Published in final edited form as:

Chemosens Percept. 2015 September ; 8(3): 138–148. doi:10.1007/s12078-015-9194-7.

The Taste and Smell Protocol in the 2011–2014 US National Health and Nutrition Examination Survey (NHANES): Test–Retest Reliability and Validity Testing

Shristi Rawal¹, Howard J. Hoffman², Mallory Honda¹, Tania B. Huedo-Medin¹, and Valerie B. Duffy¹

¹ Department of Allied Health Sciences, University of Connecticut, 358 Mansfield Road, Unit 1101, Storrs, CT 06269, USA

² Epidemiology and Statistics Program, National Institute on Deafness and other Communication Disorders (NIDCD), National Institutes of Health (NIH), Bethesda, MD, USA

Abstract

Introduction—The US NHANES 2011–2014 protocol includes a taste and smell questionnaire (CSQ) in home-based interviews and brief assessments in mobile exam centers. We report the short- and longer-term test–retest reliability and validity of this protocol against broader chemosensory measures.

Methods—A convenience sample of 73 adults (age=39.5±20.8 years) underwent the NHANES protocol at baseline, 2 weeks and 6 months. For taste, participants rated intensities of two tastants (1 M NaCl, 1 mM quinine) applied to the tongue tip and three tastants (1 M NaCl, 1 mM quinine, 0.32 M NaCl) sampled with the whole mouth. Smell function was assessed with a Pocket Smell Test™ (PST; eight-item odor identification test). The CSQ asked about chemosensory problems, distortions, and age-related changes. Broader baseline measurements were a 40-item olfactometer-generated identification task and additional whole-mouth taste intensities (1 M sucrose, 32 mM citric acid, 3.2 mM propylthiouracil).

Results—Intraclass correlations (*ICCs*) for NHANES taste measures showed moderate-to-good agreement after 2 weeks and 6 months (*ICCs* 0.42–0.71). Whole-mouth quinine intensity was significantly correlated with other taste intensities, supporting its utility as a marker for overall taste functioning. Olfactory classification from PSTs agreed for 98.5 % of participants across 2 weeks ($\kappa=0.85$; 95 % CI 0.71–0.99) and had good correspondence with the olfactometer task. CSQ items showed good-to-excellent agreement over 6 months (*ICCs* 0.66–0.90).

Valerie B. Duffy, valerie.duffy@uconn.edu.

Compliance with Ethics Requirements

Conflict of Interest Shristi Rawal declares that she has no conflict of interest.

Howard J. Hoffman declares that he has no conflict of interest.

Mallory Honda declares that she has no conflict of interest.

Tania B. Huedo-Medina declares that she has no conflict of interest.

Valerie B. Duffy declares that she has no conflict of interest.

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. Informed consent was obtained from all individual participants included in the study.

Conclusions—These findings further support that the NHANES chemosensory protocol has moderate-to-good test–retest reliability when administered to healthy, educated adults. Despite being a brief procedure with limited measures, the NHANES taste and smell assessments provided good information when compared to broader measures of taste and smell function.

Keywords

Taste perception; Smell perception; NHANES; Psychophysics; Bitter; Health status

Introduction

The National Health and Nutrition Examination Survey (NHANES) provides a picture of the health and nutritional status of adults and children in the USA through home-based health interviews followed by clinical tests, dietary interviews, and physical exams in a mobile examination center (MEC). NHANES is administered to a nationally representative sample of 5000 civilian, noninstitutionalized US persons of all ages (birth to >80 years) from 15 different, randomly-selected US counties each year. For the first time, NHANES 2011–2014 included a taste and smell protocol (CDC 2013a), designed to generate population-based estimates of variation and dysfunction in US adults 40+ years and, secondly, to assess associations between chemosensory function and other NHANES nutritional and health measures. The purpose of the present study was to examine the test–retest reliability of the NHANES taste and smell protocol and to compare this protocol to a more comprehensive assessment of taste and smell measures utilized in laboratory settings.

The NHANES taste and smell protocol was designed by an expert team (researchers, clinicians, epidemiologists) to provide the maximal amount of information on taste and smell variation within the framework of NHANES requirements that tests must be feasible, reliable, valid, and with minimal participant burden. The protocol built on the work of the NIH Toolbox project (NIH 2012), an initiative that assembled brief measurement tools in the domains of cognition, emotion, sensation, and motor function. Rapid taste (Coldwell et al. 2013b) and smell (Dalton et al. 2013) assessments selected in the NIH Toolbox were modified, adapted, and standardized for NHANES use. The NHANES taste and smell protocol implemented in 2011–2014 data cycles consisted of a taste and smell questionnaire (CSQ) administered in the respondents' homes, followed by brief taste and smell exams in the MEC.

Direct scaling of whole mouth and regional taste intensity is the measure of taste function in the NHANES protocol. The well-validated general Labeled Magnitude Scale (gLMS) (Bartoshuk et al. 2004) is used to assess perceived taste intensity, as it corresponds well with magnitude matching (Bartoshuk et al. 2004) and has been used in multiple clinical (Basson et al. 2005; Pepino et al. 2014) and population-based studies (Cruickshanks et al. 2009; McAnally et al. 2007) to study variation in taste, to examine taste-health associations (Basson et al. 2005; Dinehart et al. 2006; Fischer et al. 2014b; Hayes et al. 2010; Simchen et al. 2006), and to assess taste phenotype-genotype associations (Dias et al. 2013; Fischer et al. 2014a; Rawal et al. 2013). The NHANES taste exam begins with prescreening and orientation to the gLMS. Participants use the gLMS to rate the intensities of five LED-

generated lights (low-to-high luminescence), which serves as scale practice, allows assessment of scale understanding, and provides cross-modality comparisons for data analyses. Next, participants are asked to identify and rate intensities of two tastants (1 mM quinine hydrochloride followed by 1 M NaCl) applied with a cotton-tipped applicator to the tongue tip and three tastants (1 M NaCl, 1 mM quinine, 0.32 M NaCl) sampled with the whole mouth using a sip-and-spit procedure. The tastants were selected to capture genetic and environmentally mediated variation in taste for potential relevance to diet and health.

Smell function in NHANES is assessed with an 8-item odor identification test based on the validated and widely used 40-item University of Pennsylvania Smell Identification Test (UPSIT) (Doty et al. 1984). The PST odors were carefully selected to include food odors (strawberry, chocolate, onion, grape) relevant to diet and nutrition, two warning (smoke, natural gas) odors important for public safety, and two common household (leather, soap) odors.

A taste and smell questionnaire (CSQ) (CDC 2013b) was included in the home-interview to supplement the NHANES taste and smell exam data and to capture distortions (dysgeusia, phantosmia) and history of loss/changes in taste and smell. The CSQ was developed and standardized for NHANES from previous research (Hoffman et al. 1998; Murphy et al. 2002; Rawal et al. 2014) and standard clinical assessments. The CSQ asked perceived smell and taste problems and distortions, chemosensory losses with aging, and related risk factors and treatment of chemosensory impairments. The CSQ was content-validated and tested for response problems and cultural appropriateness. The sensitivity and specificity of smell-related questions on the CSQ, as compared to measured function, have been reported previously (Rawal et al. 2014).

In the present laboratory-based study, we report the short (2 weeks) and longer (>6 months) term test–retest reliability of the NHANES taste protocol, which extends previous short-term testing of similar taste measures in a pilot study of adults of a wide age range ($n=172$) (Coldwell et al. 2013a). In addition, we compare the NHANES taste measures with an additional set of whole-mouth taste probes. Because of time constraints, the NHANES protocol only utilizes bitter and salt tastants. Hence, this study investigated the value of the NHANES protocol in assessing taste function across all qualities. For olfactory testing, we examined the PST's test–retest reliability and concurrent validity in comparison to a 40-item olfactometer test (OLFACTID™, Osmic Enterprises, Inc.), which provides greater olfactory stimulus control. Lastly, we assessed the longer-term (6 month) test–retest reliabilities of the self-reported chemosensory measures in the CSQ and compared the CSQ responses with the taste and smell exam data.

Materials and Methods

Participants and Design

A convenience sample of 73 healthy adults, from a broad age range (18 to 87 years; mean age=39.5±20.8 years), was recruited from the university campus and surrounding community to capture a full range of chemosensory function. Exclusion criteria were pregnancy, allergies to quinine, and prescription use of thyroid medication (to avoid

individuals with Grave's Disease who might have developed a hypersensitivity to propylthiouracil, commonly known as PROP).

Data were collected during three laboratory-based sessions: baseline, with follow-ups at 2 weeks and 6 months. The University of Connecticut Institutional Review Board approved all of the procedures. Participants provided informed, written consent and were compensated for their time.

Measures

All of the measures listed below were collected at baseline. At 2 weeks, only the measures pertinent to the NHANES Taste and Smell Protocol (including the remembered light and sound intensity ratings substituted for LED-generated lights) were repeated, with exclusion of the Taste and Smell Questionnaire (CSQ). At 6 months, the 2-week measures were repeated with the addition of the CSQ.

The NHANES Taste and Smell Questionnaire (CSQ) asked questions related to perceived smell or taste dysfunction (CDC 2013b), including problems with ability to smell in the past year (yes/no), change in ability to smell since age 25 (better/worse/no change), smelling an unpleasant/bad/burning odor when nothing is there (yes/no), problems tasting sweet, sour, salty, or bitter in foods and drinks in the past year (yes/no), separate responses to change in ability to taste salty, sweet, sour, and bitter since age 25 (better/worse/no change), diminished ability to taste food flavors such as chocolate, vanilla, or strawberry since age 25 (yes/no), and in the past year, having a taste or other sensation in the mouth that does not go away (yes/no). Participants younger than age 25 ($n=19$) did not answer the questions relating to changes in smell, taste, and flavor.

Orientation to the gLMS intensity scale and light ratings was completed next using a standardized script and computerized ratings. The gLMS, a semantically labeled line scale, ranges from “no sensation” (scale score=0) at the bottom to the “strongest sensation of any kind” (100) at the top, with intermediate labels—“barely detectable” (1.4), “weak” (6), “moderate” (17), “strong” (34), and “very strong” (53)—the labeled intensities increase in quasi-logarithmic fashion. The participants were instructed to treat the top of the scale (scale score=100) as generalized, applying across all sensory domains. Following the orientation, the subjects practiced rating intensities of remembered light sensations, including “brightness of a well-lit room,” “brightness of a dimly lit restaurant,” and “brightest light ever seen.” Participants were instructed to first choose the label on the scale that best matches the intensity of the remembered sensation, within the context of all sensations, and then fine-tune their ratings by mouse clicking above/below the label. Upon clicking, the computer generated a score based on the distance along the scale from the zero point, which was then recorded by the examiner.

Intensity ratings of sound tones were made after the remembered light intensity sensations. Participants used the gLMS to rate the intensities of a series of 1000-Hz tones ranging in 12-dB steps from 50 to 98 dB (Bartoshuk et al. 1994). The tones served as a cross-modal standard and as covariates in the analysis to partition out individual variability in scale

usage. Instead of tones, LED-generated lights were used to assess scale understanding and as cross-modal standards in the NHANES.

NHANES taste measures included the intensities on gLMS for 1-mM quinine and 1 M NaCl painted on the anterior tongue with a cotton-tipped applicator (Coldwell et al. 2013b; Kveton and Bartoshuk 1994). The cotton swabs were amply saturated with the taste solutions prior to application on the tongue and were carefully “drawn” across the tongue tip to avoid distortion of taste ratings due to additional somatosensory stimulation from the cotton swabs. The procedure (with the diagram) has been described in detail elsewhere (Coldwell et al. 2013b).

After the regional taste testing, subjects sampled three tastants (1 M NaCl, 1 mM quinine 0.32 M NaCl; the order of presentation of the aqueous salts was randomized) with the whole mouth. Participants were asked to fill their mouths (sip without swallowing) with 10 ml of the taste solution, swish for approximately 5 s, and then expectorate into a sink. Each tastant was administered 30 s after the preceding tastant, and after each taste stimulus, the subjects rinsed their mouth with deionized water to remove any residual stimulus. In addition to the NHANES taste measures, participants reported whole-mouth intensity of additional prototypical tastants, including 1 M sucrose (sweet), 32 mM citric acid (sour), and 3.2 mM PROP (bitter).

NHANES Pocket Smell Test (PST™, Sensonics, Inc., Haddon Heights, NJ) was administered in two parts (four items each), for which the participants were asked to smell each odor (released when the experimenter scratched the odor strips with a pointed stylus in a “Z” pattern) and identify the odor from among the four choices provided. The forced-choice design of the PST requires the participant to choose an answer from among the four alternatives even if the participant is not sure or when no odor is perceived. The olfactory function score was calculated based on the number of correct identifications (the cumulative scores range from 0 to 8). Similar to the well-validated, eight-item San Diego Odor Identification Test (SDOIT) (Krantz et al. 2009), incorrect identification of three or more (out of eight total) odors (cumulative scores ranging from 0 to 5) was classified as olfactory dysfunction (hyposmia/microsmia or anosmia). In addition to identifying each odorant, participants also rated the respective odor intensities using the gLMS.

Olfactometer OLFACT-ID Test (Osmic Enterprises, Inc., Cincinnati, OH), administered next, was a 40-item odor identification test including all of the odorants used in the PST except for natural gas. The olfactometer was designed for computerized odor delivery via a tube in a consistent concentration and duration. Participants were asked to lean toward the odor delivery tube and smell the odorized airflow. The experimenter read the four possible choices for identification, prompting participant to select one choice, and then asked the participant to rate the odor intensity on the gLMS. Identification and intensity responses were recorded by the experimenter. Scoring was based on the number of correct identifications: 0–33 was categorized as “dysfunction” (0–18 anosmia, 19–25 severe microsmia, 26–29 moderate microsmia, 30–33 mild microsmia) and 34–40 was normal, i.e., “normosmia.” The scoring classification corresponded to the gender-blind classification based on the normative UPSIT data (Segura et al. 2013).

Statistical Analyses

Statistical analyses were accomplished using SPSS 20.0 (Chicago, IL). The significance criterion was set at $p \leq 0.05$. Means $\pm$ standard deviations (SD) are reported, unless otherwise noted; all tests were examined for age and sex differences and reported only when significant. Assumptions of data normality for relevant variables were tested with the Shapiro-Wilk statistic.

The test-retest reliabilities of the NHANES taste or smell measures, across 2 weeks and 6 months, were assessed with single measures Intraclass Correlation Coefficients (*ICC*) using the one-way random effect model. *ICCs* (single-measures, one-way random) also assessed test-retest reliabilities for intensities of cross-modal standards (remembered light sensations, sound tones) across the 2-week follow-up interval. Agreement between olfactory classifications from the NHANES Pocket Smell Tests administered at baseline and at 2-week follow-up was assessed by Cohen's kappa (κ) statistic. Cohen's κ statistic was not used to assess test-retest reliabilities of categorical responses to CSQ items across 6 months, as a large majority of participants answered "no" to questions regarding taste or smell dysfunction. Cohen's kappa statistic is highly sensitive to trait-prevalence and marginal probabilities and hence can be unstable and difficult to interpret when such skewed distributions occur (Gwet 2002). Agreement coefficient (*AC_I*), which is designed to overcome these limitations of kappa and provide a more precise and stable agreement statistic, was used instead (Gwet 2002). Our interpretation of *ICC*, κ , and *AC_I* was based on the following guidelines from Cicchetti and Sparrow (1981): <0.40 (poor agreement), $0.40\text{--}0.59$ (fair/moderate agreement), $0.60\text{--}0.74$ (good agreement), and 0.75 (excellent agreement).

In comparing the NHANES measures to broader taste or smell measures, we used Pearson correlation coefficients (r) to measure the degree of linear dependence between intensities of NHANES taste measures and additional prototypical taste measures (1 M sucrose, 32 mM citric acid, and 3.2 mM PROP), as well as the odor intensities of Pocket Smell Test (PST) and olfactometer. Paired t tests analyzed the differences in mean taste intensities between various taste measures and compared the CSQ responses with taste intensities or PST scores. Agreement between olfactory classifications by PST and olfactometer was assessed by simple percentages, by the Cohen's kappa (κ) statistic, and by sensitivity-specificity analyses. The sensitivity and specificity of the PST, as compared to the olfactometer, was calculated as follows:

$$\begin{aligned}\text{Sensitivity} &= TP / (TP + FN) \\ \text{Specificity} &= TN / (TN + FP)\end{aligned}$$

where TP, TN, FP, and FN are the number of true positive measures (judged dysfunction by olfactometer), true negative (judged normosmia by olfactometer), false positive (incorrectly classified as dysfunction by PST), and false negative (incorrectly classified as normosmia by PST), respectively.

To assess the ability of the NHANES taste test to reflect overall taste functioning, exploratory factor analyses, using the maximum likelihood estimation and promax rotations (with oblique, nonorthogonal axes), evaluated the interrelatedness and distinctiveness of the eight taste measures utilized in the study.

Results

Most subjects were female (68.8 %) and Caucasian (84.6 %). Forty-two percent of our sample had a college degree or higher; 9 % had only high school education or less. All participants described their health status as “excellent,” “very good,” or “good”; i.e., none indicated “poor” or “fair” health. All but three participants (96 % of the original sample) returned for the second visit, approximately 2 weeks (15.8 ± 1.6 days) later; 62 participants (85 % of the original sample) returned for a third session after 6 months (186.8 ± 9.5 days) from baseline. All participants rated the brightness of remembered light sensations in the correct intensity order (brightness of a dimly lit restaurant < brightness of a well-lit room < brightest light ever seen), supporting that they could use the gLMS correctly.

Taste Measures

Table 1 shows the mean $\pm$ SD for all taste measures utilized in the study. Our sample captured the expected variability in ratings for taste measures applied to the anterior tongue and sampled with the whole mouth. The taste intensity measures were normally distributed.

The *NHANES tongue tip measures* ranged from “barely detectable” to above “very strong,” with mean intensity for 1-mM quinine significantly lower than that for 1 M NaCl (Table 1; $p < 0.001$). In addition, only 1-mM quinine showed significant correlation with age ($r = -0.40$; $p < 0.001$). Among the *NHANES whole-mouth measures*, the weaker salt (0.32 M NaCl) had the lowest mean (Table 1; $p < 0.01$). As expected (Kveton and Bartoshuk 1994), the mean intensities for 1 mM quinine and 1 M NaCl were not significantly different and were significantly higher than the rest of the NHANES taste measures (Table 1). The whole-mouth intensities of 1 M NaCl ($r = 0.28$, $p < 0.01$) and 0.32 M NaCl ($r = 0.37$; $p < 0.001$) were positively correlated with age. Quinine whole-mouth intensity approached significance for sex effects, with women reporting greater intensities than men ($p < 0.06$).

NHANES Taste Measures and Broader Taste Quality Assessment

Exploratory factor analysis with all eight taste measures shown in Table 1 revealed a single factor with all factor loadings greater than 0.40 and the underlying factor explaining 58 % of the total variance in the sample ($X^2 = 32.54$, $p < 0.005$; Cronbach's $\alpha = 0.84$). In analyses controlling for age, sex, and the cross-modal standard (86-dB tone intensity), whole mouth intensity of 1 mM quinine showed the highest partial correlations with all of the taste measures evaluated (Table 1). Thus, 1-mM whole mouth quinine probe appears to be a reasonable proxy for overall taste functioning.

Test–Retest Reliability for NHANES Taste Protocol and Cross-Modal Standards

Table 2 shows the 2-week test–retest reliability coefficients for the five NHANES taste measures and the cross-modal standards used in this study (remembered light sensations,

sound tones). Six-month test–retest reliabilities for the NHANES taste measures also are reported.

Comparing baseline with 2-week testing, the intensities for NHANES taste measures showed moderate to good agreement, with intraclass correlations (one-way random, single measures), ranging from lowest for 0.32 M NaCl whole-mouth to highest for 1-mM quinine whole-mouth (Table 2). At 6 months, *ICCs* for the five NHANES taste measures were slightly lower but still showed moderate agreement. Consistent with the test–retest trend observed across 2 weeks, 0.32 M NaCl and 1 mM quinine whole-mouth ratings had the lowest and highest *ICCs*, respectively, after 6 months. In general, strong sensations tended to show better test–retest reliabilities than weak sensations, and this was true for both taste and cross-modal standards.

NHANES Pocket Smell Test

At baseline, the NHANES PST identification scores ranged from 2 to 8 (mean score 7.01 ± 1.09 SD); 8 % of the study participants were classified as having olfactory dysfunction (Table 3). At the second visit (after two weeks), the PST classified 5.7 % of the participants as having olfactory dysfunction. The PST identification scores at second visit ranged from 1 to 8, with a mean score of 7.14 ± 1.10 SD. Averaged across the two visits, chocolate had the lowest correct identification rate of 74.7 %, whereas soap had the highest correct identification rate of 98.7 %. The perceived intensities for PST odorants, averaged across the 2-week follow-up interval, showed variability, ranging from “moderate” to above “very strong.” The mean perceived intensity was highest for soap odor (36.29 ± 18.12 SD) and lowest for chocolate odor (16.05 ± 9.53 SD).

Pocket Smell Test versus Olfactometer

The olfactometer classified 14 % of the study participants as having olfactory dysfunction. Olfactometer identification scores ranged from 15 to 40, with a mean score of 35.74 ± 4.06 SD. One participant was classified with anosmia, one with severe microsmia, four with moderate microsmia, and four with mild microsmia.

As shown in Table 3, olfactory function classification from PST and olfactometer at baseline agreed for 94.5 % of participants ($\kappa=0.72$, 95 % CI 0.59–0.85), with the PST showing 60 % sensitivity (correct identification of dysfunction) and 100 % specificity (correct identification of normosmia) compared to the olfactometer. The olfactometer identified four additional participants as having dysfunction as compared to the PST, yet all four had mild microsmia (scores 30–33). For reporting the sensitivity and specificity of PST in screening for moderate–severe olfactory dysfunction, dysfunction by olfactometer was reclassified as moderate microsmia–anosmia (identifying 0–29 odors correctly), and mild microsmia–normosmia (identifying 30–40 odors correctly). This improved agreement of classification between the PST and olfactometer to 98.6 % of participants at baseline, with only one participant discordant (Table 3). With this reclassification, sensitivity and specificity of PST compared to the olfactometer improved to 100 and 98 %, respectively ($\kappa=0.90$, 95 % CI=0.80–1.00).

Correct and incorrect odor identification was consistent between the olfactometer and each PST testing (baseline and two weeks), averaging 82 % agreement (results not shown). Chocolate showed the lowest average percent agreement (68.5 %), whereas onion showed the highest average percent agreement (98 %). The perceived intensities of the seven common odorants were correlated across the PST and olfactometer, ranging from the lowest correlation for chocolate ($r=0.30$; $p<0.01$) to the highest for strawberry ($r=0.67$; $p<0.001$).

Test–Retest Reliability of NHANES Pocket Smell Test

Olfactory Function Classification—After 2 weeks, olfactory function classification by the PST agreed for 98.5 % of participants ($\kappa=0.85$, 95 % CI 0.71–0.99), with one participant classified as having dysfunction at the second visit but normosmic at the baseline (Table 3). The PST identification scores showed excellent agreement across the 2-week interval ($ICC=0.82$, 95 % CI=0.71–0.89) and good agreement across the 6-month interval ($ICC=0.69$, 95 % CI=0.55–0.79). However, since only two of the participants with olfactory dysfunction by PST at baseline returned at 6 months, we did not conduct further analyses on the 6-month test–retest reliability of the NHANES PST.

Across individual odors, correct and incorrect identification also was consistent between baseline and 2-week follow-up, averaging 87 % agreement (Fig. 1). Chocolate had the lowest percentage agreement between the two visits (67.1 %), with soap showing the highest percentage agreement (99 %). The two odors with lowest perceived intensities, chocolate and grape, showed the least agreement in identification after 2 weeks. The two odors important from a public safety point of view, natural gas and smoke, demonstrated high percentage agreements after 2 weeks (Fig. 1).

Odor Intensities—The intensity ratings of the PST odorants between baseline and at 2-week follow-up did not differ significantly (Fig. 2: $p>0.05$ for each odorant) and were in moderate-to-good agreement. ICC s for the intensities of the eight odorants in the PST ranged from 0.42 to 0.61, with chocolate and natural gas showing the lowest and highest agreement, respectively (data not shown).

CSQ Responses and Test–Retest Reliability Across 6 Months

CSQ Responses at Baseline and Comparison with Measured Function—At baseline, 11 % of the participants reported having problems with smell during the past 12 months, and only one indicated the presence of phantosmia (smelling an unpleasant/bad/burning odor when nothing is there). Eight percent of the participants indicated having problems with taste during the preceding year; only three reported experiencing dysgeusia (having a persistent bitter/metallic/bad/other taste or other sensation in mouth that does not go away). In addition, 16.4 % of participants indicated smell loss since age 25, and 13.7 % also reported that their ability to taste food flavors was not as good. Only a few subjects reported loss (diminution) in sweet, sour, salty, or bitter perception since age 25, with frequencies ranging from 2.7 to 5.5 %.

At baseline, those who indicated either recent smell problems or smell loss with age scored significantly lower on the PST (6.21 ± 0.47 SEM vs. 7.21 ± 0.10 SEM) and the olfactometer

test (32.15 ± 1.79 SEM vs. 36.59 ± 0.33 SEM) than those without these complaints. Participants who indicated taste problems at baseline tended to report lower intensities for quinine on the tongue tip (12.80 ± 2.53 SEM vs. 25.65 ± 2.01 SEM; $p < 0.08$); no significant differences were observed with other taste stimuli or taste-related questions.

CSQ Responses and Test–Retest Reliability at 6 Months—At the 6-month follow-up, 13 % of the participants reported problems with smell during the past 12 months; only two indicated experiencing phantosmia. Approximately 10 % indicated problems with taste during the past 12 months; only two indicated experiencing dysgeusia. Sixteen percent of participants reported smell loss since age 25, and almost 13 % also reported that their ability to taste food flavors was not as good. Five to eight percent of the subjects reported loss/diminution in sweet, sour, salty, or bitter perception since age 25.

Table 4 shows the longer-term test–retest reliabilities of eight CSQ items pertaining to smell and taste problems, as well as losses (diminutions) in taste, smell, or ability to perceive food flavors; questions pertaining to phantosmia and dysgeusia were not analyzed due to the low frequencies. Based on agreement coefficient measures, all eight items showed good-to-excellent agreement over the 6-month period.

Discussion

The newly added chemosensory protocol in the US NHANES will provide nationally representative data on taste and smell function as well as allow testing for chemosensory associations with a variety of health measures. The present study supports that the NHANES taste and smell protocol has moderate-to-good test–retest reliability and corresponds well with broader taste and smell assessments. The eight-item NHANES smell protocol (Pocket Smell Test™) had good concurrent validity with an olfactometer-delivered, 40-item odor identification test. The NHANES taste protocol, which only included salt and bitter stimuli, appears to be a good proxy for whole-mouth assessment with additional taste probes for sweet and sour, as well as PROP, a frequently used marker for genetic variation in taste.

Our findings are consistent with inclusion of quinine as a probe for overall taste functioning in the NIH Toolbox (Coldwell et al. 2013b). Whole-mouth intensity for 1-mM quinine showed good correlation with all taste measures utilized in our study, supporting its potential to represent broader taste function. The correlation between quinine and PROP intensity was significant but the lowest across other taste qualities tested. Although studies have reported discordance between quinine and PROP bitterness and associations with dietary behaviors (Duffy et al. 2003; Hayes and Duffy 2008), quinine appears to be an appropriate and feasible measure of taste function in population-based studies. Genetic variation in quinine intensity is tied to multiple bitter taste receptors (Reed et al. 2010) unlike PROP, which stimulates primarily one specific receptor (Kim et al. 2003). Also, quinine, unlike PROP, is an FDA-approved food additive, facilitating the consenting process for population-based studies. Finally, a recent study (Fischer et al. 2014a) showed that the associations between PROP intensity and basic taste intensities (sweet, salty, sour, bitter) differ by *TAS2R38* haplotype, further highlighting the need for alternative markers for distinguishing individuals with high taste sensitivity.

Quinine intensity on the tongue tip showed age differences in our study, supporting the need to include both regional and whole-mouth assessments of taste functioning in population studies, such as the NIH Toolbox (NIH 2012) and NHANES taste protocol (CDC 2013a). The intensity of bitter sensations on the tip of the tongue associated negatively with age, which was consistent with findings from the pilot NIH Toolbox test–retest reliability study (Coldwell et al. 2013a). However, we observed that the whole-mouth salt intensity was associated positively with age, consistent with another study (Fischer et al. 2013). One possible explanation is that the frequency and probability of exposures that cause chorda tympani taste damage, diminishing bitter sensations on the anterior tongue, may increase with age, including exposure to upper respiratory tract infections, otitis media, and head trauma (Bartoshuk et al. 1996). Reduced chorda tympani input with age, however, may disinhibit tactile sensations from the trigeminal nerve, resulting in intensified sensations of irritation from concentrated salt solutions (Duffy and Hayes 2014). More severe exposures may result in reductions of taste and irritant sensations on the tongue tip (Bartoshuk et al. 2012) and possibly whole mouth, as observed with the NIH Toolbox test–retest reliability study (Coldwell et al. 2013a).

All five NHANES taste measures showed good-to-moderate reliability over the 2-week follow-up interval in our study, consistent with findings from a pilot test–retest reliability study done with the NIH Toolbox measures (Coldwell et al. 2013a); our test–retest correlation coefficients, however, were somewhat higher. The present study is the first to evaluate and demonstrate moderate-to-good short- and longer-term test–retest reliability of the NHANES and Tool-box taste measures. In our study, remembered light sensations showed lower test–retest reliabilities than the sound tone ratings, implying that ratings of recalled stimuli are not as stable as those of measured stimuli. The NHANES protocol used LED-generated lights (CDC 2013a) as the nonoral, cross-modal standard, in order to have participants respond to uniform stimuli while learning the gLMS scaling procedure. The NHANES protocol also includes a replicate test for whole-mouth salt ratings (randomized between 1 M NaCl and 0.32 M NaCl), which will provide within session test–retest reliability of the whole-mouth taste measures (CDC 2013a).

The NHANES PST corresponded reasonably well with the 40-item OLFAC-T-ID™ test, delivered by an olfactometer, which offered greater stimulus control than the PST. The PST, however, was the most sensitive in identifying moderate-to-severe olfactory dysfunction, while missing a few cases of mild microsmia. Since the primary interest of the NHANES smell protocol is to generate prevalence estimates of moderate-to-severe dysfunction in the US population, our findings suggest that the 8-item odor identification (PST or Pocket Smell Test) is a sufficient measure of olfactory dysfunction in NHANES as well as for population-based studies with limited testing time.

The frequency of olfactory dysfunction in subjects >40 years ($n=36$), the target age group for NHANES, was similar with the olfactometer (24 %) but slightly lower with the NHANES PST (15 %) when compared to frequencies reported in population-based studies using identification tasks. For example, the frequencies of olfactory dysfunction observed in slightly older populations include (a) 24.5 % in the Epidemiology of Hearing Loss Study (53 years, (Murphy et al. 2002)), (b) 24 % in a Norwegian Study (45 years (Wehling et al.

2011)), (c) 27 % in the Blue Mountain Eye Study (60 years, (Karpa et al. 2010)), (d) 32.9 % in the Skovde Study (53 years; (Bramerson et al. 2004)), and (e) 29.6 % in the Dortmund Study (53 years, (Vennemann et al. 2008)). The National Social Life, Health and Aging Project, a national probability study of 3005 older adults (>57 years), reported a lower frequency of olfactory dysfunction (2.7 %) but had a more stringent criterion for dysfunction. The overall frequency of smell dysfunction in our study (8 %) with the NHANES PST was somewhat higher than that observed in the Beaver Dam Offspring Study (Schubert et al. 2012), which used a similar 8-item odor identification test on 2838 participants with similar age ranges (21–84 years).

The NHANES Pocket Smell Test showed good test–retest reliability after a 2-week follow-up interval for the overall identification score as well as correct identification of individual odorants. The two odors important from a public safety point of view, natural gas and smoke, showed high percentage agreements for identification across 2 weeks. In this study, we supplemented the NHANES smell protocol with an odor intensity measure. The reported intensities of the individual odorants also had good test–retest reliability after 2 weeks, added minimal time to the odor identification tasks, and demonstrated more variability than did the overall olfactory scores. Odor identification tests are designed to classify individuals into either dysfunction or “normal” (normosmics) and hence have limited ability to discriminate across the levels of olfactory function within individuals with normal sense of smell (Krantz et al. 2009). Asking participants to rate perceived odor intensity on a well-validated sensory intensity scale, such as the general Labeled Magnitude Scale (Bartoshuk et al. 2004), would allow detection of gradations in normal olfactory functioning and may also better explain variation in diet and health indices (Minski and Duffy 2009). Finally, and consistent with previous work (Rawal et al. 2014), those who reported either smell problems or smell loss (diminution) showed lower measured smell function than those who reported neither. Overall, items in the CSQ pertaining to chemosensory problems or losses also showed good-to-excellent longer-term test–retest reliability across 6 months.

Our study had several limitations. Due to lack of equipment, we did not examine the test–retest reliability of the LED-generated lights included as the cross-modal standard in NHANES. Other measures correlated with taste function, such as taste receptor density and regional intensities of sweet and sour on anterior and posterior tongue, were not examined. Additionally, we were unable to clinically verify the diagnosis for participants classified with anosmia or severe microsmia in our study. The generalizability of our findings to more diverse groups, such as those tested by NHANES, is limited since our sample was highly educated, small in size, and relatively homogeneous in terms of health status and geographical residence. The sensitivity–specificity of the NHANES PST may be more robust in the large, diverse NHANES sample with varying degrees of overall health status and potentially higher prevalence of olfactory dysfunction. On the other hand, the reliability of the NHANES taste and smell measures may be suboptimal in populations with lower formal educational attainment.

Conclusion

This study supports that the NHANES taste and smell protocol has moderate-to-good test–retest reliability, based on our study with a convenience sample of healthy adults. The NHANES taste and smell assessments also corresponded well with broader measures of taste and smell function that is usually included in laboratory or clinical-based settings. Hence, our findings combined with those from the NIH Toolbox normative study (Coldwell et al. 2013b) provide evidence that the NHANES taste and smell protocol captures appropriate population-based differences in taste and smell function and provides reliable information despite being a relatively brief procedure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by an Interagency Agreement (Y1-DC-0013) between the National Institute on Deafness and Other Communication Disorders (NIDCD), National Institutes of Health (NIH), and the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC).

References

- Bartoshuk L, Duffy V, Miller I. PTC/PROP tasting: anatomy, psychophysics, and sex effects. *Physiol Behav.* 1994; 56(6):1165–1171. [PubMed: 7878086]
- Bartoshuk LM, Duffy VB, Reed D, Williams A. Supertasting, earaches and head injury: genetics and pathology alter our taste worlds. *Neurosci Biobehav Rev.* 1996; 20(1):79–87. [PubMed: 8622833]
- Bartoshuk LM, Duffy VB, Green BG, Hoffman HJ, Ko CW, Lucchina LA, Marks LE, Snyder DJ, Weiffenbach JM. Valid across-group comparisons with labeled scales: the gLMS versus magnitude matching. *Physiol Behav.* 2004; 82(1):109–114. [PubMed: 15234598]
- Bartoshuk LM, Catalanotto F, Hoffman H, Logan H, Snyder DJ. Taste damage (otitis media, tonsillectomy and head and neck cancer), oral sensations and BMI. *Physiol Behav.* 2012; 107(4): 516–526. [PubMed: 22728784]
- Basson MD, Bartoshuk LM, Dichello SZ, Panzini L, Weiffenbach JM, Duffy VB. Association between 6-n-propylthiouracil (PROP) bitterness and colonic neoplasms. *Dig Dis Sci.* 2005; 50(3):483–489. [PubMed: 15810630]
- Bramerson A, Johansson L, Ek L, Nordin S, Bende M. Prevalence of olfactory dysfunction: the skovde population-based study. *Laryngoscope.* 2004; 114(4):733–737. [PubMed: 15064632]
- CDC. [9 Mar 2015] Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey (NHANES) Taste and Smell Examination Component Manual. 2013a. http://www.cdc.gov/nchs/data/nhanes/nhanes_13_14/Taste_Smell.pdf.
- CDC. [9 Mar 2015] Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey (NHANES). 2011–2012 data documentation, codebook, and frequencies: taste and smell disorders. 2013b. http://www.cdc.gov/nchs/nhanes/nhanes2011-2012/CSQ_G.htm.
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic.* 1981; 86:127–137. [PubMed: 7315877]
- Coldwell SE, Duffy VB, Bartoshuk L, Griffith JW, Hoffman HJ. The NIH Toolbox Brief Gustation Assessment Protocol. Abstracts from the 35th Annual Meeting of AChemS. *Chem Senses: A114.* 2013a doi:10.1093/chemse/bjt036.

- Coldwell SE, Mennella JA, Duffy VB, Pelchat ML, Griffith JW, Smutzer G, Cowart BJ, Breslin PA, Bartoshuk LM, Hastings L, et al. Gustation assessment using the NIH Toolbox. *Neurology*. 2013b; 80(11 Suppl 3):S20–S24.
- Cruickshanks KJ, Schubert CR, Snyder DJ, Bartoshuk LM, Huang GH, Klein BE, Klein R, Nieto FJ, Pankow JS, Tweed TS, et al. Measuring taste impairment in epidemiologic studies: the Beaver Dam Offspring Study. *Ann N Y Acad Sci*. 2009; 1170:543–552. [PubMed: 19686191]
- Dalton P, Doty RL, Murphy C, Frank R, Hoffman HJ, Maute C, Kallen MA, Slotkin J. Olfactory assessment using the NIH Toolbox. *Neurology*. 2013; 80(11 Suppl 3):S32–S36. [PubMed: 23479541]
- Dias AG, Rousseau D, Duizer L, Cockburn M, Chiu W, Nielsen D, El-Sohemy A. Genetic variation in putative salt taste receptors and salt taste perception in humans. *Chem Senses*. 2013; 38(2):137–145. [PubMed: 23118204]
- Dinehart ME, Hayes JE, Bartoshuk LM, Lanier SL, Duffy VB. Bitter taste markers explain variability in vegetable sweetness, bitterness, and intake. *Physiol Behav*. 2006; 87(2):304–313. [PubMed: 16368118]
- Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiol Behav*. 1984; 32(3):489–502. [PubMed: 6463130]
- Duffy VB.; Hayes, JE. Smell, taste and oral somatosensation: age-related changes and nutritional implications.. In: Chernoff, R., editor. *Geriatric nutrition: the health professional's handbook*. 4th edn.. Jones and Bartlett Publishers, Inc.; Burlington: 2014. p. 115-164.
- Duffy VB, Peterson JM, Dinehart ME, Bartoshuk LM. Genetic and environmental variation in taste: associations with sweet intensity, preference, and intake. *Top Clin Nutr*. 2003; 18(4):209–220.
- Fischer ME, Cruickshanks KJ, Schubert CR, Pinto A, Klein BE, Klein R, Nieto FJ, Pankow JS, Huang GH, Snyder DJ. Taste intensity in the Beaver Dam Offspring Study. *Laryngoscope*. 2013; 123(6): 1399–1404. [PubMed: 23625687]
- Fischer ME, Cruickshanks KJ, Pankow JS, Pankratz N, Schubert CR, Huang GH, Klein BE, Klein R, Pinto A. The Associations between 6- n -Propylthiouracil (PROP) Intensity and Taste Intensities Differ by TAS2R38 Haplotype. *J Nutrigenet Nutrigenom*. 2014a; 7(3):143–152.
- Fischer ME, Cruickshanks KJ, Schubert CR, Pinto A, Huang GH, Klein BE, Klein R, Pankow JS. The association of taste with change in adiposity-related health measures. *J Acad Nutr Diet*. 2014b; 114(8):1195–1202. [PubMed: 24882206]
- Gwet K. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. *Stat Methods Inter Rater Reliab Assess*. 2002; 2:1–9.
- Hayes JE, Duffy VB. Oral sensory phenotype identifies level of sugar and fat required for maximal liking. *Physiol Behav*. 2008; 95(1–2):77–87. [PubMed: 18538361]
- Hayes JE, Sullivan BS, Duffy VB. Explaining variability in sodium intake through oral sensory phenotype, salt sensation and liking. *Physiol Behav*. 2010; 100(4):369–380. [PubMed: 20380843]
- Hoffman HJ, Ishii EK, MacTurk RH. Age-related changes in the prevalence of smell/taste problems among the United States adult population. Results of the 1994 disability supplement to the National Health Interview Survey (NHIS). *Ann N Y Acad Sci*. 1998; 855:716–722. [PubMed: 9929676]
- Karpa MJ, Gopinath B, Rochtchina E, Jie Jin W, Cumming RG, Sue CM, Mitchell P. Prevalence and neurodegenerative or other associations with olfactory impairment in an older community. *J Aging Health*. 2010; 22(2):154–168. [PubMed: 20133956]
- Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*. 2003; 299(5610):1221–1225. [PubMed: 12595690]
- Krantz EM, Schubert CR, Dalton DS, Zhong W, Huang GH, Klein BE, Klein R, Nieto FJ, Cruickshanks KJ. Test-retest reliability of the San Diego Odor Identification Test and comparison with the brief smell identification test. *Chem Senses*. 2009; 34(5):435–440. [PubMed: 19363087]
- Kveton JF, Bartoshuk LM. The effect of unilateral chorda tympani damage on taste. *Laryngoscope*. 1994; 104(1 Pt 1):25–29. [PubMed: 8295453]

- McAnally HM, Poulton R, Hancox RJ, Prescott J, Welch D. Psychosocial correlates of 6-n-propylthiouracil (PROP) ratings in a birth cohort. *Appetite*. 2007; 49(3):700–703. [PubMed: 17766007]
- Minski KR, Duffy VB. Odor intensity, diet and nutrition-related health indices among females. Abstracts from the Thirty-first Annual Meeting of the Association for Chemoreception Sciences. *Chem Senses* 34. 2009:A37–A38. doi:10.1093/chemse/bjp032.
- Murphy C, Schubert CR, Cruickshanks KJ, Klein BE, Klein R, Nondahl DM. Prevalence of olfactory impairment in older adults. *JAMA*. 2002; 288(18):2307–2312. [PubMed: 12425708]
- National Institutes of Health (NIH), NIH Blueprint for Neuroscience Research. NIH Toolbox For the Assessment of Neurological and Behavioral Function. [9 Mar 2015] 2012. <http://www.nihtoolbox.org/WhatAndWhy/Assessments/NIH.ToolboxBrochure-2012.pdf>.
- Pepino MY, Bradley D, Eagon JC, Sullivan S, Abumrad NA, Klein S. Changes in taste perception and eating behavior after bariatric surgery-induced weight loss in women. *Obesity (Silver Spring)*. 2014; 22(5):E13–E20. [PubMed: 24167016]
- Rawal S, Hayes JE, Wallace MR, Bartoshuk LM, Duffy VB. Do polymorphisms in the TAS1R1 gene contribute to broader differences in human taste intensity? *Chem Senses*. 2013; 38(8):719–728. [PubMed: 24000232]
- Rawal S, Hoffman HJ, Chapo AK, Duffy VB. Sensitivity and Specificity of self-reported olfactory dysfunction in a home-based study of independent-living, healthy older women. *Chemosens Percept*. 2014; 7(3–4):108–116. [PubMed: 25866597]
- Reed DR, Zhu G, Breslin PA, Duke FF, Henders AK, Campbell MJ, Montgomery GW, Medland SE, Martin NG, Wright MJ. The perception of quinine taste intensity is associated with common genetic variants in a bitter receptor cluster on chromosome 12. *Hum Mol Genet*. 2010; 19(21):4278–4285. [PubMed: 20675712]
- Schubert CR, Cruickshanks KJ, Fischer ME, Huang GH, Klein BE, Klein R, Pankow JS, Nondahl DM. Olfactory impairment in an adult population: the Beaver Dam Offspring Study. *Chem Senses*. 2012; 37(4):325–334. [PubMed: 22045704]
- Segura B, Baggio HC, Solana E, Palacios EM, Vendrell P, Bargallo N, Junque C. Neuroanatomical correlates of olfactory loss in normal aged subjects. *Behav Brain Res*. 2013; 246:148–153. [PubMed: 23458742]
- Simchen U, Koebnick C, Hoyer S, Issanchou S, Zunft HJ. Odour and taste sensitivity is associated with body weight and extent of misreporting of body weight. *Eur J Clin Nutr*. 2006; 60(6):698–705. [PubMed: 16435003]
- Vennemann MM, Hummel T, Berger K. The association between smoking and smell and taste impairment in the general population. *J Neurol*. 2008; 255(8):1121–1126. [PubMed: 18677645]
- Wehling E, Nordin S, Espeseth T, Reinvang I, Lundervold AJ. Unawareness of olfactory dysfunction and its association with cognitive functioning in middle aged and old adults. *Arch Clin Neuropsychol*. 2011; 26(3):260–269. [PubMed: 21474482]

Fig. 1. Percentage agreements for individual odor identification between NHANES Pocket Tests administered at baseline and at 2-week follow-up

Fig. 2. Mean reported intensities (with error bars for SEM) on general Labeled Magnitude Scale for NHANES Pocket Smell Test odorants at baseline and at 2-week follow-up (* $p < 0.05$)

Table 1

Mean gLMS intensity ratings for all baseline taste measures and corresponding partial correlations with whole-mouth ratings for 1-mM quinine

Taste measures	Taste intensity (mean±SD) ^a	Partial correlation with 1 mM quinine whole-mouth ^b
1 mM quinine tongue tip (NHANES)	23.05±13.14	0.44 **
1 M NaCl tongue tip (NHANES)	29.84±14.13	0.53 **
1 mM quinine whole-mouth (NHANES)	49.29±20.67	1
1 M NaCl whole-mouth (NHANES)	50.81±18.97	0.77 **
0.32 M NaCl whole-mouth (NHANES)	34.73±14.20	0.60 **
1 M sucrose	43.24±20.68	0.63 **
32 mM citric acid	39.56±20.52	0.58 **
3.2 mM PROP	45.19±30.41	0.34 *

^aWord indicators on general Labeled Magnitude Scale (gLMS): “barely detectable” (1.4), “weak” (6), “moderate” (17), “strong” (34), and “very strong” (53)

^bPartial correlations with whole-mouth intensity of 1-mM quinine after controlling for age, sex, and 86-dB tone ratings

*
 $p < 0.05$

**
 $p < 0.001$

Table 2

Intraclass correlation coefficients (single measures, one-way random) for NHANES taste measures and cross-modal standards across 2 weeks and/or 6 months

	Intraclass correlation coefficients with 95 % confidence intervals ^a	Agreement
Taste measures (across 2 weeks)		
1 mM quinine tongue tip	0.58 (0.40–0.72)	Moderate
1 M NaCl tongue tip	0.51 (0.31–0.67)	Moderate
1 M NaCl whole mouth	0.55 (0.35–0.70)	Moderate
1 mM quinine whole mouth	0.71 (0.57–0.81)	Good
0.32 M NaCl whole mouth	0.47 (0.26–0.64)	Moderate
Taste measures (across 6 months)		
1 mM quinine tongue tip	0.47 (0.23–0.67)	Moderate
1 M NaCl tongue tip	0.48 (0.25–0.68)	Moderate
1 M NaCl whole mouth	0.49 (0.26–0.68)	Moderate
1 mM quinine whole mouth	0.55 (0.30–0.72)	Moderate
0.32 M NaCl whole mouth	0.42 (0.24–0.63)	Moderate
Sound tones		
50 dB	0.37(0.14–0.56)	Poor
62 dB	0.44 (0.23–0.62)	Moderate
74 dB	0.47 (0.26–0.64)	Moderate
86 dB	0.53 (0.33–0.69)	Moderate
98 dB	0.63 (0.45–0.76)	Moderate
Remembered sensations		
Dimly lit restaurant	0.44 (0.22–0.61)	Moderate
Well-lit room	0.36 (0.13–0.55)	Poor
Brightest light ever seen	0.44 (0.22–0.61)	Moderate

^a Interpretation of ICCs: <0.40 (poor agreement), 0.40–0.59 (fair agreement), 0.60–0.74 (good agreement), 0.75 (excellent agreement)

Table 3

Olfactory function classification by olfactometer at baseline and NHANES Pocket Smell test at baseline and 2-week follow-up

Pocket Score Baseline Classification	Olfactometer Classification	
	Normosmia	Dysfunction
Normosmia	63	4
Dysfunction	0	6
Pocket Score Baseline Classification	Olfactometer Classification	
	Mild microsmia-normosmia	Moderate-to-severe dysfunction
Normosmia	67	0
Dysfunction	1	5
Pocket Score Baseline Classification	Pocket Score 2nd Visit Classification	
	Normosmia	Dysfunction
Normosmia	66	1
Dysfunction	0	3

Table 4

Six-month test-retest reliability measures for NHANES Taste and Smell Questionnaire (CSQ) items

CSQ items	Agreement coefficient (AC_1) ^a	Agreement
Smell problems during past 12 months (yes/no)	0.86 (0.78–0.95)	Excellent
Smell loss since 25 years old (better/worse/no change)	0.90 (0.80–0.99)	Excellent
Taste problems during past 12 months (yes/no)	0.78 (0.64–0.92)	Excellent
Salt taste loss since 25 years old (better/worse/no change)	0.76 (0.53–0.99)	Excellent
Sour taste loss since 25 years old (better/worse/no change)	0.86 (0.75–0.97)	Excellent
Sweet taste loss since 25 years old (better/worse/no change)	0.69(0.51–0.87)	Good
Bitter taste loss since 25 years old (better/worse/no change)	0.76 (0.52–0.99)	Excellent
Food flavor loss since 25 years old (yes/no)	0.66 (0.52–0.80)	Good

^a Interpretation of AC_1 : <0.40 (poor agreement), 0.40–0.59 (fair agreement), 0.60–0.74 (good agreement), 0.75 (excellent agreement)