



Published in final edited form as:

Ann N Y Acad Sci. 2009 July ; 1170: 543–552. doi:10.1111/j.1749-6632.2009.04103.x.

Measuring Taste Impairment in Epidemiologic Studies – The Beaver Dam Offspring Study

KJ Cruickshanks¹, CR Schubert¹, DJ Snyder², LM Bartoshuk³, GH Huang⁴, BEK Klein¹, R Klein¹, FJ Nieto¹, JS Pankow⁵, TS Tweed¹, EM Krantz¹, and GS Moy¹

¹ University of Wisconsin

² Yale University

³ University of Florida

⁴ National Chiao Tung University

⁵ University of Minnesota

Abstract

Taste or gustatory function may play an important role in determining diet and nutritional status and therefore indirectly impact health. Yet there have been few attempts to study the spectrum of taste function and dysfunction in human populations. Epidemiological studies are needed to understand the impact of taste function and dysfunction on public health, to identify modifiable risk factors, and to develop and test strategies to prevent clinically significant dysfunction. However, measuring taste function in epidemiological studies is challenging and requires repeatable, efficient methods which can measure change over time. Insights gained from translating laboratory-based methods to a population-based study, the Beaver Dam Offspring Study (BOSS) will be shared. In this study, a generalized labeled magnitude scale (gLMS) method was used to measure taste intensity of filter paper disks saturated with salt, sucrose, citric acid, quinine, or 6-n-propylthiouracil and a gLMS measure of taste preferences was administered. In addition, a portable, inexpensive camera system to capture digital images of fungiform papillae and a masked grading system to measure the density of fungiform papillae were developed. Adult children of participants in the population-based Epidemiology of Hearing Loss Study in Beaver Dam, Wisconsin are eligible for this on-going study. The parents were residents of Beaver Dam and 43–84 years of age in 1987–88; offspring range in age from 21–84 years in 2005–2008. Methods will be described in detail and preliminary results about the distributions of taste function in the BOSS cohort will be presented.

INTRODUCTION

Taste or gustatory function may be an important determinant of health through possible effects on food choice and consumption. But there have been few epidemiological studies measuring taste function in adults.¹ We do not know how often, or if, taste intensity or flavor recognition changes with aging in the general population, nor if taste dysfunction is an important contributor to the risk of certain chronic diseases, frailty, or ability to recover from serious infections, surgeries, or other health problems. Most studies to address these issues have been small convenience samples or clinic-based studies relying on patient populations with severe taste disorders. Longitudinal studies of large cohorts drawn from the general population are needed to understand the natural history of taste dysfunction across the broad spectrum of

function represented outside of referral clinics. It is likely that good taste function is optimal for health, but data from epidemiological studies are needed to quantify the prevalence of taste disorders, the impact of taste dysfunction on health, and to identify the determinants of taste dysfunction.

In order to conduct epidemiological studies of taste function there must be reliable, inexpensive, portable tests with low respondent burden and agreed-upon definitions of taste function and dysfunction outcomes. As yet, these are not readily available, although efforts are underway as part of the NIH Toolbox Initiative to reach consensus on recommended measurement methods and definitions. The purpose of this paper is to report the methods used to measure taste as part of an ongoing epidemiological study of human aging sensory systems in Beaver Dam, Wisconsin.

STUDY POPULATION

The Beaver Dam Offspring Study (BOSS) is a study of age-related hearing, vision, olfactory impairments, and ocular disorders among the adult children of participants in the population-based Epidemiology of Hearing Loss Study (EHLS), a longitudinal study of aging which began in 1993.^{2,3} In 1987–88 a private census of the city and township of Beaver Dam, WI was conducted to identify all residents ages 43–84 yrs.⁴ These 5924 individuals were invited to participate in an extensive examination for a study of age-related ocular disorders, the Beaver Dam Eye Study. The EHLS was timed to coincide with the five-year follow-up for the eye study cohort and 3753 subjects (82.6%) elected to participate. During the five-year follow-up visit for the EHLS (1998–2000), participants were asked if they had living children. The parents were later re-contacted for permission to contact their children for the BOSS. As the study examination phase for the BOSS is on-going, the data presented are based on preliminary analyses of 2733 participants, ages 21–84 years, and are illustrative only. Reports of the prevalence of taste dysfunction or distributions await the complete dataset. Previously, it has been shown that the EHLS cohort is similar in age, gender, and education to residents of mid-sized cities in the U.S., although the cohort is primarily non-Hispanic white.²

Although taste testing was not part of the original scope of the study, we were approached by the National Institute on Deafness and Other Communication Disorders to include taste testing and charged with developing, standardizing, and testing methods and protocols for measuring the sense of taste in field studies, obtaining digital images of the tongue, and measuring the density of fungiform papillae. We were to use these methods to measure the distributions of taste intensity and fungiform papilla density, associations of taste intensity with fungiform papilla density and associations of taste with social, medical, and lifestyle factors and health.

MEASUREMENT METHODS

Taste intensity measurement

A generalized labeled magnitude scale (gLMS) was used to quantify the perceived taste intensity of filter paper disks (Figure 1A).⁵ Disk were impregnated with 1.0 M sodium chloride (salt), 1.8 M sucrose (sweet); 0.1M citric acid (sour), 0.001M quinine (bitter), and 6-n-propylthiouracil (PROP). Whatman #1 filter paper was soaked in room-temperature solutions of sodium chloride, sucrose, citric acid, and quinine and dried; disk diameters were 3 cm. Filter paper was soaked in a saturated solution of PROP near boiling and dried. Each paper disk contained 1.2–1.6 mg PROP. These disks were produced in the laboratory of one author (LMB) and shipped to Wisconsin in individual glassine envelopes, color-coded by tastants, in plastic zippered bags to guard against moisture.

Participants were asked to rank intensities using a scale that ranged from no sensation to strongest imaginable sensation of any kind (with a corresponding scale of 0–100 representing the distance from “no sensation”). To familiarize participants with the scale, they were asked to rate the intensity of the sound of the loudest thunder clap s/he can remember, a purring cat held in her/his lap, and a lawnmower across the street. Participants rank ordering them correctly (thunder>lawnmower>cat) proceeded on with the training while subjects who failed to order the intensities correctly were instructed. If the participants failed to rank order the intensities correctly the second time, they did not proceed with the training. Participants successfully mastering the scale (the rank order was correct), were asked to rate the intensity of the sound of snow falling on a calm night, the loudest sound imaginable, and the most intense sensation of any kind s/he could imagine. If the sound of snow was rated lower than the loudest sound imaginable then the participant was considered to have learned to use the scale and continued on; participants failing to rank snow lower than the loudest sound, failed to learn the scale and stopped.

After mastering the scale the participant was asked to rate the intensity of the brightness of the room, brightness of a dimly lit restaurant, brightest light s/he had ever seen, loudness of a whisper, loudness of a conversation, and loudest sound s/he had ever heard. Then, participants were given each filter paper disks to taste in the same order (salt, sweet, sour, bitter, PROP). For each disk, the participant was asked to report the taste quality (salt, sweet, sour, bitter, no taste, other, or unknown) and then rate the intensity of the taste using the gLMS. Between disks the participant was encouraged to sip room temperature bottled water. After tasting the PROP disk, the participant was offered a quick dissolving peppermint to mask any residual taste. Subjects who reported being pregnant or sensitive to PROP were not eligible to receive the taste disks.

Food Likes and Dislikes

To measure the intensity of food likes and dislikes a modified gLMS was used (Figure 1B), with responses ranging from –100 (strongest imaginable disliking of any kind) to 100 (strongest imaginable liking of any kind) with zero representing neutral (neither like nor dislike).⁶ In this study, participants were asked to rate the following foods: mayonnaise, whole milk, black coffee, dark chocolate, salted pretzels, grapefruit juice, sweets, strawberries, sausage, and milk chocolate. These foods were selected based on one investigator’s experience (LMB) to represent a spectrum of important food experiences corresponding to the selected taste sensations being measured as well as fatty foods. For example, pretzels represent a salty food, black coffee a bitter food, grapefruit a sour food, and sweets a sweet type.

Tongue Imaging

After completing the disk tasting, the participants proceeded to the image room where blue food coloring (McCormick) was applied with a cotton swab to the tip of the tongue to provide contrast between fungiform papillae (appear pink) and other tongue structures (coated blue). We adapted equipment and methods used for ocular examinations and ocular images to create a standardized system for obtaining digital images of tongues. An adjustable table was outfitted with a chin and forehead rest (Modified Soderberg LMP-1 Motorized Instrument Table and Shin-Nippon Forehead/Chinrest Assembly) and equipped with a digital camera system installed on the table using a column support to provide a fixed distance from the person’s face. The camera system consisted of a Canon EOS Digital Rebel XT Body fitted with a Canon EFS 60 mm f2.8 Macro lens, Canon MR-14EX Macro Ring Flash and Canon 52 mm UV filter. The camera was connected to a computer using Zoom Browser Ex to capture and store images.

The participant was asked to rest his/her forehead against the support while placing the chin on the chinrest and the examiner adjusted the table height to ensure participant comfort. The

participant was asked to stick out his/her tongue and close his/her eyes. The ring flash focusing lamp was turned on and the examiner would adjust the focus as necessary to ensure a sharp image, with minimal glare. After ensuring the contrast was sufficient, the examiner held a plastic slide on the tongue tip to the right of the midline applying a slight pressure to compress the tongue, and captured the image. Additional blue food coloring was applied as necessary for optimal contrast and, if excessive amounts were present, the participant was asked to swallow until the desired contrast was observed. After reviewing the original image, additional images could be captured, if necessary.

During the camera system development process, image quality was reviewed by two authors (LMB and DJS) to ensure comparability with existing imaging methods using specialized operating microscopes. Examiners were taught to follow the standardized testing protocol and image quality was judged acceptable on 5 practice subjects before they were considered certified to implement these procedures in the field.

Throughout the study, each examiner was observed carrying out the protocols on study subjects, image quality was reviewed, and data were monitored for deviations and drift.

GRADING FUNGIFORM PAPILLA DENSITY

Digital tongue images were transferred to Madison, WI for grading using a specially developed application (Canvas™ X, ACD Systems, Inc., Miami, FL). A standardized protocol was developed to allow graders to select the best image available for grading using a preview function and standardized criteria. The grader evaluated slide placement (the entire width must be contained in the image to calibrate the size of the measurement area), tip visibility (the entire tip should be visible), measurement area (the entire measurement area should be visible, in focus, and free of glare, bubbles, or other artifacts), and staining quality (pink circles visible on a blue background) to select the image for grading.

Once the selected image was loaded, the magnification, scale, and color were adjusted according to a standardized protocol, a standard circle was applied (equivalent to a 6mm diameter) with the right edge of the circle aligned at the midline of the tongue and the edge of the circle at the tip of the tongue. Fungiform papillae were identified by color (pink-red), appearance (mushroom-like or vascularized) and size (larger than filiform papillae) following a standardized protocol. The total number of fungiform papillae identified in the standard circle was automatically stored as the count.

Graders were trained in the procedures and certified using a standard set of ten images previously graded by an experienced researcher (DJS). In order to become certified scores must be within 5 of the standard grader's scores for both the mean difference and mean absolute difference, with at least 60% of the scores within 5 and at least 90% of scores within 10. The grader re-graded the standard set of images every three months throughout the grading period to monitor drift. Intra-grader variability was low, with a mean difference of 0.1; mean absolute difference was 3.1; 80% of gradings matched within 5 and 100% were within 10. Thus grading of digital tongue images is highly reproducible, using this standardized protocol.

LESSONS LEARNED IN BEAVER DAM

Challenges in administration

In preliminary analyses of 2733 participants, we determined that 53 (2%) were unable to participate in the taste protocol because of pregnancy, PROP sensitivity, refusals or time limitations. In field studies the risk for human subjects must be minimized in order to protect subjects from harm and encourage future participation in longitudinal studies. Although the

probable risk to pregnant women and fetuses is low, and true PROP sensitivity is rare, a conservative approach is warranted so it is important to recognize that complete data are unlikely in any epidemiological study of taste using similar methods. The total examination time for the study approached four hours, and some subjects were unable or unwilling to give sufficient time to complete the entire examination.

An additional 415 (15%) failed the practice task for the taste scale and could not proceed to the tasting of paper disks. Most participants (69%) failing to master the scale rated the intensity of the purring cat higher than or equal to the sound of the lawnmower across the street. When this problem was identified, the training protocol was revised to allow for re-instructing the participants; however, this modification did not eliminate the problem as 14.7% continued to order intensities incorrectly. Selecting other experiences for the training could increase the percent of subjects completing the taste testing.

Nonetheless, the scale did capture a range of intensity ratings for the remembered sensations as shown in Figure 2. In the sample of 2265 subjects with complete data, the median responses ranged from 5 for the loudness of a whisper to 80 for the brightest light ever seen although the ranges for each sensation were broad as indicated by the whiskers. Although respondents correctly ranked these sensations in order with light or sound groups, analyses of taste intensities may need to be adjusted for the scores assigned for the remembered sensations in order to remove variability due to individual differences in the scale range used.⁷

During the study, the author responsible for monitoring the monthly quality control reports noted that the percent of participants correctly identifying the taste quality for the sour disk declined and the average intensity rating dropped as well. Subsequent investigation revealed that the technician making the disks had inadvertently used a solution intended for another study, which had a lower concentration than that meant for the Wisconsin study. Although the time to detection of this problem was short, 226 or 8% of the subjects in this preliminary dataset received a sour disk with the incorrect concentration resulting in additional missing data for this taste. This experience highlights the importance of strict quality assurance procedures to detect changes over time.

The distributions of reported taste qualities are shown in Figure 3. Most people correctly identified salt, sweet, and bitter, with a large proportion miscalling sour as bitter and the expected variation in people identifying PROP correctly. Although the taste quality rating may be of limited analytic utility, continuing to collect this information provides important data for quality assurance, and provides a check on internal validity, namely, that the concentrations of the disks were sufficient to correctly recognize the flavor.

The distributions of taste intensity for the five disks are shown in Figure 4 also as box plots to illustrate the range of perceived intensity scores. For each taste there is a broad range of scores. The score reflects the proportional distance from “no sensation” to the “strongest imaginable sensation of any kind” recorded as a number from 0–100. There was significant digit preference with 56–72% of subjects selecting a distance reflecting an intensity score ending in zero and 81–91% of subjects selecting a distance corresponding to numbers ending in zero or five for any taste intensity score. Recording the distance on a scale with broader intervals may be sufficient as it is not known what difference in magnitude of intensity is important when comparing groups.

Participants also used a scale to quantify the intensity of liking or disliking a food. Figure 5 shows the distributions in reported intensity, with zero being neutral, negative numbers representing dislikes and positive numbers representing likes. Again, there was substantial variability in the reports, suggesting that this scale may be useful for studies of food preferences, dietary intake patterns and health.

The taste disks were easy to ship to the field site, participants did not object to the testing, although some people complained about the bitterness of the PROP disk, and the test was easy to administer.

Density measures

Our digital tongue imaging system worked well. Although there was a learning curve for applying blue food coloring without making a mess, the examiners quickly became expert in applying it neatly. The amount of dye needed varied by participant and the time it remained in the mouth also varied, so monitoring the quality of the contrast during the photography is important. The ring flash system that was selected reduced problems with glare and washout. The image management and grading systems facilitated standardized density measures by ensuring that the grading standard circle was adjusted for scale differences across images. Features of our grading system that were particularly useful were the ability to click on each fungiform papillae to incrementally add to the total score, store the graded images for later review by the epidemiologist for quality assurance purposes, and grade an image multiple times (masked to previous results) for additional quality assurance efforts (re-grading of a random sample, comparisons between and within graders). The distribution of fungiform papillae obtained in this sample is displayed in Figure 6 and shows that a broad range of densities was detectable. The grading was highly reproducible and trained graders were comparable to an experienced researcher (DJS). This relatively inexpensive imaging and grading system provided high quality images in the challenging setting of a field study.

Measuring Prevalence and Evaluating Associations

The immediate challenge when using these results to report the prevalence of taste disorders is the lack of an accepted definition of impairment. It is not known what taste intensity score represents a clinically significant problem, whether both high and low intensities represent changes in taste function (or subgroups at risk for future health problems), nor how to combine across taste qualities for a person-level outcome. In the absence of a gold standard of clinical importance, models such as used in studies of refractive error that consider both ends of the scale (myopia and hyperopia in this example) as important may be useful for studies of taste. Methods used to establish cutpoints based on young “normals” may help to identify important subgroups. Scores for young healthy subjects without olfactory disorders, who do not report problems with taste and use a broad range for scoring the remembered sensations may be used to establish cutpoints for low and high performance on taste intensity scales, which can then be applied to the study sample to classify participants into groups.

Analytic models can then move beyond simple correlations and linear relationships to explore effects of low and high sensation on other health conditions, as well as explore factors associated with low and high taste sensation. Researchers should evaluate the impact of adjusting for the scores for the remembered sensations, as this may reduce variability due to differences in the magnitude of the psychophysical scaling range used.

Additional research is needed to determine if each taste quality has a similar impact on health and similar determinants which might suggest that responses to each taste disk could be combined in some fashion to identify people with taste disorders. Analyses of grouped data can identify key factors associated with poor performance and evaluate the impact of performance on health, food likes and dislikes, and dietary intake patterns.

However, these cross-sectional data cannot be used to distinguish participants who have experienced increased or decreased function from individuals with low or high function since birth or early childhood. Longitudinal data measuring change in function over time will be

important to determine how taste function changes with age and to identify factors associated with the development of taste disorders.

Additional Issues

We have demonstrated that taste function can be measured in epidemiological studies using these simple measures of taste intensity, food likes and dislikes, and anatomic measures of the density of fungiform papillae. However, there remains a need for studies to determine the test-re-test consistency of intensity scores. Before applying these tests in longitudinal studies of change in taste function, it is important to know that the short-term variability is low. Because we used one standard order of presentation, it is not known if presentation order impacts intensity ratings.

SUMMARY

We have developed and implemented measures of taste function and fungiform papilla density in an epidemiologic study where examinations occur in a community setting. In pilot work we had considered liquid testing, but due to the difficulty of creating and maintaining stock solutions in a field site that consisted of a suite of offices without a laboratory or clean sink area, it was determined that filter paper disks would be preferable. Transporting these disks to the field site was inexpensive, and simplified testing for participants. Although spatial testing with liquids may detect clinically significant alterations in localized oral sensation, our whole mouth measures are likely to represent the usual experience of an individual. Our imaging and grading methods for evaluating lingual anatomy offer an inexpensive way to achieve high quality images and reliable estimates of fungiform papilla density which will permit studies of the complex relationships of fungiform papilla density and taste intensity. These methods may be useful to study the public health importance of differences in taste function, the magnitude of the population with taste impairments, the risk of developing taste impairments with aging, and the relationships between taste perception, food preferences, dietary intake and health.

References

1. Vennemann MM, Hummel T, Berger K. The association between smoking and smell and taste impairment in the general population. *J Neurol*. 2008;10.1007/s00415-008-0807-9[epub ahead of print]
2. Cruickshanks KJ, Wiley TL, Tweed TS, Klein BEK, Klein R, Mares-Perlman JA, Nondahl DM. Prevalence of hearing loss in older adults in Beaver Dam, WI: The Epidemiology of Hearing Loss Study. *Am J Epidemiol* 1998;148(9):879–86. [PubMed: 9801018]
3. Cruickshanks KJ, Tweed TS, Wiley TL, Klein BEK, Klein R, Chappell RJ, Nondahl DM, Dalton DS. The five-year incidence and progression of hearing loss: The Epidemiology of Hearing Loss Study. *Arch Otolaryngol Head Neck Surg* 2003;129:1041–1046. [PubMed: 14568784]
4. Klein R, Klein BE, Linton KL, De Mets DL. The Beaver Dam Eye Study: visual acuity. *Ophthalmology* 1991 Aug;98(8):1310–5. [PubMed: 1923372]
5. 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 vs magnitude matching. *Physiol Behav* 2004;82:109–114. [PubMed: 15234598]
6. Bartoshuk LM, Duffy VB, Hayes JE, Moskowitz HR, Snyder DJ. Psychophysics of sweet and fat perception in obesity: problems, solutions and new perspectives. *Philos Trans R Soc Lond B Biol Sci* 2006 Jul 29;361(1471):1137–48. [PubMed: 16815797]
7. Bartoshuk LM, Duffy VB, Chapo AK, Fast K, Yiee JH, Hoffman HJ, Ko C-W, Snyder DJ. From psychophysics to the clinic: Missteps and advances. *Food Quality and Preference* 2004;15:617–632.

Figure 1A

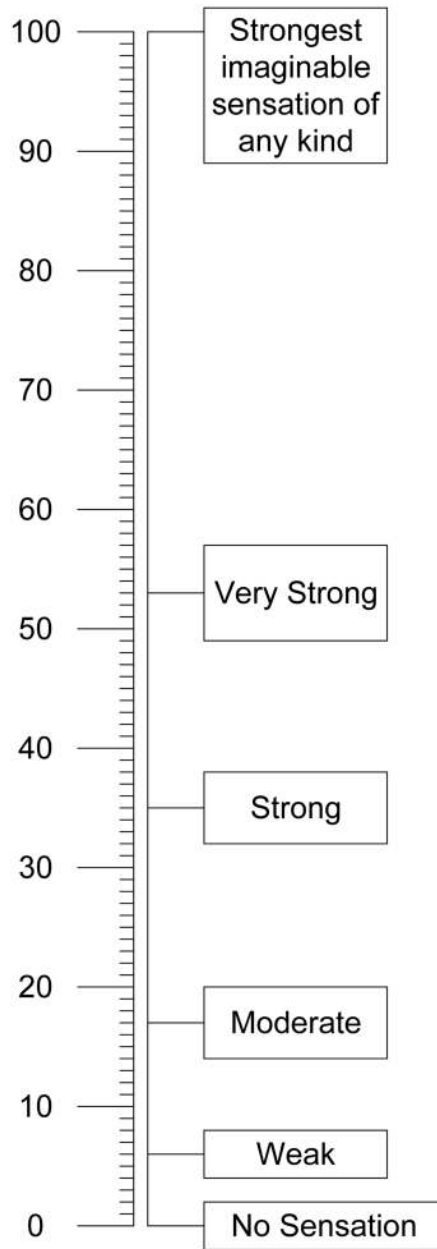


Figure 1B

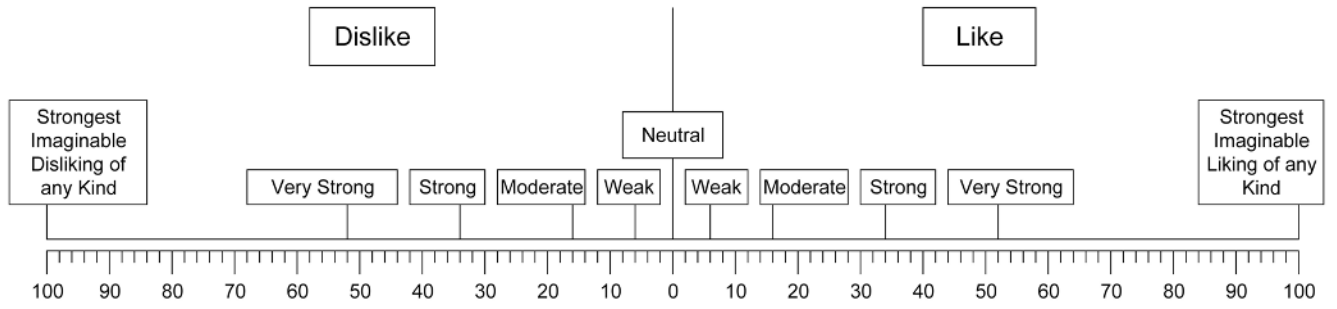


Figure 1.
1A. gLMS used to measure intensity of remembered sensations and tastes
1B. gLMS used to measure intensity of food likes and dislikes

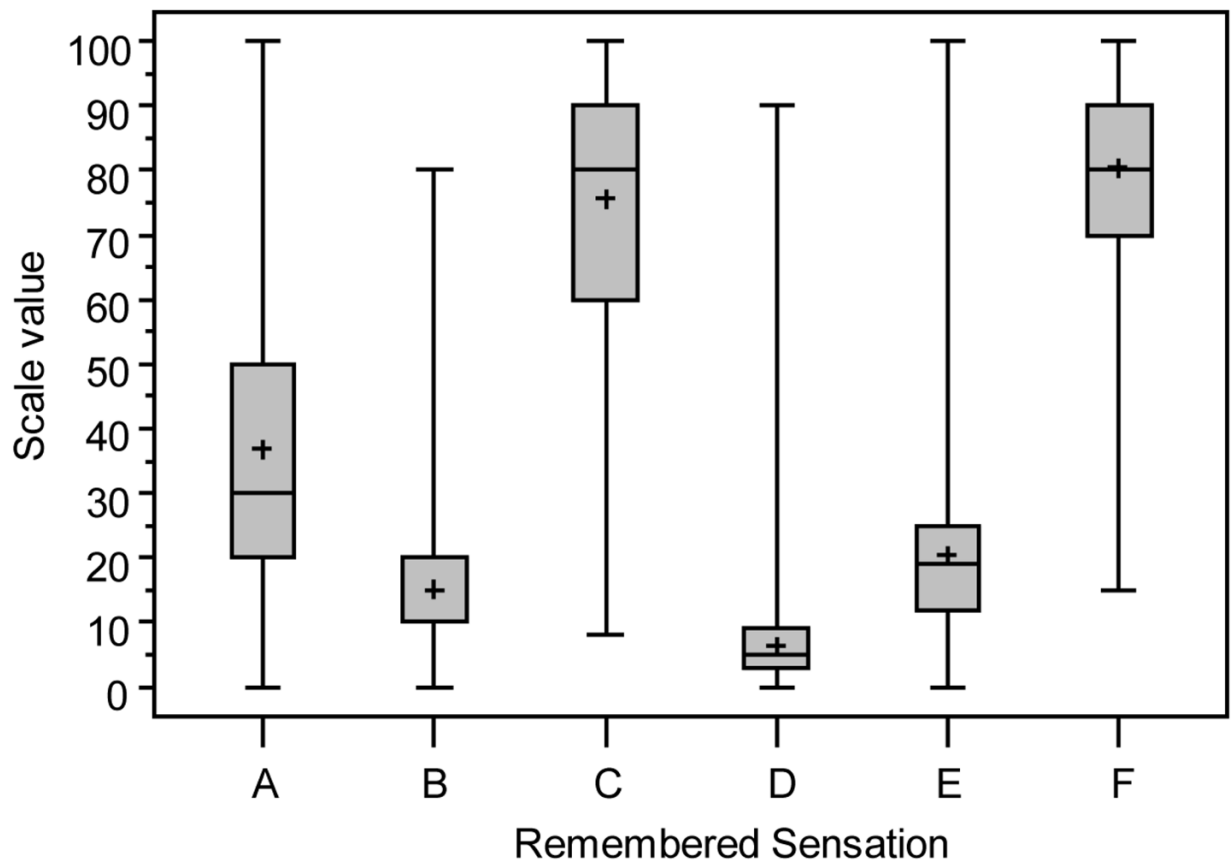


Figure 2. Distribution of Intensity of Remembered Sensations. A: Brightness of the room B: Brightness of a dimly lit restaurant C: Brightest light you have seen D: Loudness of a whisper E: Loudness of a conversation F: Loudest sound you have heard. Filled in boxes represent the interquartile range, horizontal lines within the boxes show the medians, plus symbols show the means, and whiskers represent the full range of the data.

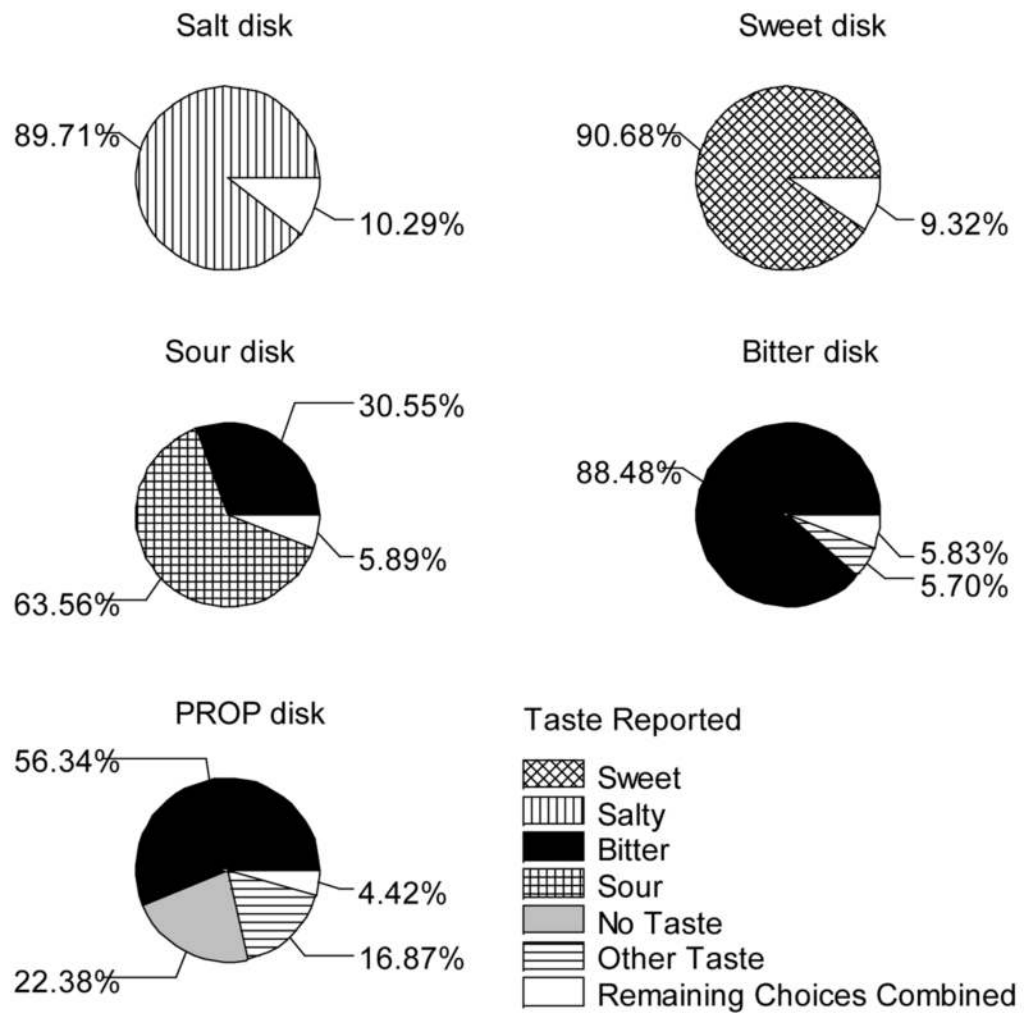


Figure 3. Distribution of Reported Taste Disk Qualities. Each patterned segment represents the proportion of participants reporting that taste quality.

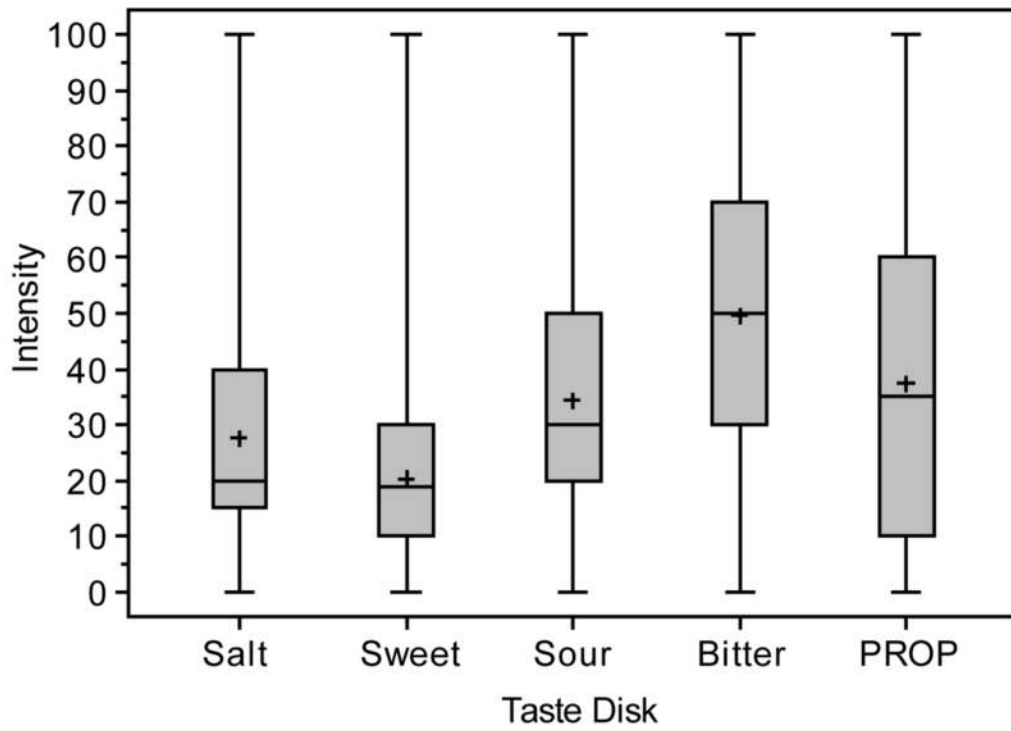


Figure 4. Distribution of Taste Disk Intensities. Filled in boxes represent the interquartile range, horizontal lines within the boxes show the medians, plus symbols show the means, and whiskers represent the full range of the data.

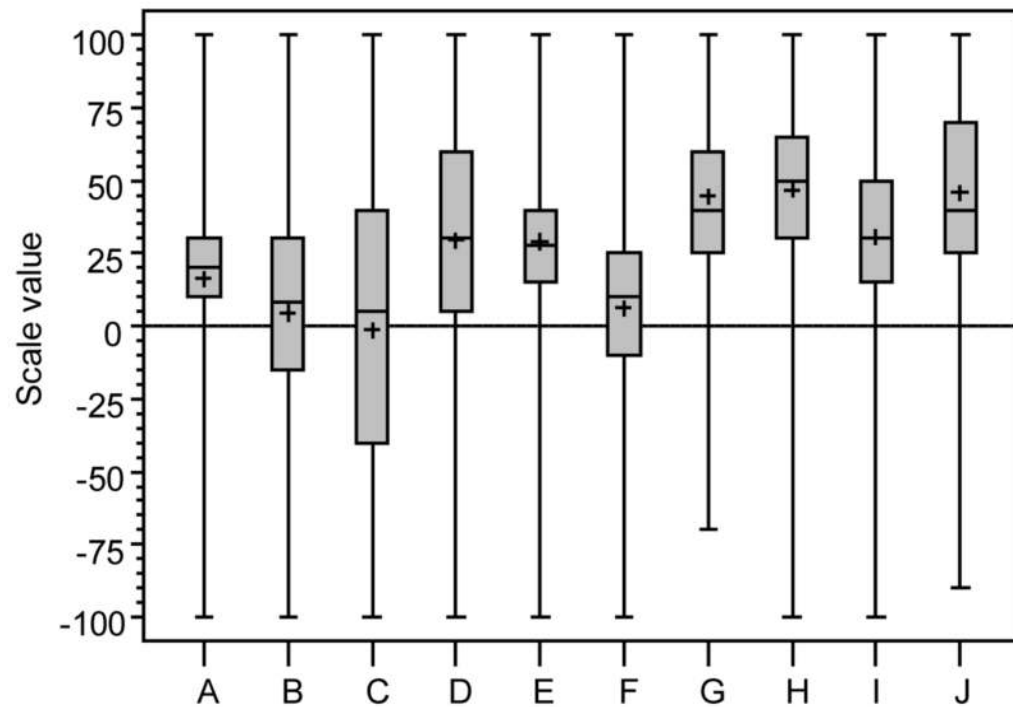


Figure 5. Distributions of Food Likes and Dislikes A: Mayonnaise B: Whole Milk C: Coffee D: Dark Chocolate E: Pretzel F: Grapefruit G: Sweets H: Strawberries I: Sausage J: Milk Chocolate. Filled in boxes represent the interquartile range, horizontal lines within the boxes show the medians, plus symbols show the means, and whiskers represent the full range of the data.

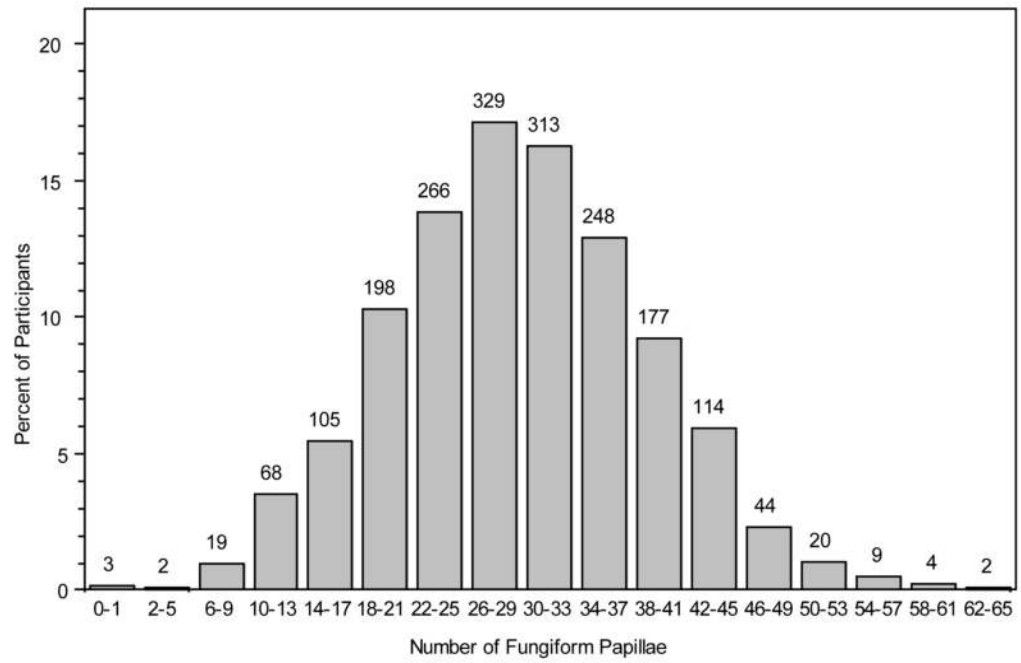


Figure 6. Distribution of Fungiform Papillae. Numbers shown above bars are the number of participants in each category.