Rapid Assessment of Olfactory Sensitivity Using the "Sniffin' Sticks"

Maria Pössel [1,2,3] · Jessica Freiherr [4,5] · Annette Horstmann [1,2,3]

Abstract

Introduction Assessment of olfactory performance is of high clinical interest in the contexts of smell loss as well as neurological diseases, and recently gained attention in obesity research. Available olfactory tests, especially for assessing olfactory sensitivity, are time-consuming and require high cognitive capacity. Therefore, we aimed to establish a short procedure for reliably testing olfactory sensitivity using a subtest of the "Sniffin' Sticks" battery. Evaluation criteria are test duration, validity, and test-retest reliability.

Methods In a preliminary study using a within-subject repeated-measures design, we measured olfactory sensitivity for n-butanol in 20 young and healthy participants. We compared sensitivity obtained with three different measures during two sessions in a pseudo-randomized order: a standard single-staircase three-alternative forced-choice procedure with seven reversals (SSP_7); an abbreviated version with five reversals (SSP_5); and an ascending presentation of 16 dilution steps from lowest to highest odor concentration (brief ascending procedure, BAP).

Results Compared to the SSP_7, the BAP was 51%, and the SSP_5 26% shorter in duration. Both the BAP and SSP_5 scores were highly correlated with the SSP_7. The test-retest reliability in all three tests was similar to that typically reported in olfactory research.

Conclusion The abbreviated tests are valid measures of olfactory sensitivity. Especially, the BAP is as reliable as the standard method, but remarkably faster and easier to perform.

Implications Thus, the short procedures bear potential for both research and clinical practice, especially for complex study designs with time constraints on olfactory testing and for patient populations with attention deficits.

Keywords Ascending procedure · Olfaction · Single staircase procedure · Smell · Threshold

Introduction

Olfaction is an integral part of the human sensorium: the recognition of smells from the surroundings is crucial for the detection of hazards such as fire and spoiled foods, but also for social communication and signaling food availability.

Measures of olfactory performance have become pivotal in both clinical and research practice. Especially, the olfactory detection threshold (ODT) oftentimes depicts differences between clinical and healthy populations (Krismer et al. 2017; Yazla et al. 2018). It measures the sensitivity to smells, that means, the lowest odor concentration that can reliably be picked up from the environment. ODTs have proven to reveal also small differences in smell perception, for example, between obese and normal-weight populations (Skrandies and Zschieschang 2015). Further, ODT and general olfactory testing is an important tool in the clinical examination of smell loss, as well as the early detection of neurological diseases like Parkinson's and Alzheimer's, as altered olfactory performance presents an early symptom and therefore a possible disease marker. For a review of available olfactory tests, see Doty (2007) and Eibenstein et al. (2005). Here, we focus on the commercially available ODT test kit from the "Sniffin' Sticks" test battery (Burghart, Wedel, Germany), which measures olfactory sensitivity to either n-butanol or phenylethyl alcohol. This test kit is well validated in Europe (Hummel et al. 2007), easy to assess by using commercially available pen-like devices, and offers explicit operating instructions

✉ Maria Pössel
poessel@cbs.mpg.de

1 IFB AdiposityDiseases, Leipzig University Medical Center, Philipp-Rosenthal-Str. 27, 04103 Leipzig, Germany

2 Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

3 CRC 1052A5 'Obesity Mechanisms', Leipzig University Medical Center, Liebigstr. 18, 04103 Leipzig, Germany

4 Diagnostic and Interventional Neuroradiology, RWTH Aachen University, Aachen, Germany

5 Fraunhofer Institute for Process Engineering and Packaging IVV, Freising, Germany

(Rumeau et al. 2016). However, like most other commonly used ODT tests, the application procedure is rather time-consuming and highly variable, rendering it demanding for patients and study participants in terms of perceptual and attentional resources. The ODT subtest of the "Sniffin' Sticks" has an implementation time of 10–25 min, showing high variability in duration depending on the patient's/participant's concentration capacity and ability to smell. To avoid sensory adaptation to the odorant, consecutive trials are usually separated by 30 s breaks, which results in a total trial length of at least 45–48 s according to Rumeau et al. (2016). Thus, shortened and less variable overall test duration would be beneficial in clinical and research routine to use patient time efficiently, allowing for complex study designs, and minimizing the patient's/participant's workload.

Few investigations tackle the shortening of the ODT subtest of the "Sniffin' Sticks" test battery. To date, the proposed short versions of the ODT test were unable to show acceptable test-retest reliability, stable test duration, and significant time-saving concordantly. Two studies present short versions based on the constant stimuli procedure (CSP) (Fechner 1860), in which the olfactory stimuli are presented once for each odor concentration in a randomized order (Kern et al. 2015; Lotsch et al. 2004). The ODT score is estimated by means of logistic regression (Linschoten et al. 2001). This method is frequently used in psychophysical threshold testing but has two major disadvantages in olfactory testing. Firstly, the interleaved presentation of high and lower odorant concentrations can lead to quick adaptations of the examinee's olfactory system. Secondly, the threshold value is estimated with logistic regression (for further details see (Linschoten et al. 2001)), but to ensure correct classification of the model, several trials for each odor concentration step would be needed. If this were considered, the test duration would be even longer compared to the standard test procedure. Furthermore, Croy et al. (2009) compared a wide step method with only 8 dilution steps to the standard procedure with 16 dilution steps in a healthy and clinical population. They showed an average time-saving of 16–30% depending on the population group (healthy vs. patient) and odor condition (n-butanol vs. PEA) when using the wide step method. The test-retest reliability of this method is compared to standard reliability of olfactory testing relatively high (.81–.86) and the test reliably differentiates patient populations from healthy volunteers. However, the wide step method cannot depict subtle differences between groups, since it has only 8 instead of 16 dilution steps of the odor.

Recently, Sijben et al. (2017) proposed an alternative short version using the ascending limits procedure (ALP) as described by Cain et al. (1988). The authors showed that thresholds obtained with the ALP are similar to those obtained with the standard single staircase procedure (SSP); however, compared to the SSP, the ALP shows comparably high variability in duration and an average time-saving of only 5 min.

Here, we evaluate shortened procedures using the "Sniffin' Sticks" ODT test kit that circumvent these limitations. To ensure stable test duration, simplify testing, and avoid sensory adaption, we use the brief ascending procedure (BAP)—an integration of both the previously discussed CSP and ALP—and compare it to the standardly used SSP and a shortened SSP version. Evaluation criteria are test duration, validity and test-retest reliability.

Materials and Methods

Subjects

A total of 20 participants (10 women; mean age 24.68 years, SD 2.6 years, range 19–30 years; mean body mass index (BMI) 22.03 kg/m^2, SD 1.66 kg/m^2, range 19.77–25.07 kg/m^2) took part in the experiment. All participants were previously screened by means of telephone interviews. Exclusion criteria included current smoking, recent history of smoking (< 3 years of abstinence), vegetarian/vegan diet, allergies, current use of medication except oral contraceptives, drug use within the last 2 months, alcoholism, current pregnancy/breastfeeding, any subjective or objective impairments of the sense of smell, nose surgery except childhood nasal polypectomy, and history of neurological or psychiatric disorders. Inclusion criteria were age between 18 and 36 years. After inclusion, participants provided written informed consent. The study was carried out in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the University of Leipzig.

Sample Size Estimation

A priori we determined the minimum number of participants required using a statistical power analysis with Gpower software (Faul et al. 2007). Based on data from Hummel et al. (1997), we performed the sample size estimation. They correlated odor thresholds that were assessed with the "Sniffin' Sticks" test battery, at two different test days. The effect size in this study was $r = .61$, considered to be large using Cohen's (1988) criteria. With an alpha = .05 and power = .80, the projected sample size needed with this effect size (GPower 3.1) is approximately $n = 16$. Thus, our proposed sample size of $n = 20$ will be more than adequate for the main objective of this study.

Study Design and Procedure

The investigation involved ODT testing on two test days in a repeated measures within-subject design. Participants were instructed to refrain from eating and drinking except for water 2 h prior to testing. In the first session, all participants were

screened for olfactory function using the short form of the olfactory identification test included in the "Sniffin' Sticks" test battery (Mueller and Renner 2006). On both test days, olfactory testing was conducted using the single staircase procedure (SSP) as described by Hummel et al. (1997) and the brief ascending procedure (BAP), in a pseudo-randomized order. The interval between test days was approximately 1 week (mean 8.45 days, SD 4.37 days, interval 7–20 days). ODT tests were conducted successively with a short break of approximately 10 min. After conducting both ODT tests, participants rated intensity (0 = very weak, 10 = very strong), pleasantness (− 5 = unpleasant, + 5 = pleasant), and familiarity (0 = unfamiliar, 10 = familiar) of the odor from the pen containing the highest concentration of n-butanol on a visual analog scale (Aitken 1969).

Materials

All odorants were presented in commercially available felt-tip pens ("Sniffin' Sticks"; Burghart Instruments, Wedel, Germany). For the screening of olfactory function, we used the short form of the olfactory identification test from the "Sniffin' Sticks" test battery (Mueller and Renner 2006). In a multiple-choice task, participants must identify the correct smell from a card with four descriptors per odorant. In total, five odorants were presented; the test confirms the presence of normosmia (≥ 4 correct answers) or hyposmia (< 4).

The ODT test kit from the "Sniffin' Sticks" test battery is performed with n-butanol, an odorant that arises from fermentation processes and is frequently used in olfactory testing. It is perceived as rather unpleasant.

Sixteen dilutions of n-butanol are prepared by stepwise diluting previous odor concentrations in a ratio of 1:2. The strongest odor concentration is 4% (pen number 1) and the weakest is 1.22 ppm (pen number 16).

The odorized pens are presented in triplets as described by Hummel et al. (1997), one containing diluted n-butanol and two containing the solvent (aqua conservans) only, serving as blanks. In this three-alternative forced-choice procedure, participants are asked to identify the pen containing the odorant. Each pen is presented for approximately 3 s at 1–2 cm distance of both nostrils. The interval between triplets is approximately 30 s. During testing, participants are blindfolded to avoid visual identification of the correct pen. We established ODTs based on the standard SSP procedure, an additionally computed threshold score with less reversals from the standard procedure, and the BAP.

In the standard SSP, odorants are presented from lowest to highest odor concentration. Two subsequent correct identifications trigger the first turning point (reversal of the staircase), thereby indicating the peri-threshold region. From there, odor concentration is increased following two correct answers in a row and decreased following an incorrect answer. Each turning point results in a reversal of the staircase. Seven reversals must be obtained in the "gold standard" SSP (Hummel et al. 1997). The short SSP follows the principle of the standard SSP but we estimated the threshold using only the first five of the seven measured reversals from the standard SSP.

In the BAP, each triplet is presented only once in an ascending order from lowest to highest odor concentration. The threshold score is defined as the point of transition between no detection and detection of the odorant, i.e., the threshold score is a value read at the boundary between correct and incorrect detection of the pen containing the odor. Based on the CSP (Fechner 1860) mentioned earlier, we presented each odor level only once. Similarly, based on the ALP (Cain et al. 1988), we defined the threshold score as being reached after five correct odor detections in a row. If the series of five correct detections begins within the five highest odor concentrations, the highest concentration level is repeated until five correct detections are reached unless the highest odor concentration is not detected, in which case the threshold value is zero.

Questionnaires and Interviews

Depressive symptoms were assessed using the Beck Depression Inventory (BDI) (Beck et al. 1961), a self-administered four-point rating scale (0 = not at all to 3 = always), which measures depressive symptoms in the past week, in order to exclude participants with depressive symptoms because depression has previously been shown to be associated with smell impairments (Croy and Hummel 2017).

Due to known effects of smoking on the olfactory system, smoking behavior was investigated using the Fagerstroem Test for Nicotine Dependence (FTQ, Fagerstroem 1978) as well as a smoking interview implemented previously in the Leipzig Life-Study containing questions about smoking behavior in the past and present, smoking onset and durations, breaks, and passive smoking hours (Loeffler et al. 2015).

To measure individual odor associations, use of the olfactory sense and the way olfaction influences decisions in daily life; we implemented the Importance of Olfaction Questionnaire (IOQ) (Croy et al. 2010).

Women were further interviewed to assess information about their menstrual cycle, because sensitivity to odors is known to be increased in follicular phase of the cycle/under oral contraceptive and decreased in luteal phase (Derntl et al. 2013; McNeil et al. 2013).

Data Analysis

JASP (version 0.8.1.1 for Mac OS X, JASP Team 2018), IBM SPSS (IBM Corp. Released 2015, IBM SPSS Statistics for Windows, Version 23.0.) and R (version 3.5.0, R Core Team 2013) were used for statistical evaluation. The α-level was set at .05.

Due to small sample size, normality of the data was ascertained using the Shapiro-Wilk test.

Age, BMI, ODT scores (SSP_7, SSP_5 on both days; BAP on the first day), perceived intensity, as well as pleasantness and familiarity of *n*-butanol on both days were normally distributed. ODT scores measured with BAP on the second day were not normally distributed. Although ANOVAs are relatively robust against violations of the assumption of normality, we nevertheless decided to perform each analysis, which included BAP threshold on day two, additionally with nonparametric testing as a precaution. As nonparametric test results did not deviate from parametric test results, we decided to report the latter here.

Data obtained with the "gold standard" SSP were analyzed twice. In a first step, we computed the standard ODT score, which is calculated by the mean of the last four of a total of seven reversals (SSP_7). A second threshold score was computed by the last two of a total of five reversals (SSP_5).

For BAP, the threshold value was estimated by identifying the point of transition between no detection and detection, which means, the point when an odorant was constantly detected five times in a row.

To compare the ODT scores obtained with the "gold standard" SSP with seven reversals (SSP_7), the short SSP with five reversals (SSP_5), the BAP, and between testing days, the data were submitted to repeated-measures analysis of variance (rm-ANOVA) using the general linear model with the within-subject factors "Method" (SSP_7/SSP_5/BAP) and "Test day" (T1/T2) and the between-subject factor "Sex" (male/female). Subsequently, we ran a Bayesian repeated measures analysis of variance (Bayesian rm-ANOVA) using the same model to ascertain that there are no significant differences between the ODT scores obtained with the different methods. While conventional statistical testing is based on the frequentist paradigm, the Bayesian approach is based on the subjective probability paradigm (van de Schoot et al. 2014). Compared to conventional statistical testing, the Bayesian approach is advantageous in that the likelihood of an outcome is considered under the null *and* the alternative hypothesis. This means that by using the Bayesian approach, we can actually estimate the probability of the null hypothesis (no differences between groups in our case), while in the conventional approach, we can only estimate the likelihood of our observations or more extreme values when the null hypothesis of no differences is true.

To examine the test-retest reliability, meaning the relationship between all ODT scores obtained with different methods and on different testing days, we used intraclass correlation (ICC). ICC estimates were calculated using SPSS statistical package version 23 (SPSS Inc., Chicago, IL) based on an absolute-agreement, two-way mixed-effects model. Additionally, we report Pearson's correlation coefficients to make our results comparable to other test-retest correlation studies in olfactory testing. To describe advantages regarding relevant time-saving of the short over the standard procedure, we used rm-ANOVA based on *p* values with the within-subject factors "Method' (SSP_7/SSP_5/BAP) and "Test day" (T1/T2) and the between-subject factor "Sex" (male/female).

To compare the differences of the interindividual variation regarding the duration of the three methods in order to find out whether the stability of the implementation time differs according to the assessed method, we first computed the three different Coefficients of Variation (CV), then adjusted the CVs for the mean of each method, and finally performed a one-way ANOVA with the dependent variable "adjusted CV."

Results

Sample Characteristics

BDI scores indicated no or only mild depressive symptoms in all subjects (mean = 6.45, SD = 4.86, range 0–17). All participants were nonsmokers (assessed with Fagerstroem scale), and none declared being ex-smokers. Passive smoking hours were the following: mean = 5.58 h/week, SD = 11.60 h/week, range 0–48 h/week. The questionnaire about the individual importance of odors showed the following results: sum-score (mean = 55.35, SD = 5.25, range 45–66), association-scale (mean = 18.10, SD = 2.59, range 13–23), application-scale (mean = 16.65, SD = 3.03, range 8–21), consequences-scale (mean = 16.60, SD = 2.33, range 11–20). No significant correlations between BDI score, passive smoking hours, or olfaction scales with ODTs were observed. Furthermore, no significant differences between men and women (SSP_7 $F = .251$, $p = .284$; SSP_5 $F = .521$, $p = .480$; BAP $F = .850$, $p = .369$; ANOVA) were observed. Perceived pleasantness, intensity, and familiarity of *n*-butanol are presented in Table 1. Odorant ratings did not differ significantly between days ($F = .004$, $p = .948$; rm-ANOVA).

Test Duration

Mean test duration is presented in Fig. 1. Test duration differed significantly between methods (rm-ANOVA, main

Table 1 Visual analog ratings for *n*-butanol

Quality	VAS-rating day 1	VAS-rating day 2	*p* value
Pleasantness	3.7 ± 2.0	3.8 ± 1.8	NS
Intensity	8.0 ± 1.4	7.4 ± 1.6	NS
Familiarity	6.6 ± 2.6	6.8 ± 2.2	NS

NS not significant

effect "Method": df = 2; F = 143.15, p < .001; post hoc t tests revealed significant differences between test duration for all three methods p < .001). The average trial number needed for threshold determination was 21.78 (SD = 2.75) trials for the standard SSP_7, 15.98 (SD = 2.71) trials for the short SSP_5, and 12.80 (SD = 1.58) trials for the BAP. The interindividual difference of test duration did not significantly differ between methods, this means, no method is more stable in terms of duration than the other (coefficient of variance for SSP_7 = 12.6%, SSP_5 = 17.0%, BAP = 12.5%; one-way ANOVA, main effect "SD of different methods": df = 2; F = 1.281, p = .286).

Validity of BAP and SSP_5

The threshold scores of the three methods did not differ (Fig. 2). The rm-ANOVA based on p values showed no significant differences between the three methods (main effect "Method": df = 2; F = 1.328, p = .278; interaction "Method" × "Testday": df = 2; F = 1.460, p = .243). Expecting no differences between the threshold scores of the three methods, we also estimated a Bayes factor using Bayesian information criteria (Wagenmakers 2007) in order to estimate the likelihood that the null hypothesis holds. The Bayesian rm-ANOVA showed moderate evidence in favor of the null hypothesis for a main effect of "Method" (BF_{01} = 3.788), that means, it is 3.8 times more likely that there is no difference between the ODT scores (null hypothesis) than that there is a difference (alternative). Furthermore, there is strong evidence in favor of the null hypothesis for the interaction "Method" × "Testday" (BF_{01} = 25.319), that is, it is 25.3 times more likely that for each method, there is no difference of ODT scores between days.

Further, correlation coefficients between all short procedures and the standard SSP were significant (Table 2) with a high positive relationship between SSP_7 and SSP_5 as well as a moderate positive relationship between SSP_7 and BAP showing the interrelation between the different methods. Moreover, the two short procedures were highly correlated (r = .696, p = .001).
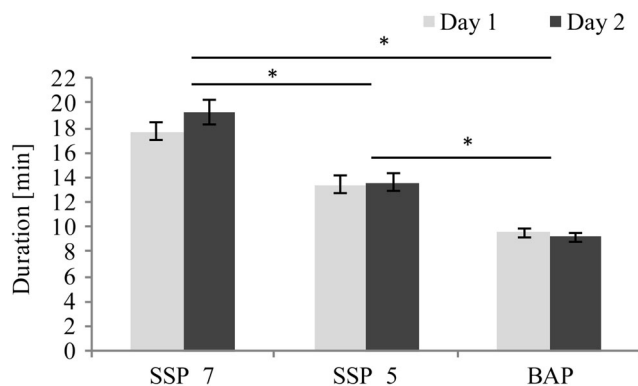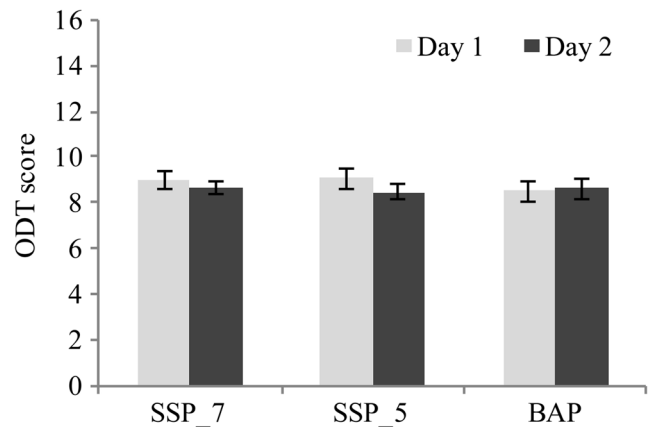

Fig. 1 Mean test duration of the three different methods

Fig. 2 Mean threshold scores of the three different methods

Test-Retest Reliability

Test-retest correlation analysis of the thresholds on different test-days showed significant correlation coefficients (Table 3, Fig. 3) with intraclass correlation as well as with Pearson's correlation. We show a moderate reliability for SSP_7 and BAP respectively and a poor reliability for SSP_5, meaning a moderate positive relationship between SSP_7 as well as for BAP between test tests and a weak positive relationship for SSP_5 between test days.

Discussion

In order to establish a brief test for measuring olfactory sensitivity, we compared two short procedures with the standard ODT test, all carried out using the "Sniffin' Sticks" ODT test kit. Our aim was to provide an ODT test that is easy to administer, requires little cognitive resources of the research participant or patient population, and shows predictable and stable test duration to be used in complex study designs and in the clinical context. We showed that both alternative ODT tests are significantly shorter than the commonly used SSP. The BAP takes only half of the time the standard SSP takes, and shows a smaller, however not significantly different variability in test duration.

Moreover, measured threshold scores do not differ between all three methods, that means, the short versions result in scores comparable to those obtained with the standard SSP.

Table 2 Correlation analysis of the short testing procedures with the standard single staircase procedure

	SSP_7	SSP_5	BAP
Correlation* with standard SSP_7 day 1	1	0.98	0.76
p value	–	< .001	< .001

*Pearson's correlation coefficient

Table 3 Test-retest reliability: intraclass correlation and Pearson's correlation coefficients

	SSP_7	SSP_5	BAP
Intraclass correlation coefficient	.64	.40	.63
p value	.001	.029	.001
Pearson's correlation coefficient	.68	0.45	0.68
p value	< .001	.050	.002

The test-retest reliability measured with intraclass correlation is similar in the standard ($r = .64$, $p = .001$) and BAP procedure ($r = .63$, $p = .001$), and smaller in the short SSP_5 ($r = .40$, $p = .029$). This means that the standard and the BAP procedure show a moderate reliability, meaning that the procedures produce scores that are relatively stable over time. The SSP_5 shows a poor test-retest reliability. However, compared to other test-retest reliability analysis in olfactory testing (measured with Pearson's correlation coefficients), all three test-retest correlation coefficients are equivalent to those normally found for odor sensitivity tests, which range from 0.43–0.85 ($p < .0001$) (Albrecht et al. 2008; Hummel et al. 1997; Lotsch et al. 2004). Moreover, the mean threshold scores did not differ between testing orders, test days and age. We did not find any differences for gender, which might be due to small group sizes. Furthermore, we did not find a correlation between olfactory performance and passive smoking hours as well as subjective importance of odors assessed via questionnaires.

To sum up, the two short procedures yield ODT scores like those obtained through the standard procedure, with similar test-retest reliability in all three procedures. Moreover, the BAP is 51% and the SSP_5 26% faster than the standard procedure.

Limitations

As mentioned earlier, the test-retest reliability in all three test procedures is comparable to that typically found for ODTs. However, the coefficients indicate a moderate (SSP_7, BAP), or even poor (SSP_5) reliability (Koo and Li 2016). The reliability in olfactory testing is generally rather low, possibly because of the susceptibility of our olfactory sense to many external factors. These include smoking (Hayes and Jinks 2012), modality of odor presentation (Sorokowska et al. 2015), hunger state (Albrecht et al. 2009; Ramaekers et al. 2016), menstrual cycle (Derntl et al. 2013; McNeil et al. 2013), climate (Katotomichelakis et al. 2007), and altered state of the nasal epithelium trough virus susceptibility according to the season (Konstantinidis et al. 2006). We here attempted to counteract those influences by controlling for several confounding factors (smoking, season, cycle phase, hunger). Nonetheless, we were unable to address all possible confounders satisfactorily, in particular hunger state. While we advised participants to refrain from eating 2 h prior to testing to avoid being either hungry or sated, we did not control food intake or quantify hunger state. For future studies, we would recommend using visual analog scales to assess feelings of hunger and satiety and, if possible, providing a standardized meal at the test institute.

Additionally, the resolution of the new BAP method is lower than the SSP, as it produces whole numbers only, whereas the thresholds computed of the SSPs give decimals. The BAP is therefore convenient in the clinical context when expecting large group differences, but might not be suitable in complex research designs to find small group differences between healthy populations for example according to different background odor stimulation or hormonal changes during pregnancy. Similarly, the BAP is more error prone than the SPPs—the impact of false positives/false negatives on the actual threshold score is higher in the BAP because each odor concentration step is presented only once. Moreover, since we derived two threshold values (SSP_7 and SSP_5) from the single staircase procedure, the true correlation is over-estimated because they are highly dependent.
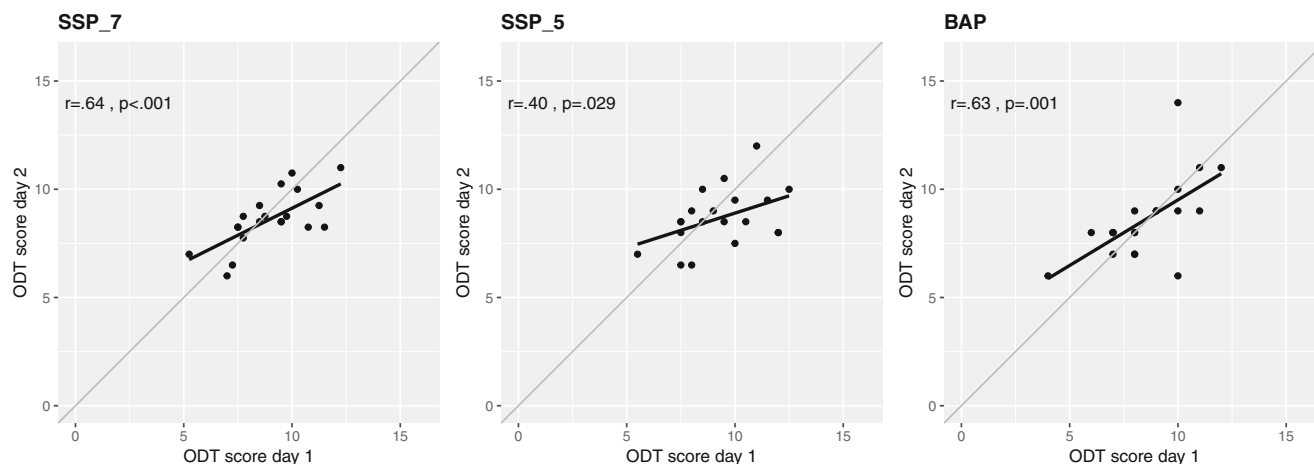

Fig. 3 Test-retest correlations for all three methods. The gray line represents the identity line ($y = x$)

Nonetheless, these limitations do not detract from the main advantages of the BAP: its brevity and more stable test duration, which can be of crucial importance in complex study designs with limited testing time. This is also an advantage for patients with limited cognitive resources, such as attention deficits. Another great advantage of the BAP compared to the SSPs is that it is easier to use. The odor concentrations are presented in an ascending order without any turning points and jumps, making the method less prone to errors from the investigator's side.

Conclusion

In this study, we assessed validity and reliability of standard and short ODT procedures to test olfactory sensitivity. We show that the short BAP is a valid and stable method and a good alternative to the standard SSP. While it is less precise and more susceptible to the influence of type one and two errors, it is also much shorter than the standard SSP. Although the task requires the same amount of effort within one trial under all three conditions regarding memory, having this demanding and exhausting task shortened 51% or 26% is very helpful for staying attentive and motivated to complete the task successfully. Moreover, especially the BAP method is also very easy to assess for the investigator and can thereby be used in the stressful daily clinical routine without further aids (computer software; paper template sheets). All three methods are easy to assess with the prefabricated, commercially available "Sniffin' Sticks."

Hence, we recommend using the BAP if only a limited time frame for testing is available or if examining patients/ participants with limited cognitive resources.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval and Informed Consent The study was carried out in accordance with the Declaration of Helsinki and was approved by the Local Ethics Committee. The research involved human participants. They provided written informed consent.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Aitken RC (1969) Measurement of feelings using visual analogue scales. Proc R Soc Med 62(10):989–993

Albrecht J, Anzinger A, Kopietz R, Schopf V, Kleemann AM, Pollatos O, Wiesmann M (2008) Test-retest reliability of the olfactory detection threshold test of the Sniffin' sticks. Chem Senses 33(5):461–467. https://doi.org/10.1093/chemse/bjn013

Albrecht J, Schreder T, Kleemann AM, Schopf V, Kopietz R, Anzinger A, Demmel M, Linn J, Kettenmann B, Wiesmann M (2009) Olfactory detection thresholds and pleasantness of a food-related and a non-food odour in hunger and satiety. Rhinology 47(2):160–165

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4:561–571

Cain WS, Gent JF, Goodspeed RB, Leonard G (1988) Evaluation of olfactory dysfunction in the Connecticut Chemosensory Clinical Research Center. Laryngoscope 98(1):83–88. https://doi.org/10.1288/00005537-198801000-00017

Cohen J (1988) Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale

Croy I, Hummel T (2017) Olfaction as a marker for depression. J Neurol 264(4):631–638. https://doi.org/10.1007/s00415-016-8227-8

Croy I, Lange K, Krone F, Negoias S, Seo H-S, Hummel T (2009) Comparison between odor thresholds for phenyl ethyl alcohol and butanol. Chem Senses 34:523–527

Croy I, Buschhuter D, Seo HS, Negoias S, Hummel T (2010) Individual significance of olfaction: development of a questionnaire. Eur Arch Otorhinolaryngol 267(1):67–71. https://doi.org/10.1007/s00405-009-1054-0

van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, van Aken MAG (2014) A gentle introduction to bayesian analysis: applications to developmental research. Child Dev 85(3):842–860. https://doi.org/10.1111/cdev.12169

Derntl B, Schopf V, Kollndorfer K, Lanzenberger R (2013) Menstrual cycle phase and duration of oral contraception intake affect olfactory perception. Chem Senses 38(1):67–75. https://doi.org/10.1093/chemse/bjs084

Doty RL (2007) Office procedures for quantitative assessment of olfactory function. Am J Rhinol 21(4):460–473

Eibenstein A, Fioretti AB, Lena C, Rosati N, Amabile G, Fusetti M (2005) Modern psychophysical tests to assess olfactory function. Neurol Sci 26(3):147–155. https://doi.org/10.1007/s10072-005-0452-3

Fagerstroem K-O (1978) Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addict Behav 3:235–241. https://doi.org/10.1016/0306-4603(78)90024-2

Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191

Fechner G. 1860. Elements of psychophysics. (H.E. Adler, trans.) New York: Holt, Rinehart & Winston, 1966

Hayes JE, Jinks AL (2012) Evaluation of smoking on olfactory thresholds of phenyl ethyl alcohol and n-butanol. Physiol Behav 107(2):177–180. https://doi.org/10.1016/j.physbeh.2012.06.020

Hummel T, Sekinger B, Wolf SR, Pauli E, Kobal G (1997) Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. Chem Senses 22(1):39–52

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(3):237–243. https://doi.org/10.1007/s00405-006-0173-0

IBM Corp. (Released 2015). IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp

JASP Team. 2018. JASP (Version 0.9)[Computer software]

Katotomichelakis M, Balatsouras D, Tripsianis G, Tsaroucha A, Homsioglou E, Danielides V (2007) Normative values of olfactory function testing using the "Sniffin' sticks". Laryngoscope 117:114–120

Kern DW, Schumm LP, Wroblewski KE, Pinto JM, Hummel T, McClintock MK (2015) Olfactory thresholds of the U.S. population of home-dwelling older adults: development and validation of a short, reliable measure. PLoS One 10(3):e0118589. https://doi.org/10.1371/journal.pone.0118589

Konstantinidis I, Haehner A, Frasnelli J, Reden J, Quante G, Damm M, Hummel T (2006) Post-infectious olfactory dysfunction exhibits a seasonal pattern. Rhinology 44(2):135–139

Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15:155–163

Krismer F, Pinter B, Mueller C, Mahlknecht P, Nocker M, Reiter E, Djamshidian-Tehrani A, Boesch SM, Wenning GK, Scherfler C, Poewe W, Seppi K (2017) Sniffing the diagnosis: olfactory testing in neurodegenerative parkinsonism. Parkinsonism Relat Disord 35:36–41. https://doi.org/10.1016/j.parkreldis.2016.11.010

Linschoten MR, Harvey LO Jr, Eller PM, Jafek BW (2001) Fast and accurate measurement of taste and smell thresholds using a maximum-likelihood adaptive staircase procedure. Percept Psychophys 63(8):1330–1347

Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R, Beutner F, Binder H, Brahler E, Burkhardt R et al (2015) The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. BMC Public Health 15:691. https://doi.org/10.1186/s12889-015-1983-z

Lotsch J, Lange C, Hummel T (2004) A simple and reliable method for clinical assessment of odor thresholds. Chem Senses 29(4):311–317. https://doi.org/10.1093/chemse/bjh034

McNeil J, Cameron JD, Finlayson G, Blundell JE, Doucet E (2013) Greater overall olfactory performance, explicit wanting for high fat foods and lipid intake during the mid-luteal phase of the menstrual cycle. Physiol Behav 112-113:84–89. https://doi.org/10.1016/j.physbeh.2013.02.008

Mueller C, Renner B (2006) A new procedure for the short screening of olfactory function using five items from the "Sniffin' sticks" identification test kit. Am J Rhinol 20(1):113–116

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/. Accessed 21 Feb 2019

Ramaekers MG, Verhoef A, Gort G, Luning PA, Boesveldt S (2016) Metabolic and sensory influences on odor sensitivity in humans. Chem Senses 41(2):163–168. https://doi.org/10.1093/chemse/bjv068

Rumeau C, Nguyen DT, Jankowski R (2016) How to assess olfactory performance with the Sniffin' sticks test((R)). Eur Ann Otorhinolaryngol Head Neck Dis 133(3):203–206. https://doi.org/10.1016/j.anorl.2015.08.004

Sijben R, Panzram C, Rodriguez-Raecke R, Haarmeier T, Freiherr J, Id, Sijben R, Hoo O (2017) Fast olfactory threshold determination using an ascending limits procedure. Chemosens Percept (pagination) 11(1):35–41. https://doi.org/10.1007/s12078-017-9239-1

Skrandies W, Zschieschang R (2015) Olfactory and gustatory functions and its relation to body weight. Physiol Behav 142:1–4. https://doi.org/10.1016/j.physbeh.2015.01.024

Sorokowska A, Albrecht E, Hummel T (2015) Reading first or smelling first? Effects of presentation order on odor identification. Atten Percept Psychophysiol 77(3):731–736. https://doi.org/10.3758/s13414-014-0811-3

Wagenmakers EJ (2007) A practical solution to the pervasive problems of p values. Psychon Bull Rev 14(5):779–804

Yazla E, Karadere ME, Kucukler FK, Karsidag C, Inanc L, Kankoc E, Donertas M, Demir E (2018) The effect of religious belief and forgiveness on coping with diabetes. J Relig Health 57(3):1010–1019. https://doi.org/10.1007/s10943-017-0504-z