Olfactory screening: validation of Sniffin' Sticks in Denmark

Fjaeldstad, A.,*† Kjaergaard, T.,* Van Hartevelt, T.J.,†‡ Moeller, A.,‡ Kringelbach, M.L.†‡ & Ovesen, T.*

*Department of Otorhinolaryngology, Aarhus University Hospital, Aarhus, Denmark †Department of Psychiatry, University of Oxford, Oxford, United Kingdom ‡CFIN/MindLab, Aarhus University, Aarhus, Denmark

Accepted for publication 20 February 2015
Clin. Otolaryngol. 2015, **40**, 545–550

Objectives: The Sniffin' Sticks 12-identification test (SIT-12) is the most commonly applied Danish olfaction screening tool; however, it has never been validated in a Danish population. The screening score depends on familiarity with descriptors, which is strongly influenced by linguistic and cultural factors, why validation is mandatory. This study aimed to validate the SIT-12 in a Danish population.
Design: Prospective controlled study.
Setting: Otorhinolaryngology department.
Participants: The SIT-12 was applied to 100 normosmic, healthy adult Danish participants.
Main outcome measures: Choice of descriptors was registered, along with nasal endoscopic examination, screening for cognitive impairment, depression and sinonasal symptoms. Descriptors of the original version of SIT-12 were evaluated in 50 participants, and misleading descriptors were identified. Modifications to these descriptors were subsequently validated in a comparable group of 50 participants.
Results: Mean odorant identification score in the evaluation group was 11.0 of a possible 12, and 11.6 in the validation group ($P < 0.0001$). Among all odorant identification errors in the evaluation group, 60% were due to two incorrect descriptors having close resemblance to the correct descriptors, lemon and cinnamon. Two additional descriptors were unfamiliar to more than half the participants. There was a significant difference in the distribution of wrong identification answers between odorants in the evaluation group ($P < 0.001$), but not in the validation group.
Conclusions: The identified systematically wrong descriptors have been modified and validated in the Danish SIT-12.

Approximately 20% of the general population suffers from an impaired sense of smell, of which 5% are anosmic.[1–3] This can affect quality of life and increase the risk of exposure to fire and gas hazards. In addition, it is a potential early clinical indicator of Parkinson's and Alzheimer's disease.[4,5] Evaluating olfactory function is essential in patients prior to nasal surgery to assess the effect of surgery and in patients with olfactory disorders to guide therapeutic intervention.

Olfactory assessment is regarded an essential part of the otorhinolaryngologic examination in patients with nasal or sinus disease[6] and can be used as a supportive diagnostic tool for Parkinson's disease.[7] In Europe, the most commonly applied smell identification test is the Sniffin' Sticks (Burghart Messtechnik GmbH, Wedel, Germany). It is based on odorants selected specifically to be applicable in the general European population[8] and contains two versions: Sniffin' Sticks 12-identification test (SIT-12) for a fast screening of olfactory function (12 odorant sticks); Sniffin' Sticks extended test for evaluation of odorant threshold, discrimination and identification (112 odorant sticks).[3,9]

Cultural components have a strong influence on odour exposure frequency and identification abilities in different countries and ethnicities. This was evident in the United Kingdom,[10] Greece[11] and Turkey,[12] where the printed 'forced multiple-choice' options showed to be inappropriate due to the odours' cultural components, resulting in a substitution of unsuitable descriptors.

In the original validation of the SIT-12, a prerequisite for the selection of the four descriptors for any given odorant was a successful identification rate of >75% in a normosmic population.[9] Accordingly, adaptation of the descriptors in a new cultural setting is considered mandatory before use.

The SIT-12 is the most widely used clinical olfactory test in otorhinolaryngological departments and clinics, and neurological departments in Denmark. However, the test has not yet been validated for clinical use in Denmark, in contrast to several other European countries (Germany,[3] United Kingdom,[10] Greece,[11] Turkey,[12] Italy,[13] Poland[14] and the Netherlands[15]). Continuing to systematically apply a non-validated test for clinical decision-making can result in unnecessary medical examinations of healthy individuals,

Correspondence: A. Fjaeldstad, Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford OX3 7JX, UK. Tel.: +45 25341568; Fax: +44 0 1865793101; e-mail: alefja@rm.dk

raise the burden on health services and make scores problematic to compare with international olfaction literature.

We hypothesised that there was no significant difference in the distribution of incorrect answers in the Danish SIT-12 and that all odorants were correctly identified by >75% in a normosmic population. The results rejected this hypothesis, and we constructed a modified list of descriptors based on participant interviews, which was subsequently validated.

Materials and methods

Participants and ethics

The Danish SIT-12 was applied to 102 subjectively normosmic, healthy Danish participants between 18 and 50 years of age: 51 participants in the evaluation group (original descriptors) and 51 participants in the validation group (modified descriptors). All participants underwent nasal endoscopic examination, screening for cognitive impairment (Mini-Mental State Examination), depression (Major Depression Inventory) and sinonasal symptoms [SinoNasal Outcome Test (SNOT-22)]. Olfactory function was evaluated using the SIT-12. The study was performed in accordance with the Declaration of Helsinki ethical principles for medical research and approved by the Danish Ethical Committee.

The Sniffin' Sticks 12-identification test

The SIT-12 consists of 12 felt-tip pen-like devices containing common odours. Participants were presented with a four-alternative forced multiple-choice test, to identify the correct odorant from a list of four descriptors. The experimenter presented the odorant by removing the cap for ≈3 seconds and placing the tip of the pen 1–2 cm in front of both nostrils of the participant, with an interval between odours of 20–30 seconds, to clear the olfactory cleft of the prior odour.[9] The participants were allowed to sample the odours twice before choosing their descriptor. The identification score between 0 and 12 was registered along with the choice of descriptor, certainty of choice and reason for any uncertainty. Normosmic identification score for the test is ≥10.

Modification process

The original Danish list of descriptors has been used clinically for several years, but has never been validated. Participants graded their certainty of each descriptor choice on a scale from 1 to 10 and elaborated on the reason for any uncertainty in a questionnaire. Familiarity of all four descriptors for each odorant was registered. The indicated reasons for uncertainty and familiarity of descriptors formed the basis for modifying descriptors, if the accumulated uncertainty for a descriptor was above 25%. Descriptor uncertainty and familiarity were reassessed in the validation group.

Statistics

Results were analysed using STATA version 12 for Mac (StataCorp LP, College Station, TX, USA). The distribution of incorrect identification of descriptors was compared between odorants using Fisher's exact test before and after the modification process. Distributions of mean identification scores and demographics in each group were compared using the student's *t*-test. Data are shown as mean values with the 95% confidence interval (95% CI) for Gaussian distributed data and as median values with interquartile ranges (IQR) for non-Gaussian distributed data.

Table 1. Demographics

	Original descriptors Evaluation group Normosmic participants (n = 50) Mean (95% CI)	Modified descriptors Validation group Normosmic participants (n = 50) Mean (95% CI)
Age (years)	27* (23–35)	29* (25–32)
Gender (male/female)	26/24	26/24
SIT-12 score	11.00 (10.76; 11.24)	11.58 (11.41; 11.74)
MMSE	29.6 (29.4; 29.8)	29.7 (29.6; 29.8)
MDI	4.5 (3.4; 5.6)	5.7 (4.4; 7.1)
SNOT-22	6.3 (4.2; 8.3)	9.3 (6.5; 12.0)

SIT-12, Sniffin' Sticks 12-identification; MMSE, Mini-Mental State Examination; MDI, Major Depression Inventory; SNOT-22, Sinonasal Outcome Test.

*Age is listed as median age with interquartile range in parenthesis, as data do not follow Gaussian distribution. All test scores are listed as mean values with 95% confidence intervals. Characteristics are based on the 100 participants with SIT-12 scores of 10–12 (normosmics).

Results

SIT-12 scores and study population

Among the 51 subjectively normosmic Danes in the evaluation group, 17 scored 12 of 12 (10 male), 16 scored 11 of 12 (seven male), 17 scored 10 of 12 (nine male) and one scored 9 of 12 (one male). In the validation group, 31 scored 12 of 12 (19 male), 17 scored 11 of 12 (five male), two scored 10 of 12 (two male), and one scored 8 of 12 (one male). As normosmia requires a score of ≥10, two participants were removed from further data analysis, resulting in 50 participants per group. The two groups had an identical distribution of sex and a median age of 27 (IQR 23–35) in the evaluation group and 29 (IQR 25–32) in the validation group.

Mean odorant identification score was significantly improved from 11.00 (95% CI: 10.76; 11.24) in the evaluation group to 11.58 (95% CI: 11.41; 11.74) in the validation group ($P < 0.001$), with no significant sex differences ($P = 0.78$) (Table 1).

Distribution and causes of descriptor errors

There was a significant difference in the distribution of wrong identification answers between odorants in the evaluation group ($P < 0.001$); among all odorant identification errors, 60% were due to two of the 12 odorants. For both odorants, the same incorrect descriptor of three incorrect possibilities was chosen. The lemon odorant was described as difficult due to the citric resemblance to grapefruit – this descriptor was subsequently replaced with the Danish word for raspberry ('hindbaer'). The sweet spicy hints in the cinnamon odorant were described as the reason for choosing the honey descriptor – this descriptor was subsequently replaced with the Danish word for a non-spicy descriptor, apple ('aeble'). The descriptor, curled mint ('krusemynte'), was by more than a quarter described as unfamiliar, but as all participants were familiar with the liquorice odour, this had no effect on identification scores. Curled mint was replaced with mint ('mynte'), as all participants were familiar with this descriptor. Cloves, the dried aromatic flower buds of an Indonesian tree, had been translated from English to Danish without taking into account that this spice is commonly referred to as the shorter Danish name for the European Dianthus flower – despite belonging to a different order of flowering plants. As only one participant knew of this reference, a botanist, the descriptor was changed to the common Danish appellation ('kryddernellike' to 'nellike') (Fig. 1).

Effect of possible confounders on odorant identification

In the evaluation group, participants with an abnormal SNOT-22 score ($n = 14$) (SNOT-22 score >7)[16] had lower SIT-12 scores than participants with normal SNOT-22 scores (mean score 10.71 *versus* 11.11, $P = 0.08$). The difference in means was non-significant, and excluding these participants did not affect the distribution of wrong answers nor bring the proportion of correct identification above 75%.

In the validation group, the difference in mean SIT-12 scores was not significantly different between participants with normal and abnormal SNOT-22 score ($P = 0.20$). Participants with an abnormal SNOT-22 score ($n = 22$) had a mean SIT-12 of 11.50 (95% CI: 11.24; 11.76), while

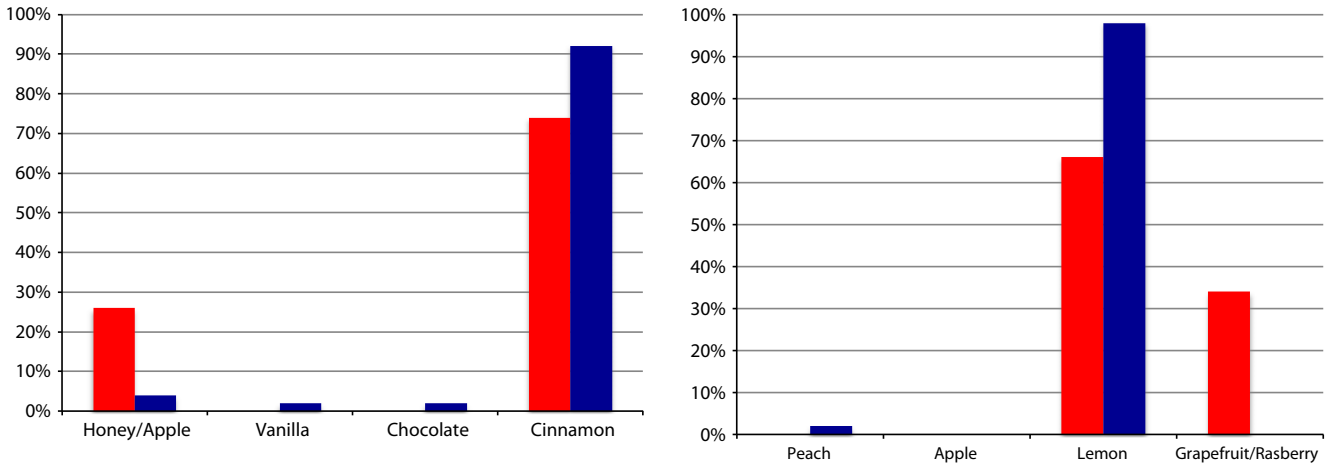

Fig. 1. Distribution of descriptor answers for the cinnamon and lemon odorants with original (red) and modified descriptors (blue). After modification of descriptors, both rates of correct identification improved from <75% to >90%.

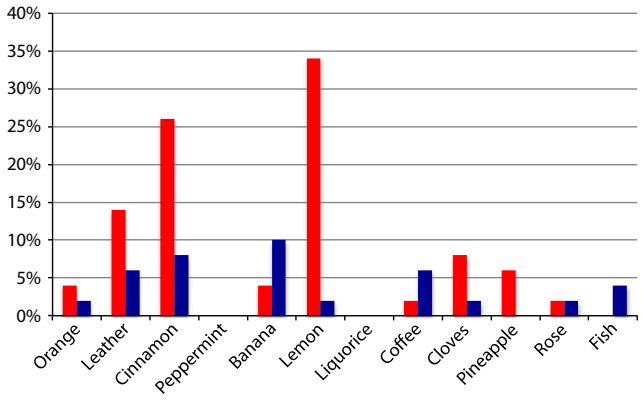

Fig. 2. Percentage of odorant identification errors before (red) and after (blue) the modification of descriptors.

participants with normal SNOT-22 had a mean SIT-12 score of 11.64 (95% CI: 11.43; 11.86).

None of the participants had pathological Mini-Mental State Examination or Major Depression Inventory scores. Nasal endoscopy revealed minor to moderate nasal mucosal inflammation in 45 of the 100 normosmic participants, and their mean SIT-12 was 11.29 (95% CI: 11.04; 11.53) compared with 11.29 (95% CI: 11.09; 11.49) in the participants without inflammation, with no difference between the two groups.

Effect of modification of descriptors on identification rate

After modification of the descriptors, all odorants had an identification rate of >75% in the normosmic validation group, all descriptors were familiar to >90% of participants, and there was no longer a significant difference in distribution of wrong answers between odorants ($P = 0.09$) (Fig. 2).

Discussion

Synopsis of key findings

We found a significant difference in the distribution of wrong answers among the 12 odorants in the evaluation group, demonstrating a systematic error in the test. Systematic errors can occur due to an unfamiliar odorant/descriptor or due to a close relation between the odorant and a wrong descriptor. The mean SIT-12 score improved significantly after modification and validation of descriptors ($P < 0.0001$).

Comparison with other studies

The presence of a systematic error is supported by comparing these Danish scores with scores from another subjectively normosmic European population; the Italian mean score was 11.5 (95% CI: 11.4; 11.6) compared to the Danish mean score of 11.0 (95% CI: 10.8; 11.2; $P < 0.0001$) in the evaluation group and 11.6 (95% CI: 11.4; 11.7; $P = 0.17$) in the validation group. The original Danish SIT-12 mean score was significantly worse than the Italian, while the modified Danish SIT-12 was not significantly different. This Italian study identified cloves as an unfamiliar Italian cultural odorant, but compared with our findings, this only had a minor influence on the mean score.[13] A Turkish validation study pointed out an identical problem with resemblance of descriptors for the lemon and cinnamon odorants, as was observed in the present Danish study. Cultural unfamiliar descriptors in the Turkish population were turpentine, liquorice and pineapple. In one case, the odorant in itself was suggested as a possible cause of error; apples come in a variety of cultivars, and the Sniffin' Sticks apple odour resembles a

Table 2. Modified list of descriptors for odorant #1–12 in Danish SIT-12

#	Descriptor 1	Descriptor 2	Descriptor 3	Descriptor 4
1	**Orange**	Blackberry	Strawberry	Pineapple
2	Smoke	Glue	**Leather**	Grass
3	(Honey) Apple	Vanilla	Chocolate	**Cinnamon**
4	Chives	**Peppermint**	Spruce	Onion
5	Coconut	**Banana**	Walnut	Cherry
6	Peach	Apple	**Lemon**	(Grapefruit) Raspberry
7	**Liquorice**	Rubber	(*Curled mint*) Mint	Cookies
8	Cigarette	**Coffee**	Wine	Smoke
9	(*Cloves*) **Cloves**	Pepper	Cinnamon	Mustard
10	Pear	Plum	Peach	**Pineapple**
11	Camomile	Raspberry	**Rose**	Cherry
12	Bread	**Fish**	Cheese	Ham

Correct odorant descriptors are in bold. Modified descriptors due to low identification rate <75% are underlined, and descriptors that modify due to linguistic unfamiliarities are in italic. Cloves were modified to a common Danish appellation.

common Turkish air freshener more than the sweet cultivars the Turks are accustomed to.[12] In a Greek study, six descriptors were modified due to a significantly decreased identification percentage. Additionally, three descriptors were replaced with descriptors more familiar to the Greek population; anise was changed to ouzo, liquorice to Greek grappa, and turpentine to painter oil.[11] The similarities of distractors were also emphasised as a possible explanation for low identification scores in the British population,[10] where the grapefruit descriptor – among others – was modified. The list of original descriptors in their study included sauerkraut – rooted in the original German list of descriptors, clearly highlighting the need for cultural adaptation.

Strengths of the study

The test requires >75% correct answers for each odorant in a normosmic population;[9] four systematic errors were identified and modified and subsequently validated in the second part of the study, where the normosmic validation group ($n = 50$) was tested in an identical setting. We controlled for cognitive impairment (Mini-Mental State Examination questionnaire), depression (Major Depression Inventory), sinonasal symptoms (Sino-Nasal Outcome Test-22) and performed endoscopic nasal examination in both groups (Table 2).

In recent European validation studies, authors have suggested modifications, with no described input from the participants being tested,[10,12] and only one study validated the proposed modified descriptors.[11] By reassessing wrongly identified odorants and descriptors with input from participants, a better modification process was possible. In the present study, a validation of the proposed modifications was made, which we propose as standard practice in future Sniffin' Sticks validation studies to avoid implementation of new systematic descriptor errors.

Conclusion

It is vital to have an accurate olfactory test to diagnose people with olfactory disorders correctly. The main purpose of the SIT-12 is a rapid screening to identify patients who need additional olfactory diagnostic evaluation; before modification, the systematic descriptor errors lowered the specificity of the test. Furthermore, these systematic errors decreased the Danish mean identification score and made it problematic to compare with international olfaction literature. Consequently, the SIT-12 has now been culturally adapted for the Danish population through a modification of four descriptors and a subsequent validation.

Author contributions

Alexander Fjaeldstad contributed to the study idea, participated in the study design, testing and examining participants, data analysis, wrote the initial draft of the paper, and edited the paper based on suggestions and comments from co-authors. Thomas Kjaergaard contributed to the study idea and study design, participated in conducting and supervising the examination of participants for consistency in clinical examination, reviewed the paper and assisted with data analysis. Tim Van Hartevelt contributed to the study idea, participated in testing participants, and reviewed the paper. Arne Moeller contributed to the study idea, improved layout of statistics and reviewed the paper. Morten Kringelbach contributed to the study idea, provided statistical assistance, and reviewed the paper. Therese Ovesen contributed to the study idea, participated in examining participants, and reviewed the paper.

Conflict of interest

None to declare.

References

1 Brämerson A., Johansson L., Ek L. *et al.* (2004) Prevalence of olfactory dysfunction: the Skövde population-based study. *Laryngoscope* **114**, 733–737
2 Mullol J., Alobid I., Marino-Sanchez F. *et al.* (2012) Furthering the understanding of olfaction, prevalence of loss of smell and risk factors: a population-based survey (OLFACAT study). *BMJ Open* **2**, e001256
3 Hummel T., Kobal G., Gudziol H. *et al.* (2007) 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.* **264**, 237–243
4 Sohrabi H.R., Bates K.A., Weinborn M.G. *et al.* (2012) Olfactory discrimination predicts cognitive decline among community-dwelling older adults. *Transl. Psychiatry* **2**, e118
5 Haehner A., Boesveldt S., Berendse H.W. *et al.* (2009) Prevalence of smell loss in Parkinson's disease – a multicenter study. *Parkinsonism Relat. Disord.* **15**, 490–494
6 Fjaeldstad A., Clausen C.H., Kjaergaard T. *et al.* (2014) Lugtesansen har stor klinisk relevans. *Ugeskr. Laeger* **176**, V04130252
7 Casjens S., Eckert A., Woitalla D. *et al.* (2013) Diagnostic value of the impairment of olfaction in Parkinson's disease. *PLoS ONE* **8**, e64735
8 Kobal G., Hummel T.H., Sekinger B. *et al.* (1996) 'Sniffin'sticks': screening of olfactory performance. *Rhinology* **34**, 222–226
9 Hummel T., Sekinger B., Wolf S.R. *et al.* (1997) 'Sniffin' Sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chem. Senses* **22**, 39–52

10 Neumann C., Tsioulos K., Merkonidis C. *et al.* (2012) Validation study of the 'Sniffin' Sticks' olfactory test in a British population: a preliminary communication. *Clin. Otolaryngol.* **37**, 23–27
11 Konstantinidis I., Printza A., Genetzaki S. *et al.* (2008) Cultural adaptation of an olfactory identification test: the Greek version of Sniffin' Sticks. *Rhinology* **46**, 292–296
12 Tekeli H., Altundağ A., Salihoğlu M. *et al.* (2013) The applicability of the 'Sniffin' Sticks' olfactory test in a Turkish population. *Med. Sci. Monit.* **19**, 1221–1226
13 Eibenstein A., Fioretti A.B. & Lena C. (2005) Olfactory screening test: experience in 102 Italian subjects. *Acta Otorhinolaryngol. Ital.* **25**, 18–22
14 Sorokowska A. & Hummel T. (2014) Polish version of the Sniffin' Sticks test - adaptation and normalization. *Otolaryngol. Pol.* **68**(6), 308–314.
15 Boesveldt S., Verbaan D., Knol D.L. *et al.* (2008) Odour identification and discrimination in Dutch adults over 45 years. *Rhinology* **46**, 131–136
16 Gillett S., Hopkins C., Slack R. *et al.* (2009) A pilot study of the SNOT 22 score in adults with no sinonasal disease. *Clin. Otolaryngol.* **34**, 467–469