Chemical Senses, 2019, Vol 44, 357–364
doi:10.1093/chemse/bjz029
Original Article
Advance Access publication 11 May 2019

Original Article

Clinical Usefulness of Self-Rated Olfactory Performance—A Data Science-Based Assessment of 6000 Patients

Jörn Lötsch[1,2] **and Thomas Hummel**[3]

[1]Institute of Clinical Pharmacology, Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany, [2]Fraunhofer Institute of Molecular Biology and Applied Ecology—Project Group Translational Medicine and Pharmacology (IME-TMP), Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany and [3]Smell and Taste Clinic, Department of Otorhinolaryngology, TU Dresden, Fetscherstrasse 74, 01307 Dresden, Germany

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

Editorial Decision 6 May 2019.

Abstract

In clinical practice, with its time constraints, a frequent conclusion is that asking about the ability to smell may suffice to detect olfactory problems. To address this question systematically, 6049 subjects were asked about how well they can perceive odors, with 5 possible responses. Participants presented at a University Department of Otorhinolaryngology, where olfactory testing was part of the routine investigation performed in patients receiving surgery at the clinic (for various reasons). According to an odor identification test, 1227 subjects had functional anosmia and 3113 were labeled with normosmia. Measures of laboratory test performance were used to assess the success of self-estimates to capture the olfactory diagnosis. Ratings of the olfactory function as absent or impaired provided the diagnosis of anosmia at a balanced accuracy of 79%, whereas ratings of good or excellent indicated normosmia at a balanced accuracy of 64.6%. The number of incorrect judgments of anosmia increased with age, whereas false negative self-estimates of normosmia became rarer with increasing age. The subject's sex was irrelevant in this context. Thus, when asking the question "How well can you smell odors?" and querying standardized responses, fairly accurate information can be obtained about whether or not the subject can smell. However, this has to be completed with the almost 30% (355 subjects) of anosmic patients who judged their ability to smell as at least "average." Thus, olfactory testing using reliable and validated tests appears indispensable.

Key words: anosmia, data science, diagnostic test performance, odor identification, olfactory testing

Introduction

Olfactory testing has entered clinical practice more broadly during the last 2 decades. However, it is achieved with costly tests that require considerable time for their application. This triggers an interest in the patients' self-rating of olfactory function. It is undoubtedly important at the stage where patients decide to consult medical professionals. Hence, a perception of researchers and physicians involved in olfactory testing is that most participants in their studies/patients are aware of their ability to smell. In clinical practice, with its time constraints, a frequent conclusion is that asking about the ability to smell may suffice to detect olfactory problems. This raises the more scientific question whether this is true and an olfactory test could be replaced by a simple question.

Indeed, self-ratings were reported to correlate with measured olfactory function to a statistically significant degree (Welge-Luessen et al. 2005; Nguyen et al. 2012). However, a clinical usefulness of olfactory self-estimates has often been doubted. Positive reports were contrasted with reports about a lack of correlation between self-ratings of olfactory abilities and measured olfactory function (Landis et al. 2003). Moreover, repeated reports point at a considerable prevalence of unawareness of absent olfactory function and seems to become more prevalent with increasing age (Nordin et al. 1995; Shu et al. 2009; Oleszkiewicz *et al.* 2018). Specifically, although anosmia is frequent with an estimated prevalence in the average population of 10–20% (Dong et al. 2017; Hummel et al. 2017), asking about the function of the sense of smell indicated that the awareness of olfactory loss varies. It can be completely absent as in a study where 0.45% of the participants were anosmic but all had maintained that their sense of smell is intact (Oleszkiewicz et al. 2018) or even turn toward exaggeration as in a study where 9% of the participants had rated their sense of smell to be impaired, although only 4% scored in the dysfunctional range when tested with the Sniffin' Sticks battery (Oleszkiewicz et al. 2019).

In the present study, a large cohort of subjects with a high prevalence of anosmia was analyzed. Subjects had been asked about the function of their sense of smell, to which 5 different standardized responses were possible, and subsequently they were tested with an established olfactory test (Hummel et al. 2001). The success of self-estimates in establishing the olfactory diagnosis was analyzed using classical performance measures of laboratory tests, which was a focus on the optimum response scenario that most accurately agreed with the clinical diagnosis.

Materials and Methods

Subjects and olfactory testing

The study followed the Declaration of Helsinki and was approved by the Ethics Committee of the Faculty of Medicine of the TU Dresden (number EK251112006) covering anonymized retrospective and pooled analyses. Informed written consent was obtained from all subjects or their caretakers. Subjects (age: range 3–105 years, mean ± standard deviation: 45 ± 17 years; sex: 3813 men, 2236 women) had presented at the Department of Otorhinolaryngology, Dept. of ORL, TU Dresden, Germany. Olfactory testing was part of the routine investigation of the sense of smell which was performed in patients receiving surgery at the clinic (for various reasons). Subjects were also asked for the possible etiology associated with their perception of a reduced olfactory performance that caused them to seek medical help.

Subjects were asked to estimate their ability to smell odors at a 5-point rating scale. Specifically, they were asked the question of "How well can you smell?" and had to choose one of the possible answers among "not at all," "impaired," "average," "good," and "very good." Subsequently, olfactory function was assessed using the 12-item odor identification (Hummel et al. 2001) test battery (Burghart). The test is composed of felt-tip pens filled with solutions of odors. For olfactory stimulation, the pens are placed, with the cap removed, for approximately 3 s at 1–2 cm beneath each nostril. The test set comprises 12 pens containing different odors, which for assessment of the subject's odor identification performance had to be recognized in a 4-alternative forced-choice task with presentation of a list of 4 possible descriptors for each pen.

The test was started with a randomly chosen nostril and at an interval of approximately 20 s, it was repeated for the contralateral nostril. Presentation of odors was guided by a specially designed software (Hummel et al. 2012) reducing possible errors during the testing procedure. For the purpose of this study, the so-called "best-performing nostril" approach was used (Betchen and Doty 1998; Frasnelli et al. 2002). Lack of olfactory function, that is, functional anosmia (further termed "anosmia") was indicated when only 0–6 odors were correctly recognized. Reduced but still present olfactory function, that is, hyposmia, was assumed when 7–9 odors were correctly identified, and from 10 correctly identified odors, normal olfactory function, that is, normosmia, was assumed.

Data analysis

Data were analyzed using the R software package (version 3.5.2 for Linux; http://CRAN.R-project.org/; R Development Core Team 2008) on an Intel Core i7 - 7500U notebook computer running on Ubuntu Linux 18.04.1 64 bit. Self-estimates of the olfactory function were transformed into an ordinal scale with numbers from 0 to 4 (question: "How well can you smell?," answer: 0 = "not at all," 1 = "impaired," 2 = "average," 3 = "good," 4 = "very good"). Subsequent analyses assessed how this self-estimated olfactory performance score related to the result of the odor identification test, expressed as the maximum number of correctly identified odors across either nostril.

First, the observed identification scores were compared between the self-estimated scores by means of χ^2 tests (Pearson 1900) for the number of subjects observed with particular identification scores and Kruskal–Wallis tests (Kruskal and Wallis 1952) followed by post hoc Dunn's tests (Dunn 1961) for the identification scores. The alpha level was set at 0.05 and corrected for multiple testing according to Bonferroni (Bonferroni 1936). The latter calculations were performed using the R library "dunn.test" (https://CRAN.R-project.org/package=dunn.test; Dinno 2017). Second, to test how well self-estimation predicted the overall olfactory diagnosis, standard measures of test performance (Altman and Bland 1994a) were calculated. Specifically, test sensitivity was calculated as "sensitivity = true positives/(true positives + false negatives)" and test specificity was calculated as "specificity = true negatives/(true negatives + false positives)" (Altman and Bland 1994b), For the olfactory diagnosis anosmia or normosmia, all possible cutoffs of self-estimates (i.e., 0, ≤1, ≤2, ≤3, and ≤4 or ≥0, ≥1, ≥2, ≥3, and 4) were assessed in an exhaustive search for the optimum cutoff for the olfactory diagnosis set at the maximum of the product of sensitivity and specificity. The balanced accuracy was calculated as the mean of sensitivity and specificity (Brodersen et al. 2010). Furthermore, Cohen's κ (Cohen 1960) was calculated as previously applied to judge the performance of a proposed short olfactory test (Hummel et al. 2010). Third, using the cutoffs identified in the second analytical step, correct and incorrect self-estimates of anosmia or normosmia were defined and their relation with age or sex was explored using correlation analyses (Spearman's ρ (Spearman 1904)) or χ^2 statistics (Pearson 1900), respectively. Fourth, to test how self-estimation, scaled [0,…,4], predicted the odor identification scores, scaled [0,…,12], the positive predictive values (PPVs) were calculated as "PPV [%] = 100 · true positive/(true negative + false negative)" for each self-estimate score versus the identification test score.

Results

Among the n = 6050 subjects, according to the odor identification test limits n = 1227 were diagnosed with anosmia (20.3%), whereas among the n = 4823 subjects with present olfactory function, n = 3113 could be labeled as normosmia (51.45%). The distribution of subjects who rated their ability whether they can smell an odor as "not at all," "impaired," "average," "good," or "very good" was n = 526, 2318, 1781, 974, and 451 cases, respectively. The odor identification scores acquired with the olfactory test ranged between 0 and 12 correctly recognized odors (Figure 1).

General associations between self-estimates and olfactory test results

Subjects who had rated their olfactory performance better had more often also higher olfactory identification scores (χ^2 = 2919.4, degree of freedom [df] = 48, $P < 2.2 \cdot 10^{-16}$). This corresponded to the significant differences in the odor identification scores among groups of subjects with a similar self-estimation of their olfactory performance (Kruskal–Wallis χ^2 = 1755.82, df = 4, $P < 2.2 \cdot 10^{-16}$). Post hoc Dunn's tests indicated that all possible paired differences were statistically significant (Table 1).

Performance of self-ratings in establishing an olfactory diagnosis

A total of n = 480 subjects (39.1% of all anosmic subjects) correctly judged their olfactory function as absent, and when adding the n = 392 subjects who estimated their sense of smell as "impaired," 71.1% of anosmic patients seemed to be aware that their sense of smell is not functioning. Indeed, setting a cutoff at the answer of not better than "impaired" provided the best overall performance (Figure 2A), detecting anosmia at sensitivity and specificity of 71.1% and 87%, respectively, and at a balanced accuracy of 79%. The value of Cohen's κ was 0.54. However, 355 patients diagnosed with anosmia (28.9%) had estimated their ability to smell odors as "average" or better. Identifying normosmia was less successful, that is, n = 337 subjects (10.1% of all normosmic subjects) correctly judged their olfactory function as "very good," and when adding

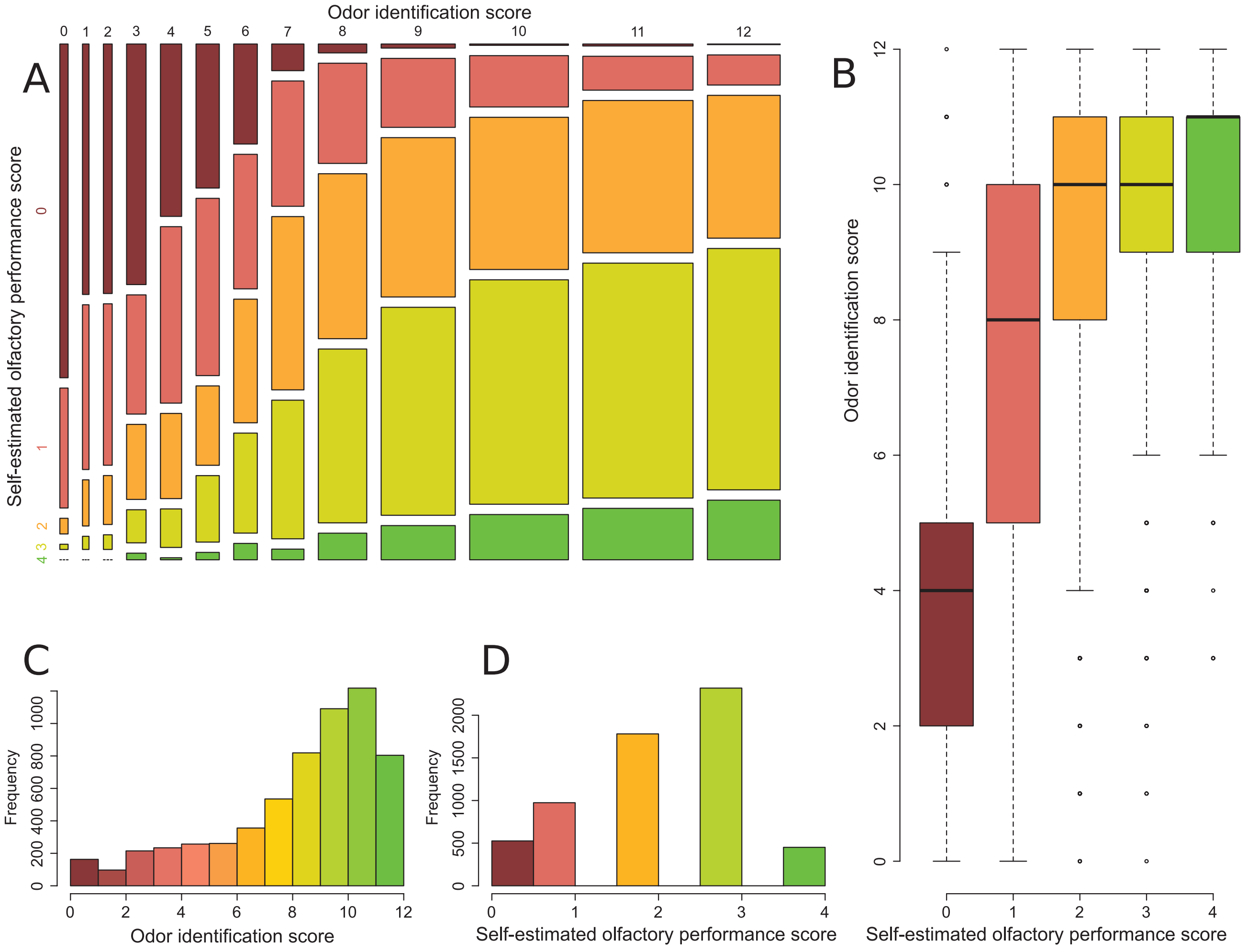
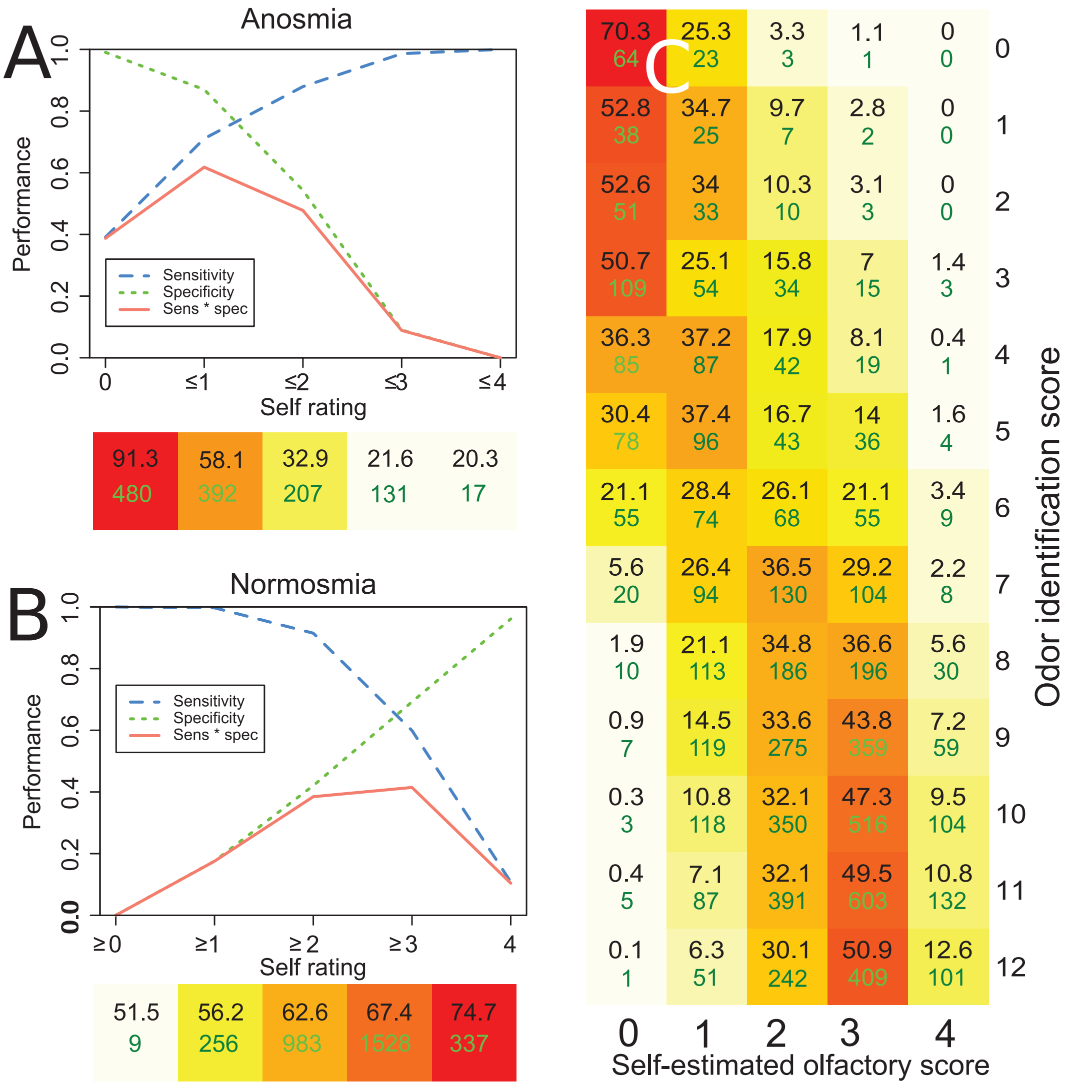


Figure 1. Distribution and frequency of the subjects' self-estimated score of their olfactory acuity for the odor identifications score or derived diagnosis obtained with a standard olfactory test. (**A**) Mosaic plot representing a contingency table of the self-ratings (question: "How well can you smell?," answer: 0 = "not at all," 1 = "impaired," 2 = "average," 3 = "good," 4 = "very good") versus odor identification test results. The size of the cells as proportional to the number of subjects included. (B) Box and whisker plots of the odor identification test score results, separately for each group of subjects sharing the same self-estimate of the olfactory acuity. They have been constructed using the minimum, quartiles, median (solid black red line within the box), and the maximum values. The whiskers add 1.5 times the interquartile range (IQR) to the 75th percentile or subtract 1.5 times the IQR from the 25th percentile and are expected to include 99.3% of the data if normally distributed. (**C**) Histogram showing the distribution (count) of the odor identification test scores. (**D**) Histogram showing the distribution (count) of the self-estimates of the subjects' olfactory performance. The plot has been created using the R software package (version 3.5.2 for Linux; http://CRAN.R-project.org/; R Development Core Team 2008).

Table 1. The observed identification scores were compared between the self-estimated scores by means of χ^2 tests (Pearson 1900) for the number of subjects observed with particular identification scores and Kruskal–Wallis tests (Kruskal and Wallis 1952) followed by post hoc Dunn's tests (Dunn 1961) for the identification scores

Self-estimate	"not at all"	"impaired"	"average"	"good"
"impaired"	−14.95($P < 0.0001$)			
"average"	−30.28($P < 0.0001$)	−17.4($P < 0.0001$)		
"good"	−35.84($P < 0.0001$)	−24.14($P < 0.0001$)	−7.25($P < 0.0001$)	

The alpha level was set at 0.05 and corrected for multiple testing according to Bonferroni (Bonferroni 1936). The latter calculations were performed using the R library. Olfactory functions were transformed into numbers from 0 to 4 (question: "How well can you smell?," answer: 0 = "not at all," 1 = "impaired," 2 = "average," 3 = "good," 4 = "very good"). Subsequent analyses assessed how this self-estimated olfactory performance score related to the result of the odor identification.



91.3 480	58.1 392	32.9 207	21.6 131	20.3 17

51.5 9	56.2 256	62.6 983	67.4 1528	74.7 337

0	1	2	3	4	Odor identification score
70.3 64	25.3 23	3.3 3	1.1 1	0 0	0
52.8 38	34.7 25	9.7 7	2.8 2	0 0	1
52.6 51	34 33	10.3 10	3.1 3	0 0	2
50.7 109	25.1 54	15.8 34	7 15	1.4 3	3
36.3 85	37.2 87	17.9 42	8.1 19	0.4 1	4
30.4 78	37.4 96	16.7 43	14 36	1.6 4	5
21.1 55	28.4 74	26.1 68	21.1 55	3.4 9	6
5.6 20	26.4 94	36.5 130	29.2 104	2.2 8	7
1.9 10	21.1 113	34.8 186	36.6 196	5.6 30	8
0.9 7	14.5 119	33.6 275	43.8 359	7.2 59	9
0.3 3	10.8 118	32.1 350	47.3 516	9.5 104	10
0.4 5	7.1 87	32.1 391	49.5 603	10.8 132	11
0.1 1	6.3 51	30.1 242	50.9 409	12.6 101	12

Figure 2. Matrix heat plots of the PPVs provided by the subjects' self-estimated score of their olfactory acuity (question: "How well can you smell?," answer: 0 = "not at all," 1 = "impaired," 2 = "average," 3 = "good," 4 = "very good") for the odor identifications score or derived diagnosis obtained with a standard olfactory test. The cells are equally sized and display the PPV in black numbers. Below the values of the PPV in black numbers, the number of subjects associated with each cell is given in green numbers. The color code ranges from light yellow indicating low PPV to red indicating high PPV. (**A**) PPV of the self-estimated scores for the olfactory diagnosis of functional anosmia (top), present olfactory function (middle), or normosmia (bottom). (**B**) Performance of different iterations of the self-estimation of olfactory function for the olfactory diagnosis. The criterion was the product of sensitivity and specificity to detect anosmia or normosmia. (**C**) Contingency table of the self-estimated olfactory score versus the olfactory identification test score. The PPV has been calculated for an odor identification test score smaller than or equal to the number given at the right of each line, that is, an ascending sequence from 0 to 12. The plot has been drawn using the "heatmap.2" function of the R library "gplots" (Warnes G. R.; https://cran.r-project.org/web/packages/gplots/index.html) and the R software package (version 3.5.2 for Linux; http://CRAN.R-project.org/; R Development Core Team 2008).

the $n = 1528$ subjects who estimated their sense of smell as "good," 59.9% of normosmic patients had indicated satisfaction with their sense of smell (Figure 2B). Here, the best sensitivity and specificity of 59.9% and 69.2% were obtained when regarding responses of at least "good" ability to smell as indication of normosmia, which was then correctly estimated at a balanced accuracy of only 64.6%. The value of Cohen's κ was 0.29. Only $n = 9$ subjects who had normal olfactory function had indicated that they cannot smell at all.

Age and sex effects on the correctness of olfactory self-estimates

Using a cutoff of at most "impaired" olfactory function for assignment of anosmia and of at least "good" olfactory function for the assignment of normosmia and analyzing how many ratings, in percent and per year of age, had been incorrect, regardless of false positive or false negative olfactory diagnoses, the fraction of false diagnoses of anosmia increased with age (Figure 3A; Spearman's $\rho = 0.48$, $P = 1.31 \cdot 10^{-6}$). This was also observed when analyzing only false negative self-estimates (Figure 3C; $\rho = 0.47$, $P = 2.69 \cdot 10^{-6}$). With normosmia, the effect of age was oppositely directed. Although false self-estimates of normal olfactory function were not significantly correlated with age (Figure 3B), false negative estimates significantly decreased with increasing age (Figure 3D; $\rho = 0.57$, $P = 3.35 \cdot 10^{-9}$). The subject's sex had no effect on the error rate of the self-estimates of poor olfactory function (χ^2 test: sex versus false anosmia: $\chi^2 = 0.76$, df = 1, $P = 0.3827$), whereas men were more often incorrect when it came to a judgment of good olfactory function (χ^2 test: sex versus false normosmia: $\chi^2 = 5.2305$, df = 1, $P = 0.02219$).

Ability to estimate the olfactory test score

When aiming at the olfactory diagnosis rather than at the odor identification test score, the cutoff estimated in earlier analytical steps provided a PPV of 58.1% for anosmia, that is, when the subject had estimated the performance to smell not better than "impaired" (Figure 2C). For identifying normosmia with self-estimates that the olfactory function was at least "average," the PPV was 67.4% (Figure 2A). Better and high self-estimates provided higher PPVs for better test results (Figure 2B). Of note, subjects who had rated their olfactory acuity as "very good" seemed to be less successful in judging their sense of smell as the PPV dropped from those found when the olfactory performance had been rated only as "good."

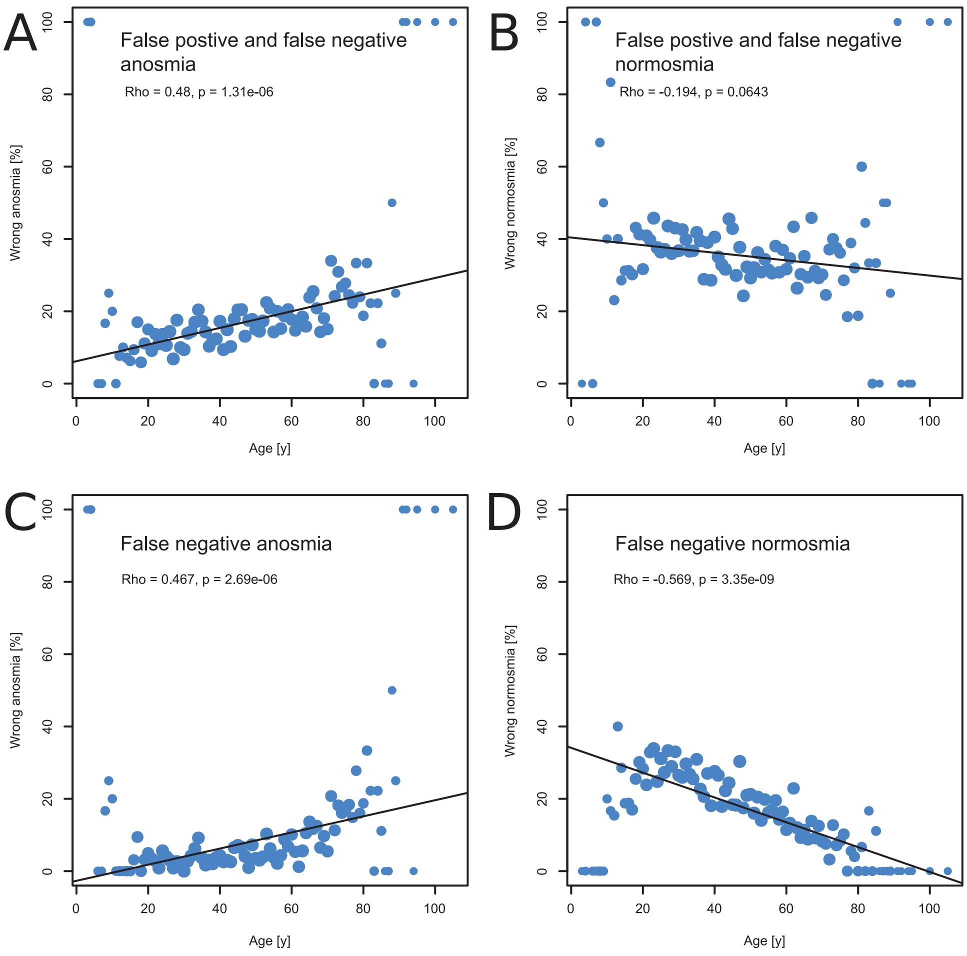


Figure 3. Relationship of the incorrect self-estimates of olfactory function with the subject's age. The dotplots show the percentages of incorrect self-estimates for each year of age. The upper 2 panels show the false positive or false negative self-estimates of (**A**) anosmia and (**B**) normosmia. The lower 2 panels show false negative self-estimates of (**C**) anosmia and (**D**) normosmia. The size of the dots is proportional to the number of subjects at the respective age. The lines have been obtained using robust linear regression and are aimed at visual enhancement of the age tendency of the false self-estimates of olfactory function. The plot has been created using the R software package (version 3.5.2 for Linux; http://CRAN.R-project.org/; R Development Core Team 2008) and the library "MASS" (https://cran.r-project.org/package=MASS; Venables and Ripley 2002).

Discussion

When asking the question "How well can you smell odors?" and querying standardized responses such as the presently proposed 5 possible estimates, roughly 80% accurate information (balanced accuracy) about whether or not the subject can smell at all can be obtained. Moreover, absence (anosmia) or presence of olfactory function can be detected at positive predictive values of >58%. However, this has to be completed with the finding that almost 30% (n = 355) of the n = 1227 anosmic patients had judged their ability to smell as at least "average." Thus, although the idea to ask the patient about olfactory function seems attractive as promising a diagnosis without applying a cost- and time-intensive special test, it can at best provide a rough diagnosis on which cannot be relied. Moreover, for scientific purposes, estimating the olfactory acuity provides imprecise information about the sense of smell. Thus, although the figures apparently supported the usefulness of asking about olfaction, as suggested by the good accuracy and high PPV, the considerable fraction of patients who were completely unaware of the absence of their olfactory function disqualifies asking about olfaction as a clinical alternative to testing it.

The unawareness of absent sense of smell has been discussed previously. For example, a study in 83 subjects suggested that there is no significant correlation between self-ratings of olfactory abilities and measured olfactory function (Landis et al. 2003). In fact, self-ratings of olfactory function were shown to be strongly confounded by nasal patency. In a study on 9139 healthy subjects (Oleszkiewicz et al. 2018), 0.45% of the subjects tested anosmic, although they maintained that their sense of smell was normal. This phenomenon was most prevalent from age 70 years upwards. In a similar vein of thoughts, it was found that in a group of 162 healthy young volunteers (age 18–35 years), 9% rated their sense of smell to be impaired, although only 4% scored in the dysfunctional range when tested with the Sniffin' Sticks battery; in the group of 435 people older than 55 years, 12% stated that their sense of smell was dysfunctional, although 36% scored in the hyposmic or anosmic range (Shu et al. 2009). Similar findings of a dissociation between rated and measured olfactory function have been observed previously (Nordin et al. 1995) when 77% of older subjects with smell loss reported normal smell sensitivity. When looking at larger epidemiological studies, this is also reflected in the generally lower prevalence of self-reported olfactory dysfunction (Hoffman et al. 1998; Murphy et al. 2002; Nordin et al. 2004) compared with that of measured olfactory dysfunction (Murphy et al. 2002; Vennemann et al. 2008).

This picture is different when it is about patients with smell loss. For example, a moderate correlation (r = 0.57) between self-rated and measured olfactory function was reported in 152 patients who recognized their olfactory deficit (Welge-Luessen et al. 2005); compare (Haxel et al. 2012). Still, in this study, 10 of 78 anosmic patients (12%) reported the presence of olfactory function. The frequency of unawareness of olfactory dysfunction was only 8% in sinusitis patients, much lower than the 77% in normal elderly subjects (Nordin et al. 1995). A different example comes from patients with idiopathic Parkinson's disease, more than 95% of whom are hyposmic at onset of the motor symptoms (Haehner et al. 2009). In this group, 14 of 37 hyposmic patients (38%) reported a normal sense of smell (Müller et al. 2002).

The present diagnosis of normosmia relied on the 12-item identification test (Hummel et al. 2001). Although this is an established olfactory test developed to simplify olfactory testing in clinical practice, it has been repeatedly discussed that a more complete olfactory test covering the olfactory threshold and the ability to discriminate odors provides additional information about a subject's sense of smell that improves the diagnosis of olfactory function (Lötsch et al. 2008; Lötsch and Hummel 2019). Therefore, basing the olfactory diagnosis only on identification could have caused false negative results, that is, some of the n = 355 subjects seemingly unaware that they cannot smell might have had at least residual olfactory function that could have been detected with a more comprehensive test battery. To explore this possibility, data of 10 713 subjects available from a recent report (Lötsch and Hummel 2019) were checked for the number of false diagnoses of anosmia when basing the diagnosis only on odor identification as compared with the diagnoses based on odor identification, discrimination, and olfactory threshold tests. Specifically, in the more comprehensive test, the olfactory diagnosis is made from sum scores of threshold, discrimination, and identification subtests, with values ≤16 indicating anosmia, values ≥30.5 in males and ≥29.5 in females indicating normosmia (Hummel et al. 2007). The limit in the performance in the 16-item odor identification test which separates hyposmia from anosmia was reported at a score of 8 (Kobal et al. 2000). In the reported cohort (Lötsch and Hummel 2019), n = 3662 subjects had the olfactory diagnosis of anosmia, which corresponds to 34.3% of the study cohort. However, using odor identification alone produced only 62 false negative diagnoses of anosmia but 2077 false positive diagnoses and found a total of n = 5677 subjects as being anosmic corresponding to 53% of the cohort. Thus, that cohort odor identification alone overestimated the number of anosmic subjects by 18.8%. Taking this back to the present, n = 355 subjects who indicated that they can smell but were tested as anosmic, 67 of them might indeed have been able to smell, responded to the question about how well they can smell odors with a positive answer but performed poorly in the odor identification test. Still, following this hypothetical correction present results would indicate that 24.8% of subjects with anosmia are unaware of it and provide an incorrect self-estimate of their olfactory function.

A further possible limitation of the present method to establish the olfactory diagnosis is the use of the 12-item short test (Hummel et al. 2001) in contrast to the 16-item odor identification subtest included in the Sniffn' Sticks test battery (Oleszkiewicz et al. 2018). A substantial impact of the reduction of the number of identification items from d = 16 to d = 12 odors would contradict the positive judgment of the test reported to differentiate anosmics, hyposmics, and normosmics at $P < 0.001$ (Hummel et al. 2001). To assess whether indeed the reduction of the number of odors to be identified compromises the detection of anosmic subjects, a further data set, comprising n = 4118 subjects, in whom the complete Sniffn' Sticks test battery, including the 16-item odor identification test, had been performed was analyzed. According to the complete threshold discrimination identification score, n = 1667 patients had anosmia, whereas n = 770 subjects had normal olfactory function. When using the sum of the 16 identification test items and the limit of d = 8 correctly identified odors for the diagnosis of anosmia, the 16-item test provided a balanced accuracy of 72.3% to detect anosmia, with a value of Cohen's k of 0.39. Surprisingly, when using only the sum of the 12 odors also included in the 12-item identification test (Hummel et al. 2001) for the diagnosis, a balanced accuracy of 83.4% and a value of Cohen's k of 0.63 were obtained for the detection of anomia. Thus, in contrast to the use of only an odor identification test rather than the more complete test battery of olfactory threshold, odor discrimination, and identification, which reduced the accuracy of the olfactory diagnosis, the use of d = 12 instead of d = 16 items seems unlikely to have introduced a significant fraction of error in the present comparison of self-estimates of olfactory function with olfactory test results. The reason why the 12-item test performed better than the 16-item test is not a subject of the present analysis; it may point at the necessity of a revision of diagnostic limits in the tests. Although the 16-item test is part of the more comprehensive Sniffn' Stick test battery, which bases the diagnosis of the sum of the performance in 3 subtests, the 12-item test had been optimized for providing a diagnosis based on odor identification. For the present analysis, this means that the preference for the stand-alone 12-item test to the use of the 16-item subtest of a more complete battery proved suitable. Of course, when judging a positive self-estimate of the ability to smell, the olfactory function should be assessed with the most sensitive test to be sure of a negative test result contradicting the self-estimate. Therefore, the use of a rather quick test in the present study may occasionally have underestimated a subject's ability to smell. The numerical consequences have been elaborated above; however, it may be worthwhile to note this again as a limitation of a generalization of the present results.

Finally, although when compared with the 12-item olfactory test, the self-estimates of olfactory function seemed to be imprecise, when compared with an earlier proposed short olfactory test based on the identification of 3 odors (Hummel et al. 2010), the poor judgment may partly be revised. Specifically, in a 3-item test using odors "cloves," "coffee," and "rose," anosmia had been identified by 0 correct identifications at a test sensitivity and specificity of 66% and 96%, respectively, with Cohen's κ = 0.655. When using a score of 1 correct identification, sensitivity and specificity to detect anosmia had been reported to be of 84% and 78%, respectively. For comparison, the present analysis indicated sensitivity and specificity of 71.1% and 87%, respectively, to detect anosmia *via* self-estimate with Cohen's κ = 0.54. Moreover, one has to keep in mind that in case of rare instances, the so-called zeroR assignment works apparently well. For example, anosmia has been reported with a prevalence of 22.3% among US black older adults and of 10.4% among US white older adults (Dong et al. 2017). When encoding anosmia with "1" and preserved olfactory function with "0," and just assigning "0" to every case, that is, zeroR or "zero rules," accuracies of 76.8% or 88.9% to "diagnose" anosmia in blacks or whites, respectively, are obtained. Balanced accuracy would be more adequate as used in the present analysis; however, accuracy seems more intuitive, which provides a possible explanation of the clinical perception suggesting that just asking about olfactory function suffices in most cases. Of note, in the present cohort, zeroR for anosmia would provide an accuracy of 79%. This apparent good result emphasizes that the correctness of a diagnosis needs to be judged more carefully rather than relied upon perceptions. It has also to be noted that parts of the present analysis used techniques of machine learning, such as the identification of the cutoffs for olfactory diagnosis, which was performed analogous to the creation of a rule-based classifier. However, as this analysis was exploratory and descriptive, stricter measures against overfitting as described elsewhere (Lötsch and Hummel 2019) had not been implemented.

Some recent publications about the same topic draw more positive conclusions about the accuracy of olfactory self-ratings. For example, pursuing the question about whether conversion to dementia can be predicted by self-reported olfactory impairment in 1529 participants with initially normal cognitive function and followed-up for 10 years (Stanciu et al. 2014), ratings of olfactory sensitivity as "worse than normal" were associated with conversion to dementia. This is not contradicted by the present results that do not imply that self-ratings of olfactory function may not be useful in other context than the replacement of careful olfactory testing. From the same cohort, a further analysis again supported that self-ratings provide additional information to olfactory testing to predict future morbidity of dementia (Ekstrom et al. 2017), which was emphasized in a further analysis of that cohort (Ekstrom et al. 2019). Although these results take a positive view on a utility of self-estimates of olfactory function in addition to olfactory testing, they do not address explicitly the agreement between both, which is the aim of the present report, and that in turn is not intended to generally dispute a utility of asking a subject about the olfactory function.

In conclusion, a major portion of patients reports their olfactory function correctly. However, because on one end of the spectrum at least one-quarter of patients with anosmia seems to be unaware of it and on the other end of the spectrum more than one-third of patients indicating olfactory loss score in the normosmic range, olfactory testing using reliable and validated tests is indispensable. This applies to more comprehensive olfactory tests, whereas a discussion of very short tests of olfactory function suggested that their utility may become subject to future reassessments when directly compared with self-estimates rather than with the full olfactory tests.

Funding

This work was supported by the Landesoffensive zur Entwicklung wissenschaftlich - ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für Translationale Medizin und Pharmakologie (J.L.), and the Deutsche Forschungsgemeinschaft [DFG HU 441/18-1 to T.H. and DFG Lo 612/10–1 to J.L.).

Acknowledgements

We would like to thank the Deutsche Forschungsgemeinschaft for support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest statement

The authors have declared that no competing interests exist.

References

Altman DG, Bland JM. 1994a. Diagnostic tests 2: predictive values. *BMJ*. 309:102.

Altman DG, Bland JM. 1994b. Diagnostic tests. 1: sensitivity and specificity. *BMJ*. 308:1552.

Betchen SA, Doty RL. 1998. Bilateral detection thresholds in dextrals and sinistrals reflect the more sensitive side of the nose, which is not lateralized. *Chem Senses*. 23:453–457.

Bonferroni CE. 1936. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*. 8:3–62.

Brodersen KH, Ong CS, Stephan KE, Buhmann JM. 2010. The balanced accuracy and its posterior distribution. Istanbul, Turkey: 20th International Conference on Pattern Recognition (ICPR), 2010. p. 3121–3124.

Cohen J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas*. 20:37–46.

Dinno A. 2017. *dunn.test: Dunn's test of multiple comparisons using rank sums*. Available from: https://cran.r-project.org/web/packages/dunn.test/index.html

Dong J, Pinto JM, Guo X, Alonso A, Tranah G, Cauley JA, Garcia M, Satterfield S, Huang X, Harris T, et al. 2017. The prevalence of anosmia and associated factors among U.S. black and white older adults. *J Gerontol A Biol Sci Med Sci*. 72:1080–1086.

Dunn OJ. 1961. Multiple comparisons among means. *J Am Stat Assoc*. 56:52–64.

Ekström I, Josefsson M, Larsson M, Rönnlund M, Nordin S, Olofsson JK. 2019. Subjective olfactory loss in older adults concurs with long-term odor identification decline. *Chem Senses*. 44:105–112.

Ekström I, Sjölund S, Nordin S, Nordin Adolfsson A, Adolfsson R, Nilsson LG, Larsson M, Olofsson JK. 2017. Smell loss predicts mortality risk regardless of dementia conversion. *J Am Geriatr Soc*. 65:1238–1243.

Frasnelli J, Livermore A, Soiffer A, Hummel T. 2002. Comparison of lateralized and binasal olfactory thresholds. *Rhinology*. 40:129–134.

Haehner A, Boesveldt S, Berendse HW, Mackay-Sim A, Fleischmann J, Silburn PA, Johnston AN, Mellick GD, Herting B, Reichmann H, et al. 2009. Prevalence of smell loss in Parkinson's disease–a multicenter study. *Parkinsonism Relat Disord*. 15:490–494.

Haxel BR, Bertz-Duffy S, Fruth K, Letzel S, Mann WJ, Muttray A. 2012. Comparison of subjective olfaction ratings in patients with and without olfactory disorders. *J Laryngol Otol*. 126:692–697.

Hoffman HJ, Ishii EK, MacTurk RH. 1998. Age-related changes in the prevalence of smell/taste problems among the United States adult population. Results of the 1994 disability supplement to the National Health Interview Survey (NHIS). *Ann N Y Acad Sci*. 855:716–722.

Hummel T, Kobal G, Gudziol H, Mackay-Sim A. 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.

Hummel T, Konnerth CG, Rosenheim K, Kobal G. 2001. Screening of olfactory function with a four-minute odor identification test: reliability, normative data, and investigations in patients with olfactory loss. *Ann Otol Rhinol Laryngol*. 110:976–981.

Hummel T, Pfetzing U, Lötsch J. 2010. A short olfactory test based on the identification of three odors. *J Neurol*. 257:1316–1321.

Hummel C, Zucco GM, Iannilli E, Maboshe W, Landis BN, Hummel T. 2012. OLAF: standardization of international olfactory tests. *Eur Arch Otorhinolaryngol*. 269:871–880.

Hummel T, Whitcroft KL, Andrews P, Altundag A, Cinghi C, Costanzo RM, Damm M, Frasnelli J, Gudziol H, Gupta N, et al. 2017. Position paper on olfactory dysfunction. *Rhinol Suppl*. 54:1–30.

Kobal G, Klimek L, Wolfensberger M, Gudziol H, Temmel A, Owen CM, Seeber H, Pauli E, Hummel T. 2000. Multicenter investigation of 1,036 subjects using a standardized method for the assessment of olfactory function combining tests of odor identification, odor discrimination, and olfactory thresholds. *Eur Arch Otorhinolaryngol*. 257:205–211.

Kruskal WH, Wallis WA. 1952. Use of ranks in one-criterion variance anaylsis. *J Am Stat Assoc*. 47:583–621.

Landis BN, Hummel T, Hugentobler M, Giger R, Lacroix JS. 2003. Ratings of overall olfactory function. *Chem Senses*. 28:691–694.

Lötsch J, Hummel T. 2019. A machine-learned analysis suggests non-redundant diagnostic information in olfactory subtests. *IBRO Rep*. 6:64–73.

Lötsch J, Reichmann H, Hummel T. 2008. Different odor tests contribute differently to the evaluation of olfactory loss. *Chem Senses*. 33:17–21.

Müller A, Müngersdorf A, Reichmann B, Strehle G, Hummel T. 2002. Olfactory function in Parkinsonian syndromes. *J Clin Neurosci*. 9:521–524.

Murphy C, Schubert CR, Cruickshanks KJ, Klein BE, Klein R, Nondahl DM. 2002. Prevalence of olfactory impairment in older adults. *JAMA*. 288:2307–2312.

Nguyen DT, Nguyen-Thi PL, Jankowski R. 2012. How does measured olfactory function correlate with self-ratings of the sense of smell in patients with nasal polyposis? *Laryngoscope*. 122:947–952.

Nordin S, Brämerson A, Bende M. 2004. Prevalence of self-reported poor odor-detection sensitivity: the Skövde population-based study. *Acta Otolaryngol*. 124:1171–1173.

Nordin S, Monsch AU, Murphy C. 1995. Unawareness of smell loss in normal aging and Alzheimer's disease: discrepancy between self-reported and diagnosed smell sensitivity. *J Gerontol B Psychol Sci Soc Sci*. 50:P187–P192.

Oleszkiewicz A, Schriever VA, Croy I, Hähner A, Hummel T. 2019. Updated Sniffin' Sticks normative data based on an extended sample of 9139 subjects. *Eur Arch Otorhinolaryngol*. 276:719–728. doi:10.1007/s00405-018-5248-1. Epub 2018 Dec 15.

Pearson K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil Mag*. Series 5 50:157–175.

R Development Core Team. 2008. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available from: http://www.R-project.org

Shu CH, Hummel T, Lee PL, Chiu CH, Lin SH, Yuan BC. 2009. The proportion of self-rated olfactory dysfunction does not change across the life span. *Am J Rhinol Allergy*. 23:413–416.

Spearman C. 1904. The proof and measurement of association between two things. *Am. J. Psychol*. 15:72–101.

Stanciu I, Larsson M, Nordin S, Adolfsson R, Nilsson LG, Olofsson JK. 2014. Olfactory impairment and subjective olfactory complaints independently predict conversion to dementia: a longitudinal, population-based study. *J Int Neuropsychol Soc*. 20:209–217.

Venables WN, Ripley BD. 2002. *Modern applied statistics with S*. New York (NY): Springer.

Vennemann MM, Hummel T, Berger K. 2008. The association between smoking and smell and taste impairment in the general population. *J Neurol*. 255:1121–1126.

Welge-Luessen A, Hummel T, Stojan T, Wolfensberger M. 2005. What is the correlation between ratings and measures of olfactory function in patients with olfactory loss? *Am J Rhinol*. 19:567–571.