How Many and Which Odor Identification Items Are Needed to Establish Normal Olfactory Function?

Jörn Lötsch¹,², Alfred Ultsch³ and Thomas Hummel⁴

¹Institute of Clinical Pharmacology, Goethe - University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany, ²Project Group Translational Medicine and Pharmacology TMP, Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Theodor-Stern-Kai 7, 60596 Frankfurt am Main, Germany, ³DataBionics Research Group, University of Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany and ⁴Department of Otorhinolaryngology, Smell & Taste Clinic, TU Dresden, Fetscherstrasse 74, 01307 Dresden, Germany

Correspondence to be sent to: Jörn Lötsch, Institute of Clinical Pharmacology, Goethe – University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany. e-mail: j.loetsch@em.uni-frankfurt.de

Accepted 18 January 2016.

Abstract

The establishment of normal olfactory function by means of a simple and reliable test is one method that could minimize olfactory test procedures in the clinic. This retrospective study analyzed the identification of 16 odors by 613 subjects (aged 18–96 years, 266 men) as a part of a complex olfactory test battery by which 183, 251, and 179 subjects were diagnosed with anosmia, hyposmia, or normosmia, respectively. Cinnamon was identified as the best scoring odor, that is, identified correctly by most normosmic subjects, but identified correctly by the fewest anosmic patients. An exact calculation of the optimum number of items needed for a diagnosis of normosmia resulted in 1 single odor identification item as being sufficient. The inclusion of more items is solely determined by the acceptable proportion of chance, which in a 4-alternative forced choice paradigm is only 1.6% with 3 odors. A proposed screening test using cinnamon, fish odor, and banana established normosmia at a sensitivity of 80.4% and a specificity of 84.3% and a negative predictive value of 91.3%. A positive test result reliably establishes normosmia providing a confidence basis to terminate olfactory assessments following the application of only 3 odor identification items.

Key words: ABC analysis, bioinformatics, clinical assessment, human olfaction, patients

Introduction

Assessment of olfactory function has become a standard diagnostic test in Neurology, Psychiatry and Otorhinolaryngology (Brewer et al. 2003; Bramerson et al. 2004, 2007; Hawkes 2006). This owes to the increasing awareness of the important role of the sense of smell for quality of life (Hummel and Nordin 2005; Rinaldi 2007; Croy et al. 2014b), recognition of olfactory dysfunction as an early symptom of neurological diseases such as Parkinson's disease (Doty et al. 1988), Alzheimer's disease (Murphy et al. 1990) multiple sclerosis (Hawkes 1996), or schizophrenia (Brewer et al. 2003), or evidence about olfactory drug effects which are among the most frequent complaints of patients (Tuccori et al. 2011).

Although comprehensive olfactory tests allow for a detailed diagnosis of a patient's olfactory function, these test batteries place a considerable workload onto physicians revealing the demand for screening tests (Simmen et al. 1999; Gilbert et al. 2002; Jackman and Doty 2003; Mueller and Renner 2006). Self-report of patients does not suffice as research tells that people are notoriously bad in rating their sense of smell (Cardesín et al. 2006; Shu et al. 2009). Therefore, several attempts at reducing the number of odor identification items have produced screening tests comprising 1–12 odors (Doty et al. 1996; Simmen et al. 1999; Hummel et al. 2001; Gilbert et al. 2002; Jackman and Doty 2005; Mueller and Renner 2006), however, it...
remains uncertain how many items are sufficient for a valid diagnosis of normosmia.

This present analysis therefore addressed the nature and the minimum number of odor identification items with a particular focus to establish normosmia as a common first clinical step in olfactory diagnostics. If this can be obtained quickly; no further action is required, while doubts about normal olfactory function should be explored with more comprehensive tests. Similarly, in scientific studies enrolment of subjects with normosmia may be an inclusion criterion, for example in studies on olfactory drug effects (Lotsch et al. 2012, 2015) or olfactory research in neuroscience (Lapid et al. 2011), and its establishment also benefits from a quick, yet valid, screening test diagnosis.

Materials and methods

Subjects and assessment of olfactory function

This study followed the Declaration of Helsinki and was approved by the Ethics Committee of the Faculty of Medicine of the TU Dresden (EK251112006). Detailed olfactory test results were analyzed from 613 subjects (age: range 18–96 years, mean ± standard deviation: 56.2 ± 17.3 years; sex: 266 men, 347 women) who had presented at the Smell & Taste Clinic, Department of ORL, TU Dresden, either with the symptom “olfactory loss” or as healthy controls recruited via flyers. The patients provided informed written consent into study procedures and data acquisition. A sufficiently large fraction of all olfactory diagnoses was identified in the study sample, that is, functional (general) anosmia (further on termed “anosmia”; n = 183), hyposmia (n = 251), and normosmia (n = 179) (Table 1). This deviates from a random sample in the average population by a larger fraction of subjects with less-than-normal olfactory function (Hummel et al. 2007). The main causes of olfactory disturbances were also represented, that is, suspected causes were: 138 idiopathic, 83 sinonasal disease, 93 viral, 119 head trauma, 17 neurodegenerative diseases, 6 congenital, 8 other causes, and in addition 149 healthy subjects. Smoking status, which has been reported to affect olfactory function (Venstrom and Amoore 1968; Frye et al. 1990; however, see also Gudziol et al., 2013), was not identified in this cohort.

Following rhinological examination including nasal endoscopy and a detailed structured history (Welge-Luessen et al. 2013) the subjects’ olfactory acuity was quantified by means of the clinically established “Sniffin’ Sticks” test battery (Burghart; Kobal et al. 1996; Hummel et al. 1997). This test is based on felt-tip pens that contain a solution of an odorant instead of liquid dye. The pen’s cap was removed by the experimenter for approximately 3s and the pen’s tip was placed 1–2 cm in front of the nostrils, in the case of triplet pen presentation at an interval of approximately 3s. Specifically, “odor thresholds” were obtained for the rose-like odor phenylethyl alcohol presented in 16 successive 1:2 dilution steps starting from a 4% solution in mineral oil and using a 3 alternatives forced choice (3-AFC) and a staircase paradigm, obtaining the odor threshold as the mean of the last 4 out of 7 staircase reversals (normal values > 6; Hummel et al. 2007). “Odor discrimination” was determined with 16 triplets of pens, 2 containing the same odorant and the third a different, “target” one (i.e., [target/nontarget] butanol/phenylethyl alcohol, isoamylacetate/anethole, anethole/eugenol, limonene/fenchone, (−) carvone/(+)-carvone, eugenol/cinnamon aldehyde, dihydrorosavoxenoxide/methyl, acetdehyde/isoamylacetate, citronellallinalool, pyridine/limonene, limonene/citronellall, eucalyptol/dipryridyl, dipryridyl/cyclopentadecanoate, butanol/fenchone, octylacetate/cinnamon aldehyde, carvone/acetdehyde). The discrimination performance was assessed employing a 3-AFC task (normal score: ≥11 correct discriminations).

“Odor identification” was determined with 16 odors (i.e., orange, leather, cinnamon, peppermint, banana, lemon, liquorice, turpen-tine, garlic, coffee, apple, clove, pineapple, rose, anise, and fish odor) using a 4-alternative forced-choice task with presentation of a list of 4 descriptors (Supplementary Table 1) for each pen (normal score: ≥12 correct identifications). Finally, the olfactory diagnosis was derived from the composite “Threshold Discrimination Identification score” (“TDI score”) that is the sum of the scores from the 3 subtests (Wolfensberger et al. 2000). Pathologic olfactory function is indicated by TDI ≤30.5, with the separation of hyposmia (30.5 ≥ TDI > 16.5) from functional anosmia at TDI ≤ 16.5 (Hummel et al. 2007).

Table 1. Olfactory diagnoses and likely underlying causes of olfactory disturbances as recorded for the subjects of the presently analyzed data set

<table>
<thead>
<tr>
<th>Etiology/diagnosis</th>
<th>Anosmia</th>
<th>Hyposmia</th>
<th>Normosmia</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiopathic</td>
<td>48</td>
<td>81</td>
<td>9</td>
<td>138</td>
</tr>
<tr>
<td>Sinonasal</td>
<td>41</td>
<td>38</td>
<td>4</td>
<td>83</td>
</tr>
<tr>
<td>Viral</td>
<td>18</td>
<td>63</td>
<td>12</td>
<td>93</td>
</tr>
<tr>
<td>Traumatic</td>
<td>66</td>
<td>47</td>
<td>6</td>
<td>119</td>
</tr>
<tr>
<td>Degenerative</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Congenital</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Healthy</td>
<td>1</td>
<td>11</td>
<td>137</td>
<td>149</td>
</tr>
<tr>
<td>Sum</td>
<td>183</td>
<td>251</td>
<td>179</td>
<td>613</td>
</tr>
</tbody>
</table>

Data analysis

Data was analyzed using the R software package (version 3.0.2 for Linux; http://CRAN.R-project.org/). Data was explored to find suitable odor identification items based on the observations in the study cohort. Ideally, such an item should (i) be correctly identified by a large fraction (percentage, i.e., %CorrectAnosmics) of subjects diagnosed with normosmia, however, (ii) it should not be correctly identified by a large fraction (percentage, i.e., %CorrectAnosmics) of subjects who were previously diagnosed with anosmia. Thus, the difference between condition (i) and condition (ii) should be small. This was numerically implemented as a so-called “odor specificity score,” OSS, which was calculated as “OSS = %CorrectNormosmics − %CorrectAnosmics.” This aimed at selecting odors that qualified for the intended test by their specificity to be recognized by nomosmics. To establish their suitability statistically, the number of correctly identified items by normosmics versus the number of correctly identified items by non-normosmic subjects was analyzed by means of χ² tests, which calculation of 95% confidence intervals of the test statistics in 1000 Bootstrap (Efron and Tibshirani 1993) resampling runs. The obtained candidate items (odors), that is, those with the largest OSS provided that the 95% confidence interval of the χ² test statistics did not include the critical value of 3.84 for a P-value of 0.05 for one degree of freedom in the χ² distribution, were stepwise analyzed with respect to the fraction of normosmic subjects who had correctly identified them. This was repeated, starting from the best scoring item, until all normosmics had been found, noting for each item the number of normosmic subjects additionally captured from those already captured by the items with the better OSS.

The list was used to obtain the optimum number of identification items required to establish a diagnosis of normosmia by means of a computed ABC analysis (Ultsch and Lotsch 2015), which is an inventory...
categorization technique originally developed for studies in economics (Juran 1975; Pareto 1909) to search for the minimum possible effort that gives the maximum yield, and was used presently as a selective inventory category technique that can be used to classify odors into separate categories. The analysis aims at dividing a set of positive data into three disjoint subsets of which subset “A” comprised the odors of interest representing those which, when correctly identified, capture most of the normosmic subjects. Its results are visualized in an ABC curve (Gastwirth and Glauberman 1976) representing the cumulative distribution function; calculations were done using our R package “ABCAnalysis” (http://cran.r-project.org/web/packages/ABCAnalysis/index.html).

Finally, the total number of odors to be included in the screening test was calculated based on probability considerations of acceptable normosmia diagnoses due to chance. The sensitivity and specificity of the test to detect normosmia were calculated using standard equations (Altman and Bland 1994b), that is, “sensitivity [%] = 100 true positives/(true positives + false negatives) and specificity [%] = 100 true negatives/(true negatives + false positives)” (Altman and Bland 1994a). Furthermore, the negative predictive value indicating how likely was that the subject is indeed not normosmic given a negative test result was calculated as “NPV [%] = 100 true negatives/(true negative + false negative)” (Altman and Bland 1994) and the test accuracy was obtained as “accuracy [%] = number of correct diagnoses/total number of diagnoses” (Metz 1978).

Results
Peppermint, clove, and fish odor had been identified by the largest number of normosmic subjects, in decreasing order (Figure 1). However, the first 2 odors were also among those relatively often recognized by subjects who were diagnosed as anosmic. This reduced their suitability for a short olfactory test. The highest OSS, indicating correct identification by a large fraction of normosmic but not of anosmic subjects, were reached by cinnamon, banana and fish odor. However, $\chi^2$ statistics on the raw frequencies of correctly identified items by normosmetics versus other olfactory diagnosis groups suggested cinnamon as the best suitable item outperforming all others in the diagnosis of normosmia versus non-normosmia and in particular of normosmia versus hyposmia where the Bootstrap confidence intervals of the $\chi^2$ values crossed the critical value of 3.84 for a $P$-value of 0.05 with orange, leather, turpentine, coffee, and rose (Figure 2).

Cinnamon, as the best scoring odor in the above analysis, had been correctly identified by 136 subjects (87.2%) with normal olfactory function. Of the further 2 odors, fish odor had been correctly identified by the remaining 23 normosmic subjects (12.8%). More positive associations of successful odor identification with the diagnosis of normosmia were not needed, which was substantiated in the results of the ABC analysis (Figure 3) that assigned only a single item to subset “A.” However, in a 4-AFC design, the correct identification of an odor without having an intact sense of smell (false positive) can occur at a chance of 25%. False positives were observed for cinnamon at a frequency of only 14% (Figure 1) whereas false negatives, that is, incorrect identifications by normosmetics or hyposmetics were observed at frequencies of 12.8% and 62.2%, respectively, although with hyposmics, nonidentification is not necessarily “false” negative. By theory, each odor added decreases this chance, calculated as Chance = $\frac{1}{\text{No. of alternatives}}$ resulting in 25%, 6.25%, 1.56%,
0.39%, and 0.097% for \( n = 1–5 \) items, respectively. Three items (cinnamon, fish odor, and banana) providing a chance of false correct diagnosis of normosmia of 1.56% were chosen as a convenient and traditionally favored item number. The test provided the diagnosis of normosmia, indicated by correct identifications of all 3 odors, at a sensitivity and specificity of 80.4% and 84.3%, respectively, a negative predictive value of 91.3% and a cross-validated (1000 Bootstrap runs) diagnostic accuracy for normosmia in the present data set of 83.2% (95% confidence interval 80.3–86.3%).

Discussion

The analysis emphasized that the diagnosis of olfactory function follows “the winner takes it all” principle, however, for a valid diagnosis of normosmia 3 identification items seem to suffice when accepting a chance of 1.56%. The selection of cinnamon as the primary odorant for this purpose owed to its comparatively best performance in providing the correct diagnosis in the present data set, however, alternatives might be possible as the score levels of further odors came close. A brute force approach on all possible combinations of 3 odors confirmed that cinnamon should definitely be included in the test suite, because all top ranked combinations contained this odor. Further suitable odors suggested by this analysis were fish odor and banana. Although their specificity was slightly exceeded by anise and pineapple (Figure 2) which provided a test alternative with a specificity of 91.5%, the respective replacement resulted in a low sensitivity of only 62%, supporting therefore the present selection.

Although cinnamon might not be the only choice for a short olfactory test, it emerged as the best odor distinguishing normosmia from reduced olfactory function as based on the OSS (Figure 2). Therefore, cinnamon is advised from the present analysis as best suited for a short olfactory test aimed at establishing normosmia. By contrast, some of the other odors were clearly disqualified, including peppermint that provided a low OSS because it was correctly identified by anosmic subjects at a fraction of 33%, which clearly exceeds the chance of 25%. This probably owes to a trigeminal component acting via activation of menthol and cold-activated transient receptor potential cation channels, subfamily M member 8 (TRPM8) (Peier et al. 2002; Bautista et al. 2007). Although cinnamon also possesses a trigeminal component, that reportedly excited TRPA1 and TRPM8 following topical application at a concentration of 10% (Namer et al. 2005), it has been correctly recognized by only 14.75% of the anosmic patients. This might owe to a concentration below a trigeminal threshold, that is, while the identification item in the “Sniffin Sticks” contains approximately 12% cinnamaldehyde, its volatile concentration emerging from the tip of the pen is probably only a small fraction of that and therefore, it most likely evokes only an olfactory sensation, which would agree with the observation that trigeminal stimuli must be at much higher threshold detection concentrations when compared with olfactory stimuli (Cometto-Muñiz and Cain 1990). Finally, the selection of cinnamon also agrees with the use of this odor in several international smell tests (Doty et al. 1996; Croy et al. 2014a). The other 2 odors, that is, fish odor and banana, are also commonly used odors in many different odor identification tests, either as distractors or target odors and

![Figure 2](http://chemse.oxfordjournals.org/)

Figure 2. Bar plots showing the \( \chi^2 \) values obtained from assessing the raw frequencies of correctly identified items by normosmics versus other olfactory diagnosis groups (A: versus anosmics, B: versus pathological olfactory function, C: versus hyposmia) were performed. The error bars indicate 95% confidence intervals obtained in 1000 Bootstrap (Efron and Tibshirani 1993) resampling runs. The horizontal dashed lines indicate the limit if statistical significance in the \( \chi^2 \) tests, that is, the critical value of 3.84 for a \( P \) value of 0.05 for one degree of freedom in the \( \chi^2 \) distribution.
one could assume that they are widely known all around the globe, although there is no proof yet of this assumption.

Although the first criterion may be claimed for the presently selected odors (table 1 in Hummel et al. 1997), a report that cinnamon and banana display age dependency of the identification (Konstantinidis et al. 2006) needs attention. OSS stratified for the age groups used in the “Sniffin’ Sticks” normative data (Hummel et al. 1997), a report that cinnamon and banana display age dependency of the identification was concluded from correlation coefficients of $r = -0.15$ for cinnamon and $r = -0.18$ for banana (Konstantinidis et al. 2006), the correlation may retrospectively be disputable (Dawson and Trapp 2004).

The present study compares favorably with alternative tests. Numerical information about test sensitivity and specificity has been published for the Connecticut Smell Test (CST), derived from the comprehensive Connecticut Chemosensory Clinical Research Center (CCCRC) test (Cain 1989), which was originally reported with a sensitivity and a specificity to detect olfactory loss of 88% and 77%, respectively (Toledano et al. 2005) and in a subsequent evaluation the respective values were 93.3% and 76% (Toledano et al. 2009). The numbers were obtained in only 40 patients who had an endoscopically proven grade II nasal polyposis, In a further small sample of 60 patients, the Pocket Smell Test (PST), derived from the comprehensive University of Pennsylvania Smell Identification Test (UPSIT; Doty et al. 1994), discriminated Alzheimer's dementia from other forms of dementia at a sensitivity of 100% and a specificity of 92.5% (Duff et al. 2002), or from major depression at a sensitivity of 95% and a specificity of 100% (McCaffrey et al. 2000). Both settings may have excluded smaller degrees of olfactory disturbance leading to an all-or-nothing setting where the tests performed very well. For further short tests sensitivity and specificity were not reported.

The nature and the minimum number of odor identification items, needed to establish normal olfactory function, was obtained using contemporary strategies that increasingly exploit computational aid in the analysis of complex biomedical data. The present test development strategy put an emphasis on the quick diagnosis of normosmia, and not on the detailed detection of decreased olfactory functions, based on the perception that the first clinical step is usually the establishment of normal olfactory function while for the diagnosis of an olfactory pathology the patient can expect to receive a more sophisticated assessment than a simple short olfactory test. Indeed, the workflow proposed 16 years ago for the “q-Sticks” test required “no further olfactory assessment” when normosmia was diagnosed from the correct identifications of all three odors; in all other cases “detailed olfactory testing” was advised (Nehlig 1999). A positive test result reliably establishes normosmia and provides a confidence basis to terminate olfactory assessments following the application of only 3 odor identification items, which provided a more than 90% confidence (NPV) that a negative test result truly indicates a reduced olfactory function.

**Supplementary material**

Supplementary material can be found at [http://www.chemse.oxfordjournals.org/](http://www.chemse.oxfordjournals.org/)

**Funding**

Data analytics work has been supported by the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für Translationale Medizin und Pharmakologie (J.L.). T.H. was funded by a grant from the Deutsche Forschungsgemeinschaft (DFG HU441/18-1). The funders had no role in method design, data selection and analysis, decision to publish, or preparation of the manuscript.

**Conflict of interest statement**

The authors have declared that no competing interests exist.

**References**


