80.0	.889	.847
61-70 years	12/23	9/16	22	26	82.6	87.5	91.7	88.9	.897	.917
≥71 years	11/14	9/14	24	27	100.0	92.9	81.8	77.8	.987	.929

Sample size for male and female subgroups is presented as, HC/PD.

Table 4. Item analysis of the UPSIT Spanish version in PD.

Item	Label	TP	TN	FP	FN	Diagnostic Accuracy	Sensitivity	Specificity	PPV	NPV	Chi-square/P value
1	Pizza	76	33	32	21	67.3	78.4	50.8	70.4	61.1	**14.85/<.001**
2	Bubble gum	35	56	9	62	56.2	36.1	86.2	79.5	47.5	9.73/.002
3	Menthol	56	54	11	41	67.9	57.7	83.1	83.6	56.8	**26.73/<.001**
4	Cherry	66	22	43	31	54.3	68.0	33.8	60.6	41.5	.06/.802
5	Motor oil	58	47	18	38	65.2	60.4	72.3	76.3	55.3	**16.65/<.001**
6	Mint	61	57	8	35	73.3	63.5	87.7	88.4	62.0	**41.54/<.001**
7	Banana	42	56	9	54	60.9	43.8	86.2	82.4	50.9	**16.01/<.001**
8	Clove	63	54	11	34	72.2	64.9	83.1	85.1	61.4	**36.18/<.001**
9	Leather	38	57	8	59	58.6	39.2	87.7	82.6	49.1	**13.82/<.001**
10	Coconut	41	61	4	56	63.0	42.3	93.8	91.1	52.1	**25.30/<.001**
11	Onion	23	63	2	74	53.1	23.7	96.9	92.0	46.0	**12.69/<.001**
12	Fruit juice	51	46	19	46	59.9	52.6	70.8	72.9	50.0	8.64/.003
13	Talcum powder	27	64	1	70	56.2	27.8	98.5	96.4	47.8	**18.82/<.001**
14	Cheese	55	38	27	42	57.4	56.7	58.5	67.1	47.5	3.58/.058
15	Cinnamon	50	56	8	47	65.8	51.5	87.5	86.2	54.4	**25.51/<.001**
16	Gasoline	55	57	8	42	69.1	56.7	87.7	87.3	57.6	**32.27/<.001**
17	Strawberry	57	46	19	40	63.6	58.8	70.8	75.0	53.5	**13.63/<.001**
18	Cedar	64	29	36	33	57.4	66.0	44.6	64.0	46.8	1.85/.174
19	Chocolate	31	61	4	66	56.8	32.0	93.8	88.6	48.0	**15.30/<.001**
20	Apple	51	53	12	46	64.2	52.6	81.5	81.0	53.5	**19.06/<.001**
21	Lilac	37	62	3	60	61.1	38.1	95.4	92.5	50.8	**23.53/<.001**
22	Turpentine	66	41	24	30	66.5	68.8	63.1	73.3	57.7	**15.93/<.001**
23	Peach	44	49	16	53	57.4	45.4	75.4	73.3	48.0	7.18/.007
24	Rubber tire	61	44	21	36	64.8	62.9	67.7	74.4	55.0	**14.56/<.001**
25	Pickle	76	17	48	21	57.4	78.4	26.2	61.3	44.7	.40/.507
26	Pineapple	51	56	9	46	66.0	52.6	86.2	85.0	54.9	**25.04/<.001**
27	Raspberry	32	58	7	65	55.6	33.0	89.2	82.1	47.2	**10.51/.001**
28	Orange	51	41	24	46	56.8	52.6	63.1	68.0	47.1	3.84/.050
29	Walnut	73	15	50	24	54.3	75.3	23.1	59.3	38.5	.06/.808
30	Watermelon	40	57	8	57	59.9	41.2	87.7	83.3	50.0	**15.62/<.001**
31	Paint thinner	58	58	7	39	71.6	59.8	89.2	89.2	59.8	**38.93/<.001**
32	Grass	55	55	10	42	67.9	56.7	84.6	84.6	56.7	**27.65/<.001**
33	Smoke	42	64	1	55	65.4	43.3	98.5	97.7	53.8	**34.81/<.001**
34	Pine	59	44	21	38	63.6	60.8	67.7	73.8	53.7	**12.66/<.001**
35	Grape	65	41	23	32	65.8	67.0	64.1	73.9	56.2	**15.02/<.001**
36	Lemon	67	26	39	29	57.8	69.8	40.0	63.2	47.3	1.65/.199
37	Soap	59	41	23	38	62.1	60.8	64.1	72.0	51.9	9.56/.002
38	Natural gas	43	57	8	54	61.7	44.3	87.7	84.3	51.4	**18.50/<.001**
39	Rose	60	46	19	37	65.4	61.9	70.8	75.9	55.4	**16.58/<.001**
40	Peanut	44	57	8	53	62.3	45.4	87.7	84.6	51.8	**19.51/<.001**

Items marked in blue showed no significant differences in the percentage of correct answers

between HC and PD patients. FN: false negative; FP: false positive; HC: healthy controls;

NPV: negative predictive value; PD: Parkinson's disease; PPV: positive predictive value; TN: true negative; TP: true positive.

Pearson's chi-squared tests were used to compare the percentage of HC and PD patients who correctly identified each item. P-values that survive Bonferroni correction for multiple comparisons (P≤.00125) are highlighted in bold.

Figure 1

UPSIT sensitivity and specificity for PD diagnosis considering age group.

UPSIT: *University of Pennsylvania Smell Identification Test;* UPSIT-40 (cut-off ≤18): normative cut-off; UPSIT-40 (cut-off ≤25): optimal cut-off for this study sample; UPSIT-40 (cut-off per age group): optimal cut-off score for each age group.
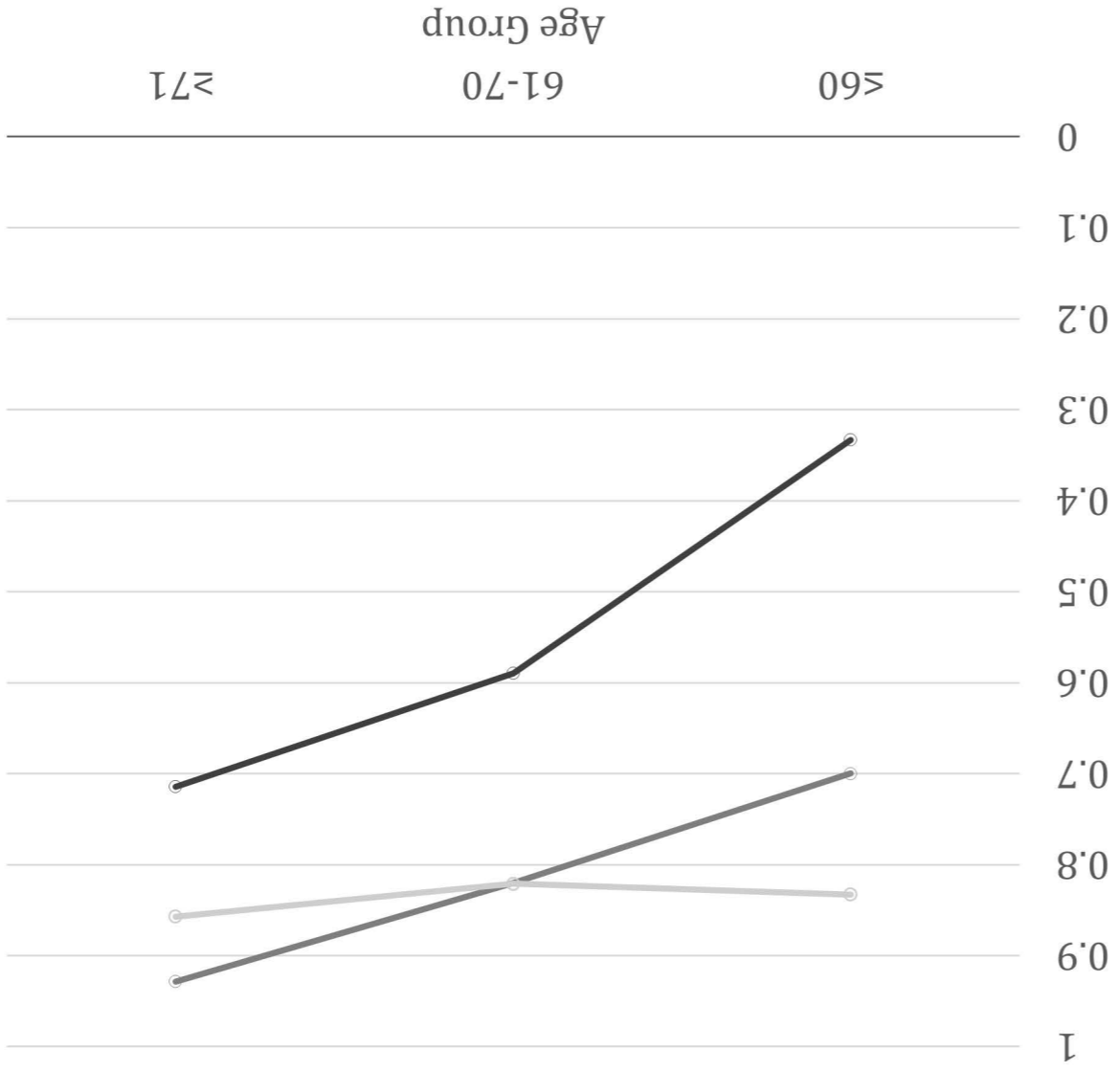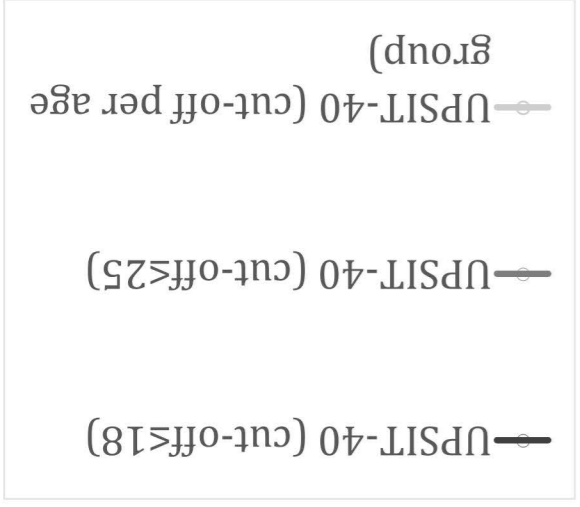
Age groups and corresponding group sizes: ≤60 years (PD=30; HC=24), 61-70 years (PD=39; HC=21), ≥71 years (PD=28; HC=20).
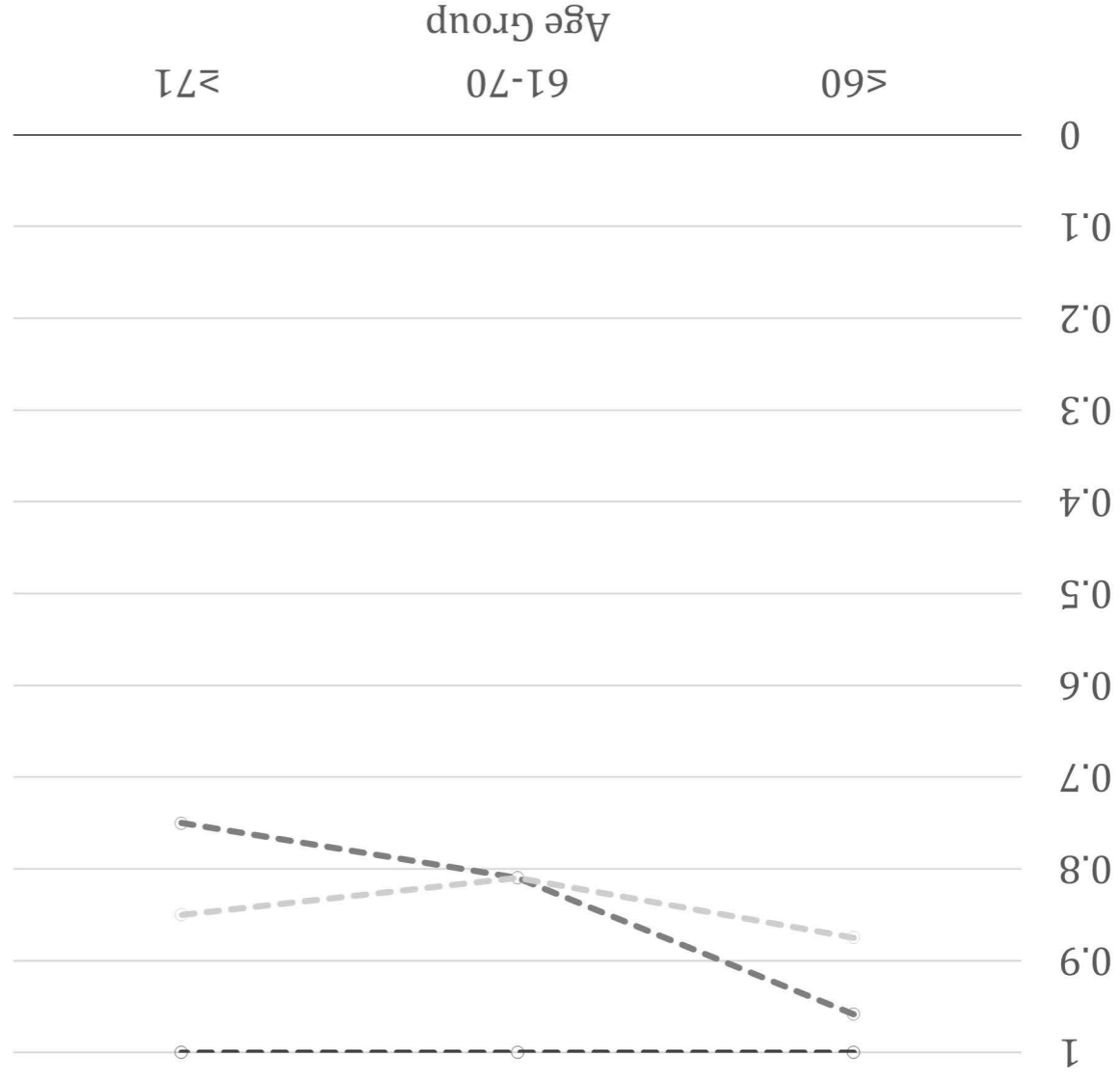
Figure 2

Percentage of subjects who correctly identified each item of the UPSIT.

* Items that did not show significant differences between groups in the chi-squared test after Bonferroni correction

HC: healthy controls; PD: Parkinson's disease.

Age Group

≥71 61-70 ≤60

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Sensitivity

UPSIT-40 (cut-off per age group)

UPSIT-40 (cut-off≤25)

UPSIT-40 (cut-off≤18)

Age Group

≥71 61-70 ≤60

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Specificity
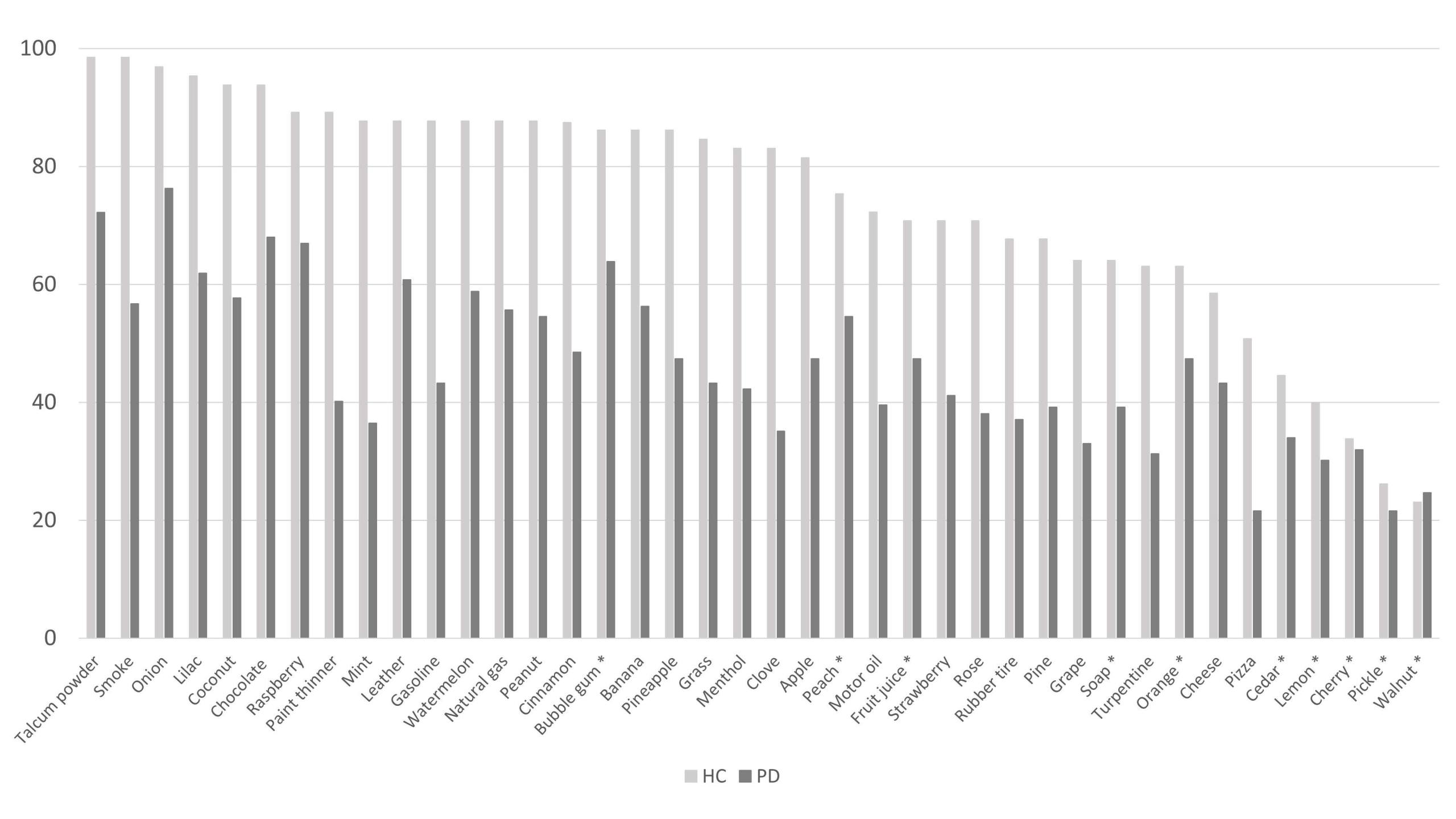
HC   PD

Supplementary Table 1. Demographic and clinical characteristics of PD patients and HC

	PD (n=97)	HC (n=65)	Test stats/P-value
Age, yrs., mean (SD)	65.6 (9.6)	64.2 (11.1)	1.28/.437[a]
Education, yrs., mean (SD)	11.4 (5.3)	12.1 (4.5)	.64/.429[a]
Vocabulary subtest, mean (SD)	45.2 (9.1)	45.2 (7.9)	.04/.977 [a]
Sex, male, n (%)	62 (63.9)	32 (49.2)	3.45/.063 [b]
Smoking status (nonsmoker/ex-smoker/smoker)	60/6/30	24/5/10	1.77/.413
Disease duration, yrs., mean (SD)	8.7 (5.9)	NA	NA
Age of onset, yrs., mean (SD)	56.7 (11.3)	NA	NA
UPDRS-III, mean (SD)	17.4 (9.5)	NA	NA
Hoehn & Yahr stage, n 1/1.5/2/2.5/3/4	8/2/51/5/24/0	NA	NA
LEDD, mg, mean (SD)	765.9 (452.2)	NA	NA

Abbreviations: HC: healthy controls; LEDD: L-DOPA equivalent daily dose; NA: not applicable; PD: Parkinson's disease; UPDRS part III: Unified Parkinson's Disease Rating Scale motor section. Data are presented as mean (SD), except for sex (number of cases and percentage) and Hoehn and Yahr staging (number of cases).

a. Independent Student's t-tests were used for continuous variables when comparing HC with all PD patients.
b. Pearson's chi-squared tests were used for categorical variables.

Supplementary table 2. Olfactory measures in PD and HC.

	PD (n=97)	HC (n=65)	Test stats/ P-value	Effect size
UPSIT-40	18.6 (6.8)	29.8 (4.2)	11.96/<.001 [a]	0.69
Olfactory status according to UPSIT-40 normative data, **n (%)**				
Normosmia	2 (2.1)	9 (13.8)		
Mild hyposmia	4 (4.1)	25 (38.5)		
Moderate hyposmia	12 (12.4)	22 (33.8)	80.97/<.001[b]	
Severe hyposmia	26 (26.8)	9 (13.8)		
Anosmia	53 (54.6)	0 (0)		

Abbreviations: HC, healthy controls; PD, Parkinson's disease; UPSIT-40, *University of Pennsylvania Smell Identification Test.* Data are presented as mean (SD) or n (%).

a. Independent t-tests were used for continuous variables when comparing HC with all PD patients.

b. Pearson's chi-squared tests were used for categorical variables.