The use of olfactory testing when diagnosing Parkinson's disease – a systematic review

Tine Nielsen[1], Martin Bang Jensen[1], Egon Stenager[1, 2] & Andreas Dammann Andersen[1, 2]

ABSTRACT

INTRODUCTION: The diagnosis of Parkinson's disease (PD) is typically based on the presence of motor symptoms, but in the early phase of the disease, the diagnostic process can be challenging. Examination of non-motor symptoms in patients suspected of PD has gained growing attention. Olfactory tests have shown promising results as ancillary diagnostic tests. The aim of this study was to investigate how olfactory tests may be used clinically in diagnostic process in PD.
METHODS: A systematic search was conducted in PubMed for relevant literature on 8 March 2017. A total of 358 articles were found. Our screening process left 27 articles, which were included for further analysis.
RESULTS: In all, 20 of the included studies analysed the diagnostic value of olfactory testing by comparing patients with PD to healthy controls. Sensitivities varied from 61% to 95% and specificities from 66% to 99%. Ten studies used olfactory tests to distinguish between PD and diseases that mimic PD. The sensitivities varied from 62% to 92% and the specificities from 65% and 96%.
CONCLUSIONS: Olfactory test can be a valuable ancillary tool in the diagnostic process in PD. In a clinical setting, the identification part from Sniffin' Sticks 16 is the most usable because it may be conducted quickly and independently of disease duration and severity. Before using an olfactory test in a clinical setting, it is necessary to adjust the odours to the patient population, and to establish the optimal specificity-adjusted cut-off.

Parkinson's disease (PD) is the second-most common neurodegenerative disease. It is characterised by tremor at rest, bradykinesia, rigidity and postural instability [1]. In addition to the motor symptoms, PD patients suffer from a wide range of non-motor symptoms, one of the most frequent of which is olfactory dysfunction. Studies have reported that olfactory dysfunction is present in approximately 90% of early-stage PD cases and that it may precede the onset of motor symptoms by years [2].

The diagnostic process of PD can be challenging, especially in the early phase of disease. Atypical Parkinsonian syndrome, including progressive supranuclear palsy (PSP), multiple system atrophy (MSA) and corticobasal degeneration (CBD), often mimic PD in the first years after onset [1]. Essential tremor (ET) can also

be a challenge to differentiate from PD in the early stages [3]. Studies suggest that olfactory function is mildly impaired in MSA, CBD and PSP [4-6] and is normal among patients with ET [7].

Several tests have been developed to examine the olfactory function. The psychophysical test is the easiest and fastest test. It examines the ability to identify, detect, discriminate and memorise odours [2].

In general, two psychophysical tests are widely used [2]. In the USA, the University of Pennsylvania Smell Identification Test (UPSIT) [8] is the most common, and in Europe the Sniffin' Sticks 16 (SS-16) [9] is the most common [10]. The UPSIT consists of 40 odorants at the supra threshold level. For each odorant, the patient is required to choose the correct answer among four alternatives. Several versions of the UPSIT have been developed, e.g. the Brief Smell Identification Test, which is a short version with only 12 items developed for cross-cultural use [11]. The SS-16 consists of three parts: threshold, discrimination and identification. In each part, the subject scores 0-16 points, and the final TDI score, which is the sum of all three scores, is calculated [9]. The identification part of SS-16, called Sniffin' Sticks odour Identification Test (SIT), is often used separately. The test consists of 16 pens filled with common odorants, and the patient is required to choose between four alternatives [10]. Many of the tests are modified and translated to the culture in which they are used as cultural differences may influence which odours people know [12-14], and for some cultures new tests have been developed, e.g. the Odour Stick Identification Test for Japanese (OSIT-J) [15].

Several studies have suggested that olfactory testing is a valuable tool in the process of diagnosing PD [16-


⚠ KEY POINTS

- Olfactory dysfunction is one of the most frequent non-motor symptoms in Parkinson's disease
- Olfactory testing can be a valuable ancillary tool in diagnosing Parkinson's disease
- Three out of five studies found a specificity > 80% in distinguishing Parkinson's disease from a pool of conditions that mimic Parkinson's disease
- The Sniffin' Stick identification test is the most usable olfactory test in a busy clinical setting.

Flow chart presenting the article selection process.
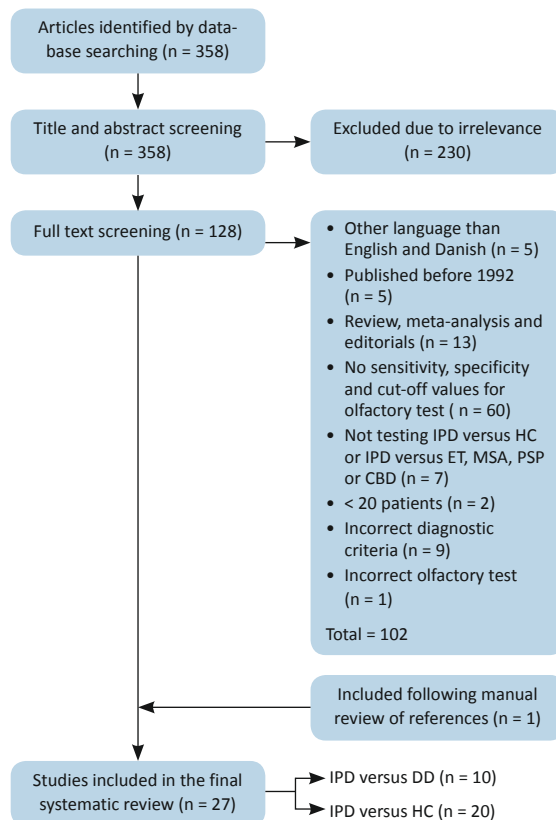
CBD = corticobasal degeneration; DD = differential diagnosis; ET = essential tremor; HC = healthy controls; IPD = idiopathic Parkinson's disease; MSA = multiple system atrophy; PSP = progressive supranuclear palsy.

18]. According to the new MDS criteria, a test can be used as an ancillary diagnostic tool if it has a specificity > 80% in differentiating PD from other parkinsonian conditions in most studies. Currently olfactory test and meta-iodobenzylguanidine scintigraphy meet these criteria [19].

The aim of this study was to investigate how olfactory tests can be used clinically in the process of diagnosing PD by focusing on the following four parameters:

1. The ability of the olfactory tests to distinguish between idiopathic Parkinson's disease (IPD) and healthy controls (HC)
2. The ability of the tests to distinguish between IPD and diseases that mimic IPD
3. How the tests could be used and in which settings
4. The quality of the individual tests.

METHODS

This article is based on a systematic review of the literature published before 8 March 2017. The search was conducted in PubMed on 8 March 2017 with the following search terms: (olfactory dysfunction OR olfactory deficiency OR hyposmia OR anosmia OR olfaction) AND (Parkinson* OR Parkinson's disease OR Parkinson disease OR neurodegeneration) AND (diagnostic test OR olfactory test OR odour test).

A total of 358 articles were found, and titles and abstracts were screened using the online screening tool Covidence. A total of 230 articles were excluded due to irrelevance. A full text screening was conducted on the remaining 128 articles. In order to be included in the final study, the articles had to fulfil the following eight criteria: I) English or Danish written study, II) Published after 1st of January 1992, III) Not a meta-analysis, systematic review or editorial, IV) Inform of sensitivity, specificity and cut-off for an olfactory test, V) Testing IPD vs HC or IPD vs ET, MSA, PSP or CBD, VI) > 20 patients in the PD group, VII) The PD patients must meet The UK Parkinson's Disease Society Brain Bank criteria [1], VIII). The olfactory test should either be SS-16, UPSIT, OSIT-J or a modified version of one of these. A total of 102 articles did not meet these criteria and were excluded.

The reference lists of the remaining 26 studies were analysed for potential inclusion of studies not found in the initial search; this led to addition of one extra article. Accordingly, 27 articles were included in the final study for further analysis (see **Figure 1**). MBJ and TN separately screened all articles. In case of disagreement about the potential inclusion of an article, the article in question was debated in plenum and a final decision was made.

RESULTS

Identification of Parkinson's disease from healthy controls

Twenty of the included studies analysed olfactory test by comparing IPD to HC, and predicted the diagnostic value. Sensitivities varied from 61% [20] to 95% [16], and specificities from 66% [21] to 96% [22]. Calculation of the mean, without weighting the studies, yielded a sensitivity of 83% and a specificity of 84%. **Table 1** summarizes the characteristics, sensitivity and specificity for each study, and **Figure 2** illustrates sensitivities and specificities.

Two studies used SS-16 to differentiate PD from HC. Krismer et al [16] found a sensitivity of 95% and a specificity of 95.1%. They found the area under the curve (AUC) for the receiver-operating characteristic curve for SS-16 to be 0.96, whereas the individual AUC for identifi-

cation, discrimination and threshold was 0.94, 0.87 and 0.84, respectively. Boesveldt et al [22] found a sensitivity of 81% and a specificity of 96%. They found AUC to be 0.94, identification 0.91, discrimination 0.83 and threshold 0.90. They found that identification combined with threshold significantly improved the AUC compared with identification alone.

The remaining studies used identification tests alone, seven used the UPSIT and nine used the SIT. Several studies gave multiple results at different cutoffs. Four studies used a shorter version of the SIT with 12 odours (SIT-12), and one used a shorter version of the UPSIT. Casjens et al [20] analysed each odour from the SIT, and found that peppermint, anise and coffee best differentiated PD from HC, and calculated sensitivity and specificity for these three odours. Bohnen et al [23] did the same based on the UPSIT, and found banana, liquorice and dill pickle to best differentiate PD from HC.

Gender was a significant independent predictor in six studies [6, 24-28] with increased olfactory score in females compared with males, whereas six studies found no significant difference [7, 20, 21, 29-31].

Progression of olfactory dysfunction

Fourteen studies found increased age to be a significant independent predictor of decreased olfactory score [4, 6, 20, 21, 24-29, 32-35], whereas three studies found that it was not significant [7, 30, 31].

Olfactory dysfunction as a marker of progression

Casjens et al [20] and Rodriguez-Violante et al [34] found Hoehn & Yahr stage to be a significant independent predictor of the olfactory score, and Rodriguez-Violante et al [34] also found disease duration to be a significant predictor. Five studies did not find disease duration and severity to be significant independent predictors [17, 26, 28, 32, 36], and four studies found that only disease duration was not a significant independent predictor [16, 18, 23, 31]. Boesveldt et al [28] found the ability to discriminate odours to be reduced when the disease duration increases, but identification remains the same.

Separating Parkinson's disease from atypical Parkinson's disease

Five studies analysed the ability of the tests to distinguish PD from a pool of several diagnoses that mimic PD [4, 16, 18, 29, 32] (see **Table 2**), and found sensitivities from 75% to 92% and specificities from 70% to 85%. Mahlknecht et al [18] applied a lower cut-off and found an increase in both sensitivity and specificity.

Several studies analysed the ability of the olfactory tests to distinguish PD from one specific type of atypical PD. Three studies analysed the ability to distinguish PD from MSA [16, 18, 37], and found sensitivities from 74% to 92% and specificities from 78% to 87%. Three studies analysed the ability to distinguish PD from PSP [6, 16, 18] and found sensitivities from 75% to 92% and specifi-

---

⟋ FIGURE 2

Overview of sensitivities (dark blue) and specificities (light blue) of the individual studies testing idiopathic Parkinson's disease against healthy controls. The studies are listed by olfactory test and decreasing specificity. The dotted line illustrates the 80% limit.
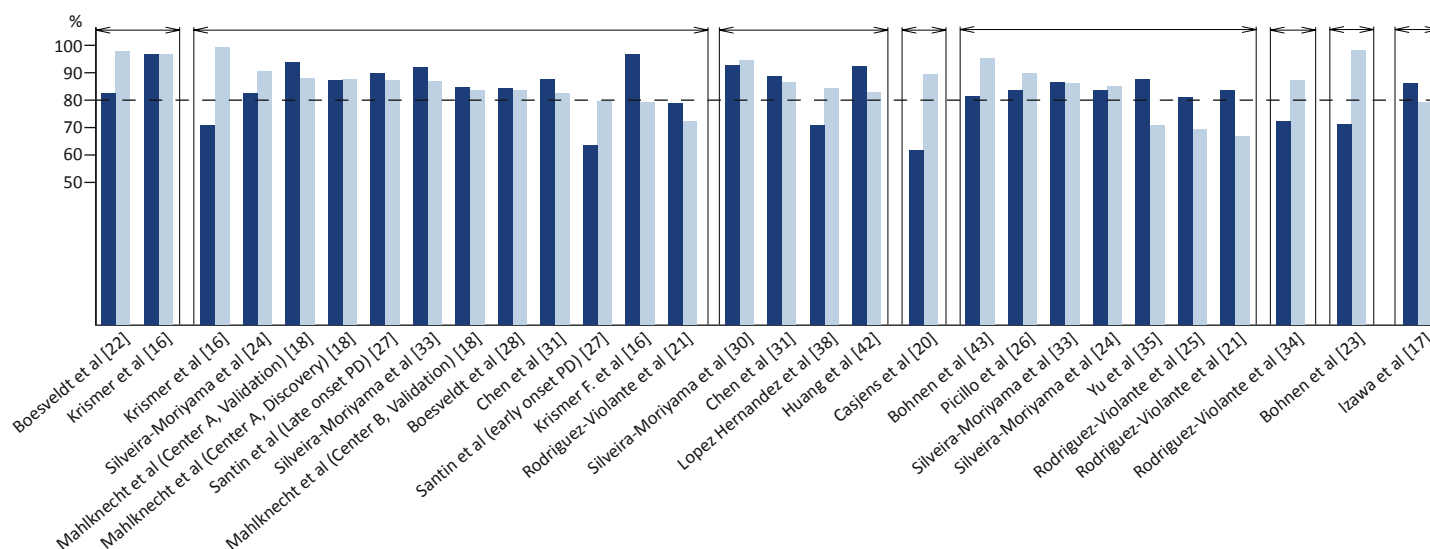
B-SIT = Brief Smell Identification Test; OSIT-J = Odour Stick Identification Test for Japanese; SI = Sniffin' Sticks Odour Identification Test; SS-16 = Sniffin' Sticks 16; UPSIT = University of Pennsylvania Smell Identification Test.

TABLE 1

Overview of the included studies testing the ability to distinguish Parkinson's disease from healthy controls. The studies are sorted by olfactory test, and the references are subdivided by decreasing specificity.

Olfactory test Reference	Specificity, %	Sensitivity, %	PD patients, n (♂/♀)	Controls, n (♂/♀)	H&Y-stage: 1 + 2, 3, ≥ 4, % (mean)	Disease duration, mean, yrs	Cut-off	Comments
SS-16								
Boesveldt et al, 2009, Netherlands [22]	96	81	52 (29/23)	50 (27/23)	92, 8, 0	6, 7	Z-score: −0.736	-
Krismer et al, 2017, Austria [16]	95.1	95	20 (13/7)	41 (23/18)	(2.3)	5, 9	23.875	Similar results were found by analysing patients with disease duration < 3 yrs
SIT								
Krismer et al, 2017, Austria [16]	97.6	70	20 (13/7)	41 (23/18)	(2.3)	5, 9	≤ 8	Cut-off value with optimal specificity
	78	95					≤ 11	Cut-off value with optimal sensitivity
Silveira-Moriyama et al, 2008, Brazil [24]	89	81.1	106 (71/35)	118 (103/15)	-	11, 4	≤ 9: < 60 yrs ≤ 8: ≥ 60 yrs	Subjects were tested by either the SIT, UP-SIT or both tests It was not stated how many subjects were tested by each test Age was a significant variable
Mahlknecht et al, 2016, Austria & Italy [18]	86.5	92.1	134 (84/50)	336 (156/180)	(2.4)	6, 2	≤ 10	Validation cohort, Centre A Similar results were found by analysing patients with disease duration < 3 yrs Followed for 24 mo.s to validate the diagnosis
	86.1	85.9	134 (84/50)	336 (156/180)	(2.4)	6, 2	≤ 10	Discovery cohort, Centre A
	82	83.3	400 (249/151)	150 (87/63)	(2.6)	11, 4	≤ 10	Validation cohort, Centre B
Santin et al, 2010, Brazil [27]	85.7	88.2	51 (29/22)	70 (38/32)	(2.2)	5, 9	≤ 9	Analysed late onset: > 45 yrs Age and gender were significant variables
	78.6	63.2	19 (9/10)			9, 7	≤ 10	Analysed early onset: < 45 yrs
Silveira-Moriyama et al, 2009, Brazil [33]	85.5	90.4	193 (115/78)	157 (93/64)	-	10, 2	≤ 10	The SIT tested on 115 PD and 83 HC of the study population
Boesveldt et al, 2008, Netherlands [28]	82	83	404 (253/151)	150 (87/63)	-	-	10.5	Age and gender were significant variables Disease duration and disease severity were not significant variables
Chen et al, 2012, China [31]	81	86	110 (66/44)	110 (66/44)	84, 16, 0	4, 3	9.5	Age and gender were not significant variables
Rodríguez-Violante et al, 2014, Mexico [21]	71.2	77.8	99 (64/35)	99 (64/35)	70, 25, 5	7, 3	≤ 9	Age was a significant variable Gender was not a significant variable
SIT-12								
Silveira-Moriyama et al, 2009, Sri Lanka [30]	93	91	89 (50/39)	100 (52/48)	-	6, 4	≤ 8	Used the 12 odours from the SIT for which > 50% of the HC had answered correctly Age and gender were not significant variables
Chen et al, 2012, China [31]	85	87	110 (66/44)	110 (66/44)	84, 16, 0	4, 3	7.5	Used the SIT and excluded the 4 odours without a significant difference between the score of PD and HC
López Hernández et al, Spain 2015 [38]	83	70	30 (22/8)	47 (26/21)	-	2, 1	< 8	-
Huang et al, 2016, China [42]	81.5	90.7	54 (43/11)	54 (43/11)	81, 19, 0	2, 7	7.5	Followed for ≥ 24 mo.s to validate the diagnosis
SIT-3								
Casjens et al, 2013, Germany [20]	88	61	148 (78/70)	148 (81/67)	-	-	≤ 1	Used the 3 odours with the highest sensitivity from the SIT Age, disease duration and severity were significant variables Gender was not a significant variable

Continues

TABLE 1 CONTINUED

Olfactory test Reference	Specificity, %	Sensitivity, %	PD patients, n (♂/♀)	Controls, n (♂/♀)	H&Y-stage: 1 + 2, 3, ≥ 4, % (mean)	Disease duration, mean, yrs	Cut-off	Comments
UPSIT								
Bohnen et al, 2008, USA [43]	93.3	80	45 (31/14)	45 (31/14)	98, 2, 0	3, 5	27	-
Picillo et al, 2014, Italy [26]	88.2	82	68 (40/28)	61 (27/34)	(1.6)	4, 8	≤ 21	Age and gender were significant variables Disease duration and severity were not significant variables
Silveira-Moriyama et al, 2009, Brazil [33]	84.6	85	193 (115/78)	157 (93/64)	-	10, 2	≤ 23: ≥ 60 yrs ≤ 28: < 60 yrs	Age was a significant variable UPSIT examined on 167 PD and 130 HC of the study population
Silveira-Moriyama et al, 2008, Brazil [24]	83.5	82.1	106 (71/35)	118 (103/15)	-	11, 4	♀ ≤ 29: < 60 yrs ≤ 25: ≥ 60 yrs ♂ ≤ 19: < 60 yrs ≤ 16: ≥ 60 yrs	Age and gender were significant variables
Yu et al, 2014, Taiwan [35]	70	86	30 (15/15)	161 (81/80)	100, 0, 0	< 2	29.5	Age was a significant variable
Rodríguez-Violante et al, 2014, Mexico [25]	68.5	79.7	138 (81/57)	175 (85/90)	73, 22, 5	7, 4	≤ 25	Age and gender were significant variables
Rodríguez-Violante et al, 2014, Mexico [21]	66	82	100 (55/45)	100 (55/45)	73, 20, 7	7, 4	≤ 25	-
B-SIT								
Rodríguez-Violante et al, 2011, Mexico [34]	85.7	71.4	70 (41/29)	70 (41/29)	61, 34, 5	6, 7	≤ 9: < 60 yrs ≤ 7: ≥ 60 yrs	Age, disease duration and severity were significant variables
UPSIT-3								
Bohnen et al, 2007, USA [23]	96.3	70.3	27 (20/7)	27 (20/7)	96, 4, 0	2, 5	≤ 1	Used the 3 odours from the UPSIT with the highest sensitivity Disease duration was not a significant variable
OSIT-J								
Izawa et al, 2012, Japan [17]	78.1	84.8	33 (17/16)	32 (17/15)	-	4, 6	≤ 8	Disease duration and severity were not significant variables

B-SIT = Brief Smell Identification Test; H&Y = Hoehn & Yahr; HC = healthy controls; OSIT-J = Odour Stick Identification Test for Japanese; PD = Parkinson's disease; SIT = Sniffin' Sticks Odour Identification Test; SS-16 = Sniffin' Sticks 16; UPSIT = University of Pennsylvania Smell Identification Test;

cities from 65% to 74%. ET was also included, and four studies analysed the ability of the tests to differentiate PD from ET [7, 18, 36, 38], and found sensitivities between 62% and 92% and specificities between 67% and 96%.

DISCUSSION
All included studies found a significant correlation between olfactory function and IPD, but there was a wide range of sensitivities and specificities.

Identification of Parkinson's disease from healthy controls
Krismer et al [16] found the best ability to distinguish IPD from HC, but they only included 20 patients. The study was based on two independent cohorts, one which found specificity-weighted cut-off and one which determined the accuracy. Patients were followed for 24 months to ensure diagnosis. Mahlknecht et al [18] and Boesveldt et al [28] both had populations counting more than 400 patients, and both found sensitivities and specificities for SIT of approximately 83% and 82%. Mahlknecht et al [18] also made a prospective cohort at three independent centres, while Boesveldt et al [28] made a case-control study.

The study with largest patient population and with the most powerful study design also reported results closest to the mean of all included studies, which supports these results.

Rodríguez-Violante et al [21] performed a case-

TABLE 2

Overview of the included studies testing the ability to distinguish Parkinson's disease from other parkinsonian conditions: the ability to distinguish from multiple system atrophy, the ability to distinguish from progressive supranuclear palsy, the ability to distinguish from essential tremor, and the ability to distinguish from a pool of conditions that mimic idiopathic Parkinson's disease.

Reference	Differential diagnosis	PD n (♂/♀)	PD H&Y-stage, mean (%)	PD duration, mean, yrs	DD n (♂/♀)	DD H&Y-stage, mean	DD duration, mean, yrs	Test	Cut-off	Sens, %	Spec, %	Comments
Differentiating from MSA												
Mahlknecht et al, 2016, Austria/Italy [18]	MSA	134 (84/50)	2.4	6.2	23 (11/12)	3.3	4.2	SIT	≤ 9	84.1	87	Validation cohort, Centre A. Similar results were found by analysing patients with a disease duration < 3 yrs. Followed for 24 mo.s to validate the diagnosis
									≤ 10	92.1	78.3	Validation cohort, Centre A
Kikuchi et al, 2011, Japan [37]	MSA	42 (18/24)	-	2.7	42 (24/18)	-	2.6	OSIT-J	8.5	73.81	85.71	Final diagnosis was confirmed after 2 yrs of follow-up
Krismer et al, 2017, Austria [16]	MSA	47 (31/16)	2.5	7	23 (11/12)	3	4.2	SS-16	23.875	78.7	78.3	Similar results were found by analysing patients with a disease duration < 3 yrs
Differentiating from PSP												
Krismer et al, 2017, Austria [16]	PSP	47 (31/16)	2.5	7	23 (16/7)	3	3.2	SS-16	23.875	78.7	73.9	-
Mahlknecht et al, 2016, Austria/Italy [18]	PSP	134 (84/50)	2.4	6.2	23 (16/7)	3	3.1	SIT	≤ 9	84.1	73.9	Validation cohort, Centre A
									≤ 10	92.1	65.2	
Silveira-Moriyama et al, 2010, United Kingdom [6]	PSP	86 (49/37)	-	10.4	36 (20/16)	-	4.8	UPSIT	18	75	65.1	Age and gender were significant variables
Differentiating from ET												
Chen et al, 2012, China [36]	ET	37 (26/11)	(1 + 2: 76; > 2: 24)	3.6	26 (14/12)	-	12	SIT	9.5	62.2	96.2	Disease duration and severity were not significant variables
Shah et al, 2007, United Kingdom [7]	ET/ healthy controls	64 (44/20)	-	4.8	304 (111/193)	-	Familial ET: 24.2 Non-familial ET: 12.4	UPSIT	< 25	83	94	Gender was not a significant variable. PD patients was diagnosed with tremor dominant PD
Mahlknecht et al, 2016, Austria/Italy [18]	ET	134 (84/50)	2.4	6.2	29 (12/17)	-	18.7	SIT	≤ 9	84.1	89.7	-
									≤ 10	92.1	82.8	
López Hernández et al, 2015, Spain [38]	ET	30 (22/8)	-	2.1	21 (15/6)	-	6.3	SIT	< 8	70	66.7	-

Continues

control study and found the lowest specificity. The study included 100 patients, and used the Spanish UPSIT on a Mexican population, which could partly explain the low specificity because the odours were not familiar to the Mexican population.

Two studies [16, 22] used the entire SS-16, which includes odour identification, discrimination and threshold. The results were among the best with regard to sensitivities and specificities. Odour identification is the olfactory feature that is most frequently impaired in PD,

and the combination of several features only improved the quality of the SS-16 a little [16, 28]. Only Boesveldt et al [22] found that a combination of threshold and identification improved the AUC compared with the identification test alone, but the study was small and they did not follow the patients over time to ensure diagnosis. Only two studies used the SS-16. More studies are therefore required, but the results indicate that the SS-16 is only slightly better than identification alone. Identification test is significantly less time consuming

TABLE 2 CONTINUED

Reference	Differential diagnosis	PD n (♂/♀)	H&Y-stage, mean (%)	duration, mean, yrs	DD n (♂/♀)	H&Y-stage, mean	duration, mean, yrs	Test	Cut-off	Sens, %	Spec, %	Comments
Differentiating from several DD												
Wenning et al, 1994, England [4]	MSA/PSP/CBD	118 (58/60)	-	-	51 (34/17)	-	-	UPSIT	25	77	85	Age was a significant variable
Mahlknecht et al, 2016, Austria/Italy [18]	MSA/PSP/ET	134 (84/50)	2.4	6.2	75 (39/36)	3.2	8	SIT	≤ 9	84.1	84	-
Wolz et al, 2014, Germany [29]	ET, MSA and other tremor disorders	167 (116/51)	-	-	132 (73/59)	-	-	SIT	11	79	80.3	Age was a significant variable Gender was not a significant variable If there was doubt about the diagnosis, DaTSCAN was used to determine the final diagnosis
								SS-16	25	76	78.8	-
Krismer et al, 2017, Austria [16]	PSP/MSA	47 (31/16)	2.5	7	46 (27/19)	3	3.7	SS-16	23.875	78.7	76.1	-
Mahlknecht et al, 2016, Austria/Italy [18]	MSA/PSP/ET	134 (84/50)	2.4	6.2	75 (39/36)	3.2	8	SIT	≤ 10	92.1	76	-
Busse et al, 2012, Germany [32]	APS/VP/ET/MDD	385 (257/128)	3	9. 9	132 (67/65)	2	6	SIT-12	< 8	75	70	Age was a significant variable Disease duration and severity were not significant variables

APS = atypical parkinsonian syndromes; CBD = corticobasal degeneration; DD = differential diagnosis; ET = essential tremor; H&Y = Hoehn & Yahr; MDD = major depressive disorder with motor slowing; MSA = multiple system atrophy; OSIT-J = Odour stick identification test for Japanese; PD = Parkinson's disease; PSP = progressive supranuclear palsy; ; Sens = sensitivity; Spec = specificity; SIT = Sniffin' Sticks Odour Identification Test; SS-16 = Sniffin' Sticks 16; UPSIT = University of Pennsylvania Smell Identification Test; VP = vascular parkinsonism.

than the entire SS-16 [16]; and in a busy clinical setting, the use of an identification test would be preferable.

According to Figure 2, the sensitivities between the SIT and the UPSIT do not vary much, but the figure suggests a slightly lower specificity for the UPSIT. A few studies used smaller versions of the tests, and these studies indicate no inferior ability to diagnose PD, but this issue remains controversial due to the low number of studies.

Casjens et al [20] and Bohnen et al [23] analysed three odours they found to be the best to discriminate PD from HC. The tested subjects were from Germany and USA. None of the odours was the same in the two studies, probably because the familiar odours are different in each country. The low sensitivities in both studies may be the result of only using three odours. According to our study, the UPSIT or the SIT is the most useful olfactory test. Before using the olfactory test in the clinic, the odours should be tested and validated in the population in which they are intended to be used, and odours with a low recognisability in the healthy population should be substituted with more familiar odours.

Comparing the included studies, we find no correlation between the cut-off scores and the sensitivity and specificity. When choosing a low cut-off, we would expect a higher proportion of subjects to be classified as healthy, resulting in a higher specificity. The test should be used as a diagnostic tool and the cut-off should be adjusted to a specificity-balanced cut-off score.

The influence of gender on the olfactory scores was not consistent. Silveira-Moriyama et al [24] was the only study with gender-adjusted cut-off, and they did not find a better diagnostic value than the studies with a general cut-off value.

Progression of olfactory dysfunction
With increasing age comes a decrease in olfactory function in healthy people [39]. Most included studies found a negative correlation between age and olfactory function. The three studies [24, 33, 34] that used different cut-offs for different age groups did not find better diagnostic value than the studies using a general cut-off. The lowest sensitivity was found by Santin et al [27] who tested subjects below 45 years of age. The same

Testing olfactory function in a clinical setting.

study tested subjects above 45 years of age and found a markedly increased sensitivity and specificity (see Table 1). It seems that different cut-offs for different age groups do not increase the ability of the olfactory tests to diagnose PD, but the tests are more precise among older people because of a non-parallel reduction in olfactory score.

Olfactory dysfunction as a marker of progression
Of the included studies only Rodriguez-Violante et al [34] found a significant correlation between both disease stage and disease duration. Furthermore, Casjens et al [20] found disease stage to be correlated with olfactory score. Both were case-control studies with 70 and 148 patients, respectively. Five studies found no correlation between the olfactory identification score and the disease stage or duration [17, 26, 28, 32, 36]. Busse et al [32] was the only prospective cohort study where 632 patients with clinical parkinsonian symptoms were followed. The study found no correlation with either disease stage or duration, but a mild progression of hyposmia was found when comparing baseline and five-year follow-up data.

Boesveldt et al [28] indicated that the ability to discriminate odours was reduced by increased disease duration, also demonstrated by Tissingh et al [40]. These studies indicate that the identification test is correlated neither with disease stage nor with duration, and therefore is not a good prognostic factor; however, more prospective studies that analyse the olfactory function over time are needed, and special attention should be devoted to analysing other aspects than identification. The

identification test, however, can be used as a diagnostic tool independently of disease duration and severity, also early in the disease.

Separating Parkinson's disease from atypical Parkinson's disease
In clinical use, an important feature is the ability of a test to distinguish IPD from other parkinsonian conditions.

In MSA and PSP, the olfactory function is mildly impaired [4] and we should thus expect a lower specificity. Only three studies analysed MSA [16, 18, 37] and PSP [6, 16, 18]. None of them used the same olfactory test, and they all used different cut-offs; therefore, the frame of reference is weak. Overall, the ability to differentiate IPD from MSA varied, but the results indicate that the tests are useful in separating the two groups. Regarding PSP, the sensitivities and specificities indicated that the ability of the tests to distinguish IPD and PSP is weaker. Olfactory function in ET is maintained [7], and we expected results similar to those reported in studies comparing PD to HC. The four studies that analysed ET reported varying results, which made it difficult to draw a definite conclusion about the ability to differentiate between patients with PD and ET.

Five studies compared PD against two or more conditions that mimic PD (see Table 2). None of these studies were similar in terms of the differential diagnoses with which PD was compared, and the number of differential diagnoses in the pool also varied. The largest study in this group was Busse et al [32] which is described above. Patients were examined with a short version of SIT, and a low specificity of 70% was found. Mahlknect et al [18] found a specificity of 76% with a cut-off ≤ 10. Lowering the cut-off to ≤ 9 increased the specificity to 84%.

To meet the MDS criteria for the olfactory test to be an ancillary diagnostic tool, the test should provide a specificity > 80% in most studies, with a minimum of three studies from different centres [19]. According to our study, three out of five studies found a specificity > 80% (Table 2), but the study with the strongest study design and the largest population only had a specificity of 70%. This could indicate that with the full UPSIT or the SIT and with a specificity-adjusted cut-off, the olfactory test can be used according to the MDS criteria.

According to the MDS, the test is an ancillary test among many others. If the olfactory score is below the cut-off, the positive predictive value (PPV) will depend on the setting in which the test is used. In general practice, the prevalence of IPD is low, which will result in a low PPV. However, in a specialised neurologic clinic, the test could be a valuable tool since the prevalence of IPD is higher.

A strength in our study is that the screening and exclusion processes were accomplished by two authors, which can minimize the risk of personal bias and opinion. Furthermore, two authors independently went through all the included studies, which also increases the re-test ability.

A limitation of this study was that the number of studies using SS-16 and UPSIT was limited, which might lead to uncertain results. The same applies for studies that compare PD to other parkinsonian conditions. The studies had to be published after 1992 to be included in the final analysis. UPSIT has been used since 1984 [41], and with a bigger time frame, the results could have been stronger. We chose studies published after 1992 because the UK PD Brain Bank Criteria were published that year [1]. This ensures that the population in our study was diagnosed according to consistent criteria. Hughes et al [1] studied patients diagnosed with PD, and found that the UK PD Brain Bank misdiagnosed 24% [1]. Thus, when comparing PD with parkinsonian syndromes, some of the PD patients might have had atypical PD, and the ability to distinguish PD from atypical PD might be higher than the one we found. Olfactory test as well as other tools may potentially assist early identification of PD, but this is beyond the scope of this review.

CONCLUSIONS

Olfactory test can be a valuable ancillary tool in the diagnostic process of PD. The diagnostic value becomes more precise by increasing age. SS-16 seems to be the most precise test, but SIT is the most usable test in a busy clinical setting because it can be conducted quickly and independently of disease duration and severity.

Sensitivities and specificities in distinguishing PD from MSA, PSP or ET indicate that the olfactory identification test is a valuable ancillary tool in the diagnostic process of PD, but the results are weak due to the low number of studies. According to our study, three out of five studies found a specificity > 80% in distinguishing PD from a pool of conditions that mimic PD, which is required for an ancillary test according to the new MDS criteria.

Before using an olfactory test in a clinical setting, it is necessary to adjust the odours to the population it is intended for, and to find the optimal specificity-adjusted cut-off.

CORRESPONDENCE: *Tine Nielsen.* E-mail: tine_nielsen@hotmail.dk

LITERATURE

1. Hughes AJ, Daniel SE, Kilford L et al. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992;55:181-4.
2. Doty RL. Olfactory dysfunction in Parkinson disease. Nat Rev Neurol 2012;8:329-39.
3. Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. Lancet Neurol 2006;5:75-86.
4. Wenning GK, Shephard B, Hawkes C et al. Olfactory function in atypical parkinsonian syndromes. Acta Neurol Scand 1995;91:247-50.
5. Muller A, Mungersdorf M, Reichmann H et al. Olfactory function in Parkinsonian syndromes. J Clin Neurosci 2002;9:521-4.
6. Silveira-Moriyama L, Hughes G, Church A et al. Hyposmia in progressive supranuclear palsy. Mov Disord 2010;25:570-7.
7. Shah M, Muhammed N, Findley LJ et al. Olfactory tests in the diagnosis of essential tremor. Parkinsonism Relat Disord 2008;14:563-8.
8. Doty RL, Shaman P, Kimmelman CP et al. University of Pennsylvania Smell Identification Test: a rapid quantitative olfactory function test for the clinic. Laryngoscope 1984;94:176-8.
9. Kobal G, Hummel T, Sekinger B et al. "Sniffin' sticks": screening of olfactory performance. Rhinology 1996;34:222-6.
10. Eibenstein A, Fioretti AB, Lena C et al. Modern psychophysical tests to assess olfactory function. Neurol Sci 2005;26:147-55.
11. Doty RL, Marcus A, Lee WW. Development of the 12-item Cross-Cultural Smell Identification Test (CC-SIT). Laryngoscope 1996;106:353-6.
12. Fjaeldstad A, Kjaergaard T, Van Hartevelt TJ et al. Olfactory screening: validation of Sniffin' Sticks in Denmark. Clin Otolaryngol 2015;40:545-50.
13. Konstantinidis I, Printza A, Genetzaki S et al. Cultural adaptation of an olfactory identification test: the Greek version of Sniffin' Sticks. Rhinology 2008;46:292-6.
14. Tekeli H, Altundag A, Salihoglu M et al. The applicability of the "Sniffin' Sticks" olfactory test in a Turkish population. Med Sci Monit 2013;19:1221-6.
15. Kobayashi M, Reiter ER, DiNardo LJ et al. A new clinical olfactory function test: cross-cultural influence. Arch Otolaryngol Head Neck Surg 2007;133:331-6.
16. Krismer F, Pinter B, Mueller C et al. Sniffing the diagnosis: olfactory testing in neurodegenerative parkinsonism. Parkinsonism Relat Disord 2017;35:36-41.
17. Izawa MO, Miwa H, Kajimoto Y et al. Combination of transcranial sonography, olfactory testing, and MIBG myocardial scintigraphy as a diagnostic indicator for Parkinson's disease. Eur J Neurol 2012;19:411-6.
18. Mahlknecht P, Pechlaner R, Boesveldt S et al. Optimizing odor identification testing as quick and accurate diagnostic tool for Parkinson's disease. Mov Disord 2016;31:1408-13.
19. Postuma RB, Berg D, Stern M et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30:1591-601.
20. Casjens S, Eckert A, Woitalla D et al. Diagnostic value of the impairment of olfaction in Parkinson's disease. PLoS One 2013;8:e64735.
21. Rodriguez-Violante M, Gonzalez-Latapi P, Camacho-Ordonez A et al. Comparing the accuracy of different smell identification tests in Parkinson's disease: relevance of cultural aspects. Clin Neurol Neurosurg 2014;123:9-14.
22. Boesveldt S, de Muinck Keizer RJ, Knol DL et al. Extended testing across, not within, tasks raises diagnostic accuracy of smell testing in Parkinson's disease. Mov Disord 2009;24:85-90.
23. Bohnen NI, Gedela S, Kuwabara H et al. Selective hyposmia and nigro-striatal dopaminergic denervation in Parkinson's disease. J Neurol 2007;254:84-90.
24. Silveira-Moriyama L, Carvalho Mde J, Katzenschlager R et al. The use of smell identification tests in the diagnosis of Parkinson's disease in Brazil. Mov Disord 2008;23:2328-34.
25. Rodriguez-Violante M, Gonzalez-Latapi P, Camacho-Ordonez A et al. Low specificity and sensitivity of smell identification testing for the diagnosis of Parkinson's disease. Arq Neuropsiquiatr 2014;72:33-7.
26. Picillo M, Pellecchia MT, Erro R et al. The use of University of Pennsylvania Smell Identification Test in the diagnosis of Parkinson's disease in Italy. Neurol Sci 2014;35:379-83.
27. Santin R, Fonseca VF, Bleil CB et al. Olfactory function and Parkinson's disease in Southern Brazil. Arq Neuropsiquiatr 2010;68:252-7.
28. Boesveldt S, Verbaan D, Knol DL et al. A comparative study of odor identification and odor discrimination deficits in Parkinson's disease. Mov Disord 2008;23:1984-90.
29. Wolz M, Hahner A, Meixner L et al. Accurate detection of Parkinson's disease in tremor syndromes using olfactory testing. Eur Neurol 2014;72:1-6.
30. Silveira-Moriyama L, Sirisena D, Gamage P et al. Adapting the Sniffin' Sticks to diagnose Parkinson's disease in Sri Lanka. Mov Disord 2009;24:1229-33.
31. Chen W, Chen S, Kang WY et al. Application of odor identification test in Parkinson's disease in China: a matched case-control study. J Neurol Sci 2012;316:47-50.
32. Busse K, Heilmann R, Kleinschmidt S et al. Value of combined midbrain sonography, olfactory and motor function assessment in the differential

diagnosis of early Parkinson's disease. J Neurol Neurosurg Psychiatry 2012;83:441-7.

33. Silveira-Moriyama L, Petrie A, Williams DR et al. The use of a color coded probability scale to interpret smell tests in suspected parkinsonism. Mov Disord 2009;24:1144-53.

34. Rodriguez-Violante M, Lees AJ, Cervantes-Arriaga A et al. Use of smell test identification in Parkinson's disease in Mexico: a matched case-control study. Mov Disord 2011;26:173-6.

35. Yu CY, Wu RM. Application of the University Of Pennsylvania Smell Identification Test (traditional Chinese version) for detecting olfactory deficits in early Parkinson's disease in a Taiwanese cohort. J Parkinsons Dis 2014;4:175-80.

36. Chen W, Tan YY, Hu YY et al. Combination of olfactory test and substantia nigra transcranial sonopraphy in the differential diagnosis of Parkinson's disease: a pilot study from China. Transl Neurodegener 2012;1:25.

37. Kikuchi A, Baba T, Hasegawa T et al. Differentiating Parkinson's disease from multiple system atrophy by [123I] meta-iodobenzylguanidine myocardial scintigraphy and olfactory test. Parkinsonism Relat Disord 2011;17:698-700.

38. López Hernández N, Garcia Escriva A, Shalabi Benavent M. Diagnostic value of combined assessment of olfaction and substantia nigra hyperechogenicity for Parkinson's disease. Neurologia 2015;30:496-501.

39. Hummel T, Kobal G, Gudziol H et al. Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. Eur Arch Otorhinolaryngol 2007;264:237-43.

40. Tissingh G, Berendse HW, Bergmans P et al. Loss of olfaction in de novo and treated Parkinson's disease: possible implications for early diagnosis. Mov Disord 2001;16:41-6.

41. Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. Physiol Behav 1984;32:489-502.

42. Huang SF, Chen K, Wu JJ et al. Odor identification test in idiopathic REM - behavior disorder and Parkinson's disease in China. PLoS One 2016;11: e0160199.

43. Bohnen NI, Studenski SA, Constantine GM et al. Diagnostic performance of clinical motor and non-motor tests of Parkinson disease: a matched case-control study. Eur J Neurol 2008;15:685-91.