Although a large number of head trauma patients evidence significant alterations in the ability to taste and smell, quantitative assessment of their dysfunction is rarely performed. This is unfortunate, since such assessment is critical for establishing the validity of the complaint, characterizing the specific nature of the problem, objectively monitoring changes in function over time, including those resulting from medical interventions or treatments, detecting malingering, and establishing compensation for disability. In this article, a discussion of the theory and methods for objectively testing the senses of taste and smell is presented, along with a number of examples of the application of various chemosensory tests in the clinical setting.

DECREASED ABILITY to taste or smell is common in individuals who have suffered head injury. Indeed, in a recent study of 750 consecutive patients presenting to the University of Pennsylvania Smell and Taste Center with chemosensory dysfunction, nearly 20% had olfactory dysfunction as a result of head trauma. Although most trauma patients who complain of decreased taste function are, in fact, experiencing decreased flavor sensation as the result of loss of olfactory stimulation via the retronasal route, subtle alterations in the ability to perceive sweet, sour, bitter, and salty stimuli can be detected in a number of these individuals using modern psychophysical taste tests.

This article discusses practical and theoretical issues related to the clinical assessment of taste and smell and provides the practitioner with information helpful in patient assessment. First, procedures for presenting olfactory and gustatory stimuli are described. Next, the concepts of test reliabil-

Richard L. Doty, PhD
Director
Smell and Taste Center
Hospital of the University of Pennsylvania
Associate Professor
Department of Otorhinolaryngology
Head and Neck Surgery
School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania

This work was supported by Grant No. P01 DC 00161 from the National Institute on Deafness and Other Communication Disorders and Grant No. ROI AG 08148 from the National Institute on Aging.

J Head Trauma Rehabil 1992;7(1):47-65
ity and validity are discussed, with the goal of sensitizing the reader to their importance in chemosensory evaluation. Psychophysical methods used to measure taste and smell function are then described, followed by a discussion of the utility of unilateral and bilateral testing procedures and strategies for detecting malingering. Finally, the psychometric properties of currently available tests, including the degree to which they measure independent components of chemosensation, are reviewed.

STIMULUS CONTROL AND PRESENTATION

Many types of equipment have been described for presenting chemosensory stimuli to patients. Fortunately, the exact number of molecules entering the nose or reaching the taste buds need not be known for a clinical test to be useful, so long as the stimuli are presented in a reliable manner and norms are available to establish whether the obtained responses are normal or abnormal. Thus, accurate clinical assessment of chemosensory function can be made using surprisingly simple stimulus-presentation equipment coupled with sound psychophysical procedures.

In the case of olfaction, equipment traditionally used to present stimuli to patients includes

- the draw-tube olfactometer of Zwaardemaker,
- glass sniff bottles,
- glass rods, wooden sticks, or strips of blotter paper dipped in various concentrations of odorants,
- plastic squeeze bottles,
- air-dilution olfactometers (which can be very complex),
- microencapsulated "scratch and sniff" odorized strips, and
- bottles from which blasts of saturated air are presented.

The first six of these are illustrated in Figs 1 and 2. The strengths and weaknesses of most of these stimulus-presentation procedures are discussed elsewhere.

---

Diagnostic Tests and Assessment 49

Fig 2. A subject sniffing an odorant released from a microencapsulated strip of the University of Pennsylvania Smell Identification Test, a widely used modern clinical test. This 40-odorant test consists of four 10-page booklets. Each page contains a different "scratch and sniff" scented strip and an associated multiple-choice question. The stimuli are released using a pencil tip.

Gustatory stimuli are usually presented in deionized water. "Whole-mouth" procedures require the patient to sip a small amount of stimulus solution from a paper cup, swish the solution in the mouth, and expectorate. Rinsing with deionized or spring water between trials eliminates residuals. "Regional" testing of the anterior tongue (chorda tympani branch of cranial nerve [CN] VII), posterior tongue (CN IX), and in some instances, the palate (greater superficial petrosal branch of CN VII) can be performed in the clinic using

- small amounts of stimulus solution pipetted to the selected areas,
- Q-tip application of stimulus solution,
- microampere levels of current (usually anodal) applied to the tongue surface (termed electrogustometry).

Intravenous administration of odorants and tastants can also be used to produce noticeable chemosensory sensations. This procedure has been used primarily in Japan for clinically assessing olfactory function (see Takagi for review). The clinical paradigm is to inject thiamine propyldisulfide (Alinamin) slowly into the median cubital vein and to record the duration and latency of the onset of a garlic-like sensation experienced by the patient. The degree to which the sensation is due to exposure of the receptors to odorized lung air or diffusion of the stimulus from nasal capillaries to olfactory receptors is unknown, although studies using tracheotomized rats and turtles suggest that both probably occur and that the phenomenon is receptor mediated.

TEST RELIABILITY AND VALIDITY: BASIC CONSIDERATIONS

The utility of a sensory test depends upon the degree to which it is reliable (consistent, dependable, or stable) and valid (accurately measures what it portends to measure). Related to its validity are its sensitivity (ability to detect abnormalities when present) and specificity (ability to detect abnormalities with a minimum of false positives). Although empirical indices of these characteristics are available for a number of medical and psychologic tests, this is unfortunately not the case for the vast majority of chemosensory tests.

A test's reliability can be determined in several ways. First, and most commonly, the same test is administered on two occasions to a group of subjects and a correlation coef-
ability is a necessary but not sufficient condition for validity.

Three concepts related to a test's validity have been particularly emphasized by epidemiologists: (1) sensitivity, (2) specificity, and (3) positive predictive value. A test's sensitivity reflects the proportion of individuals within a population who have a particular disorder (e.g., anosmia) that score positively on the test (in effect, the "hit rate" of the test). A test's specificity reflects the proportion of persons without the disorder who have a negative test result (in effect, the "correct rejection" rate of the test); whereas the test's positive predictive value is the proportion of all positive tests that are true positives.

**CLINICAL TESTS OF CHEMOSENSORY FUNCTION**

Numerous tests of chemosensory function have been developed over the years (for reviews, see 21-23,31-36). A subset of these tests has made its way into the clinic, where time is at a premium and the use of complex stimulus-presentation equipment is not practical. Such tests can be divided into two types: (1) psychophysical and (2) psychophysiologic. In the latter category are stimulus-induced cortical-evoked potentials and autonomic measures of reactivity, such as changes in heart rate. Only psychophysical tests are described in this article; the reader is referred to the literature on taste- and odor-evoked potentials and to the literature on autonomic changes associated with odorant stimulation.21-23,33-36

**Detection- and recognition-threshold tests**

The most popular class of procedures used to assess chemosensory function establishes an operational measure of the lowest concentration of a stimulus discernible to a sub-
Diagnostic Tests and Assessment

Object, that is, the absolute threshold measure. In the case of olfaction, a qualitative odor sensation is rarely perceived at very low odorant concentrations (eg, "rose-like"), and subjects only experience the faint presence of something. The lowest concentration where such a presence is noted is termed the detection threshold, whereas the lowest concentration where odor quality is discernible is the recognition threshold. It is now common on a detection-threshold task to ask a subject to indicate, on a given trial, which of two or more stimuli (eg, an odorant and one or more blanks) smells strongest (the forced-choice procedure), rather than to present just the odor stimulus and ask whether it is perceived or not. Forced-choice procedures are less susceptible to response biases (ie, the conservatism or liberalism in reporting the presence of an odor under uncertain conditions) and are more accurate and reliable than non-forced-choice procedures.  

Recognition thresholds can be obtained in a manner similar to detection thresholds by having the patient indicate the lowest concentration at which the target odor or taste quality is perceived. Unfortunately, criterion biases are difficult, if not impossible, to control in recognition threshold paradigms. Thus, even in a forced-choice situation, guesses are not randomly distributed among alternatives, thereby biasing the threshold value. The classic example of this is in taste psychophysics, where there is a tendency for some subjects to report "sour" more often than the other primary qualities when no stimulus is clearly discernible, resulting in a spuriously low sour-taste recognition-threshold measure.  

Two threshold procedures have received the widest usage in the clinic: (1) the ascending method of limits and (2) the single staircase. In the ascending method of limits, stimuli are presented sequentially from low to high concentrations, and the point of transition between detection and no detection is determined to provide an estimate of threshold concentration. Forced-choice responses are required on each trial. In the single staircase method (a variant of the method of limits technique), the stimulus concentration is increased, following trials on which a subject fails to detect the stimulus, and decreased, following trials where correct detection occurs.  

An example of a clinical use of the method of limits-threshold procedure is provided by Cain. This investigator used specially constructed 60-mL glass sniff bottles to present, in a two-alternative forced-choice situation, either diluent (water) or odorant (n-butanol dissolved in water) to 43 patients with varying degrees of olfactory dysfunction using four repeated ascending series. Testing was performed for each nasal chamber separately. This procedure, which took approximately half an hour per patient, was found to clearly discern between anosmic, hyposmic, and normosmic subjects and demonstrated that olfactory dysfunction is typically bilateral.  

An example of the use of a modified two-alternative forced-choice staircase procedure comes from a study of patients with early Alzheimer's disease. In this study, a trial consisted of the presentation of two 100-mL glass sniff bottles to the subject in rapid succession. One contained 20 mL of a given concentration of phenyl ethyl alcohol (PEA) dissolved in USP-grade light mineral oil, whereas the other contained mineral oil alone. The subject's task was simply to report which stimulus was strongest. The first trial commenced at the -6.50 log (liquid volume/volume) concentration step. If a miss occurred on any trial before five successive blank/odorant pairs were completed, the next trial began at the concentration one log step higher. When five consecutive correct trials occurred at a given concentration, the
staircase was "reversed," and the subsequent pair of trials was presented at a concentration 0.5 log step lower. However, from this point on, only one or two trials were presented at each step (ie, if the first trial was missed, the second was not given, and the staircase was moved to the next higher 0.5 log step concentration). If both trials were correct, the staircase was moved down one 0.5 log-unit step. The geometric mean of the last four of seven staircase-reversal points was used as the threshold estimate. Compared to age-matched control subjects, the Alzheimer's disease patients had significantly higher threshold values.

It should be noted that threshold values produced by the aforementioned staircase procedure are not mathematically equivalent to those produced by most other methods since, after the initial reversal point, two correct trials are required for the staircase to descend and only one incorrect trial is required for the staircase to ascend. As discussed by Wetherill and Levitt, the threshold concentration determined by this method represents the value chosen correctly by the subject 71% of the time.

**Signal-detection tests**

Signal-detection theory is an approach to sensory measurement that differs radically from the approach used in threshold testing, in that it does not accept the classic concept of a threshold and focuses on the importance of subject expectancies and rewards in influencing the detection decision. Unlike classic threshold procedures, methods based on this theory provide a measure of response bias, as well as a measure of sensitivity theoretically unconfounded by response bias. However, with rare exception, sensitivity procedures based on signal-detection theory have found little applicability in the clinical setting, since they require large numbers of trials and are most applicable in studies where repeated measures on the same subject over time are obtained. For these and other reasons, they are not discussed in detail here.

**Difference-threshold tests**

In classic psychophysics, the concentration by which a stimulus must be changed to make it perceptibly stronger or weaker is termed a *just noticeable difference* (JND). According to Weber's law, the size of the increment in odorant concentration (ΔI) required to produce a JND increases as the comparison concentration (I) increases, with the ratio being a constant; that is, ΔI/I = K. Although K, in fact, is rarely a constant (being influenced by the size of I at extremes of the continuum and by such factors as the stimulus used), it is a rough index of a sensory system's sensitivity; the smaller the K value, the greater the sensory discrimination.

Eichenbaum et al devised a relatively brief clinical test to assess the discrimination ability of a group of neurologically normal subjects and the famous patient HM (whose amygdala, uncus, and anterior two-thirds of the hippocampus and parahippocampal gyrus had been removed bilaterally to control cerebral seizures). In this test, 10 binary dilutions (in water) of each of 4 stimuli (acetone, ethanol, almond extract, and lemon extract) were used. Subjects were initially presented with the highest and lowest concentrations of a given odorant and were required to choose the stronger stimulus. Successively stronger stimuli were paired with the strongest stimulus until, on the last of the 10 trials, the 2 samples were identical. The difference threshold was defined as the lowest concentration for which discrimination up to and including that dilution was errorless.

**Suprathreshold-scaling procedures**

Odors and tastes can be described in terms of a number of psychologic attributes, including strength, pleasantness, and quality.
Since sensation magnitude varies as a function of stimulus concentration, ratings or other measures of stimulus intensity have proved useful in evaluating taste and smell function. In general, the intensity of a stimulus is related to the number of neurons that are recruited and the frequency at which they fire; therefore, perceived intensity is presumed to be a sensitive index of the extent of neural damage in afferent pathways. Suprathreshold tests have an advantage over threshold tests in that they are often brief and easy to administer; furthermore, they are less susceptible to subtle stimulus contamination that can occur at low odorant and tastant concentrations.

Although psychophysicists and psychometrists seek to develop psychologic scales with ruler-like properties (i.e., where distances along the scale can be thought of as ratios and a true zero point is present), the degree to which they have been successful in the chemosensory sciences is debatable. For the purposes of clinical testing, however, the exact form of underlying psychologic scales is not particularly important, so long as responses on standardized scaling tasks are reliable and can differentiate among persons with various degrees of dysfunction. Thus, the fact that judgments of the intensity of both odors and tastes are relative and influenced by contextual factors (e.g., a moderately intense odor is judged as more intense when presented with weak comparison stimuli than when presented with strong comparison stimuli) is not of great concern in the development or administration of a clinical test, so long as the test and the test-taking procedures are standardized.

Rating scales are widely used to assess the relative amount of a psychologic attribute perceived by a subject. In chemosensory assessment, two types are popular: (1) category scales, where the subject signifies the relative amount of a sensation by indicating which of a series of discrete categories best describes the sensation, and (2) line scales (also termed visual analog or graphic scales), where the magnitude of the sensation is signified by placing a mark along a line that has descriptors (termed "anchors") located at its extremes (e.g., very weak—very strong). The reader is referred elsewhere to discussions regarding the optimal number of categories and other theoretic aspects of rating scales.

Intensity matching procedures have also been used in clinical studies. Since intramodal intensity matching procedures have never been used in a clinical context for either taste or smell assessment, only cross-modal intensity matching procedures are considered here.

In cross-modal matching, the relative magnitude of the sensory attribute is signified for each member of a stimulus set by using some other sensory modality or cognitive domain. A key distinction between this procedure and rating-scale procedures is that the subject or patient attempts to estimate the ratio relations among the intensities of the different concentrations of stimuli that are presented, without being constrained by categories or a short response line. Continua commonly used in this task include distance (e.g., pulling a tape measure a distance proportionate to the odor's intensity) and number (e.g., assigning numbers proportionate to the odor's intensity), the latter procedure being termed magnitude estimation. When subjects judge intensities of sensations from two or more modalities on a single common scale, the procedure is termed the method of magnitude matching. Since magnitude estimation and magnitude matching are among the most commonly used cross-modal matching procedures, they are discussed in detail here.

In magnitude estimation, the patient or subject assigns numbers relative to the magnitude of the sensations. For example, if the number 50 is used to indicate the intensity of
one concentration of a tastant, a concentration that tastes 4 times as intense would be assigned the number 200. If another concentration is perceived to be half as strong as the initial stimulus, it would be assigned the value 25. The examinee can assign any range of numbers to the stimuli, so long as they reflect the relative magnitudes of the perceived intensities. In some cases, a standard for which a number has been preassigned (often the middle stimulus of the series) is presented to the subject in an effort to make the responses more reliable. In other cases, this is not done (the "free-modulus" method), and the individual is free to choose any number system desired, so long as the numbers are made proportional to the magnitude of the attribute. For example, one patient may choose to assign the first stimulus the number 200, whereas another may choose to assign this same stimulus the number 10. If a second stimulus is perceived to be 10 times stronger than the first by each of these individuals, the first one would assign the number 2000, whereas the second one would assign the number 100. The important point is that the absolute value of the number is not important, only the ratios between the numbers.

Intensity magnitude estimation data are most commonly analyzed by plotting the log-magnitude estimates on the ordinate and the log-odorant concentrations on the abscissa and fitting the best-fit line to the data using linear regression. The resulting function, log \( P = n \log \sigma + \log k \) (where \( P \) = perceived intensity, \( k \) = the Y intercept, \( \sigma \) = stimulus concentration, and \( n \) = the slope), can be represented in its exponential form as a power function, \( P = k \sigma^n \), where the exponent \( n \) is the slope of the function on the log-log plot. In olfaction, \( n \) varies in magnitude from odor to odor but is generally less than one, reflecting a negatively accelerated function on linear-linear coordinates. As discussed elsewhere, modifications have been made to these equations in attempts to take into account threshold sensitivity and adaptation.\(^{31,57}\)

It is important to note that magnitude estimation, perhaps more so than a number of other sensory procedures, can be biased or influenced in systematic ways by procedural and subject factors.\(^{31,58}\) The task is comparatively complex in that accurate responses to a stimulus require a good memory for the prior stimulus and, ideally, for a number of the stimuli that are being judged. If too much time lapses between the presentation of stimuli, the memory of the prior stimulus fades. On the other hand, if the trials are spaced too closely together, adaptation can distort the relationship. There is evidence that not all subjects consistently provide ratio estimates of the stimuli, and a number do not understand the concept of producing ratios.\(^{39,60}\) Furthermore, the magnitude of the exponent is dependent upon the choice of the stimulus scale (i.e., the units in which the stimulus concentration is expressed), although in olfaction this is probably of minor consequence.\(^{61}\) However, these and other potential shortcomings need not hinder the use of this procedure clinically, so long as the instructions, test procedures, and test stimuli are carefully standardized and monitored, and reliable responses are obtained. Nevertheless, comparative assessments of nine-point rating scales, line scales, magnitude estimation scales, and a hybrid of category and line scales suggest that, for relatively untrained or mathematically unsophisticated subjects, category scales and line scales may be superior to magnitude estimation when such factors as variability, reliability, and ease-of-use are considered.\(^{62,63}\)

Although investigators typically view the slope of the magnitude estimation function (on log-log coordinates) as a key indicator of sensory function, this is not the case with the
intercept and other parameters associated with the function’s position relative to the ordinate. The intercept is not independent of the slope and, more importantly, is dependent upon differences in the use of numbers by subjects.

The method of magnitude matching, unlike that of magnitude estimation, provides information about the perceived overall intensity of stimuli from the absolute position of the magnitude estimation function and corrects, to some degree, for differences among subjects in number usage (for a discussion of strengths and weakness of this procedure, see Marks et al.46). In the most common application of this method, judgments of the intensity of sensations from two modalities (e.g., loudness, odor intensity) are made on a common magnitude estimation scale.46 Under the assumption that subjects experience stimuli on one of the continua (e.g., loudness) in a similar manner, one can assume that differences among their loudness ratings reflect differences in number usage. The odor intensity continuum can then be adjusted accordingly. Such normalization allows, theoretically, for a direct comparison of scale values across subjects; thus, if the adjusted odor intensity magnitude value for 1 subject is 10 and another subject is 20 at the same concentration level, the 2nd subject is presumed to experience twice the odor intensity as the 1st subject.

An example of cross-modal matching in a study of electric taste is presented by Salata et al.46 In this study, 12 subjects estimated, during the same test sessions, the intensity of one-second duration white-noise pulses (at 1000 Hz ranging from 40 to 60 dB SPL in 5 dB increments) presented binaurally and half-second electric current pulses (40, 64, 100, 160, and 250 microamperes) delivered to the tongue tip, anterior tongue side, posterior tongue side, the posterior medial tongue, and the soft palate. Since the data from the two sides of the tongue were similar, they were combined. As shown in Fig 3, it is apparent that the tip of the tongue is significantly more sensitive than the other areas to electrical stimulation, as evidenced both by the slope and the absolute position of the psychophysical functions.

Quality-discrimination tests

The most straightforward chemosensory quality-discrimination test requires individuals to decide whether two stimuli have the same or different quality. In the case of odors, a series of same-odorant and different-odorant pairs are presented, and the measure of discrimination is calculated as the proportion of pairs that are correctly differentiated.46,48 Variants on this theme include...
tests in which a stimulus differs from a set of other stimuli that, in turn, are equivalent, and the task is to select the "odd" stimulus (eg, the so-called triangle test). A variant on this theme is the stimulus-matching task, in which a set of stimuli are provided and the patient is required to match the stimuli, one by one, to those of a set of identical stimuli. Abraham and Matha provide an example of this test, in which four odorants are contained in eight vials (two vials per odor), and the patient is asked to pair up the equivalent two-vial containers. The number of pairs correctly matched on each of two administrations of the test is used as the test score. If a discrepancy arises in the two tests, a third administration is given, and the data are averaged to provide the final score.

Two general classes of quality-recognition tests can be envisioned. In the first class, a series of odorants or tastants are presented, and the subject is simply asked on each presentation whether or not a stimulus is recognized. Identification is not required. This relatively crude procedure, which is analogous to testing vision by shining a flashlight in the eye and asking the examinee if light was seen, is unfortunately one of the most common tests used by medical practitioners to ascertain CN I function. In the second class, a patient is presented with a "target" stimulus and is subsequently asked to select the target from a larger set of stimuli. This type of test uses the number of correct responses in a series of several such presentations as the test score.

Among the most popular procedures for assessing taste and smell function are those that require stimulus-quality identification. Such tests can be divided into three major types: (1) naming tests, (2) yes/no identification tests, and (3) multiple-choice identification tests. The respective responses required, on a given trial, in these three types of tests are

- to provide a name for the stimulus;
- to signify whether or not the stimulus smells like an object named by the examiner (eg, does this smell like a rose?); and
- to identify the stimulus from a list of names.

Odor-naming tests, although they have been used clinically, have limited usefulness, since many normal individuals have difficulty in naming or identifying even familiar odors without cues. This problem is less acute in taste-naming tests, although a considerable proportion of individuals in the general population exhibit sour-bitter and bitter-salty confusions.
Yes/no identification tests require a patient only to report whether or not each of a set of stimuli smells (or tastes) like a particular substance named by the experimenter. Two trials with each stimulus are usually given, with the correct alternative provided on one trial and an incorrect one on the other (eg, lemon odor is presented and the subject is asked on one trial whether the odor smells like lemon and on another trial whether the odor smells like peppermint). Although this test requires the subject to keep the percept in memory long enough to compare it with the target word (which, of course, must also be recalled from memory), some of its proponents argue that it is less influenced by cognitive and memory demands than multiple-choice identification tests (see following). However, since chance performance on this type of test is 50% compared to 25% on a 4-alternative multiple-choice identification test, its range of discriminability is lower and, therefore, more trials are needed to obtain the same statistical power.

A number of multiple-choice odor-identification tests have been described in the clinical literature. These tests are conceptually similar and, in the few cases that have been examined, strongly correlated. The most widely used of these tests (the University of Pennsylvania Smell Identification Test or UPSIT, commercially termed the Smell Identification Test™, Sensonics, Inc, Haddonfield, NJ) examines the ability of subjects to identify, from sets of 4 descriptors, each of 40 "scratch and sniff" odorants (Fig 2). The number of correct items out of 40 serves as the test measure; this value is compared to norms and a percentile rank is determined, depending upon the age and gender of the patient. This test has several unique features, including amenability to self-administration and a means for detecting malingering.

Wright has recently developed an odor identification confusion matrix applicable to clinical settings (see also Koster). In this test, each of 10 stimuli are presented to a patient in counterbalanced order 10 times apiece. The response alternatives are the names of the 10 stimuli: (1) ammonia, (2) Clorox, (3) licorice, (4) mothballs, (5) peppermint, (6) roses, (7) turpentine, (8) vanilla, (9) Vicks vapor rub, and (10) vinegar. No feedback is given to the examinee concerning the correctness or incorrectness of the responses. The data are presented as the percentage of responses given to each alternative for each odorant and displayed in a rectangular matrix (stimuli making up rows and response alternatives making up equivalently ordered columns). Responses that fall along the negative diagonal represent correct responses, whereas those that fall away from the diagonal represent "confusions." The percentage of responses that are correct serves as the primary test measure, although Wright argues that the "off-diagonal" responses (confusions) may provide meaningful clinical information.

Memory tests

A basic odor or taste-recognition memory test requires a subject to taste or smell a stimulus (termed the target or inspection stimulus) and then to select, after an interval of time (eg, 30 seconds), that stimulus from a set of foils. Repeated trials are often performed at one or more retention intervals for each of several stimuli. In an effort to minimize the rehearsal of verbal labels for delay intervals longer than 10 seconds, the examinee is sometimes asked to perform an unrelated task, such as counting backwards, during the retention period. The proportion of trials where correct performance occurs is commonly used as the test measure.

An example of a clinical application of an odor recognition task is described by Jones, Moskowitz, and Butters. These authors presented 20 pairs of odorants at 2 delay inter-
vals (0 and 30 seconds) to 14 alcoholic Korsakoff’s psychosis patients, 14 alcoholic controls, and 14 nonalcoholic controls. On a given trial, the task of each subject was to report whether the second stimulus of a pair of odors was the same or different than the first. In the 30-second delay interval, the subjects counted backwards by 3s from a 3-digit number given by the experimenter after the presentation of the first stimulus of the pair. The Korsakoff’s psychosis patients performed significantly more poorly than did the control groups at both the 0- and 30-second retention intervals.

Campbell and Gregson developed a test of short-term odor memory in which a patient was given four odors in a row and asked if the fourth odor, which was the same as one of the first three, was equivalent to the first, second, or third presented odorant. No delay interval, per se, was defined between the presentation of the stimuli, but presumably the trials were presented soon after one another. Seven 3-odor combinations of 12 inspection stimuli were given. If patients had difficulty with this task, they were then given two-odor combinations. The test score value was the number of odors that were consistently recognized by the subject. This test has been shown to be sensitive to deficits observed in schizophrenia, Kallmann’s syndrome, and Korsakoff’s psychosis.

OTHER CONSIDERATIONS

Unilateral and bilateral olfactory testing

Most patients evidence olfactory dysfunction bilaterally. When unilateral losses are present, they often go unnoticed and appear to be of little significance to the patient. Since bilateral testing requires less time and results in the detection of clinically meaningful deficits, it is often selected for assessing olfactory function. However, there are a number of occasions when unilateral olfactory testing is of considerable value, for example in head injury and the detection of rare tumors. A complete workup of a patient should include unilateral, as well as bilateral, olfactory testing.

Unilateral testing is straightforward. Although it is possible to present a stimulus to one naris and obtain unilateral stimulation, the possibility of the crossing of odorant to the contralateral side within the nasopharynx upon exhalation cannot be excluded. Thus, it is prudent to close the contralateral naris without distorting the septum (eg, by using a piece of Microfoam™ tape [3M Corporation, Minneapolis, Minn] cut to fit tightly over the borders of the naris). This precaution serves to make it difficult for air to enter the blocked nasal chamber via the retronasal route.

Furukawa et al report that 7 of 94 patients they examined evidenced significant unilateral threshold deficits, despite the fact that none evidenced a deficit on bilateral testing. A similar phenomenon was observed in 6 of 12 patients who had received brain surgery. A recent assessment of 82 consecutive nonanosmic patients presenting to our center with chemosensory dysfunction found 14 (ie, 17%) whose unilateral detection-threshold values were discrepant from one another by at least 3 orders of magnitude (Doty, unpublished data). Interestingly, 9 of these 14 individuals appeared to be anosmic on one side of the nose, even though only 3 had bilateral detection-threshold values that were considered abnormal.

Removal of temporal lobe tissue for control of intractable seizure activity typically produces greater olfactory dysfunction on the side ipsilateral to the operation, reflecting the fact that most olfactory projections from the olfactory bulb are ipsilateral. Interestingly, head injuries that produce damage to the right hemisphere sometimes result in unilateral neglect to the left side of the body that
Detection of malingering

Malingering on chemosensory tasks is not uncommon, largely because considerable compensation can be available in accident cases for alterations in the ability to taste or smell. It is frequently suggested in the medical literature that if a patient cannot readily perceive the vapors from an irritating substance presented to the nose, the patient is malingering. However, this is not a surefire means for detecting malingering. Thus, individuals who, on other grounds, are believed to be feigning anosmia usually have difficulty in denying experiencing the effects of ammonia or other irritants, particularly since these stimuli often produce eye watering, coughing, and reflexive responses that are manifested overtly. Furthermore, there appears to be considerable variability in trigeminal responsiveness to such stimuli within the normal population.

A more valid approach for detecting malingering is to examine response strategies of patients on forced-choice tests. For example, in a taste-threshold test in which a blank and a taste stimulus are presented in random order on each trial, an individual with no ability to detect the stimuli should be correct, by chance alone, on approximately 50% of the stimulus trials. However, malingerers often avoid the correct response more often than expected on the basis of chance and, for this reason, can be detected. This is well illustrated in olfaction by responses to the UPSIT. Since this is a four-alternative forced-choice test, one would expect that approximately 25% of the test items (ie, 10) should be correctly answered, on average, by an anosmic. The probability of scoring 5 or less on the UPSIT and not having at least some ability to smell is less than 5 in 100. The probability of scoring zero on the UPSIT and having no sense of smell is approximately 1 in 100,000.

PSYCHOMETRIC PROPERTIES OF CLINICAL CHEMOSENSORY TESTS

As indicated in the first section of this article, an effective clinical test must be reliable and valid. A consummate test battery would be made up of tests that measure different aspects of the function of a sensory system, ideally ones that reflect the location and degree of the physical pathology itself. However, despite the large number of olfactory and gustatory clinical tests that have been devised, there are only a handful of tests for which reliability or validity information is available. With the exception of grossly localizing pathologies via regional tongue testing or unilateral olfactory testing, there is presently no convincing evidence that the physical location of an olfactory or gustatory pathology can be established solely on the basis of psychophysical testing. Nevertheless, most of the olfactory and gustatory tests available today have demonstrable clinical utility, and research is progressing in this field.

It is well documented that forced-choice odor identification tests, such as the UPSIT, the Connecticut Chemosensory Clinical Research Center's odor identification test, and
Wright’s confusion matrix test, evidence high reliability (e.g., both the test-retest and split-half rs of the UPSIT are above 0.90). Since olfactory detection threshold values vary considerably among individuals and evidence considerable day-to-day fluctuations within the same individuals, one might expect their reliability to be suspect. However, with noticeable exceptions, this appears not to be the case. Thus, in a study of 40 subjects, Koelega reported test-retest reliability coefficients (Spearman r) for a four-alternative forced-choice n-amyl acetate threshold test to be 0.65, 0.51, and 0.59 for bilateral, right nostril, and left nostril presentations, respectively. More recently, in a study of 32 subjects ranging in age from 22 to 59 years, Cain and Gent established correlations between the left and right sides of the nose for detection threshold values of butanol, phenyl ethyl methyl carbinol (PEMEC), isoamyl butyrate, and pyridine. The high correlations (0.68, 0.96, 0.86, 0.83, respectively) were interpreted as reflecting split-half reliability coefficients. Recently, we evaluated the test-retest reliability of 10 olfactory tests in the same set of 57 subjects aged 24 to 84 years. Included in this evaluation were four detection threshold tests (the PEA single-staircase threshold, the butanol single series ascending method of limits threshold [BT], the PEMEC single-ascending method of limits threshold, and the Japanese T&T olfactometer-threshold series [T&T]). Most of the olfactory tests, including the threshold tests, evidenced relatively high test-retest reliability coefficients (i.e., coefficients > 0.65). A factor analysis of the correlations among the tests supported the notion that most of the tests evaluated measure, in large part, a common sensory domain. Whether this is true for subject samples containing individuals with olfactory dysfunction as a result of defined clinical conditions, such as those arising from head injury or degenerative diseases, is not known.

The literature on gustatory testing suggests that, as in the case of olfaction, considerable intraindividual variability is present. In a relatively extensive study of 12- to 25-year-olds, Mattes assessed, over a 29-day period, test-retest reliabilities of five measures of sucrose taste perception. These measures included sweet recognition thresholds, exponents of power functions fit to suprathreshold intensity magnitude estimates, the most preferred sucrose concentration, reaction times to the onset of a sweet sensation following sucrose presentation to the tongue, and the duration of such a sweet sensation. Tests were conducted separately for sucrose solutions dissolved in three different media: (1) water, (2) cherry Kool-Aid, and (3) chocolate milk. As can be seen in Table 1, the threshold and preference measures were more reliable than the magnitude estimation measures. Although the reaction time measure was quite reliable for some test intervals, it is not known to what extent the high-reliability coefficient is simply a reflection of the reliability of reaction times, per se.

As noted in a recent review of the epidemiology of olfactory dysfunction, formal studies on the comparative sensitivity, specificity, and positive predictive value of chemosensory tests do not exist. Thus, while it is clear from the present review that a number of chemosensory tests are highly reliable, research is still needed to establish their relative efficacy.

In the present article, both theoretical and practical aspects of psychophysical chemosensory testing have been presented. It is apparent that basic psychometric parameters such as test reliability and validity are not
available for the vast majority of tests that have been used to assess taste and smell function. Nevertheless, estimates of the reliability of the most widely used clinical tests are now becoming available. Furthermore, it is clear that most such tests detect alterations in chemosensory function associated with head trauma and other pathologic conditions. Importantly, a number of types of olfactory tests appear to correlate well with one another, at least in "normal" populations consisting of both young and elderly individuals. Pending further research, it would appear that the basis for choosing one set of tests over another set of tests for routine clinical assessment should reflect such factors as the ease and cost of administration, the availability of normative data, and the degree to which there is a need to detect small amounts of dysfunction.

In regards to the latter point, research is sorely needed to establish the relative sensitivity of olfactory tests since, in disorders such as Parkinson's disease, early detection of subtle dysfunction may be of considerable diagnostic and prognostic value. Early detection of this disorder may lead to more effective treatment with monoamine oxidase inhibitors and other agents possibly associated with retardation of disease progression and thereby mitigate the severity of the symptomatology. Whether detection of subtle chemosensory alterations in head injury might lead to early treatment strategies

<table>
<thead>
<tr>
<th>Test measure and dilution medium</th>
<th>Intertest interval in days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (N=12)</td>
</tr>
<tr>
<td>Recognition threshold</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>0.86*</td>
</tr>
<tr>
<td>Kool-Aid</td>
<td>0.59</td>
</tr>
<tr>
<td>Chocolate milk</td>
<td>0.63</td>
</tr>
<tr>
<td>Intensity exponent</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>0.66</td>
</tr>
<tr>
<td>Kool-Aid</td>
<td>0.54</td>
</tr>
<tr>
<td>Chocolate milk</td>
<td>0.29</td>
</tr>
<tr>
<td>Preferred concentration</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>0.57</td>
</tr>
<tr>
<td>Kool-Aid</td>
<td>0.88*</td>
</tr>
<tr>
<td>Chocolate milk</td>
<td>0.86*</td>
</tr>
<tr>
<td>Reaction time</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>0.37</td>
</tr>
<tr>
<td>Persistence time</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>0.72*</td>
</tr>
</tbody>
</table>

*P<0.01.
in order to minimize damage to the olfactory system and other neural pathways is not presently known, but it is certainly a possibility.

Health care professionals should be aware of the variety of diagnostic tests available for the assessment of taste and smell function. A number of the tests discussed in this article are available at specialized smell and taste centers located at university medical centers. Chemosensory tests can be of particular value in head injury, since they can provide a quantitative assessment of sensory impairment, establish the validity of chemosensory complaints, detect malingering, and objectively monitor changes during rehabilitation.

REFERENCES


57. Overbosch P. A theoretical model for perceived in-


90. Doty RL, Agrawal U, Frye R. Evaluation of the internal consistency reliability of the fractionated and whole University of Pennsylvania Smell


