Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Olfactory Training for COVID-19–Related Smell Loss: The SMELL Trial

Beatrice Heim

beatrice.heim@i-med.ac.at

Medical University of Innsbruck

Nicolas De Cleene

Medical University of Innsbruck

Frank Jagusch

Medical University of Innsbruck

Marina Peball

Medical University of Innsbruck

Clancy Cerejo

Medical University of Innsbruck

Joachim Schmutzhard

Medical University of Innsbruck

Judith Löffler-Ragg

Medical University of Innsbruck

Georg Göbel

Medical University of Innsbruck

Atbin Djamshidian

Medical University of Innsbruck

Günter Weiss

Medical University of Innsbruck

Klaus Seppi

Medical University of Innsbruck

Research Article

Keywords: COVID-19, SARS-CoV-2, olfactory dysfunction, anosmia, hyposmia, olfactory training, Sniffin' Sticks, randomized controlled trial

Posted Date: May 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9750365/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Persistent olfactory dysfunction (OD) after SARS-CoV-2 infection is common and disabling, but evidence for effective treatments remains limited. We aimed to evaluate the efficacy and safety of 12-week olfactory training (OT) in adults with persistent post-COVID-19 OD.

Methods

In this single-centre, controlled, assessor-blinded trial at the Medical University of Innsbruck, adults aged 18 years or older with OD persisting for at least 3 months after confirmed SARS-CoV-2 infection and impairment on Sniffin' Sticks identification testing were randomly assigned (1:1) to 12 weeks of OT or no intervention (natural history control). Because the protocol was developed during the COVID-19 pandemic under contemporaneous infection-control recommendations for reusable odour pens, the threshold subtest was omitted and a prespecified modified Sniffin' Sticks total score (mSST; identification plus discrimination, range 0–32) was used as the primary outcome. The primary endpoint was change in mSST from baseline to 12 weeks. Analysis was per protocol.

Findings

Following trial registration, 77 participants were screened through October 31, 2024. Of these, 70 were randomly assigned, and 68 completed the randomised phase and were included in the per-protocol analysis (OT n=38; control n=30). OT improved mSST compared with natural course, with a significant time-by-group interaction ($F=20.72$, $p<0.001$; partial $\eta^2=0.24$). Significant interactions were also observed for identification ($F=18.46$, $p<0.001$) and discrimination ($F=11.14$, $p=0.001$). At 12 weeks, more participants in the OT group were rated as improved on the Clinical Global Impression of Improvement (CGI-I) and reported improvement on the Patient Global Impression of Improvement (PGI-I) than in the control group (CGI-I: 68.4% vs 31.6%; $p=0.004$; PGI-I: 71.1% vs 28.9%; $p<0.001$). No serious adverse events occurred.

Interpretation

Twelve weeks of structured OT produced clinically meaningful improvement in olfactory outcomes compared with natural course in adults with persistent post-COVID-19 OD. OT appears to be a safe, low-risk, and scalable treatment option for this condition.

Funding

Austrian Science Fund (FWF, KLI1049).

Research context

Evidence before this study

Before undertaking this study, we performed a focused search of PubMed and the Cochrane Library using combinations of the terms "COVID-19", "SARS-CoV-2", "olfactory dysfunction", "anosmia",

“hyposmia”, and “olfactory training”. We aimed to identify clinical studies and reviews addressing olfactory training in adults with persistent smell loss after SARS-CoV-2 infection. At that time, evidence specific to chronic post-COVID-19 olfactory dysfunction was sparse. Olfactory training was biologically plausible, low risk, and supported by previous work in post-infectious olfactory dysfunction. Therefore, uncertainty remained regarding the magnitude of benefit and the patients most likely to respond.

Added value of this study

This study provides randomised controlled evidence that a standardised 12-week olfactory training programme improves objective olfactory performance and patient-reported recovery compared with natural course in adults with persistent post-COVID-19 olfactory dysfunction. The study also shows consistent improvement in participants from the original control group after delayed initiation of training in an exploratory extension phase, supporting the clinical usefulness of olfactory training even when started late.

Implications of all the available evidence

Taken together with prior work, our findings support olfactory training as a safe and practical first-line intervention for persistent post-COVID-19 olfactory dysfunction. Future multicentre trials should define the optimal training duration and regimen, identify predictors of response, and assess longer-term patient-centred outcomes.

Role of the funding source

The study was funded by the Austrian Science Fund (FWF, grant doi 10.55776/KLI1049). The funder had no role in data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Trial registration

European Database on Medical Devices (EUDAMED), CIV-AT-25-10-054874; closed to new participants. Trial registration number: NCT05421221.

Introduction

Olfactory dysfunction (OD) is among the most characteristic symptoms of COVID-19. Objective testing indicates that a large proportion of SARS-CoV-2-positive individuals experience measurable smell loss, even when self-reported deficits are lower. While recovery occurs within weeks in many, a clinically relevant proportion continues to report OD months to years after infection (1–4). Persistent OD affects safety, nutrition, and social functioning and has been linked to reduced quality of life and mood symptoms, underscoring the need for effective, low-risk, scalable interventions (1–3).

Pathophysiological models propose that SARS-CoV-2 primarily affects the olfactory system through infection-related dysfunction of non-neuronal sustentacular cells in the olfactory epithelium, resulting in local inflammation, down-regulation of olfactory receptors and signal transduction molecules, and downstream reduction of synaptic input to the olfactory bulb with potential secondary effects in higher-order cortical regions (3,5–7). Neuroimaging has provided complementary evidence for central correlates of olfactory dysfunction (7,8). Similar mechanisms have been described in other post-viral olfactory disorders (9–11).

Olfactory training leverages the intrinsic regenerative capacity and neuroplasticity of the olfactory system. In post-viral and post-traumatic OD, repeated structured exposure to selected odors improves olfactory performance and may induce functional and structural changes in olfactory pathways (12–17). However, controlled evidence in the specific context of persistent post-COVID-19 OD remains limited. The SMELL trial was designed to evaluate the effectiveness and safety of a standardized 12-week OT protocol using a commercially available four-odor Sniffin' Sticks set in adults with persistent post-COVID-19 OD. Participants assigned to the natural history control arm during the randomized controlled trial (RCT) were offered olfactory training after study termination for a subsequent 12-week period in an exploratory extension study.

Methods

Study design

The SMELL trial was a prospective, single-center, investigator-initiated, parallel-group, randomized, controlled, single-blinded (rater-blinded) trial conducted at the Medical University of Innsbruck (Department of Neurology) between June 13, 2022 and October 31, 2024. The randomized phase comprised a baseline visit (BL) and a 12-week follow-up visit (V1). The study was conducted in accordance with the Declaration of Helsinki, ISO 14155, and applicable national regulations. The study was approved by the local ethics committee (Ethics Committee approval No. EK 1273/2020, October 30, 2020). Written informed consent was obtained from all participants prior to any study procedures. The trial was registered in clinicaltrials.gov (Trial registration number NCT05421221; URL: https://clinicaltrials.gov/study/NCT05421221; Trial registration date: 13 June 2022). Detailed prespecified methods were published previously in the trial protocol (18). Patients and the public were not involved in the design, conduct, reporting, or dissemination plans of this research. The study was funded by the Austrian Science Fund (FWF, grant doi 10.55776/KLI1049).

Participants

Eligible participants were adults aged 18 years or older with confirmed prior SARS-CoV-2 infection at least three months before screening, documented either by a positive PCR test in the medical history or by positive serum anti-nucleocapsid antibodies, who reported persistent post-infectious OD for at least three months and demonstrated impairment on an age- and sex-adapted Sniffin' Sticks identification (SSI) test.

Exclusion criteria were olfactory dysfunction preceding COVID-19 because of other causes, such as traumatic brain injury or neurodegenerative disease; nasal pathology unrelated to COVID-19 on systematic evaluation including endoscopy of the olfactory cleft (if not already performed as part of routine work-up); and pregnancy.

Participants meeting all eligibility criteria were randomized 1:1 to OT or natural history control using permuted-block randomization (RandList v1.2). The trial was single-blinded with blinded assessment of olfactory testing (rater-blinded) at follow-up with respect to group allocation.

Participants were recruited at the Medical University of Innsbruck. Sex was self-reported at baseline during the structured interview.

Randomisation and masking

Participants were randomly assigned (1:1) to olfactory training or natural history control (=no training) using permuted-block randomisation generated with RandList version 1.2.

Outcome assessment at follow-up was performed by one rater blinded to group allocation.

The trial was assessor-blinded. Outcome assessors performing follow-up olfactory testing were masked to group allocation. Participants and personnel delivering the intervention were not masked because of the nature of the intervention. Missing primary efficacy data led to exclusion from the per-protocol analysis; for other outcomes, analyses used available data for the respective variable.

Procedures

Participants randomized to OT received a commercially available four-odor training set (19) consisting of citronellal (lemon), eucalyptol (eucalyptus), eugenol (clove), and phenyl-ethyl alcohol (rose) and performed OT twice daily for 12 weeks. Each training session consisted of attentive smelling of each odor for approximately 15 seconds (held about 1–2 cm under the nostrils) with deliberate inhalation, repeated across all four odors, for a total session duration of approximately 15 minutes. Adherence was documented using a patient diary.

Participants randomized to the natural history control group were instructed to continue usual daily life without structured OT during the randomized 12-week period.

After completion of the 12-week visit, control-group participants were offered the same olfactory training protocol for a subsequent 12-week exploratory extension phase.

Follow-up assessments were done at baseline and at 12 weeks in both randomised groups. Participants entering the extension phase underwent a further assessment at 24 weeks from baseline.

Outcomes

Olfactory function was assessed using the Sniffin' Sticks discrimination test (SSD; 16 triplets with two different odours; maximum score 16) and the Sniffin' Sticks identification test (SSI; 16 common odours presented in a four-alternative forced-choice procedure; maximum score 16).

The threshold subtest was prespecified to be omitted because the study protocol had been designed during the COVID-19 pandemic under contemporaneous infection-control recommendations of the German Society of Otorhinolaryngology, Head and Neck Surgery, which discouraged direct presentation of reusable odor pens under the patient's nose and recommended single-use systems or odor presentation on disposable paper (20). Under these conditions, a standardized corona-safe administration of the threshold subtest was not feasible. We therefore used a modified Sniffin' Sticks total score (mSST; SSI+SSD, range 0–32) as the objective olfactory outcome. This approach remains consistent with current position-paper recommendations that psychophysical olfactory assessment should include validated tests of odor threshold and/or identification or discrimination, ideally two or three subcomponents (21). In addition, prior validation studies have shown no significant differences between standard and paper-based administration for Sniffin' Sticks identification and discrimination, whereas threshold testing appears less robust under paper-based self-administration (22, 23).

Because the mSST is not identical to the canonical TDI score, established TDI-based diagnostic cut-offs were not applied; analyses therefore focused on within-subject change and between-group differences in the prespecified objective olfactory outcomes (mSST, SSI, and SSD).

The prespecified primary outcome was change in the modified Sniffin' Sticks total score (mSST) from baseline to 12 weeks. The mSST was defined as the sum of the Sniffin' Sticks identification score (SSI; range 0–16) and discrimination score (SSD; range 0–16), yielding a total range of 0–32.

Subjective olfactory function was assessed on a visual analogue scale (VAS) ranging from 1 to 10, with 10 indicating excellent olfactory function.

Patient progress, global improvement, and treatment response over time were assessed at follow-up using the Clinical Global Impression–Improvement (CGI-I) and Patient Global Impression–Improvement (PGI-I) scales, which are 7-point measures of overall change since randomization ranging from 1 (very much improved) to 7 (very much worse), with scores of 1–3 indicating improvement, 4 no change, and 5–7 worsening.

Mood and affect were assessed using the Hospital Anxiety and Depression Scale (HADS), a validated self-report questionnaire consisting of 14 items divided into two 7-item subscales measuring anxiety (HADS-A) and depression (HADS-D), with higher scores indicating greater symptom severity. Emotional well-being was additionally assessed using the Befindlichkeitsskala (Bf-S), a validated self-rating scale of current mood and subjective well-being, with higher scores indicating better emotional state.

Health-related quality of life was assessed using the 36-Item Short Form Health Survey (SF-36), a widely used self-report instrument that evaluates health-related quality of life across eight physical and mental

health domains, with higher scores indicating better perceived health status. As recommended previously (24), the 8 subscale scores were aggregated into two summary measures: Physical Functioning, Role Physical, Bodily Pain, and General Health constituted the Physical Component Summary (PCS) score, whereas Vitality, Social Functioning, Role Emotional, and Mental Health contributed to the Mental Component Summary (MCS) score. Consistent with each domain, the subscales were scored on a 0–100 scale, with higher values indicating better health status and a score of 50 representing the normative population mean.

Prespecified secondary outcomes in the randomised phase were change from baseline to 12 weeks in SSI, SSD, and VAS scores; CGI-I and PGI-I at 12 weeks; HADS-A and HADS-D; Bf-S; and the physical and mental component summary scores of the SF-36.

Clinically relevant improvement was defined a priori as an increase of at least 12.5% (25) for the modified SST score (mSST) as well as SSI and SSD scores, corresponding to a ≥4-point increase for the mSST score (maximum score 32) and a ≥2-point increase for both the SSI and SSD scores (maximum score 16 for both the SSI and SSD).

Harms were assessed at follow-up visits and by spontaneous reporting throughout the study. Serious adverse events, device-related adverse effects, and product defects were recorded.

The extension phase was exploratory and had no prespecified primary endpoint. In that phase, changes in olfactory performance and patient-reported outcomes across baseline, week 12, and week 24 were analysed within participants from the original control group who initiated delayed olfactory training after completion of the randomised phase.

Statistical analysis

Sample size calculation was based on the primary endpoint, defined as a clinically relevant improvement in total olfactory score (≥12.5% or ≥4 points out of a maximum of 32 points on the mSST score) after 12 weeks. Using assumptions derived from a previous randomised trial (25), and allowing for an anticipated dropout rate of 15%, the planned sample size was 100 participants.

Demographic and clinical data are presented as frequencies (for categorical variables) or means (median) ± standard deviations (for continuous variables). Gaussian distribution was confirmed by visual interpretation of the Q–Q (quantile–quantile) plots and the Kolmogorov–Smirnov test. Since all continuous variables—except for the duration of smell impairment—were not normally distributed, a natural logarithm (ln) transformation was applied to normalize the data before conducting statistical analyses.

Analysis of the primary endpoint utilized a repeated-measures mixed analysis of variance (ANOVA) model with Sniffin Sticks scores as the dependent variable, and an interaction between group assignment (OT vs. control) and time (baseline and follow-up) as independent variables.

The primary outcome was analysed per protocol, including participants who completed the 12-week randomised phase.

Continuous secondary outcomes (Bf-S, HADS, SF-36 domains) were analysed using the same approach. Inner-group significances were only assessed, if there was a significant effect of the interaction between group assignment (OT vs. control) and time (baseline and follow-up) as independent variables on the dependent variable. Changes in scales and questionnaires including the Sniffin Sticks scores were reported as mean differences with 95% confidence intervals (CI). Effect sizes and p-values from the statistical analyses are reported based on the ln-transformed data.

The CGI-I and PGI-I scores, assessed only at the 12-week visit, were compared between groups using the non-parametric Mann–Whitney U test. Dichotomized CGI-I and PGI-I ratings (improvement vs. no improvement) were compared using the chi-square test. Between-group differences in the proportion of participants showing clinically relevant improvement were also analysed using the chi-square test, defined as an increase of at least 12.5% (25) in the mSST score (≥4 points) and in SSI and SSD scores (≥2 points each). Statistical significance was set at a two-sided alpha of 5%.

Because participants randomized to the OT arm had non-significantly higher baseline Sniffin' Sticks scores, we additionally performed a sensitivity analysis using univariate analysis of covariance (ANCOVA), with baseline Sniffin' Sticks scores included as a covariate. The ANCOVA was further extended to adjust for age, sex, and duration of olfactory impairment.

Effect sizes were calculated using partial η^2 for continuous endpoints, η^2 coefficients for CGI-I and PGI-I, and the φ coefficient for dichotomized comparisons. Interpretations of effect size followed established conventions: partial η^2 of 0.01, 0.06, and 0.14; η^2 of 0.02, 0.13, and 0.26; and φ coefficients of 0.1, 0.3, and 0.5, corresponding to small, medium, and large effects, respectively.(15-17, 19)

The extension phase of olfactory training was exploratory in nature and not powered; therefore, no primary endpoint was defined and no sample size calculation was performed for this phase. Changes in olfactory performance in the natural history control group during the extension phase were analysed using a one-way repeated-measures ANOVA, with time (baseline, post–RCT follow-up V1, and extension follow-up V2) as the within-subject factor. Changes in scale and questionnaire scores are reported as mean differences with 95% confidence intervals (CIs).

Subgroup analyses of Sniffin' Sticks endpoints were performed in participants showing clinically relevant improvement, defined as an increase of at least 12.5% (25) in the mSST score (≥4 points) and in SSI and SSD scores (≥2 points each), during both study periods, the RCT without olfactory training and the extension phase. Effect sizes and p-values were derived from the ln-transformed data. The CGI-I and PGI-I scores assessed after the RCT (V1) and after the extension phase (V2) were compared using the non-parametric Wilcoxon signed-rank test. Dichotomized CGI-I and PGI-I ratings (improvement vs. no improvement) at V1 and V2 were compared using McNemar's test. Within-group differences in the proportion of participants showing clinically relevant improvement were also analysed using the

McNemar's test, defined as an increase of at least 12.5% (25) in the mSST score (≥4 points) and in SSI and SSD scores (≥2 points each).

Effect sizes are reported as partial η^2 for continuous endpoints, r (defined as $Z/\sqrt{N}$, where N denotes the number of paired observations) for CGI-I and PGI-I scores, and odds ratios (ORs), calculated from discordant pairs with corresponding 95% confidence intervals, as well as Cohen's g for dichotomized outcomes. Effect size interpretation followed established conventions: partial η^2 values of 0.01, 0.06, and 0.14; r values of 0.10, 0.30, and 0.50; ORs of 1.5, 3.0, and 5.0; and Cohen's g values of 0.05, 0.15, and 0.25, corresponding to small, medium, and large effects, respectively.

Statistical analyses were performed using IBM SPSS Statistics, version 30.0 (IBM Corp., Armonk, NY, USA).

Role of the funding source

The funder of the study (Austrian Science Fund, FWF) had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The SMELL trial was conducted between June 13, 2022, and October 31, 2024. Seventy-seven study participants were screened, as illustrated in the study flowchart (figure 1). All recruited study participants were Caucasian.

The planned sample size was 100 participants. Recruitment was closed before the planned sample size was reached; therefore, all randomized participants who completed the 12-week randomized phase and had available primary outcome data were included in the per-protocol analysis.

Seven participants were screening failures; reasons for this were: pre-existing olfactory dysfunction prior to COVID-19 based on medical history (n = 2), and withdrawal of consent due to no longer being interested in study participation (n = 5). Ultimately, 70 study participants entered the trial, and were randomized in a 1:1 fashion. Two participants withdrew consent prior to the 12-week follow-up visit (one per group). Consequently, 68 participants completed the randomized phase and were included in the per-protocol analysis.

Baseline characteristics are shown in Table 1. Participants had a mean age of 54.3 years (SD 14.4), 44 of 68 were women and 24 were men, and mean duration of olfactory dysfunction was 19.4 months (SD 11.0). Baseline characteristics were broadly similar between groups.

Changes in primary and secondary outcome measures from baseline (BL) to the 12-week follow-up visit (V1) during the RCT are presented in Table 2. Baseline characteristics of participants entering the extension phase (former natural history control group) are shown in Table 3. Changes in outcome

measures across the exploratory extension phase, from BL to the post–randomized controlled trial follow-up (V1) and to the 24-week follow-up after the extension phase (V2), are reported in Table 4.

The primary endpoint was successfully met. Mean mSST increased from 17.13 (SD 5.08) at baseline to 22.79 (4.58) at 12 weeks in the olfactory training group and from 18.93 (4.42) to 19.80 (4.14) in the control group. The time-by-group interaction for mSST was significant, indicating effectiveness of OT versus control ($F = 20.72$, $p < 0.001$, partial $\eta^2 = 0.24$). There was a significant improvement in mSST scores over time in the OT group ($F = 55.84$, $p < 0.001$, partial $\eta^2 = 0.60$), but not in the control group ($F = 2.55$, $p = 0.121$, partial $\eta^2 = 0.08$). In sensitivity analyses, the treatment effect remained significant after adjustment for baseline mSST ($F = 31.52$, $p < 0.001$, partial $\eta^2 = 0.33$). Adjusting additionally for age, sex and duration of smell impairment still yielded significant results ($F = 29.65$, $p < 0.001$, partial $\eta^2 = 0.32$). Clinically relevant improvement in mSST occurred in 31 of 38 participants (81.6%) in the OT group, and 4 of 30 (13.3%) in the control group ($\varphi = 0.68$, $p < 0.001$).

Olfactory training also improved both mSST subcomponents. SSI increased from 7.74 (SD 3.06) to 11.05 (2.78) in the olfactory training group and from 9·07 (2.77) to 9·57 (2.58) in the control group, with a significant time-by-group interaction ($F = 18.46$, $p < 0.001$, partial $\eta^2 = 0.22$). SSD increased from 9.39 (2.69) to 11.74 (2.44) in the OT group, and from 9.87 (2.10) to 10.23 (1.94) in the control group, showing also a significant time-by-group interaction ($F = 11.14$, $p = 0.001$, partial $\eta^2 = 0.14$). The ANCOVA models examining changes in SSI and SSD scores over time were also significant. After adjustment for baseline scores, significant effects were observed for both SSI ($F = 20.49$, $p < 0.001$, partial $\eta^2 = 0.24$) and SSD scores ($F = 34.09$, $p < 0.001$, partial $\eta^2 = 0.34$). After additional adjustment for age, sex, and duration of smell impairment, changes in SSI and SSD scores over time remained significant ($p < 0.001$), with large effect sizes for both SSI ($F = 19.52$, partial $\eta^2 = 0.24$) and SSD scores ($F = 32.32$, partial $\eta^2 = 0.34$). A substantially higher proportion of participants achieved a clinically relevant improvement in SSI in the OT group (84.2%) compared with controls (30.0%; $\varphi = 0.55$, $p < 0.001$), as well as in the SSD (57.9% vs. 23.3%; $\varphi = 0.34$, $p = 0.005$).

Key secondary endpoints, including subjective improvement assessed via VAS, CGI-I, and PGI-I, were similarly met (summarized in table 2). A significant interaction between time and group for VAS ratings ($F = 7.52$, $p = 0.008$, partial $\eta^2 = 0.10$) was observed. Significantly greater improvement occurred in the OT group on CGI-I ($\eta^2 = 0.16$) and PGI-I ($\eta^2 = 0.21$) compared to controls. More OT-group participants reported overall improvement at follow-up (CGI-I: 68.4%; PGI-I: 71.1%) compared to the control group (CGI-I: 31.6%, $p = 0.004$, $\varphi = 0.35$; PGI-I: 28.9%; $p < 0.001$, $\varphi = 0.44$). Additionally, no participant in the natural history control group rated themselves as "much improved" or better, compared to participants in the OT group (CGI-I: 15.8%, p=0.022; PGI-I: 13.2%, $p = 0.036$). Other secondary endpoints (table 2) yielded non-significant differences.

No serious adverse events occurred during the randomized phase, and no device-related adverse effects or product defects were reported. The intervention was well tolerated, and diary records indicated good adherence, with no participant showing <80% training compliance.

Of the 30 participants initially allocated to and completing the natural history control arm without olfactory training during the RCT, n=25 (83%) subsequently decided to further participate in the extension study and received olfactory training over the following 12 weeks. One-way repeated-measures ANOVA revealed significant effects of time for the mSST ($F = 25.43$, $p < 0.001$, partial $\eta^2 = 0.51$), SSI ($F = 16.93$, $p < 0.001$, partial $\eta^2 = 0.40$), and SSD scores ($F = 20.32$, $p < 0.001$, partial $\eta^2 = 0.46$). These effects reflected significant improvements during the extension phase, whereas no meaningful changes were observed during the randomized controlled period in which participants did not receive OT. A substantially higher proportion of study participants achieved clinically relevant improvement in mSST (64%), SSI (76%), and SSD scores (60%) following the extension phase (V2) compared with the RCT period (V1), with the mSST score improving from 12% to 64% ($p = 0.002$, OR = 7.50), the SSI score from 28% to 76% ($p = 0.008$, OR = 5.00), and the SSD score from 24% to 60% ($p = 0.049$, OR = 3.25). The CGI-I and PGI-I scores improved significantly at V2 compared with V1, with large effect sizes for CGI-I ($Z = -3.45$, $p < 0.001$; $r = 0.69$) and PGI-I ($Z = -3.18$, $p = 0.001$; $r = 0.64$).

Consistent with these findings, a higher proportion of participants reported overall improvement following the extension phase (V2) (CGI-I: 80%; PGI-I: 88%) than after the RCT (V1) (CGI-I: 36%, $p = 0.007$, OR = 6.50; PGI-I: 28%, $p < 0.001$, OR > 15). VAS scores improved during the extension phase, whereas no meaningful changes were observed during the randomized controlled period in which participants did not receive OT, and this pattern was confirmed by one-way repeated-measures ANOVA ($F = 13.51$, $p < 0.001$, partial $\eta^2 = 0.36$). Across the entire study period from BL to V2, objective and subjective olfactory outcomes demonstrated significant improvement, supported by post hoc analyses of the repeated-measures ANOVAs. The remaining endpoints showed non-significant changes during the extension phase compared with the randomized controlled period.

No serious adverse events or device-related adverse effects were observed during the extension phase.

In participants who received OT during the RCT phase, a repeated-measures ANOVA was performed, with change in mSST as the dependent variable and sex as the independent variable, to assess the influence of sex on the effect of OT. A significant main effect of time on olfactory performance was observed ($F = 57.17$, $p < 0.001$), whereas no time × sex interaction was detected ($F = 1.41$, $p = 0.243$).

Similarly, in participants who did not receive OT during the RCT phase and underwent OT as part of the extension study, a corresponding repeated-measures ANOVA was conducted, again with change in mSST as the dependent variable and sex as the independent variable. A significant main effect of time was also observed in this group ($F = 26.53$, $p < 0.001$), with no evidence of a time × sex interaction ($F = 0.13$, $p = 0.721$). Taken together, both analyses showed that olfactory training was effective independent of sex.

Male participants showed poorer SSI scores ($t = -2.757$, $p = 0.009$). In addition, older age was linked to poorer SSD performance ($B = -0.264$, $SE = 0.097$, $\beta = -0.315$, $t = -2.733$, $p = 0.008$), and this association remained after adjusting for sex ($B = -0.264$, $SE = 0.094$, $\beta = -0.315$, $rp = -0.324$, $rsp = -0.315$, $t = -2.804$, $p = 0.007$).

Sex and age were not related to any other clinical outcomes or to the duration of smell impairment.

Discussion

In this randomised, controlled, assessor-blinded trial, a structured 12-week olfactory training programme improved objective and subjective olfactory outcomes compared with natural course in adults with persistent post-COVID-19 olfactory dysfunction. The primary endpoint was met with a significant treatment effect on the prespecified mSST score, and responder analyses showed substantially higher rates of clinically meaningful improvement with OT than with the natural course. Improvements were also observed for SSI and SSD subscores, indicating that OT benefits multiple components of olfactory performance.

The main deviation from the planned protocol was that the prespecified target sample size of 100 participants was not reached. Recruitment was closed after 70 participants had been randomized. This occurred before final analysis of the primary endpoint. No changes were made to the primary outcome, intervention, follow-up duration, or main statistical analysis because of the achieved sample size.

The patient-reported findings were consistent with the psychophysical results. Participants assigned to olfactory training reported greater subjective recovery on the VAS and more favourable global improvement on CGI-I and PGI-I than those assigned to the control group. This concordance between objective and subjective outcomes strengthens the clinical relevance of the treatment effect.

The exploratory extension phase complements the randomised evidence. Participants initially assigned to the control group showed little change during the no-training period but improved after delayed initiation of olfactory training. Although this phase was not powered for confirmatory inference and remains vulnerable to expectation effects and regression to the mean, the within-participant pattern supports the clinical usefulness of olfactory training even when started relatively late after infection.

Our findings are consistent with meta-analytic evidence demonstrating multidimensional benefits of olfactory training in chronic post–COVID-19 olfactory dysfunction (26) and is consistent with prior evidence supporting olfactory training as an effective intervention for postinfectious and other acquired olfactory dysfunction.(15–17, 19) In contrast, placebo-controlled randomized trials have failed to show superiority over sham interventions.(27,28)

By directly comparing OT with the non-treatment control, our study isolates the effect of active training from natural recovery, suggesting that the structured implementation of OT itself may be the key determinant of clinical benefit, in line with studies showing no added advantage of intensified or alternative training protocols.(29,30)

The within-participant pattern of minimal change during the non-training period followed by marked improvement during the extension phase with OT supports the clinical usefulness of OT even when initiation is delayed. However, the extension phase is inherently susceptible to expectation effects and

regression to the mean and was exploratory and not powered for confirmatory analyses; therefore, its findings should be interpreted as supportive rather than definitive.

Olfactory training was effective independent of sex.

Broader patient-reported outcomes, such as mood, affect, and generic health-related quality of life, did not show significant between-group differences, suggesting that baseline burden on these instruments may have been limited, that these measures were not sufficiently sensitive to olfactory dysfunction–specific disability, or that longer follow-up may be required for olfactory improvements to translate into measurable psychosocial change.(2,4)

OT showed an excellent safety profile with no serious adverse events, no device-related adverse effects, and good tolerability, supporting its role as a low-risk, scalable first-line intervention for persistent post-COVID smell loss.

Several limitations should be noted. First, this was a single-center trial, which may limit generalizability. Second, the trial did not reach the prespecified target sample size of 100 participants. Although the primary outcome showed a statistically significant and clinically meaningful between-group difference, the smaller-than-planned sample size limits the precision of effect estimates and reduces power for secondary outcomes, subgroup analyses, and safety conclusions.

Third, the threshold subtest was omitted because of hygiene constraints at the beginning of the COVID-19 pandemic, when the protocol was designed and ethically approved in 2020. Fourth, the primary analysis was per protocol rather than intention-to-treat, although dropout was minimal. Fifth, the study population had limited demographic diversity, which restricts assessment of generalizability to broader populations. Sixth, broader patient-reported outcomes, such as mood and health-related quality of life, did not differ meaningfully between groups, possibly because these instruments were not sufficiently sensitive to smell-specific burden or because longer follow-up is needed.

Future multicenter studies with more diverse cohorts, longer follow-up, and olfactory dysfunction–specific quality-of-life instruments are warranted to evaluate durability of treatment effects, predictors of response, and patient-centered benefits beyond olfactory test scores.

Conclusion

Twelve weeks of structured olfactory training with a four-odor Sniffin’ Sticks set significantly improved objective olfactory function and subjective recovery compared with the natural course in adults with persistent post-COVID-19 olfactory dysfunction, with a favourable safety profile. The extension phase supports clinically meaningful improvement after delayed initiation of training and reinforces OT as a practical therapeutic option in routine care.

Given the lack of pharmacological interventions, this approach provides a low risk and useful therapeutic option which can be easily implemented in clinical practice.

Declarations

Contributors

Conceptualisation: BH, KS.

Data curation: NDC, BH, KS.

Formal analysis: NDC, KS.

Funding acquisition: BH, KH.

Investigation: all authors.

Methodology: BH, KS.

Project administration: NDC, BH, KS.

Supervision: BH, KS.

Validation: NDC, BH, KS.

Writing – original draft: NDC, BH, KS.

Writing – review & editing: all authors.

Data sharing

De-identified individual participant data will be made available beginning 3 months after publication. The study protocol has already been published (18).

Acknowledgments

This study was funded by the Austrian Science Fund (FWF, grant doi 10.55776/KLI1049.). We thank the participants and the SMELL study group for their contribution to this work. No medical writer was involved in the preparation of this manuscript.

Financial and other conflicts of interest of the manuscript authors

This study was funded by the Austrian Science Fund (FWF, grant doi 10.55776/KLI1049; Beatrice Heim). There are no financial or other conflicts of interest of the manuscript authors.

Trial registration: EUDAMED CIV-ID number: CIV-AT-25-10-054874; closed to new participants.

Funding: Austrian Science Fund (FWF) grant doi 10.55776/KLI1049.

Trial registration number: NCT05421221; URL: https://clinicaltrials.gov/study/NCT05421221

Trial registration date: 13 June 2022

References

1. Hannum ME, Ramirez VA, Lipson SJ, et al. Objective Sensory Testing Methods Reveal a Higher Prevalence of Olfactory Loss in COVID-19-Positive Patients Compared to Subjective Methods: A Systematic Review and Meta-Analysis. *Chemical Senses*. 2020;45(9):865–874. https://doi.org/10.1093/chemse/bjaa064
2. Boscolo-Rizzo P, Guida F, Polesel J, et al. Self-Reported Smell and Taste Recovery in Coronavirus Disease 2019 Patients: A One-Year Prospective Study. *Eur Arch Oto-Rhino-Laryngol*. 2022;279:515–520. doi:10.1007/s00405-021-06839-w
3. Tsukahara T, Brann DH, Datta SR. Mechanisms of SARS-CoV-2-associated anosmia. *Physiol Rev*. 2023;103(4):2759–2766. https://doi.org/10.1152/physrev.00012.2023
4. Rass V, Tymoszuk P, Sahanic S, et al. Distinct smell and taste disorder phenotype of post-acute COVID-19 sequelae. *Eur Arch Otorhinolaryngol*. 2023;280(11):5115–5128. doi:10.1007/s00405-023-08163-x
5. Hernandez-Clavijo A, Gonzalez-Velandia KY, Rangaswamy U, et al. Supporting Cells of the Human Olfactory Epithelium Co-Express the Lipid Scramblase TMEM16F and ACE2 and May Cause Smell Loss by SARS-CoV-2 Spike-Induced Syncytia. *Cell Physiol Biochem*. 2022;56(3):254–269. doi:10.33594/000000531
6. Verma AK, Zheng J, Meyerholz DK, Perlman S. SARS-CoV-2 infection of sustentacular cells disrupts olfactory signaling pathways. *JCI Insight*. 2022;7(24):e160277. doi:10.1172/jci.insight.160277
7. Lu Y, Li X, Geng D, et al. Cerebral micro-structural changes in COVID-19 patients – an MRI-based 3-month follow-up study. *EClinicalMedicine*. 2020;25:100484. doi:10.1016/J.ECLINM.2020.100484
8. Han P, Zang Y, Akshita J, Hummel T. Magnetic Resonance Imaging of Human Olfactory Dysfunction. *Brain Topogr*. 2019;32(6):987–997. doi:10.1007/s10548-019-00729-5
9. Li YC, Bai WZ, Hashikawa T. The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. *J Med Virol*. 2020. doi:10.1002/jmv.25728
10. Seiden AM. Postviral olfactory loss. *Otolaryngol Clin North Am*. 2004;37(6):1159–1166. doi:10.1016/j.otc.2004.06.007
11. Suzuki M, Saito K, Min WP, et al. Identification of viruses in patients with postviral olfactory dysfunction. *Laryngoscope*. 2007;117(2):272–277. doi:10.1097/01.mlg.0000249922.37381.1e
12. Merabet LB, Pascual-Leone A. Neural reorganization following sensory loss: the opportunity of change. *Nat Rev Neurosci*. 2010;11(1):44–52. doi:10.1038/nrn2758
13. Kollndorfer K, Kowalczyk K, Hoche E, et al. Recovery of olfactory function induces neuroplasticity effects in patients with smell loss. *Neural Plast*. 2014;2014:140419. doi:10.1155/2014/140419
14. Wilson DA, Best AR, Sullivan RM. Plasticity in the olfactory system: lessons for the neurobiology of memory. *Neuroscientist*. 2004;10(6):513–524. doi:10.1177/1073858404267048

15. Altundag A, Cayonu M, Kayabasoglu G, et al. Modified olfactory training in patients with postinfectious olfactory loss. *Laryngoscope*. 2015;125(8):1763–1766. doi:10.1002/lary.25245
16. Hummel T, Rissom K, Reden J, Hahner A, Weidenbecher M, Huttenbrink KB. Effects of olfactory training in patients with olfactory loss. *Laryngoscope*. 2009;119(3):496–499. doi:10.1002/lary.20101
17. Kollndorfer K, Fischmeister FP, Kowalczyk K, et al. Olfactory training induces changes in regional functional connectivity in patients with long-term smell loss. *Neuroimage Clin*. 2015;9:401–410. doi:10.1016/j.nicl.2015.09.004
18. De Cleene N, Jagusch F, Schmutzhard J, et al. Protocol for olfactory training in persisting COVID-19-associated loss of smell (SMELL): a monocentric randomised controlled trial conducted in Innsbruck. BMJ Open. 2025 May 27;15(5):e094027. doi: 10.1136/bmjopen-2024-094027
19. Li Z, Richter L, Krueger T, Haehner A, Hummel T. Effects of olfactory training on patients with parosmia. Neuroscience. 2025 Nov 19;588:112-118. doi: 10.1016/j.neuroscience.2025.10.026
20. Deutsche Gesellschaft für Hals-Nasen-Ohren-Heilkunde, Kopf- und Hals-Chirurgie. Handlungsempfehlungen HNO (Elektiveingriffe/Corona) – Hygieneempfehlungen zur sicheren Durchführung. 28 Apr 2020. Available at: https://smelltest.eu/wp-content/uploads/2020/04/Handlungsempfehlungen_HNO-Elektiveingriffe_Corona_28-4-20.pdf
21. Whitcroft KL, Altundag A, Balungwe P, et al. Position paper on olfactory dysfunction: 2023. Rhinology. 2023 Oct 1;61(33):1-108. doi: 10.4193/Rhin22.483
22. Mueller CA, Grassinger E, Naka A, Temmel AF, Hummel T, Kobal G. A self-administered odor identification test procedure using the "Sniffin' Sticks". Chem Senses. 2006 Jul;31(6):595-8. doi: 10.1093/chemse/bjj064
23. Besser G, Liu DT, Renner B, et al. Self-Administered Testing of Odor Threshold and Discrimination Using Sniffin' Sticks—Reviving the "Odor-Curves-On-Paper" Method. Chem. Percept. 13, 71–77 (2020). https://doi.org/10.1007/s12078-019-09263-x
24. Andersen JR, Breivik K, Engelund IE, et al. Correlated physical and mental health composite scores for the RAND-36 and RAND-12 health surveys: can we keep them simple? Health Qual Life Outcomes. 2022 Jun 3;20(1):89. doi: 10.1186/s12955-022-01992-0
25. Hummel T, Rissom K, Reden J, et al. Effects of olfactory training in patients with olfactory loss. Laryngoscope. 2009 Mar;119(3):496-9. doi: 10.1002/lary.20101
26. Chen CH, Shih CF, Hummel T, Chao YT. Multidimensional benefits of olfactory training for chronic COVID-19-related olfactory dysfunction: a systematic review and meta-analysis. Rhinology. 2025 Aug 1;63(4):431-440. doi: 10.4193/Rhin24.355
27. Mogensen DG, Aanaes K, Andersen IB, Jarden M, Backer V. Effect of Olfactory Training in COVID-19 Related Olfactory Dysfunction-A Placebo-Controlled Trial. Laryngoscope. 2025 Dec;135(12):4849-4857. doi: 10.1002/lary.32275
28. Serrano TLI, Antonio MA, Giacomin LT, Morcillo AM, Dirceu Ribeiro J, Sakano E. Olfactory training for the treatment of COVID-19 related smell loss: a randomised double-blind controlled trial. Rhinology. 2025 Jun 1;63(3):325-333. doi: 10.4193/Rhin24.081

29. Masson V, Nguyen-Thi PL, Gallet P, Jankowski R, Rumeau C, Nguyen DT. Assessment of two olfactory training methods for post-COVID-19 loss of olfaction: Classical and intensive. Eur Ann Otorhinolaryngol Head Neck Dis. 2025 Nov;142(6):294-300. doi: 10.1016/j.anorl.2025.04.007
30. Pires ÍAT, Steffens ST, Mocelin AG, Shibukawa DE, Leahy L, Saito FL, Amadeu NT, Lopes NMD, Garcia ECD, Albanese ML, De Mari LF, Ferreira IM, Veiga CA, Jebahi Y, Coifman H, Fornazieri MA, Hamerschmidt R. Intensive Olfactory Training in Post-COVID-19 Patients: A Multicenter Randomized Clinical Trial. Am J Rhinol Allergy. 2022 Nov;36(6):780-787. doi: 10.1177/19458924221113124

Tables

Table 1: Demographics and Results at Baseline

	Total (n=68)	OT group (n=38) (=direct training)	Control group (n=30) (=no direct training)	p-value
Age (years)	54.25 ± 14.35 (57.00)	53.95 ± 15.80 (57.00)	54.63 ± 12.53 (57.00)	0.706
Women:Men	44:24	24:14	20:10	0.764
Asthma	3 (4.4%)	3 (7.9%)	0	0.115
Pollinosis	15 (22.1%)	7 (18.4%)	8 (26.7%)	0.416
Current smoker	3 (4.4%)	3 (7.9%)	0	0.115
Ex smoker	6 (8.8%)	3 (7.9%)	3 (10%)	0.761
Diabetes mellitus II	2 (2.9%)	2 (5.3%)	0	0.202
BMI (kg/m^2)	25.30 ± 4.02 (24.55)	25.39 ± 4.31 (25.11)	25.18 ± 3.70 (24.21)	0.835
Duration of OD	19.41 ± 11.01 (18.00)	19.39 ± 9.51 (19.00)	19.43 ± 12.84 (17.50)	0.989
mSST	17.93 ± 4.85 (18.50)	17.13 ± 5.08 (18.50)	18.93 ± 4.42 (18.50)	0.178
SSD	9.6 ± 2.44 (10.00)	9.39 ± 2.69 (9.50)	9.87 ± 2.10 (10.00)	0.377
SSI	8.32 ± 2.99 (9.00)	7.74 ± 3.06 (8.00)	9.07 ± 2.77 (9.00)	0.101
VAS	3.34 ± 2.27 (3.00)	3.26 ± 2.31 (3.00)	3.43 ± 2.25 (3.00)	0.696
HADS-A	4.26 ± 3.26 (3.50)	4.84 ± 3.63 (4.00)	3.53 ± 2.60 (3.00)	0.114
HADS-D	3.13 ± 2.97 (2.00)	3.24 ± 2.95 (3.00)	3.00 ± 3.03 (2.00)	0.531
BF-S	10.22 ± 9.59 (7.00)	10.05 ± 9.72 (6.00)	10.43 ± 9.58 (10.50)	0.611

SF-36-PCS	72.04 ± 24.70 (82.30)	73.73 ± 22.25 (82.88)	69.90 ± 27.74 (78.13)	0.386
SF-36-MCS	74.12 ± 2.59 (83.13)	73.31 ± 20.68 (81.88)	75.14 ± 22.42 (87.38)	0.863

Data are presented as mean ± standard deviation (median) for continuous variables and number (percent) for categorical variables. Higher Score values indicate better outcome in all scales and questionnaires. OD, olfactory dysfunction; HADS-A/D, Hospital anxiety and depression scale-Anxiety/Depression; BF-S, Befindlichkeitsskala; SF36, Short Form 36 Health Survey; SF36-PCS, Physical Component Summary; SF36-MCS, Mental Component Summary; SSD/SSI, discrimination/identification score of the Sniffin' sticks test; mSST, modified total score of the Sniffin' sticks test; NA, not applicable.

Table 2: Change in primary and secondary endpoint scores between baseline (BL) and follow-up (V1)

		baseline	follow-up	p-value (F) [1]	p-value (F) [2,a]	Effect size [1,b]	Effect size [2,b]
mSST	C	18.93 ± 4.42 (18.50)	19.80 ± 4.14 (20.00)	0.121 (2.55)	< 0.001 (20.72)	0.08	0.24
	OT	17.13 ± 5.08 (18.50)	22.79 ± 4.58 (24.00)	< 0.001 (55.84)		0.60	
SSI	C	9.07 ± 2.77 (9.00)	9.57 ± 2.58 (9.00)	0.422 (0.66)	<0.001 (18.46)	0.02	0.22
	OT	7.74 ± 3.06 (8.00)	11.05 ± 2.78 (11.50)	<0.001 (41.55)		0.53	
SSD	C	9.87 ± 2.10 (10.00)	10.23 ± 1.94 (10.00)	0.236 (1.46)	0.001 (11.14)	0.05	0.14
	OT	9.39 ± 2.69 (9.50)	11.74 ± 2.44 (12.00)	<0.001 (32.64)		0.47	
CGI-I	C	NA	3.80 ±0.66	NA	0.001 (-3.27)	NA	0.16
	OT	NA	3.16 ±0.79	NA		NA	
PGI-I	C	NA	3.83± 0.59	NA	< 0.001 (-3.76)	NA	0.21
	OT	NA	3.11 ±0.86	NA		NA	
CGI-I improved	C	NA	12 (31.6)	NA	0.004 (8.29)	NA	0.35
	OT	NA	26 (68.4)	NA		NA	
PGI-I improved	C	NA	11 (28.9)	NA	< 0.001 (13.22)	NA	0.44
	OT	NA	27 (71.1)	NA		NA	
VAS	C	3.43 ± 2.25 (3.00)	4.10 ± 2.26 (4.00)	0.039 (4.67)	0.008 (7.52)	0.14	0.10

	OT	3.26 ± 2.31 (3.00)	5.13 ± 2.06 (5.00)	<0.001 (34.60)		0.48	
HADS-A	C	3.53 ± 2.60 (3.00)	3.77 ± 3.45 (3.00)	0.867 (0.03)	0.097 (2.83)	0.00	0.04
	OT	4.84 ± 3.63 (4.00)	4.16 ± 3.80 (3.00)	0.006 (8.33)		0.18	
HADS-D	C	3.00 ± 3.03 (2.00)	2.83 ± 3.24 (2.00)	0.362 (0.86)	0.625 (0.24)	0.03	0.00
	OT	3.24 ± 2.95 (3.00)	2.63 ± 2.48 (2.50)	0.037 (4.66)		0.11	
BF-S	C	10.43 ± 9.58 (10.50)	10.53 ± 9.60 (11.00)	0.823 (0.05)	0.971 (0.00)	0.00	0.00
	OT	10.05 ± 9.72 (6.00)	11.00 ± 9.52 (9.00)	0.701 (0.15)		0.00	
SF36 PCS	C	69.90 ± 27.74 (78.13)	74.21 ± 26.12 (83.50)	0.020 (6.06)	0.744 (0.11)	0.17	0.00
	OT	73.73 ± 22.25 (82.88)	77.64 ± 20.59 82.00)	0.046 (4.27)		0.10	
SF36 MCS	C	75.14 ± 22.42 (87.38)	77.33 ± 20.56 (83.50)	0.440 (0.61)	0.649 (0.21)	0.02	0.00
	OT	73.31 ± 20.68 (81.88)	77.10 ± 19.08 (85.50)	0.138 (2.30)		0.06	

Data of continuous variables are presented as mean ± standard deviation (median) (endpoint scores at baseline and follow-up), mean (95% CI) (change of endpoint scores) or n (%) for dichotomized variables.

[1] Within group comparison. [2] Between group comparison. [a] F-test for all variables except for CGI-I and PGI-I (z-score; Mann Whitney U test) as well as dichotomized variables (Pearson Chi-Square value). [b] Effect size according to partial η^2 for all variables except for the CGI-I and PGI-I (η^2) as well as dichotomized variables (φ coefficient).

For all p-values, significance level was set at $p \leq 0.05$.

Higher Score values indicate better outcome in all scales and questionnaires but in the HADS-A/D, CGI-I and PGI-I.

Abbreviations: C, control group; OT, olfactory training group; CGI-I/PGI-I, Clinical/Patient Global Impression – Improvement Scale; SSD/SSI, discrimination/identification score of the Sniffin' sticks test; mSST, modified total score of the Sniffin' sticks test; VAS, Visual Analogue Scale; mm, millimeters; NA, not applicable

Table 3: Demographics and Results at start of extension study of the control group

	Total (n=25)
Age	54.92 ± 12.66 (57.00)
Women:Men	17:8
Asthma	0
Pollinosis	8 (32%)
Current smoker	3 (12%)
Ex smoker	2 (12%)
Diabetes mellitus II	0
BMI	25.28 ± 3.96 (24.22)
Duration of OD (months)	19.88 ± 13.66 (19.00

Data are presented as mean ± standard deviation (median) for continuous variables and number (percent) for categorical variables.

Table 4: Change in endpoint scores between Baseline (BL), Follow up after 12 weeks (V1), and follow-up after 24 weeks (V2) for the control group subjects participating in the extension phase (n=25).

		Mean ± SD (Median)	p-value (F) [a]	Effect size [b]	post-hoc p-values [c]
mSST	BL	19.04 ± 3.48 (18.0)	<0.001 (25.43)	0.51	0.380/<0.001/ <0.001
	V1	20.12 ± 3.37 (20.0)			
	V2	24.72 ± 3.56 (24.0)			
SSI	BL	9.16 ± 2.44 (9.00)	<0.001 (16.93)	0.40	0.552/<0.001/ <0.001
	V1	9.72 ± 1.88 (9.00)			
	V2	12.04 ± 2.19 (12.0)			
SSD	BL	9.88 ± 1.45 (10.0)	<0.001 (20.32)	0.46	0.488/<0.001/ <0.001
	V1	10.40 ± 1.83 (10.0)			
	V2	12.20 ± 1.68 (12.0)			
CGI-I	V1	3.72 ± 0.61 (4.00)	<0.001 (-3.45)	0.69	NA
	V2	3.00 ± 0.65 (3.00)			
PGI-I	V1	3.80 ± 0.58 (4.00)	0.001 (-3.18)	0.64	NA
	V2	3.16 ± 0.62 (3.00)			
CGI-I improved (n)	V1	9 (36%)	0.007	0.44	NA
	V2	20 (80%)			
	OR	6.50 (1.47, 28.80)			
PGI-I improved (n)	V1	7 (28%)	<0.001	0.60	NA
	V2	22 (88%)			
	OR	>15.00			
VAS	BL	3.72 ± 2.25 (3.0)	<0.001 (13.51)	0.36	0.162/0.024/ <0.001
	V1	4.36 ± 2.23 (5.0)			
	V2	5.72 ± 2.09 (6.0)			
HADS-A	BL	3.36 ± 2.55 (3.0)	0.227 (1.53)	0.06	1.00/0.272/ 0.578
	V1	3.40 ± 2.71 (3.0)			

	V2	2.96 ± 3.06 (2.0)				
HADS-D	BL	2.48 ± 2.24 (2.0)	0.398 (0.94)	0.04	1.00/0.914/	
	V1	2.40 ± 2.35 (2.0)			0.694	
	V2	2.36 ± 2.77 (2.0)				
BF-S	BL	9.56 ± 9.28 (9.0)	0.412 (0.90)	0.04	1.00/0.493/	
	V1	8.88 ± 8.33 (9.0)			0.803	
	V2	7.56 ± 8.80 (4.0)				
SF36-PCS	BL	71.43 ± 25.34 (77.5)	0.315 (1.18)	0.05	0.105/1.00/	
	V1	75.90 ± 23.54 (82.5)			1.00	
	V2	75.54 ± 25.61 (84.5)				
SF36-MCS	BL	76.92 ± 19.76 (87.2)	0.749 (0.29)	0.01	1.00/1.00/	
	V1	79.43 ± 16.70 (82.2)			1.00	
	V2	81.63 ± 18.38 (88.8)				

Data of continuous variables are presented as mean ± standard deviation (median), mean (95% CI) (change of endpoint scores), n (%) and OR (95% CI for dichotomized variables.

For all p-values, significance level was set at $p<0.05$ with Bonferroni correction [c].

Higher Score values indicate better outcome in all scales and questionnaires but in the HADS-A/D, CGI-I and PGI-I

[a] F-test for all variables except for CGI-I and PGI-I (z-score; Wilcoxon signed rank test).

[b] Effect size according to partial η^2 for all variables except for the CGI-I and PGI-I (r, defined as $Z/\sqrt{N}$) as well as dichotomized variables (Cohen's g).

[c] p- -values for post hoc comparisons of changes from BL to V1/ from V1 to V2/ from BL to V2.

Abbreviations: BL, baseline; V1, post–randomized controlled trial follow-up; V2, extension follow-up; OR, odds ratio (calculated from discordant pairs); CGI-I/PGI-I, Clinical/Patient Global Impression – Improvement Scale; CGI-I improved / PGI-I improved: number of participants rated as improved on the CGI-I/PGI-I since the start of the RCT (i.e., since baseline); HADS-A/D, Hospital anxiety and depression

scale- Anxiety/Depression; BF-S, Befindlichkeitsskala; SF36, Short Form 36 Health Survey; SF36-PCS, Physical Component Summary; SF36-MCS, Mental Component Summary; SSD/SSI, discrimination/identification score of the Sniffin' sticks test; mSST, modified total score of the Sniffin' sticks test; VAS, Visual Analogue Scale; mm, millimeters; NA, not applicable

Figures

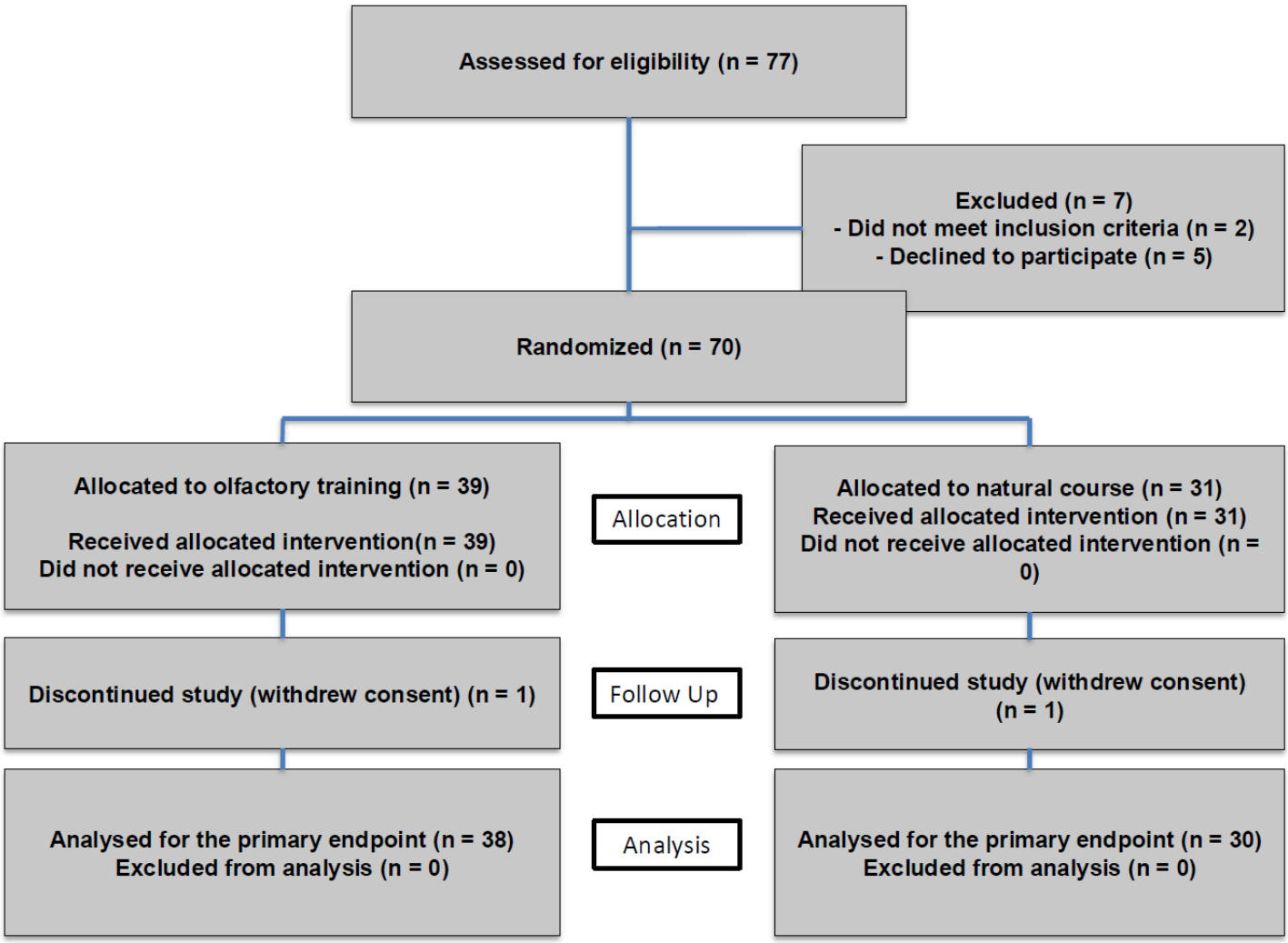


Figure 1

CONSORT flow diagram of participant recruitment, randomization, and analysis.

Participants were assessed for eligibility (n = 77); 7 were excluded (did not meet inclusion criteria, n = 2; declined to participate, n = 5). Seventy participants were randomized to the control arm (natural course/usual care; n = 31) or to olfactory training (n = 39). In the control arm, 31 received the allocated intervention; 1 discontinued the study (withdrew consent); 30 were analyzed for the primary endpoint. In the olfactory training arm, 39 received the allocated intervention; 1 discontinued the study (withdrew

consent); 38 were analyzed for the primary endpoint. No participants were excluded from the primary endpoint analysis in either arm.