RESEARCH GUIDELINES FOR COOKERY, SENSORY EVALUATION, AND INSTRUMENTAL TENDERNESS MEASUREMENTS OF MEAT

Second Edition

Version 1.0

March, 2015



American Meat Science Association

Marmon Foodservice Technologies, Inc. Ex. 2010
Duke Manufacturing v. Marmon Foodservice Technologies IPR2026-00406

Research Guidelines for Cookery, Sensory Evaluation, and Instrumental Tenderness Measurements of Meat

American Meat Science Association

Copyright © 2015 by the American Meat Science Association

All rights reserved

Writing Committee:

Co-Chairs

Tommy L. Wheeler, USDA–ARS, U.S. Meat Animal Research Center
Linda S. Papadopoulos, LP & Associates
Rhonda K. Miller, Texas A&M University

Committee Members

Keith E. Belk, Colorado State University
Michael E. Dikeman, Kansas State University
Chris R. Calkins, University of Nebraska
D. Andy King, USDA–ARS, U.S. Meat Animal Research Center
Mark F. Miller, Texas Tech University
Steven D. Shackelford, USDA–ARS, U.S. Meat Animal Research Center
Bridget Wasser, National Cattlemen's Beef Association
Lawrence D. Yates, USDA–AMS

The recommendations in these Research Guidelines are intended to be the most appropriate available at the time of this writing. Recommendations for improvements, new technology, and errors needing correction, however, should be sent to AMSA to be addressed by the Research Protocol Committee. Thus, the Guidelines can be continuously updated and improved as technology and meat science advances.

March, 2015: Second edition (version 1.0)

American Meat Science Association
201 W Springfield Ave, Suite 1202
Champaign, Illinois USA 61820
800-517-2672
information@meatscience.org
http://www.meatscience.org

ACKNOWLEDGEMENTS

Thanks to Dan Hale, Hannah Laird, and Tanner Luckemeyer for assistance with the video of Warner-Bratzler Shear Force. Thanks to Patty Beska and Adria Grayson for assistance with the slice shear force and thermocouple wiring videos and the photos of the sensory sample sizer and beef patty sectioning. Thanks to Darin Doerscher for contributions to instrumental tenderness measurement.

The development of this guide was supported in part by the American Meat Science Association Educational Foundation.



Contents

I. INTRODUCTION 6
II. SAMPLE COLLECTION/PREPARATION 10
A. Selection of Samples 10
B. Time Postmortem for Processing into Cuts 11
C. Steak, Chop, and Patty Variables 11
D. Freezing and Frozen Storage 12
III. PRE- AND POST-COOKING PROCEDURES 13
A. Physical State of the Cut Prior to Cooking 13
B. Methods for Monitoring Temperature Changes 13
C. Evaluating Degree of Doneness 14
IV. COOKERY METHODS 15
A. Roasting 16
B. Broiling 17
C. Panbroiling 17
D. Impingement Oven 18
E. Clam Shell 19
F. Summary of Recommendations for Cookery Methods 19
V. SENSORY TESTING FACILITIES 21
A. Sensory Testing Environment 21
B. Additional Options for Testing Locations 21
VI. PREPARATION AND PRESENTATION OF SAMPLES TO THE PANEL 23
A. Preparation of Sensory Samples 23
1. Trained panel evaluations 23
2. Consumer panel evaluations 24
B. Sample Presentation 24
1. Order of sample presentation 25
2. Number of samples per session 26
C. Sensory Panel Participants' Informed Consent 26
VII. SENSORY EVALUATION METHODS 29
A. Discrimination Methods 30

1. Overall difference tests 30
2. Specific attribute tests 34
B. Descriptive Analysis Methods 34
1. Flavor profile method 34
2. Texture profile method 35
3. Quantitative descriptive analysis 37
4. Spectrum™ descriptive attribute analysis 38
5. Short-version Spectrum™ descriptive method for quality assurance and shelf-life studies 38
6. Meat descriptive attribute evaluation 44
7. Magnitude estimation 44
C. Training Sensory Panelists for Discriminative or Descriptive Testing 47
1. Scope 47
2. Selection of potential panelists 47
3. Training 58
4. Performance evaluation for descriptive attribute panels 71
5. Monitoring panelist performance 72
6. Panel maintenance 72
D. Consumer Sensory Panel 73
1. Scope 73
2. Determining the test type 73
3. Protocol design 74
4. Ballot development 75
VIII. INSTRUMENTAL MEASURES OF TENDERNESS AND TEXTURAL PROPERTIES 82
A. Whole-Muscle Steaks/Roasts/Chops 82
1. Instruments/measurements 82
2. When NOT to use shear force 85
3. Instrument calibration 85
B. Ground Beef/Patties 86
a. Shear force measures 86
b. Compression measures 86
IX. DATA ANALYSES 88
A. Data Preparation 88

B. Panelist Effects 91
C. Analysis Options 92
1. Univariate Analyses 92
2. Multivariate Analyses 93
X. LITERATURE CITED 100

I. INTRODUCTION

In 1978, the American Meat Science Association (AMSA) first published *Guidelines for cookery and sensory evaluation of meat* (AMSA, 1978). A three-year effort by several committees was responsible for that publication. During the next 17 years, these AMSA "Guidelines" were very useful to both AMSA members and nonmembers involved in meat cookery/sensory evaluation. Interpretation of published reports was much easier when the AMSA Guidelines were used to guide the research. Research that utilized the AMSA Guidelines has greatly assisted in determining key factors responsible for differences in sensory, instrumental texture, and cooking properties of meat. In addition, the AMSA Guidelines provided greater consistency in multi-institutional projects.

In 1995, a much-needed update was published titled *Research guidelines for cookery, sensory evaluation, and instrumental tenderness measurements of fresh meat* by the AMSA and the National Live Stock and Meat Board (AMSA, 1995). A 12-member AMSA committee expanded and updated the information from the 1978 version to bring greater consistency, accuracy, and relevancy to cookery and sensory research. Numerous changes had occurred in cooking equipment and meat products since the Guidelines were first published. As a result of diet/health concerns, meat products were leaner and often smaller in portion size. Precooking followed by reheating and the use of microwave cookery was more prevalent than in 1978, and there was much more variety in meat entrees. Certainly, food safety concerns were greatly elevated from producers throughout the entire processing and marketing chain, including consumers. Many changes had occurred in cooking equipment and instruments available for measuring various meat properties.

Much of the information in the previous version of the Guidelines is applicable today; however, much has changed. This version is more comprehensive, as much additional and updated information are provided in this revision. As was stated in the previous Guidelines, this manual is not a "standard" to which everyone will be expected to adhere for every research study. It is, as the title suggests, "Research Guidelines." The researcher must decide the most appropriate methods to use to answer the question at hand. The methods and approaches described herein, however, are accepted and recommended as the most appropriate for most circumstances. These also are not *consumer* cookery guidelines, but *research* guidelines. Thus, recommended research methods may not always produce the highest level of consumer acceptance. The methods and approaches recommended in these Guidelines, however, are designed to control unwanted variability, to determine the most accurate answer to the questions being addressed with the most relevant methods possible, and, when feasible, to allow for valid comparative interpretation of published research.

Information is included on recommendations for collecting and preparing appropriate samples for sensory and/or tenderness evaluation for fresh beef, pork, and lamb steaks/chops, roasts, and ground patties; but it also may be applicable to certain enhanced, cured, or comminuted products. Additional topics covered include product handling, cookery methods, sensory panel methods, and a data analyses overview. In addition, more information is now given on instrumental approaches to measuring meat tenderness and consumer evaluation. Specific

approaches and procedures, especially for meat cookery, have been provided in selected situations. This does not mean that other suitable procedures/equipment, etc., are not available. Rather, it is the consensus of the committee that the procedures presented are reliable and appropriate.

The development of sensory evaluation as a science has undergone tremendous expansion in the last 25 years. The efforts of ASTM International (formerly American Society for Testing and Materials) Committee E-18, the Society of Sensory Professionals, and the Institute of Food Technologists (IFT) have led to numerous recent publications and annual workshops on sensory evaluation. Thus, although more comprehensive details are included in this revision than in the previous version, more detailed approaches and procedures can be found in the references provided. The reference section has been expanded to include important sources of information (especially for sensory evaluation) that have occurred mainly in the past 25 years.

Before initiating an experiment, the cooking and handling procedures, sensory method, and testing parameters should be determined. Factors to consider in method selection include following:

- What is your hypothesis?
- What questions are you trying to answer (test objectives)?
- How will the results be used?
- How large of a difference are you trying to detect?
- How much variability is there within and between samples?

The diagram shown in Figure 1 may be useful in selecting the most appropriate test method. If, based on the preliminary work, the sensory differences among treatments are not expected to be detectable, the lack of significant differences can be verified using discrimination or descriptive analysis methods. If, however, the sensory differences are expected to be detectable, consumer testing methods would be more appropriate. It would be important to determine if the differences are detectable to consumers, and if detectable, how they affect consumer acceptability.

Quantitative sensory methods can be grouped into three primary categories: (1) discrimination, (2) descriptive analysis, and (3) consumer. Discrimination methods can use either trained panelists or untrained consumer panelists, depending on the test objectives. If the test objectives are to determine with a high degree of certainty if treatment differences are significant, trained panelists are suggested. Trained panelists are carefully selected, highly trained, and hypercritical as compared to average consumers. When using consumers for discrimination tests, consumers are not as critical and may not detect differences. The selection of a testing method should be based on the objectives of the study. Data should be interpreted based on the sample population used for the study.

Descriptive methods use trained panelists. Panelists can be defined as trained (6 to 10 training sessions) to highly trained (6 months or more of training) or expert (10 or more years of experience). As the amount of training and experience increases, panelists can detect smaller differences in attributes between samples. The amount of training should be noted and data presented based on the panelists' level of training. Descriptive tests are used to quantify the

level of an attribute within the samples. Depending on the testing method selected, scales and attributes may vary, but each test is used to determine if samples differ in sensory attributes.

Consumer evaluation can be either qualitative or quantitative. Qualitative consumer methods provide input from consumers on their opinions on a product(s) based on more loosely structured questions that provide opportunities for consumers to give their opinions and inputs. These data are very valuable, but they do not give quantitative results that can be statistically analyzed. Consumer quantitative testing provides an avenue to measure consumer opinions using questions on a ballot with a scale that is or can be converted to numerical values for statistical analyses. This provides a method of quantifying consumers' responses and determining differences. Test-booth conditions, coded containers, and scoring methods used in central location tests (CLTs) are certainly not typical of normal conditions of food consumption. Consumer home placement or home use tests (HUTs) provide the opportunity to include in-home preparation, family opinions and environment, and use of normal containers.

When selecting the test to use, sensory professionals should always remember that consumer panels differ from trained panels. Although members of a trained panel are consumers, their opinions and preferences may not be representative of the general population. Trained panelists should therefore never be asked to respond with their opinions on preference or liking. More information on the testing methods and considerations is provided in the appropriate sections. Meat scientists are strongly encouraged to understand each testing method and the strengths and weaknesses of each method when determining the method that provides the best procedures to meet the test hypothesis and objectives.

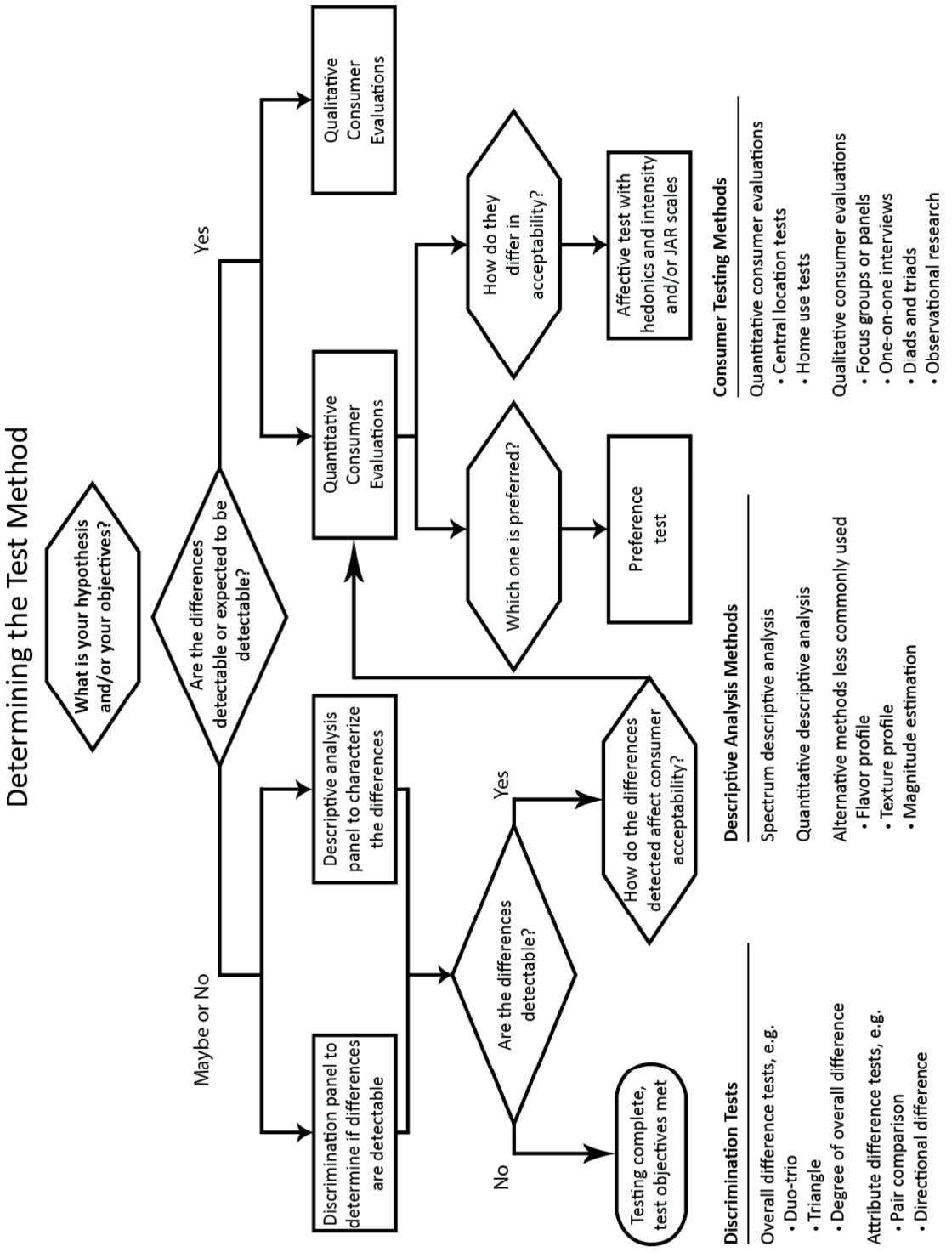


Figure 1. Flowchart for determining the appropriate test method(s) for sensory testing.

II. SAMPLE COLLECTION/PREPARATION

Various postmortem processing procedures, such as rate of chilling, aging, fabrication, and freezing, can vary and also might affect cookery, sensory, and instrumentally measured textural properties of meat. The Guidelines provide recommendations for these procedures, but it is realized that the overall objectives of projects may dictate variation in these factors. All cooperating institutions on joint projects, however, should standardize their methods to ensure uniformity during each experiment.

The previous Guidelines stated that steaks, chops, and roasts should not be sized unless preslaughter treatment or postmortem processing affects intended size. There is merit, however, in "sizing" meat cuts if variation in cooking procedures is the major focus of the project. Certainly, consideration should be given to removal of bone and connective tissue, degree of subcutaneous fat removal, thickness, weight, and shape of the cut. Ground beef patties, being thin, should have very close controls on weight and thickness. The variation in patty manufacturing parameters should be considered and standardized. All of these factors should be controlled and/or standardized to the extent necessary to collect relevant and accurate data.

A. Selection of Samples

Researchers are strongly advised to consult with statisticians and/or perform appropriate statistical tests for determination of sample size for a research study. A number of parameters including an estimate of variability in tenderness (shear force or sensory), flavor, juiciness, or other traits of interest are needed in order to determine sample size. Many statistics textbooks provide guidelines for estimating the sample size needed for a specific experiment. It will be more fiscally and scientifically sound to have adequate sample size than to use marginal sample size and not be able to detect differences that might truly exist. These types of tests also are important for estimating the number of sensory sessions and funding to complete a project. As a rule, all samples within a study should be representative of the products or processes under study. Proper sampling within muscles is of critical importance. Steaks or chops within a muscle that are to be assigned to different treatments should be statistically randomized (if there is no known location variation) or blocked (if there is known location variation) to alleviate bias. If each carcass or cut represents one replicate of one treatment, steak location within a muscle should be standardized. For example, the first steak from the rib end of the short loin could be assigned to sensory analysis and the second steak to shear force evaluation tests.

Most research that has the goal of determining production, antemortem, or postmortem treatment effects on tenderness and other palatability traits utilizes the longissimus muscle because it has the highest total value in carcasses and typically is sold as steaks or chops for dry heat cookery. Greater emphasis, however, recently has been placed on characterizing and marketing other muscles; thus, it may be appropriate to evaluate treatment effects on multiple muscles. When doing so, keep in mind that location effects are very important in some muscles. In addition, muscle selection should be made to ensure the proper answers are obtained for the experiment's questions. Available data show that the relationships in tenderness among

muscles in the same animal vary from relatively low to moderately high; thus, results in one muscle may not be representative of other muscles.

B. Time Postmortem for Processing into Cuts

Cuts should not be removed before rigor mortis is complete. The timing of rigor mortis completion may vary among species and with chilling conditions. The earliest that carcasses should be ribbed and fabricated into cuts is 24 hours postmortem. The timing is dependent on the objectives of the research and the schedules of commercial facilities if product is obtained commercially. For additional postmortem aging time, cuts should be vacuum packaged. The following postmortem times are recommended across species:

- Beef: between 10 and 28 days, with 14 days being optimum for most research
- Pork: between 5 and 14 days, with 10 days being optimum for most research
- Lamb: between 7 and 14 days, with 7 days being optimum for most research

These times are optimum for detecting differences in breeds, management systems, antemortem treatments, and postmortem treatments, but they should not preclude unique research with unique postmortem treatments. Most of the effects of aging on tenderness are complete by the above recommended times, but the industry average is considerably longer with wide variation both within and between retail and foodservice product.

C. Steak, Chop, and Patty Variables

Following are recommended thicknesses of steaks, chops, and patties:

- Beef steaks (dry heat): 2.54 cm
- Beef steaks (moist heat): 1.9 to 2.54 cm
- Lamb chops (dry heat): 2.54 cm
- Pork chops (dry heat): 2.54 cm
- Beef patties: not < 0.95 cm or > 1.10 cm for 91.5-g patties
- Beef patties: not < 1.10 cm or > 1.27 cm for 113.5-g patties

A cutting guide should be used so that steaks and chops are uniform in thickness. When product must be frozen, very uniform thicknesses can be obtained by freezing and sawing on a bandsaw.

Because it is impractical to list all of the various roast cuts from all species, the reader is referred to the latest editions of the *Meat Buyer's Guide* (NAMI 2015) and the *Institutional meat purchase specifications* (USDA, 2014) and encouraged to use common sizes found in retail and foodservice products that meet experimental objectives. The following are minimum weights and thicknesses for roasts from the three species:

- Beef: 1.5 kg, 5.0 cm thick
- Pork: 1.0 kg, 5.0 cm thick
- Lamb: 0.5 kg, 5.0 cm thick

The small amount of external fat currently present on retail cuts suggests that most if not all external fat should be removed from cuts before cooking. It is recommended that products be

vacuum packaged or in packaging materials with very low oxygen and moisture transmission properties.

D. Freezing and Frozen Storage

For Warner-Bratzler or slice shear force determinations, steaks or chops should be cooked and sheared without freezing whenever possible because freezing and thawing has been shown to decrease shear force proportionately to the fresh/never-frozen level of shear force (Grayson & King, 2004) (Grayson et al., 2014). When the experimental design is such that sample size or treatment protocol require it (multiple days of aging, etc.), however, freezing is acceptable as long as it is done under standardized conditions across the experiment and its potential effect is acknowledged. Freezing should be accomplished so that all samples freeze in the same amount of time (i.e., steaks spread out in one layer rather than placing a full box into the freezer). Freezing should be done rapidly at very cold temperatures, preferably < -20°C to limit damage from ice crystal formation (Hiner, Madsen, & Hankins, 1945). It is recommenced that pork and ground beef patties be evaluated within 3 months of frozen storage; and beef and lamb steaks, roasts, and chops within 6 months of frozen storage.

III. PRE- AND POST-COOKING PROCEDURES

A. Physical State of the Cut Prior to Cooking

Whenever possible, samples should be evaluated fresh to avoid introducing spurious variation due to freezing and thawing rates and length of frozen storage. Steaks or chops should be vacuum packaged in oxygen-impermeable film to avoid desiccation and oxidation. Internal temperature of the meat at the initiation of cooking also can affect tenderness (Berry & Leddy, 1990; Wheeler, Shackelford, & Koohmaraie, 1996). Likewise, if the temperature is < 2°C, the cut is prone to surface charring. The internal temperature should be recorded just prior to placing meat in the cooking environment.

It is acceptable to cook ground beef patties fresh, frozen, or thawed, depending on the objectives of the experiment. While foodservice frequently cooks patties from the frozen state, consumers usually cook patties from the fresh or thawed state.

- Thaw steaks, chops, and patties (vacuum packaged) at 2 to 5°C until internal temperature reaches 2 to 5°C (≥ 24 hours may be necessary, depending on sample size).
- Thaw roasts (vacuum packaged) at 2 to 5°C for 48 to 72 hours or until the internal temperature reaches 2 to 5°C.
- Frozen, thawed, and cooked weights should be collected for the determination of losses at different stages from the frozen to the cooked state.

B. Methods for Monitoring Temperature Changes

- Only thermocouples connected to recorders or hand-held digital readout thermocouple thermometers should be used.
- Infrared temperature-measuring guns may be used to monitor the surface temperature of grills or meat cuts during cooking, or to rapidly scan a cut surface immediately after cutting.
- Use iron/constantan or copper/constantan thermocouple wires with a diameter of < 0.05 cm and special limits of error of < 2°C.
- Use multichannel recorders with rapid readout and movement between channels or multiple hand-held digital thermometers when procedures necessitate several cuts being cooked at one time.
- Do not use thermocouples placed inside metal sheaths because they may conduct heat into the meat and give erroneous readings of internal temperature.
- To place the thermocouple wire into the geometric center of the steak, chop, or roast, use a metal probe inside a 16-gauge spinal needle. In the case of steaks and chops, with the product lying on a flat surface, insert the probe and needle into the side of the meat at the end of the steak or chop and push completely through the cut to the other side. Remove the probe from the needle and insert the end of the thermocouple wire into the needle; then pull the needle back out of the meat, which pulls the thermocouple wire through the section of meat and out the other side. Next, reinsert the probe into the needle and push them back through the steak or chop from the opposite side. Remove the probe from the needle, loop the end of the thermocouple wire around, insert it into the needle again, and pull the needle back through the meat with the wire.

Remove the needle from the wire and then pull the thermocouple wire back just enough to place the end of the wire at the center of the meat. Tighten the loop of the thermocouple wire, which helps hold the thermocouple in place while turning the sample during cooking. For a demonstration of inserting the thermocouple wire into a steak, watch the video at http://www.meatscience.org/sensory

- The end of the thermocouple should not be touching bone or fat.
- For patties, follow the Food Safety and Inspection Service instructions (FSIS, 2013) for verifying the internal temperature of the patties. If patties are thick enough, the thermocouple wire can be used in the same way as described for steaks/chops above. If patties are too thin, either a probe-type thermocouple or a thermocouple in the tip of a hypodermic needle is used. Insert the thermocouple into the center of the patty during the last phases of cooking. Place the sensitive part of the thermocouple as close to the geometric center of the patty as possible. Do not help the thermocouple along by tapping the patty or wiggling the probe. Do not pick the patty up off of the cooking surface to insert the probe.

C. Evaluating Degree of Doneness

Because of concern regarding food safety and for consistency in cooking that provides the greatest likelihood of detecting treatment differences, all meat in an experiment should be cooked to the same temperature endpoint, unless effects of endpoint temperature are part of the experiment. Researchers should collect information regarding cooking times, peak/endpoint temperature, and cooking yields. Because of differences in cooking rate, the method of measuring "cooked temperature" is important. Some researchers prefer cooking to an endpoint temperature and then removing the samples from the heat, whereas others prefer to remove samples from the heat before the target endpoint is reached and then monitor temperature rise and record the peak temperature (usually for cooking methods with a high cooking rate). It is recommended that if cooking at a lower rate such that the post-cooking temperature rise is less than 5°C, remove meat from the heat at the endpoint temperature. If cooking at a high rate such that the post-cooking temperature rise is greater than 5°C, then determine at what temperature or cooking time samples should be removed from the heat in order to obtain the target endpoint temperature after the post-cooking temperature rise. Appropriate thermocouples and recorders should be used for measuring and recording internal temperature.

Changes in color, interpreted as degree of doneness upon completion of cooking, might be influenced by many factors including animal maturity, type of muscle, fat content, added ingredients, method of cooking, pH, length of cooking, and internal temperature. Color variation occurring in beef patties, such as "premature browning" or "hard-to-cook hamburgers" when cooked to constant final temperatures, poses a unique problem regarding "safe" final internal temperatures. Because of the possibility of color variation in beef patties, apparent degree of doneness based on cooked color cannot be used as a reliable indicator of the selected temperature endpoint; thus, actual temperature should be measured.

IV. COOKERY METHODS

Cookery method can markedly affect the palatability and cooking losses of meat. A cooking method should be selected with the same care one would use in selecting any other research method and likely varies depending on the research objective. Questions that should be addressed include: (1) Is the cooking method similar to common consumer practice? (2) Can the cooking temperature be controlled in a precise and consistent way? (3) Might the cooking method significantly affect palatability and perhaps overshadow treatment effects? (4) Will the cooking method cause excessive cooking losses? (5) But, most importantly, will the cooking method provide results relevant to the research question?

The following guidelines are based upon results from numerous scientific meat cookery and palatability studies and on the availability and affordability of cooking equipment. These recommendations: (1) are a reflection of methods commonly tested and used by meat scientists, (2) are those in which research has demonstrated that they consistently detect treatment differences, (3) do not greatly extend cooking time or cooking losses, and (4) do not mask flavors or impart unusual flavors. These guidelines are not intended to prevent research on new cooking methods as technology advances or to prevent methods of preparation that might enhance palatability and nutrient retention. Therefore, the guidelines are not intended to stifle creativity in investigating new methods. They are to give direction for the proven accurate, consistent, and commonly used methods to provide more continuity among various research institutions and accuracy in all data.

The terminology and definitions for cooking methods vary among investigators. These guidelines attempt to clarify the terminology and resolve confusion that appears in the literature. Cooking steaks in ovens has been described as “oven roasting,” “modified oven roasting,” or “modified broiling.” There is some confusion as to whether oven broiling is the same as oven roasting. For intact muscle cuts, two methods of cooking with dry heat (roasting, broiling) are defined and suggested. In regard to moist heat cookery, many researchers are not particularly pleased with the results in palatability and cooking loss obtained from braising. Thus, the guidelines do not include braising, and it is recommended that further research be conducted to provide the changes necessary for this or a similar cooking procedure to be acceptable.

In the past, open-hearth electric grills have been used rather extensively by researchers as a broiling method. These units are no longer manufactured, however—they have been shown to have a wide range in cooking temperature (Berry and Dikeman, 1994) and result in low repeatability of measures of tenderness. Gardner, Nelson, Dolezal, Morgan, and Novotny (1996) compared impingement oven and clam-shell cookery and demonstrated them to provide acceptable cooking. Clam-shell cookery equipment generally has been shown to provide a consistent cooking environment, rapid cooking times, and significantly reduced cooking losses compared to open-hearth grills. Conveyor belt grills have been shown to be very consistent in providing a consistent cooking environment. Wheeler, Shackelford, and Koohmaraie (1998) compared belt-grill cooking with open-hearth electric grill cooking and the belt grill was superior for Warner-Bratzler Shear Force (WBSF) and trained sensory panel trait repeatability. Belt grill instruments, however, are no longer manufactured.

Microwave and high-velocity forced-air, convection-oven cookery, while popular with consumers and foodservice, are *not* suggested as cooking procedures in these guidelines because of the inconsistencies among microwave ovens and in cooked product characteristics. Many studies show that high-velocity, forced-air convection oven cookery does not meet most of the criteria for selecting research cooking equipment and is not recommended.

Researchers should use a cooking method that best combines consistency and relevance with the goals of minimal cooking losses, relatively rapid cooking times, consistent cooking temperatures, and minimal masking of normal flavors or imparting of abnormal flavors. Lack of attention to details in selecting cookery equipment and controlling the cooking process can lead to variability among data sets and interpretation of data among institutions.

A. Roasting

Roasting is a method by which heat is transmitted to the meat by convection, either by normal or forced air, in a closed, preheated oven. As indicated above, forced-air convection-oven cookery is not recommended. For roasting, meat is placed on a rack either in or over a shallow pan to catch drippings. The oven door is closed and the meat is not turned during cooking.

Roasting procedures

- Roast meat at 163°C. Preheat oven to 163°C (higher, if necessary, to control temperature drop when the door is opened).
- Take roast(s) directly from refrigerator and record weights. Place roast(s) on a rack in the center of a roasting pan.
- Insert thermocouple into the geometric center of the meat. Record internal temperature. Roasts with the highest internal temperature should be cooked first.
- Place another thermocouple near the center of the oven adjacent to the meat to record oven temperature.
- Minimize the number of roasts in each oven. If the door is opened to remove a roast, the oven temperature may drop 10–30°C, depending on the type of oven.
- Roast to desired internal temperature; the recommended degree of doneness for all species is 71°C.
- Control the oven temperature to within ± 5°C with whatever means available. Some research ovens are insulated and equipped with devices to control the temperature to within ± 2°C. Most household ovens do not have sensitive temperature controls, so oven temperature should be constantly monitored. By use of the thermocouple in conjunction with oven thermostats and vents, the rise and fall of the oven temperature can be controlled.
- Leave roasts on racks in pan during cooling and record the weights necessary to determine cooking losses.
- Record endpoint or peak temperature as described above.
- Evaluate samples as soon as possible. Calculate cooking losses.

B. Broiling

Broiling is a method by which meat is cooked by direct radiant heat or "broiler." The meat is placed above or below the heat source. The heat usually radiates from one direction, so the meat must be turned during cooking. In these guidelines, this method will be referred to as "broiling." Many investigators prefer broiling because it closely resembles the method commonly used by consumers. Others prefer the "oven-roasting" technique because the cooking conditions can be more easily controlled. The results of Cross, Stanfield, Elder, and Smith (1979) comparing the roasting and broiling techniques on beef steaks indicated either would be acceptable.

Broiling procedures

- Set oven or broiler on "broil" long enough before initiating cooking for broiler to reach its ultimate temperature. Open-air broilers may have a temperature range of 190–232°C at the cooking surface. In ovens having multiple broiler settings, use "high" to be comparable to open-hearth broilers.
- Record weights.
- Insert thermocouples into the geometric center of the steaks, chops, or patties as described in the section on monitoring temperature. Record initial temperatures.
- Place the meat and thermocouple on a rack or in a broiler pan. Arrange meat and rack 8 to 15 cm below or above the heat. Preliminary testing may be necessary to obtain the optimal temperature setting and distance from the heat source. Broiling temperatures may vary widely and should be monitored and reported in the experimental section of the manuscript.
- Preferably, steaks and chops should be turned once with tongs and not forks during broiling. This may be at half the anticipated cooking time or when the internal temperature has reached half the increase to the final internal temperature. The proper point to turn cuts depends on the type of equipment used and should be determined in preliminary studies. Additional turning may be necessary if uneven cooking is apparent from both sides of the cut or excess browning or crust formation occurs.
- Remove the meat when it reaches the desired internal temperature (71°C is the standard for cuts of all species). Record the cooking time.
- Immediately record cooked weights for determination of cooking losses.
- Evaluate steaks and chops as soon as possible.

C. Panbroiling

Panbroiling is a method by which small, thin cuts such as patties are cooked by direct heat of conduction. Patties are placed in preheated frying pans or electric griddles and are cooked without added fat or water. Patties are turned frequently if needed to prevent excess surface browning or crust formation and to allow for more even cooking. The pan or griddle should not be covered during panbroiling.

Panbroiling procedures for patties

- Preheat skillet or electric griddle to 163°C. Monitor surface temperature to know what stage of the temperature cycle the griddle is in when placing patties.
- Record patty weights.
- Place fresh, frozen, or thawed patties on griddle or skillet. Loosen the patties after 30 seconds with a spatula to prevent sticking, especially if patties are low in fat.
- Turn patties every 2 minutes until 4 minutes of cooking has occurred. After 4 minutes, more frequent turning may be needed to prevent sticking and excess surface crust formation. Patties weighing 91.5 g cooked from the frozen state frequently need 6 to 9 minutes total cooking time. Fresh or thawed patties frequently need 4 to 7 minutes total cooking time.
- From preliminary studies, determine time when patties are approximately 5°C from endpoint. Insert thermocouples at this time as described above.
- Remove patties when they reach 71°C.
- Lightly blot surface moisture and fat, and immediately record cooked weights for determination of cooking losses.
- Evaluate patties for palatability and color as soon as possible.

D. Impingement Oven

Impingement ovens use air heated by gas or by electricity and forced into a chamber or chambers (in large commercial systems multiple chambers are used to control the cooking method) and the product is carried via a belt, usually a wire mesh belt, through the chamber. Some systems have humidity controls within each chamber. These systems have been used and have replaced belt grills for some institutions.

Impingement oven procedures

- Many commercial models are available for laboratory sensory or instrumental tenderness analyses; one chamber systems are usually used. Some manufacturers market ovens that are convection conveyor ovens as well as impingement ovens.
- Key aspects are control of the service or chamber temperature and length of time or dwell time of the meat product in the chamber to achieve the final internal temperature desired for the steak, chop, or patty. Preliminary tests need to be conducted to determine dwell time with standard service or chamber temperature. Dwell times should be monitored during cooking, especially during more than 2 hours of use of the oven to adjust dwell times to most accurately obtain internal temperature endpoint. Usually, service or chamber temperature is standard within a study and dwell time is altered to standardize cooking.
- Steaks, chops, or patties can be thermocoupled and temperature monitored at the initiation and at the end of the cook cycle. Thermocouple wires that are insulated and can withstand the oven service temperature should be used.
- When samples are within 2 to 3 inches of exiting the cooking chamber, thermocouple wires can be obtained and temperature measured. If the steaks or patties are close to the desired internal temperature, they can be pulled at the exact internal temperature.

If the steak or patty needs additional dwell time, it can be pushed back into the chamber to increase internal temperature and it can then be removed at the appropriate time. If standard dwell time is used without individually monitoring the internal temperature to define cook time as described above, increased variation in internal temperature, cook yield, and sensory characteristics may occur.

- Initial internal temperature, initial weight, and time in should be recorded. Total dwell time, final internal temperature, cook chamber temperature, and final cooked weight should be recorded after cooking.

Humidity, room temperature, steak or patty thickness, and factors influencing heat transfer within the product affect dwell time to reach a defined internal temperature endpoint and contribute to different dwell times across days and within a day.

E. Clam Shell

Clam shell is a method of cooking to simulate grilling that uses either a flat griddle or raised grill-mark heated surface with a lid that closes down on top of the sample that also is heated so that samples cook on top and bottom simultaneously. This is a popular method of cooking that has largely replaced open-hearth electric grills.

Clam shell procedures

- George Forman, Pro Star Max, and other brands are grills with a lid that closes on the meat and heats from top and bottom.
- Plug in the grill and allow to preheat at least 10 minutes or according to manufacturer's instructions.
- Record sample weights.
- Insert thermocouples into the geometric center of the steaks, chops, or patties as described above. Record initial temperatures.
- Place the meat on the grill surface so that the thermocouples are accessible with the lid closed. Close the lid. Grill temperatures may vary and should be measured and reported in the experimental section of the manuscript.
- Steaks and chops should not need to be turned during grilling because they are heated from top and bottom.
- Remove the meat when it reaches the desired internal temperature (71°C is the standard for cuts of all species). Record cooking time.
- Immediately record cooked weights for determination of cooking losses.
- Evaluate steaks and chops as soon as possible.

F. Summary of Recommendations for Cookery Methods

These guidelines do not specify which cooking method or manufacturer's equipment should be used because the cooking method chosen may depend on the objectives of the research. The cookery method selected, however, should provide acceptable repeatability levels ($R > 0.70$) for tenderness measures and minimal cooking losses, and it should be relatively rapid. Because no repeatability minimum is set for flavor, juiciness, or other sensory traits, this does not imply that juiciness and flavor are not important. Rather, the repeatability of tenderness is much higher

than for juiciness or flavor traits. For determining the variability imparted by cooking method, however, it is important to control the non-cooking factors influencing tenderness as much as possible. Scientists are encouraged to investigate cooking methods that would improve repeatability of tenderness, juiciness, and flavor; that minimize cooking losses; and that do not mask flavor or impart off-flavors.

In selecting appropriate cooking equipment, researchers are encouraged to work closely with manufacturers. Information is not readily available regarding the suitability of one particular type, model, or manufacturer over another. In communicating with equipment manufacturers, it is suggested that information be requested from the company regarding the following:

- Sensitivity, accuracy, and precision of temperature control
- Variation in temperature swings
- Frequency of temperature swings
- “Hot”- and “cold”- spot variation

These items could be included as a part of the purchase specifications. Before considerable investment is made, it might be advisable to test equipment in the manufacturers’ test facilities or consult someone who has used the equipment in the past.

Methods other than those described that are demonstrated to be acceptable may be used.

V. SENSORY TESTING FACILITIES

A. Sensory Testing Environment

The methods used for preparing and presenting samples to the panel require decisions about the testing environment, number of sessions, and physical condition of the samples. The validity of sensory panel results is partially dependent on control of various factors within the testing environment. General recommended procedures for panel room location and layout, lighting, odor control, and comfort of panelists are presented in ASTM MNL26-2nd (1996). Detailed recommendations on these items as well as sensory panel booth design can be found in ASTM MNL60-2nd (2008).

The sensory testing facilities should be in an accessible location with sufficient space, temperature and humidity control, and freedom from noise and odors. There should be a slight positive pressure within the sensory room to prevent cooking odors from entering the room.

Sensory panel booths are commonly used for product evaluations as they maximize the ability to control the testing environment (lighting, temperature, food odors, noise, etc.). If space is at a premium, these booths can be portable and capable of quick assembly. Incandescent, fluorescent, or both incandescent and fluorescent lighting can be used in the testing area. Whichever approach is chosen, the lighting should be uniform. The use of a dimmer switch is desirable because it allows for a variety of intensities ranging from 70 to 80 foot candles, which is typical in an office area, to higher levels of 110 foot candles.

In situations where treatments or cooking methods create color variability, red filtered lights can be used to mask the color differences. When testing with consumers, however, colored lights should only be used if absolutely necessary because colored lights can cause artifacts or non-typical responses from consumer panelists; the colored lights might make them suspicious and more attentive to smaller differences in the samples that they would not normally detect using standard lighting. It is best to select consumer panelists whose preferences match the degree of doneness that they will be evaluating when treatments or cooking methods create color variability.

For descriptive analysis, panelists can be seated around a table, usually a round table. Panelists who have extensive training and experience should be able to independently evaluate products using a round table. The panel leader should assess the situation and only place panelists around a table when they are assured that this environment will not result in communication and influence of one panelist to others.

B. Additional Options for Testing Locations

While a sensory lab is the best location for testing with trained panelists because of the ability to control the testing environment, there are several alternative options for testing locations when conducting consumer tests, depending on the test objectives and the compromises that researchers are willing to make. The ability to control sample preparation and test administration, which includes noise and distractions, varies greatly depending on the type of testing location that is chosen.

- Field agency facility—This can be similar to a sensory laboratory in the ability to control sample preparation and test administration. In order to test more efficiently, panelists often are seated at individual tables in a classroom-style setting with all panelists facing the same direction.
- School cafeterias and churches—These are easy to recruit and demographics can be controlled by the selection of the locations and groups in each city. There is less ability to control the atmosphere and distractions, however, than in the booths. Testing in school cafeterias during lunchtime should be avoided because of the excessive amount of noise and distractions.
- Grocery stores and restaurants—These are easy to recruit but contain more distractions, and the ability to control the testing environment is very limited. The consumer sample size should be increased to accommodate the less-controlled testing environment.
- In-home testing—This type has the least amount of testing control but provides actual use conditions. This would be important in testing packaging features and product preparation, as well as getting feedback from other household members.

VI. PREPARATION AND PRESENTATION OF SAMPLES TO THE PANEL

A. Preparation of Sensory Samples

Selection of sample preparation method and serving size should be determined based on project objectives and the amount of variation between and within treatments. It is critical that each panelist receive a standardized amount of each sample. Standardization of samples should be not only by weight or dimensions but also by temperature.

1. Trained panel evaluations

In order to account for the moderate to sometimes high degree of variability between and within treatments, meat samples often are cut into cubes, and each panelist receives two to three cubes from different locations within the piece of meat. For steaks, chops, and roasts, cubes that are 1.27 cm × 1.27 cm × the thickness of the cooked cut are suggested. If cooking procedures result in variation in cut thickness and charred surfaces, however, the thickness dimension should be standardized and cooked surfaces removed from cuts. After the hot sample is trimmed of all bone and epimysia connective tissue, it is placed in the plexiglass sample sizer (Figure 2). The sample sizer should have dimensions of 14 cm long × 12 cm wide × 4 cm deep to accommodate large cuts. On each side, the slots are spaced 1.27 cm apart and have an opening of 3 mm to allow the knife to cut the sample in each direction. For beef patties (depending on the size of beef patties being evaluated [91.5 or 113.5 g]), cooked patties can be cut into six or eight pie-shaped samples as shown in Figure 3. Even with thicker or larger-sized patties, cutting patties into cubes might result in breakage and the inability to obtain equal-sized pieces to serve the panelists. Cutting patties into pie-shaped or wedge samples is recommended.

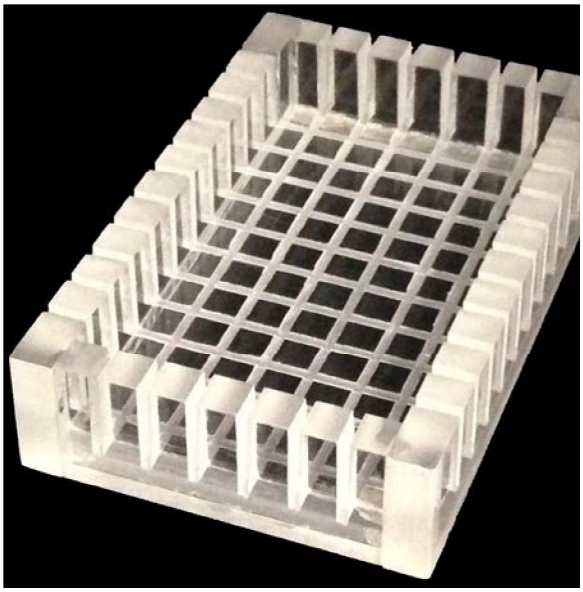


Figure 2. Sensory sample sizer for steaks or chops



Figure 3. Sectioning beef patties for sensory evaluation.

2. Consumer panel evaluations

During the normal eating experience, consumers assess the juiciness of the meat visually as well as get an initial impression of the tenderness as they cut the sample into bite-sized pieces. Therefore, when conducting consumer tests, it is best to serve samples that are large enough for the panelist to cut in order to provide a more accurate representation of the actual consumer eating experience. The serving size should be standardized. Location effects within a sub-primal should be randomized. The effect of location can then be included in the analysis of variance (ANOVA) model, usually as a random effect.

Depending on the objectives and circumstances of the study, there will be times when it is advisable to serve the smaller 1.27-cm cubes. By randomizing the selection of the cubes within a steak, the variability in tenderness within a sample can be better accommodated. This results, however, in a poorer simulation of the normal eating experience when panelists receive the bite-sized cubes. Researchers should understand that they are giving up aspects of the eating experience that influence consumer perception when serving 1.27-cm cubes, and therefore, the interpretation and use of the results might be affected. The experimental design might need to be altered (i.e., increase the number of consumers per treatment) if cubes are used.

B. Sample Presentation

Standard presentation procedures need to be followed to insure that all panelists receive samples at the most appropriate and consistent temperature for the attributes being measured. The minimum recommended serving temperature for meats is 60°C (ASTM E1871, 2010) and under most conditions should be adhered to. At this temperature, especially for poultry, maximum volatile aromatics will be detected. For consumer studies designed with endpoint temperatures to meet the consumer preferences, however, lower serving temperatures may be needed. Under these conditions, it is recommended that meat be held at 49°C for no more than 20 minutes in a glass, covered container to not affect the sensory properties for steaks and roasts. If samples are at 49°C when cut, placed in the serving container, and served, the sample should be about 38°C. To humans, the sample will be warm, but not cold. For each study, procedures for holding, cutting, and serving should be determined ahead of the study. Temperature should be monitored to assure that the samples are a standard temperature when serving and that the temperature is not too high or low. Variation in temperature of samples when the panelists receive them should be known. As flavor and texture are affected by temperature and holding time, not only should the serving temperature be consistent, but the holding time of the cooked samples should be consistent as well. When cooking larger cuts to constant temperature endpoints, appropriate time spacing between samples may be difficult to achieve due to different rates of cooking. Some samples may need to be held in a warm environment. Panelists should be informed so that they understand that time between samples (a minimum time should be defined and adhered to) may vary and be longer than the minimum.

Ideally, the samples will be served immediately after being cut, with panelists receiving cubes from various locations within the steak or, in the case of larger portions for consumer panelists, the same steak portion for all treatments. If samples need to be held prior to serving, preliminary tests need to be conducted to assure that holding methods do not affect color,

tenderness, juiciness, or flavor of the samples. Several procedures for maintaining the temperature of the samples have been used. These include the use of the following:

- Covered pans or glass dishes that are placed in a preheated container of sand or on a warming plate or heated oven (i.e., 49°C)
- Double boilers on electric hot plates
- Wrapping the sample in aluminum foil or placing sample in Pyrex or glass baking dish with lid and storage in a heated oven
- Using preheated yogurt-maker glass dishes placed in the yogurt maker or similar apparatus

Care should be taken in selecting the serving containers and utensils. They should be neutral in color (unless tint is needed to mask color differences) and must be made of inert materials that are nonreactive and odor free so that they do not impart any odors or flavors to the samples (ASTM E1871, 2010). Glass or white-glazed china that has been washed in an unscented detergent, followed by baking at 93°C for several hours, is best but not very practical for large studies. In these cases, plastic containers and utensils can be used. It should be noted, however, that some plastic materials are less inert, more susceptible to temperature changes, and less odor free than others, so they should be pretested prior to their use. The use of wooden toothpicks could transfer flavors to the meat and disrupt textural integrity of the sample and should, therefore, be avoided when possible.

Samples should be coded with three-digit random numbers. Materials used to mark containers with three-digit random numbers should be odor free. For example, if Sharpie® pens are used, the codes should be placed on the containers 24 hours prior to use and the container allowed to air for 24 hours prior to use. Avoid an alphabetical or numerical order to minimize bias.

1. Order of sample presentation

Every sample should be served in each serving order an equal number of times to reduce any bias related to serving position. Furthermore, every sample should be served before and after every other sample an equal number of times in order to nullify any bias related to carryover effects. William's Square designs are one way to achieve both types of balance. For some studies with a large number of treatments, this may not be possible. Order should be randomized using a random number generator, and order should be analyzed as a random effect in the model.

For trained descriptive panels, first sample bias and variation associated with evaluations conducted on different days can be an issue. A standardized warm-up sample—usually a sample that represents the typical product being evaluated in the study—should therefore be served to the panelists at the initiation of the sensory session and then discussed. This warm-up sample will assist in standardization among panelists, improve panelists' concentration, increase panelists' confidence, and increase calibration of panelists before initiating the sensory session. The sensory leader can use the warm-up sample as a tool to address panel drift and lack of motivation, to increase panelist confidence, and to remove prior environmental factors that can influence the sensory verdict on a given day.

2. Number of samples per session

The number of samples that should be presented in a given session is a function of the following:

- Product characteristics
- Experience of the panelists
- Sensory and mental fatigue
- Number of attributes to be measured per sample

Readers are encouraged to review Bohnenkamp and Berry (1987) regarding the effects of sample numbers/session and sessions/day on panelist performance in evaluating beef patties.

Great care should be taken in setting the number of samples to be served per session in consumer testing. Because consumers can differ greatly in their food preferences, panelists are the greatest source of error in the statistical model. It is, therefore, best for all panelists to evaluate all products. In order to minimize taste bud fatigue and loss of interest or concentration among the panelists, however, the number of samples served per session should be limited based on the product type and the number of questions on the ballot. For unseasoned meat products, panelists can easily evaluate six to eight samples per 1-hour session. Spicy or highly seasoned products, such as marinated pork tenderloin with a spicy barbecue flavor, should be limited to four samples per session. Palate cleansers such as room-temperature distilled water and unsalted crackers should also be used to minimize taste bud fatigue and flavor carryover. Additionally, fat-free ricotta cheese is an effective palate cleanser for spicy products, whereas warm water or seltzer water is effective for high-fat products. Preliminary studies should be conducted with various palate cleansers to ensure that the palate is thoroughly cleansed without contributing to taste bud fatigue or influencing attributes within the product.

In situations where the number of samples in the test is greater than what each person can evaluate in one session, it is best to conduct the test over multiple days, with each panelist evaluating a subset of samples each day. If multiday sessions are not an option, then a partially balanced, incomplete-block design can be used, but the overall number of panelists should be increased to achieve the desired number of observations per sample. Consult a statistical text for tables on partially balanced, incomplete-block designs. These designs ensure that each sample or treatment appears with each other sample or treatment an equal number of times in any given session.

Because the length of the questionnaire also should be considered when determining the maximum number of samples to serve as well as the session length, it is a good practice to do a pretest with the ballot and the actual test products in order to verify that the panelists will be able to complete the test without loss of sensory acuity. Consumer testing generally results in large “halo effects,” and thus ballots should be as short as possible.

C. Sensory Panel Participants’ Informed Consent

The Belmont Report entitled *Ethical principles and guidelines for the protection of human subjects of research* was created by the Department of Health, Education, and Welfare in 1979

to safeguard human subjects used for research, including sensory panels (NIH, 1979). This report protects human subjects used in research by fulfilling three fundamental ethical principles:

- Respect for persons—protecting the autonomy of all people and treating them with courtesy and respect and allowing for *informed consent*
- Beneficence—maximizing benefits for the research project while minimizing risk to the research subjects
- Justice—ensuring that reasonable, non-exploitative, and well-considered procedures are administrated fairly

In the United States, federally funded research and research conducted at state and federally supported institutions involving human subjects such as sensory panels must obey ethical rules that include obtaining participants' informed consent and supervision by an institutional review board (IRB). The form prepared for obtaining participants' consent should precisely convey all the information about the sensory panel to the participants. The consent form should include the purpose of study and study design, who can participate, who will be conducting the research, what participants will be asked to do, possible risk and discomforts, possible benefits/compensation, and who to see with problems or concerns. Figure 4 shows an example of a sensory panel informed-consent form.

IRB#
Date Approved: Valid Until:

SENSORY PANEL CONSENT FORM

Title of Protocol: Evaluation of Meat Palatability Traits

INVITATION TO PARTICIPATE

You are invited to participate in a taste panel. The panel will be conducted in a university or commercial facility.

BASIS FOR SUBJECT SELECTION

You received this invitation because you met the demographic criteria required for this study. By signing this consent form, you will be selected for this study. Individuals with colds, sinus conditions, or allergy to a specific ingredient cannot participate. The general adult population is used for testing. Participants must be 19 years old or older and a nonsmoker.

PURPOSE OF THE STUDY

The purpose of this study is to evaluate meat palatability characteristics. The products used will be normal, unadulterated foods containing only approved food ingredients, produced under U.S. Department of Agriculture inspection.

EXPLANATION OF PROCEDURES

Samples will be provided for you to place in your mouth for evaluation of palatability characteristics. These samples will be whole muscle meat products that have been cooked and cut into bite-sized pieces for your evaluation. Samples need not be consumed and can be discarded in provided containers following evaluation. You will be provided water and crackers to cleanse your palate between samples. Testing will occur in a university or commercial facility and will take approximately one-half hour. You are requested to refrain from smoking, eating, or drinking fluids other than water for one hour prior to each session.

POTENTIAL RISKS AND DISCOMFORTS

There will be no risks other than those normally associated with eating of meat products. The food will be prepared under sanitary conditions.

POTENTIAL BENEFITS

Your recognition of the importance of sensory panels, and your contribution to them, is one benefit. Society in general benefits from the production of meat products with improved consumer acceptance.

ASSURANCE OF CONFIDENTIALITY

Any information obtained in connection with this project and that could be identified with you will be kept confidential. Summary results and statistical data may be reported in scientific journals or presented at scientific meetings; however, individual panelist responses will be maintained in confidence.

WITHDRAWAL FROM THE STUDY

Participation in this study is voluntary. Your decision whether or not to participate will not affect your present or future relationship with the investigator or the university. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without penalty.

COMPENSATION FOR PARTICIPATION

You will be given a small, commercially available, wrapped candy treat; a coupon for credit at the Meat Sales Store; an entry into a drawing for a small cash prize; or payment for your time.

OFFER TO ANSWER QUESTIONS

If you have any questions, please do not hesitate to ask. If you think of questions later, please feel free to contact (Investigator Name), at (Investigator Phone Number). If you have any additional questions concerning the rights of research subjects, you may contact the university Institutional Review Board (IRB) by telephone at (Institutional Review Board Phone Number).

YOU ARE VOLUNTARILY MAKING A DECISION WHETHER OR NOT TO PARTICIPATE IN THIS RESEARCH TODAY. YOUR SIGNATURE CERTIFIES THAT YOU HAVE DECIDED TO PARTICIPATE HAVING READ THE INFORMATION PRESENTED. YOUR SIGNATURE ALSO CERTIFIES THAT YOU HAVE HAD AN ADEQUATE OPPORTUNITY TO DISCUSS THIS STUDY WITH THE INVESTIGATOR AND YOU HAVE HAD ALL YOUR QUESTIONS ANSWERED TO YOUR SATISFACTION. YOU WILL BE GIVEN A COPY OF THIS CONSENT FORM TO KEEP.

_____ _____

SIGNATURE OF SUBJECT DATE

IN MY JUDGEMENT THE SUBJECT IS VOLUNTARILY AND KNOWINGLY GIVING INFORMED CONSENT AND POSSESSES THE LEGAL CAPACITY TO GIVE INFORMED CONSENT TO PARTICIPATE IN THIS RESEARCH STUDY.

Figure 4. Example Consent Form for Sensory Panelist **Reference:** NIH. 1979. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. http://ohsr.od.nih.gov/guidelines/belmont.html.

VII. SENSORY EVALUATION METHODS

Sensory evaluation uses both subjective and objective procedures. Untrained consumer studies are designed to determine the responses of "typical" consumers. Yet, through the use of sophisticated panel training and method selection, sensory evaluation also can provide accurate and repeatable objective data. Additional details of the various sensory techniques can be found in ASTM and IFT publications as well as Stone and Sidel (2004), Meilgaard, Civille, and Carr (2007), and Lawless and Heymann (2010). The IFT Guidelines for the preparation and review of papers reporting sensory evaluation data (1995) should be carefully read by researchers planning to publish in the *Journal of Food Science*. Approval by the institution's committee for use of human subjects in research will be needed prior to the initiation of any sensory work.

Scaling is an important aspect of sensory testing methods. Scaling uses words or numbers to express the intensity or degree of an attribute. For many research projects in which sensory traits of meat are evaluated, rating scales are most appropriate. The following types of scales may be used:

- **Graphic or line scales**. Either a simple line or one marked off into segments. Intensity or degree of the characteristic evaluated must be shown on each end of the scale.
- **Verbal scales**. A series of brief written statements, usually the name of the sensory attribute, with appropriate adverbial or adjectival modifiers, that are written out in appropriate order.
- **Numerical scales**. A series of numbers, ranging from low to high, that are understood to represent successive levels of quality or degrees of a characteristic.

A. Discrimination Methods

Discriminative, or difference testing, methods are used to determine if there is a detectable difference between samples. Trained or untrained consumer panelists can be used, depending on the test objectives. If consumers are used, it is understood that consumers are not as sensitive to differences between products; whereas when trained panelists are used, smaller differences may be discernible.

Discriminative testing methods are divided into two broad categories: (1) testing for overall difference; or (2) testing for differences in specific attributes. Sensory professionals select the test based on the test objectives. A brief description of the five most commonly used discriminative tests used in meat research is shown below. Other methods are used and discussed in Meilgaard et al. (2007), Stone and Sidel (2004), and Lawless and Heymann (2010).

1. Overall difference tests

a. Triangle test

The Triangle test is an overall difference test and is believed to be very sensitive. The test does not, however, give directional or attribute information. There is a guessing rate of 33% that is accounted for in the statistical tables used to determine if a difference exists or not. The number of panelists required must be determined prior to conducting the test, and it is dictated by the level of sensitivity required by the study objectives and criteria. Test sensitivity is a function of the levels of α, β, and P_d (proportion of distinguishers). The α risk is the probability of finding a difference when there is not a difference, the β risk is the probability of missing a difference that is there, and P_d is the maximum allowable proportion of the population that could discriminate. Table 17.7 from Meilgaard et al. (2007) can be used to determine how many panelists are needed with the defined levels for α, β, and P_d. The following guidelines can be used in setting the levels for these factors (ASTM E1885-04, 2004):

- For α risk, a statistically significant result at 10% to 5% (0.10 to 0.05) indicates “slight” evidence that a difference is apparent, 5% to 1% (0.05 to 0.01) indicates “moderate” evidence, 1% to 0.1% (0.01 to 0.001) indicates “strong” evidence, and < 1% (< 0.001) indicates “very strong” evidence.
- For β risk, the strength of evidence that a difference is not apparent is assessed using the same criteria as above but substituting “is not apparent” for “is apparent.”
- For P_d, the maximum allowable P_d falls into three categories:
- P_d < 25% represents small values
- 25% < P_d < 35% represents medium-sized values
- P_d > 35% represents large values

If the study objective is to determine whether or not a detectable difference exists between treatments, the value selected for α is typically smaller than the value selected for β. For example, with α = 0.05, β = 0.20, and P_d = 30%, 40 panelists would be required. If the study objective is to determine if samples are sufficiently similar to be used interchangeably, however, the value selected for β is typically smaller than the value selected for α. For example, with an α = 0.20, β = 0.05 and P_d = 30%, 39 panelists would be required.

To conduct the test, panelists are presented samples either singly or all at once. The panelist is asked to evaluate each sample in the specified order and to identify the sample that is different. The instructions define that two samples are the same and one is different (Figure 5). Each sample should be presented with a three-digit code, and order should be randomized by panelist. There are six potential serving orders where A and B represent the two products: AAB ABA BAA ABB BAB BBA. Results are analyzed by comparing the number of correct responses to the total number presented using the predetermined α, such as Table 17.8 in Meilgaard et al. (2007).

TRIANGLE TEST

INSTRUCTIONS:

Taste the meat samples from left to right. Two are identical. Determine which is the different sample.

Please cleanse your palate with a sip of water and a bite of cracker in between samples.

If no difference is apparent, you must guess.

Set of Three Samples	**Indicate Different Sample**	**Comments Why sample was different**
831 731 248		

Figure 5. Sample ballot for Triangle testing. Adapted from Mielgaard et al. (2007)

b. Duo-Trio test

The Duo-Trio test is an overall difference test and is easier for consumers to understand than the Triangle test. There is a 50% guessing rate with the Duo-Trio test, however, and that is why researchers need more respondents to conduct this test than the Triangle test. Panelists are served a control. After evaluation of the control, panelists are served two samples; one is the same as the control and the other is different. They are asked to identify the sample that is the same as the control (Figure 6). Again, directional and attribute information is not obtained. As previously defined, the researcher defines the α, β, and P_d for the test. This determines the number of respondents as in Table 17.9 of Meilgaard et al. (2007). The recommended number of panelists is 69 with $\alpha = 0.05$, $\beta = 0.20$, and $P_d = 30\%$. Usually, the two samples are rotated because the control and the order of presentation of the two samples rotates as well. Based on this, the serving order with the first sample defined as the control is ABA AAB BAB BBA. After completion of the test, the total number of responses and the number of correct responses are counted. Using the α defined, the number of correct responses needed to determine a difference is identified as in Table 17.10 in Meilgaard et al. (2007). Additional information on how to conduct this test can be found in ASTM E2610-08 (2008).

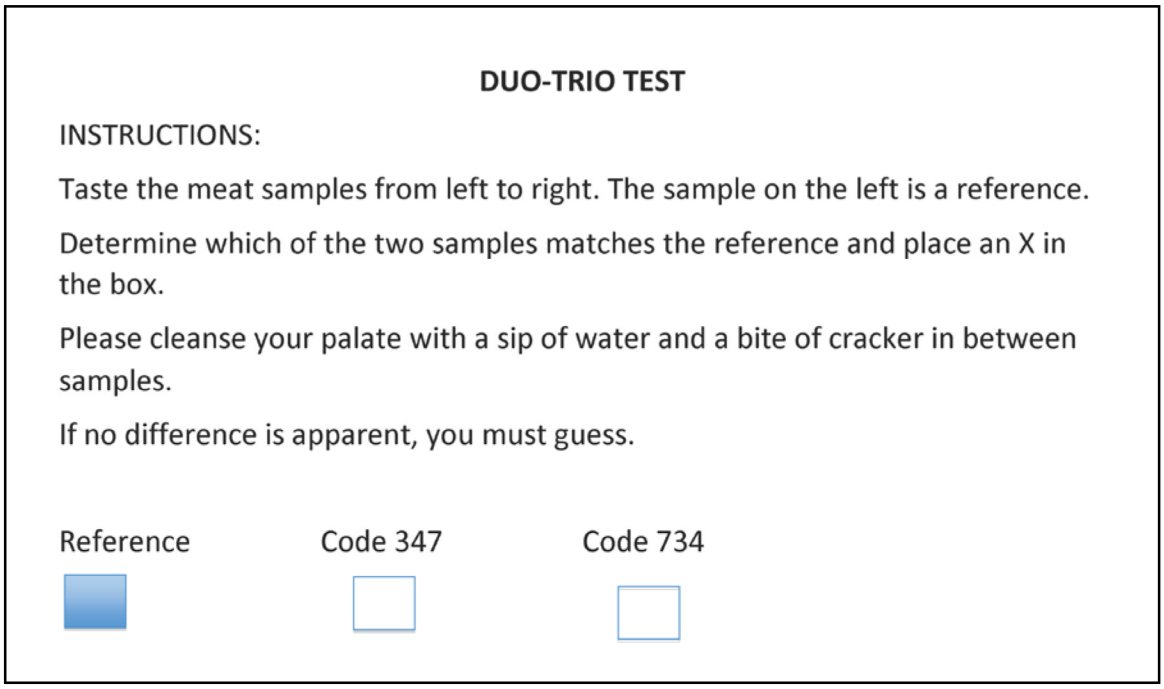


Figure 6. Sample Ballot for Duo-Trio Testing. Adapted from Meilgaard et al., (2007)

c. Degree of Difference

This method, also known as difference from control, can be used to determine not only if a detectable difference exists but also to measure the amount of differences among a set of samples. Degree of difference (DOD) testing is also recommended when products are highly variable. This method can be expanded to measure not only the degree of overall difference but also differences in specific attributes, such as beef flavor strength or tenderness.

In DOD testing, panelists receive the control sample twice—one time as a labeled reference sample and a second time served blind as a coded sample along with the other coded test samples. Panelists are instructed to taste the reference first, then each coded sample, and indicate the degree of difference from each coded sample using a scale from "none" to "very large difference," as shown in Figure 7. The scales are typically 7–10 points and can be line scales or category scales that are fully anchored or only end anchored—i.e., 0 = no difference and 7 = very large difference. Twenty to 50 panelists are generally recommended (Meilgaard et al., 2007). The number of panelists required, however, will be determined by the amount of variability within samples and the test objective. Is the emphasis on being able to detect a difference (emphasis on Type I error) or verify similarity (emphasis on Type II error)? Testing for similarity will require more sample presentations (panelists) than testing for difference.

This method can also easily accommodate multiple treatments in a study. With multiple treatments, it is best to present all samples at the same time, with clear instructions on the order in which samples are to be evaluated. If, however, it is not possible to present all samples at the same time due to issues such as taste bud fatigue, present only one test sample at a time along with the reference.

DEGREE OF DIFFERENCE TEST

Name _____ Date _____

You have received a set of samples—a "control" sample labeled R and three test samples, each labeled with a three-digit random code. One of the test samples might be a duplicate control.

1. Taste the control sample first, evaluating for appearance, flavor, and texture.
2. Next, taste each test sample in the order indicated, evaluating for appearance, flavor, and texture.
3. Indicate the size of the overall difference of each test sample, relative to the control, using the scale below.

No Difference	Very Slight Difference	Slight/Moderate Difference	Moderate Difference	Moderate/Large Difference	Large Difference	Very Large Difference
1	2	3	4	5	6	7

	Rating			
Sample Code	Flavor Difference	Texture Difference	Overall Difference	Comments / Describe differences
281				
440				
795				

Figure 7. Sample ballot for Degree of Difference testing

Results are analyzed by calculating the mean difference between the identified control and each test sample vs. the difference between the identified control and the blind control. The mean differences are then compared, using analysis of variance if there are multiple test samples in the study or paired t-test if there is only one test sample. Additional details on this method and data analysis procedures can be found in Meilgaard et al. (2007).

2. Specific attribute tests

a. Paired comparison tests

Paired comparison tests are the most common method used to determine the differences in one attribute between two samples. The advantage of this method is that it is simple to administer and for respondents to understand. Two samples are presented to the respondents and they identify whether the samples are the same or different. There is a 50% guessing rate with this test. Paired comparison tests can either be one-sided or two-sided tests. How the question is asked determines if the test is one sided or two sided. For a one-sided test, the question is whether A is more (or less) than B. For one-sided tests, the α, β, and P_d are determined using the same tables as for Duo-Trio tests. A two-sided test asks if A is different than B. When a test is two sided, the distribution is bilateral and is also defined as a Two-sided Directional Difference test. To determine the number of respondents, the α, β, and P_d should be predetermined and the number of respondents determined as in Table 17.11 of Meilgaard et al. (2007). After the two-sided test is completed, to determine if a difference was present between the samples, Table 17.12 of Meilgaard et al. (2007) can be used applying the predetermined α.

b. Ranking

Ranking tests are used to determine differences in several samples for a specific attribute. This test is different from the previously mentioned tests in that this testing method provides the ability to answer the question "How does attribute X differ among samples?" The respondents are provided a set of samples in a balanced, random order and asked to rank the samples from lowest to highest for the intensity of the attribute of interest. It is recommended that all samples be presented at one time, but presentation of samples singly also can be done. Respondents are trained; no fewer than 8 respondents should be used. If more than 16 respondents are used, the sensitivity of the test is increased. The data are analyzed using the Friedman's test. Additional information on this method can be found in Meilgaard et al. (2007) and ASTM MNL26-2 (1996).

B. Descriptive Analysis Methods

Descriptive sensory analysis involves sensory approaches that discriminate and describe both qualitative and quantitative properties using trained panelists. Considerable time and effort are involved in training and maintaining descriptive sensory panels. All sensory attributes of a meat product can be subjected to descriptive analysis or limited to selected criteria such as flavor and texture. The four prevalent methods are described below. For a more complete discussion and understanding of descriptive analysis methods, see ASTM MNL13-EB (1992).

1. Flavor profile method

In the flavor profile method, potential panelists are screened according to their abilities to discriminate aroma and flavor differences. This method uses descriptive terms to characterize

the flavor of a product as well as provide the intensities and order of appearance of the various aromas, flavors, and aftertastes detected. After the four to six members of a panel individually evaluate a sample, the results are submitted to the panel leader, who leads a discussion to arrive at a general consensus of the sample. Reference samples can be used to present flavor attributes. Data can be presented in tabular, graphic, or verbal format. With this method, panelist effects are not accounted for.

2. Texture profile method

The texture profile method is based on the same concept as flavor profiling in that overall texture of a product is comprised of a number of different texture attributes. Panelists define the procedures and terminology to use in the textural evaluation. Among the modifications in the texture profile approach (as outlined in ASTM MNL13-EB, 1992) are the following:

- Development of more precise definitions and evaluation procedures
- Development of new reference scales
- Use of various scaling procedures
- Collection of individual scores without consensus
- Statistical analysis of data

Texture attributes in the texture profile method can be classified as to the following:

- Mechanical
- Geometrical
- Those related to moisture and fat content

Mechanical characteristics are revealed as the meat samples react to stress, such as chewing. Geometrical characteristics relate to size, shape, and orientation of the product before and during breakdown. The contributions of moisture and fat are determined through mouth feel. Texture references have been developed and can be used for training. An example of a texture lexicon for ground beef patties was published by AMSA (1983). An example ballot that includes flavor and texture attributes for ground beef is presented in Figure 8. This ballot uses beef flavor attributes from the Beef Flavor Lexicon (Adhikari et al., 2011), as discussed below, and texture attributes from AMSA (1983). The texture attributes can be easily defined and scaled using solid oral texture attributes from Meilgaard et al. (2007). Note that not all texture attributes for ground beef were included on the ballot; only those important to the hypothesis of the study were included.

SAMPLE ID #:	warm-up				
AROMATICS:					
Cooked Beef Identity					
Cooked Beef Fat					
Serumy/Bloody					
Brown/Roasted					
Cardboardy					
Painty					
Fishy					
Livery					
Soured					
Burnt					
Green/Hay-like					
Metallic					
Other (describe)					
BASIC TASTES:					
Salt					
Sour					
Bitter					
Sweet					
AFTERTASTES:					
Fat Mouthfeel					
Bitter					
Browned/Burnt					
Sour					
Sweet					
Other (describe)					
TEXTURE:					
Initial Juiciness					
Sustained Juiciness					
Springiness					
Hardness					
Cohesiveness					
Cohesiveness of Mass					
Toothpacking					

Figure 8. Example of a flavor and texture descriptive attribute ballot for ground beef based on AMSA (1983) and Meilgaard et al. (2007); panelists provide ratings for each attribute from each sample provided.

3. Quantitative descriptive analysis

Quantitative descriptive analysis (QDA) was developed to provide a stronger statistical treatment of data than data from profile-type methods. While training is provided regarding methods and terminology, to some degree panelists are free to develop their own approach to scoring. Overall discussions of the evaluations after a session are not usually conducted. Data are reported in the form of a spider web with a branch or spoke for each attribute. An example of this form of data reporting is shown in Figure 9.

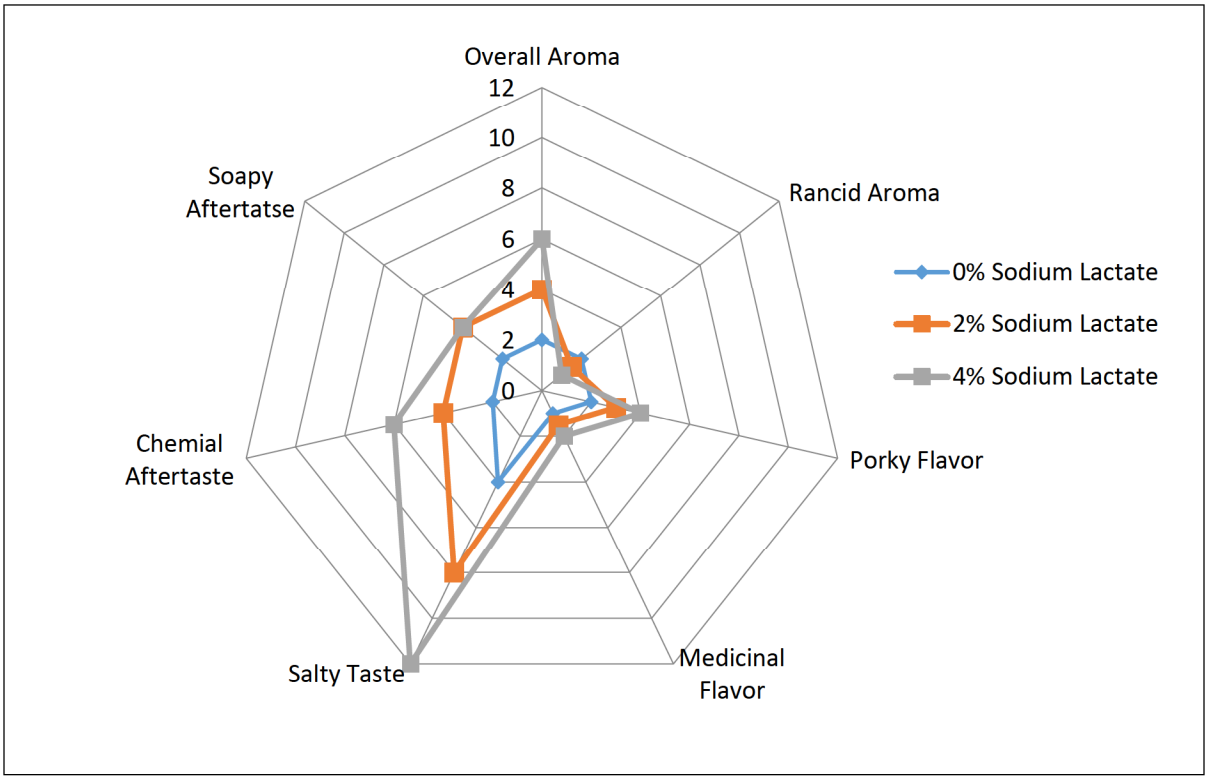


Figure 9. Spider web from QDA where three levels of sodium lactate were added to pork roasts.

4. Spectrum™ descriptive attribute analysis

The Spectrum™ procedure (Meilgaard et al., 2007) is a custom-design approach providing detailed information on the sensory attributes including aroma, flavor, and texture, as well as their intensities using absolute or universal scales. Many times a lexicon, or dictionary of attributes, references, and examples, is used as a basis to present specific attributes within a product. With this method, line scales or the universal 16-point intensity scale is used. Munoz and Civille (1998) discussed the use of the universal scale and lexicon development. Also, extensive panelist training is conducted to assure that each panelist understands each attribute, can scale each attribute for intensity, and can consistently accomplish this task across products. Use of established references and panel training is a necessity. The Beef Flavor Lexicon developed by Adhikari et al. (2011) is presented in Table 1 as an example of a list of terms that can be used to define beef flavor. This lexicon defines major aroma and flavor attributes found in whole-muscle beef. Table 2 has the Pork Lexicon developed by Chu (2015) that can be used similar to the beef lexicon but for whole muscle pork. For other whole-muscle meats, these two lexicons can be used as a base. When evaluating lamb, lamb identity would replace beef or pork identity, and attributes that might be product specific could be added based on ballot development sessions and referenced in ASTM DS72 (2011).

With this method, data are analyzed using Analysis of Variance to determine differences across products within an attribute. Additionally, multivariate analyses can be used to understand how multiple attributes affect a treatment of product. Analyses of these data are presented in the statistical analysis section.

5. Short-version Spectrum™ descriptive method for quality assurance and shelf-life studies

This method is based on the Spectrum™ method and applies the same principles when a full lexicon of attributes is not necessary. The sensory professional determines, either through lexicon development or product knowledge, the main attributes that need to be evaluated for a product. These attributes are defined, references are defined, panelists are trained on the references and scaling exercises, and the products are evaluated. See Munoz et al. (1992) for a more comprehensive understanding of how to apply this method.

Table 1. The Beef Flavor Lexicon defined by Adhikari et al. (2011).

Attribute	Definition	References
Animal hair	Aromatics perceived when raw wool is saturated with water	Caproic acid (hexanoic acid) = 12.0 (aroma)
Beef identity*	Amount of beef flavor identity in the sample	Swanson® Beef Broth = 5.0 (aroma and flavor) 80% lean ground beef = 7.0 (aroma and flavor) Beef brisket = 11.0 (aroma and flavor)
Bitter*	Fundamental taste factor associated with a caffeine solution	0.01% caffeine solution = 2.0 (flavor) 0.02% caffeine solution = 3.5 (flavor)
Bloody/serumy*	Aromatics associated with blood on cooked meat products; closely related to metallic aromatic	USDA Choice strip steak = 5.5 (aroma and flavor) Beef brisket = 6.0 (aroma and flavor)
Brown/roasted*	Round, full aromatic generally associated with beef suet that has been broiled	Beef suet = 8.0 (aroma and flavor) 80% lean ground beef = 10.0 (aroma and flavor)
Burnt	Sharp/acrid flavor note associated with over-roasted beef muscle, something over-baked or excessively browned in oil	Alf's Puffed Red Wheat® = 5.0 (aroma and flavor)
Chemical	Aromatics associated with garden hose, hot Teflon pan, plastic packaging, and petroleum-based product like charcoal lighter fluid	Ziploc® sandwich bag = 13.0 (aroma) Clorox® in water = 6.5 (flavor)
Cocoa	Aromatics associated with cocoa beans, powdered cocoa, and chocolate bars; brown, sweet, dusty, often bitter aromatics	Hershey's® cocoa powder in water = 3.0 (flavor) Hershey's® chocolate kiss = 7.5 (aroma), 8.5 (flavor)
Cooked milk	Combination of sweet, brown flavor notes and aromatics associated with heated milk	Mini Babybel® original swiss cheese = 2.5 (flavor) Dillon's whole milk = 4.5 (flavor)
Dairy	Aromatics associated with products made from cow's milk, such as cream, milk, sour cream, or butter milk	Dillon's reduced fat milk (2%) = 8.0 (flavor)
Fat-like*	Aromatics associated with cooked animal fat	Hillshire Farms Beef Lit'l Smokies® = 7.0 (aroma and flavor) Beef suet = 12.0 (aroma and flavor)
Green	Sharp, slightly pungent aromatics associated with green/plant/vegetable matters such as parsley, spinach, pea pod, fresh cut grass, etc.	Hexanal in propylene glycol (5,000 ppm) = 6.5 (aroma) Fresh parsley water = 9.0 (flavor)
Green-hay	Brown/green dusty aromatics associated with dry grasses, hay, dry parsley, and tea leaves	Dry parsley in medium snifter = 5.0 (aroma) Dry parsley in ~30-mL cup = 6.0 (flavor)
Leather	Musty, old leather (like old book bindings)	2,3,4-Trimethoxybenzaldehyde = 3.0 (aroma)
Liver-like	Aromatics associated with cooked organ meat/liver	Beef liver = 7.5 (aroma and flavor) Braunschweiger liver sausage = 10.0 (aroma and Flavor—must taste and swallow)
Metallic*	Impression of slightly oxidized metal, such as iron, copper, and silver spoons	0.10% potassium chloride solution = 1.5 (flavor) USDA choice strip steak = 4.0 (aroma and flavor) Dole® canned pineapple juice = 6.0 (aroma and flavor)

Table 1. The Beef Flavor Lexicon defined by Adhikari et al. (2011). *(Continued)*

Attribute	Definition	References
Overall sweet*	Combination of sweet taste and sweet aromatics; the aromatics associated with the impression of sweet	Post Shredded Wheat®, spoon size = 1.5 (flavor) Hillshire Farms Beef Lit'l Smokies® = 3.0 (flavor) SAFC ethyl maltol 99% = 4.5 (aroma)
Rancid	Aromatics commonly associated with oxidized fat and oils; may include cardboard, painty, varnish, and fishy	Microwaved Wesson® vegetable oil (3 min at high) = 7.0 (flavor) Microwaved Wesson® vegetable oil (5 min at high) = 9.0 (flavor)
Salty*	Fundamental taste factor of which sodium chloride is typical	0.15% sodium chloride solution = 1.5 (flavor) 0.25% sodium chloride solution = 3.5 (flavor)
Sour*	Fundamental taste factor associated with citric acid	0.015% citric acid solution = 1.5 (flavor) 0.050% citric acid solution = 3.5 (flavor)
Sour aromatics*	Aromatics associated with sour substances	Dillon's buttermilk = 5.0 (flavor)
Sour dairy	Sour, fermented aromatics associated with dairy products such as buttermilk and sour cream	Laughing Cow® light swiss cheese = 3.0 (aroma), 7.0 (flavor) Dillon's buttermilk = 4.0 (aroma), 9.0 (flavor)
Spoiled	Presence of inappropriate aromatics and flavors that are commonly associated with the products; a foul taste and/or smell that indicates the product is starting to decay and putrefy	Dimethyl disulfide in propylene glycol (10,000 ppm) = 12.0 (aroma)
Sweet*	Fundamental taste factor associated with sucrose	2.0% sucrose solution = 2.0 (flavor)
Umami*	Flat, salty, somewhat brothy; taste of glutamate, salts of amino acids, and other molecules called nucleotides	0.035% accent flavor enhancer solution = 7.5 (flavor)
Warmed-over	Perception of a previously cooked and reheated product	80% lean ground beef (reheated) = 6.0 (flavor)
Other attributes	**These attributes are either minor or may not be detected depending on the samples used in the study.**	
Smoky-charcoal	An aromatic associated with meat juices and fat dripping on hot coats which can be acrid, sour, burned, etc.	Wright's Natural Hickory seasoning in water = 9.0 (Smelled)
Smoky-wood	Dry, dusty aromatic reminiscent of burning wood.	Wright's Natural Hickory seasoning in water = 7.5 (Smelled)
Buttery	Sweet, dairy-like aromatic associated with natural butter	Land O'Lakes Unsalted Butter tasted = 7.0 (flavor & aroma)
Refrigerator stale	Aromatics associated with products left in the refrigerator for an extended period time and absorbing a combination of odors (lack of freshness/flat)	80/20 ground beef, cooked, left chilled overnight = 6.0 (F), 8.0 (A)
Soapy	An aromatic commonly found in unscented hand soap	0.12 oz Clorox Wipe Liquid in 4 oz Water= 3.0 (A) 0.5g Ivory Bar Soap in 100mL water = 6.5 (A)

Table 1. The Beef Flavor Lexicon defined by Adhikari et al. (2011). *(Continued)*

Attribute	Definition	References
Barnyard	Combination of pungent, slightly sour, hay-like aromatics associated with farm animals and the inside of a horn.	White pepper in water = 4.5 Tinture of civet = 6.0
Heated oil	The aromatics associated with oil heated to a high temperature	Wesson Oil, microwaved 3 min = 7.0 (F&A) Lay's Potato Chips = 4.0 (A)
Asparagus	The slightly brown, slightly earthy green aromatics associated with cooked green asparagus.	Asparagus water = 7.5 (Smelled) Asparagus water = 6.5 (Tasted)
Cumin	The aromatics commonly associated with cumin and characterized as dry, pungent, woody and slightly floral.	McCormick or Shilling Ground Cumin = 10.0 (Smelled) McCormick or Shilling Ground Cumin = 7.0 (Tasted)
Floral	Sweet, light, slightly perfume impression associated with flowers	0.12 oz Clorox Wipe Liquid in 4 oz Water= 8.0 (A) Geraniol = 7.5 (A) 1:1 White Grape Juice to Water = 5.0 (F&A)
Beet	A dark damp-musty-earthy note associated with canned red beets.	Food club sliced beets = 6.0 (Tasted) Food club sliced beets = 4.0 (Tasted) 1 part juice to 2 parts water served ½ oz in 1 oz cups.
Petroleum-like	A specific chemical aromatic associated with crude oil and its refined products that have heavy oil characteristics.	Vaseline petroleum jelly = 3.0 (Smelled) 1/2 teaspoon of Vaseline in 1 oz cups.

*major attributes that should be present at some level in all samples.

Table 2. The Pork Flavor Lexicon adapted from Chu (2015).

ATTRIBUTE	DEFINITION	REFERENCE
Astringent	The chemical feeling factor on the tongue or other skin surfaces of the oral cavity described as a puckering/dry and associated with tannins or alum	Lipton Tea, 1 bag = 6.0 (F) Lipton Tea, 3 bags = 12.0 (F)
Boar taint	Aromatic associated with boar taint; hormone-like; sweat, animal urine	0.1g 3-methylindole = 13.0 (A) Androstenone = 15.0 (A)
Bitter	The fundamental taste factor associated with a caffeine solution	0.05% caffeine in water = 2.0 (F) 0.08% caffeine in water = 5.0 (F)
Bloody/ serumy	An aromatic associated with blood on cooked meat products; closely related to metallic aromatic	Boneless Pork Chop, 135°F = 2.0 (F& aroma)
Brown/ roasted	A round, full aromatic generally associated with pork suet that has been broiled	Pork Fat, cooked = 3.0 (F), 4.0 (A)
Burnt	The sharp/acrid flavor note associated with over roasted pork muscle, something over baked or excessively browned in oils	Arrowhead Puffed Barley Cereal® = 5.0 (A & F)
Cardboardy	Aromatic associated with slightly oxidized fats and oils, reminiscent of wet cardboard packaging	Dry cardboard = 5.0 (F), 3.0 (A) Wet cardboard = 7.0 (F), 6.0 (A)
Chemical	Aromatic associated with garden hose, hot Teflon pan, plastic packaging and petroleum-based products such as charcoal lighter fluide	1 drop Clorox in 200 mL water = 6.5 (F) Ziploc Bag = 13.0 (aroma)
Fat-like	Aromatics associated with cooked animal fat	Pork fat, cooked = 10.0(F); 7.0(A)
Floral	Sweet, light, slightly perfume impression associated with flowers	0.12 oz Clorox Wipe Liquid in 4 oz Water= 8.0 (A) Geraniol = 7.5 (A) 1:1 White Grape Juice to Water = 5.0 (F & A)
Heated oil	The aromatics associated with oil heated to a high temperature	Wesson Oil, microwaved 3 min = 7.0 (F& A) Lay's Potato Chips = 4.0 (A)
Liver-like	Aromatics associated with cooked organ meat/liver	Pork Liver, cooked = 15.0 (F); 12.0(A)
Metallic	The impression of slightly oxidized metal, such as iron, copper, and silver spoons	Dole Pineapple Juice = 6.0 (A & F) 0.10% KCl solution= 1.5 (A & F)
Nutty	Nutty characteristics are: sweet, oily, light brown, slightly musty and/or buttery, earthy, woody, astringent, bitter, etc.	Diamond Shelled Walnut, ground for 1 min= 6.5 (F)
Pork identity	Amount of pork flavor identity in the sample	Boneless Pork Chop, 175°F = 7.0 (F), 5.0 (A) 80/20 Ground Pork, cooked = 6.0 (F); 5.0 (A)
Refrigerator stale	Aromatics associated with products left in the refrigerator for an extended period time and absorbing a combination of odors (lack of freshness/flat)	80/20 Ground Pork, cooked, left chilled overnight = 6.0 (F), 8.0 (A)
Salty	The fundamental taste factor of which sodium chloride is typical	0.2% Salt in Water = 2.5 (F) 0.35% Salt in Water = 5.0 (F)

Table 2. The Pork Flavor Lexicon adapted from Chu (2015). *(Continued)*

ATTRIBUTE	DEFINITION	REFERENCE
Soapy	An aromatic commonly found in unscented hand soap	0.12 oz Clorox Wipe Liquid in 4 oz Water= 3.0 (A) 0.5g Ivory Bar Soap in 100mL water = 6.5 (A)
Sour	The fundamental taste factor associated with citric acid solution	0.05% citric acid in Water = 2.0 (F) 0.08% Citric Acid in Water = 5.0 (F)
Spoiled/putrid	The presence of inappropriate aromatics and flavors that is commonly associated products. It is a foul taste and/or smell that indicates product is starting to decay and putrefy.	Boneless Pork Chop, 175°F, left out for 24 hours then refrigerate for 6 days = 3.0 (A) 80/20 Ground Pork, cooked, same as above = 5.0 (A)
Sweet	The fundamental taste factor associated with a sucrose solution	0.05% Sugar in Water = 2.0 (F) 0.08% Sugar in Water = 5.0 (F)
Umami	Flat, salty, somewhat brothy. The taste of glutamate, salts of amino acids and other molecules called nucleotides.	0.035% Accent flavor = 7.5 (F)
Vinegary	Aroma notes associated with vinegar	1.1g Vinegar in 200g water = 6.0 (F); 4.0 (A)
Warmed-over	Perception of a product that has been previously cooked and reheated	80/20 Ground Pork, cooked, left chilled overnight and reheated = 5.0 (A & F)

6. Meat descriptive attribute evaluation

In the first Guidelines in 1978, a unified method of evaluating meat palatability attributes was introduced. This method concentrated on juiciness, connective tissue amount, muscle fiber tenderness, and overall tenderness as the major attributes of meat palatability. The two previous versions of the Guidelines also included overall flavor intensity. This attribute, however, cannot be uniformly referenced and scaled and is not repeatable across panels. When possible, it should be replaced with the species-specific flavor lexicons where each flavor attribute can be referenced and scaled. If it is not practical to train for the entire flavor lexicon, or if it is not necessary based on the experiment objectives, one or two flavor notes, such as beef flavor identity, could be used (Meilgaard et al., 2007; Munoz, Civille, and Carr, 1992). Selection of attributes to include in the lexicon should be based on the study objectives. This method uses an 8- or 9-point, verbal anchored scale (Figure 10), or it can be modified to a 15- or 16-point scale to more easily incorporate flavor attributes (Figure 11). Regardless of scale, the method requires training on each attribute, panel performance validation, and use of correct cooking and product sampling techniques.

7. Magnitude estimation

Magnitude estimation involves the assignment of numbers to indicate intensity in relation to the first sample or reference sample. All subsequent samples are evaluated in proportion to the first sample rating.

SAMPLE ID #:	warm-up				
AROMATICS Major Notes:					
Beef Flavor Identity					
Brown/Roasted					
Bloody/Serumy					
Fat-like					
Metallic					
Cardboardy					
Painty					
Fishy					
Liver-like					
Putrid, Sulphur-like					
Overall Sweet					
Sour Milk/Sour Dairy					
Aftertaste					
Other					
BASIC TASTES:					
Sweet					
Sour					
Salty					
Bitter					
Umami					
Meat Descriptive Attributes					
Juiciness					
Muscle Fiber Tenderness					
Connective Tissue Amount					
Overall Tenderness					

0 * 1 * 2 * 3 * 4 * 5 * 6 * 7 * 8

Slight | Moderate | Strong

Other Flavor Attributes:

Asparagus	Apricot	Barnyard	Beet	Buttery	Burnt
Chemical	Chocolate/Cocoa	Cooked Milk	Cumin	Dairy	Floral
Green-haylike	Heated Oil	Refrigerator Stale	Rancid	Warmed-over	

Juiciness	**Muscle Fiber Tenderness and Overall Tenderness**	**Connective Tissue Amount**	**Flavor Intensity**
8 Extremely Juicy	8 Extremely Tender	8 None	8 Extremely Intense
7 Very Juicy	7 Very Tender	7 Practically None	7 Very Intense
6 Moderately Juicy	6 Moderately Tender	6 Traces	6 Moderately Intense
5 Slightly Juicy	5 Slightly Tender	5 Slight	5 Slightly Intense
4 Slightly Dry	4 Slightly Tough	4 Moderate	4 Slightly Bland
3 Moderately Dry	3 Moderately Tough	3 Slightly Abundant	3 Moderately Bland
2 Very Dry	2 Very Tough	2 Moderately Abundant	2 Very Bland
1 Extremely Dry	1 Extremely Tough	1 Abundant	1 None

Figure 10. Example ballot for beef descriptive flavor and texture attributes using 8- and 9-point scales (Miller, 2013).

SAMPLE ID #:	warm-up				
AROMATICS Major Notes:					
Beef Flavor Identity					
Brown/Roasted					
Bloody/Serumy					
Fat-like					
Metallic					
Cardboardy					
Painty					
Fishy					
Liver-like					
Putrid, Sulphur-like					
Overall Sweet					
Sour Milk/Sour Dairy					
Aftertaste					
Other					
BASIC TASTES:					
Sweet					
Sour					
Salty					
Bitter					
Umami					
Meat Descriptive Attributes					
Juiciness					
Muscle Fiber Tenderness					
Connective Tissue Amount					
Overall Tenderness					

0 * 1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10 * 11 * 12 * 13 * 14 * 15

Slight | Moderate | Strong

Other Flavor Attributes:

Asparagus	Apricot	Barnyard	Beet	Buttery	Burnt
Chemical	Chocolate/Cocoa	Cooked Milk	Cumin	Dairy	Floral
Green-haylike	Heated Oil	Refrigerator Stale	Rancid	Warmed-over	

Juiciness	**Muscle Fiber Tenderness and Overall Tenderness**	**Connective Tissue Amount**
15 Extremely Juicy	15 Extremely Tender	15 None
13 Very Juicy	13 Very Tender	13 Practically None
11 Moderately Juicy	11 Moderately Tender	11 Traces
9 Slightly Juicy	9 Slightly Tender	9 Slight
7 Slightly Dry	7 Slightly Tough	7 Moderate
5 Moderately Dry	5 Moderately Tough	5 Slightly Abundant
3 Very Dry	3 Very Tough	3 Moderately Abundant
1 Extremely Dry	1 Extremely Tough	1 Abundant

Figure 11. Example ballot for beef descriptive flavor and texture attributes using 15- and 16-point scales (Miller, 2013).

C. Training Sensory Panelists for Discriminative or Descriptive Testing

1. Scope

Before initiating training of panelists, the researcher must determine which sensory test method is most appropriate to fulfill the objectives of the study. Various test methods are available and it is not the intent of these Guidelines to fully describe these procedures. It is suggested that the various ASTM and IFT publications and Meilgaard et al. (2007) be reviewed to determine the most appropriate procedure. A short discussion of the most common methods used was presented above.

2. Selection of potential panelists

a. Recruitment

Trained panelists can be recruited from the surrounding community (external panelists) or they can be company employees (internal panelists). There are advantages and disadvantages to each type of panelists.

The advantage of using external trained sensory panelists is that they are independent and do not have institutional or product knowledge, or product loyalty. Therefore, product-specific attributes that could induce expectation errors with internal panelists are easily avoided. Additionally, a large selection pool is available when using external panelists. Because external panelists have only one obligation within the organization, they may have fewer daily conflicts. The main disadvantage of using external panelists is that they are not on site. Parking and ease of travel to the testing facility have to be considered.

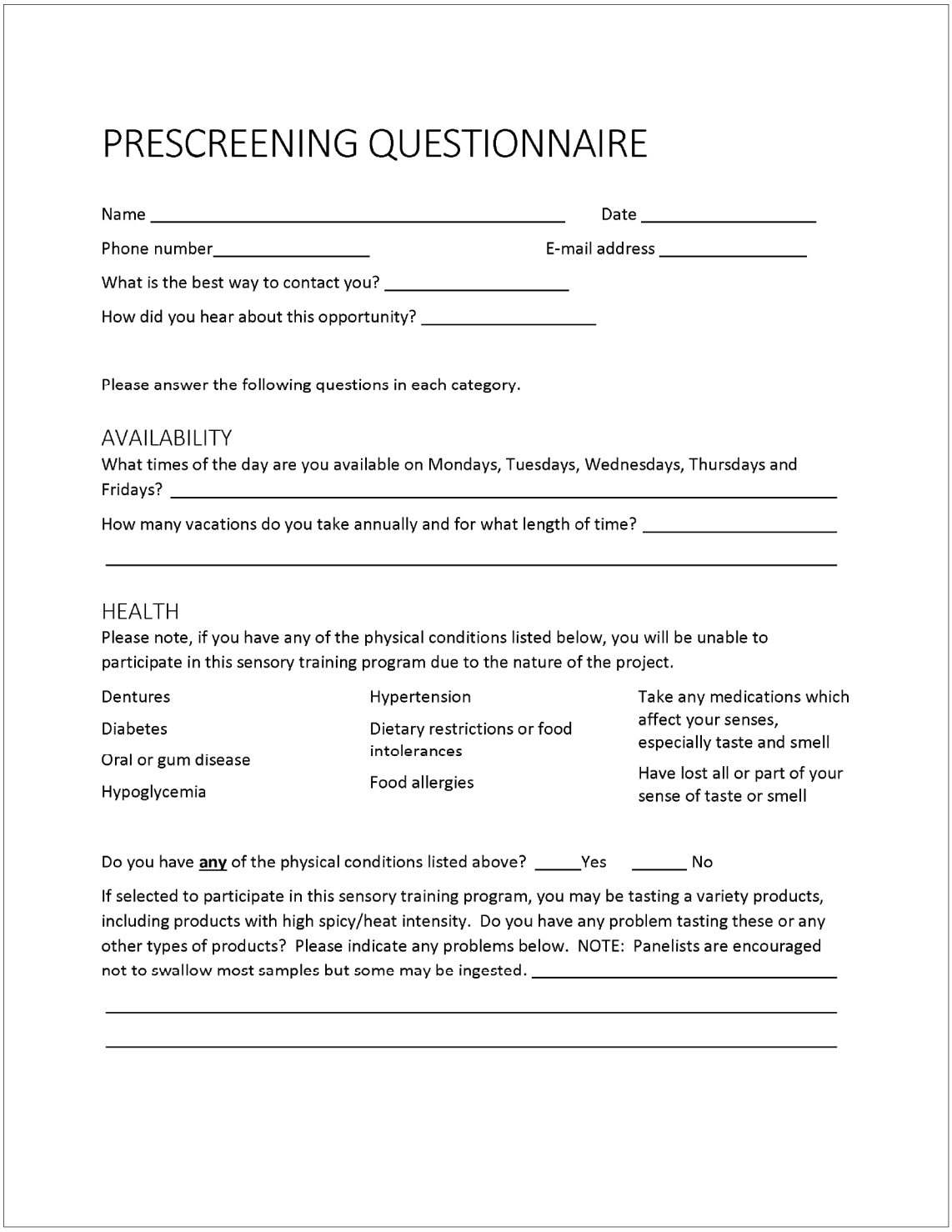
Internal panelists are easily recruited within a company or research entity, and the prospect of excused time from daily responsibilities can be a motivating factor. The time requirement, however, needs to be communicated effectively to potential panelists and supervisors and supported by upper management. The selection pool for qualified candidates might be limited, and the tendency to accept marginal panelists to increase the number of panelists should be avoided. A key advantage of using internal panelists, such as Research & Development and Quality Assurance/Quality Control personnel and/or graduate students, is that their scientific knowledge and understanding makes them easier to train and they tend to have the ability to be able to make more concise judgments. Furthermore, this training in sensory evaluation enhances their ability to detect and describe flavor and/or texture attributes, which makes them better future Research & Development employees.

Recruitment of internal and external candidates may occur through advertisements, posted announcements and formal and informal verbal methods. When written recruitment methods are used, frequently there are word limitations, and thus it is critical that the important factors are clearly identified. For external candidates, the advertisement should plainly indicate the time commitment required and whether or not it is a paid position. The wording should focus on seeking those who have an interest in food, while clearly stating the importance of the panel's effort, especially regarding the need for training.

b. Prescreening candidates

Potential candidates are first prescreened to determine their level of interest, availability, dependability, health (including dentures, allergies, use of medication), work experience, gender, age, smoking/tobacco use status, and food likes/dislikes (Figure 12). Background information on prospective panelists is valuable in selecting those individuals who have the greatest potential to become effective panelists.

During the prescreening process, a candidate's ability to follow directions or make concise judgments also should be determined because panelists who are not able to do these tasks will not be successful panelists. A panelist's ability to make judgments and follow directions is best determined through the use of logic tests discussed below. Examples of these tests are provided in Meilgaard et al. (2007) and in Figure 13. A panel leader will have to continually monitor panelists who struggle to follow directions or make judgments and this might result in frustration in other panelists. The end result can be a decrease in motivation and positive attitude of acceptable panelists. When grading, a 70% or above is considered acceptable on a logic test.



PRESCREENING QUESTIONNAIRE

Name _____ Date _____

Phone number_____ E-mail address _____

What is the best way to contact you? _____

How did you hear about this opportunity? _____

Please answer the following questions in each category.

AVAILABILITY

What times of the day are you available on Mondays, Tuesdays, Wednesdays, Thursdays and Fridays? _____

How many vacations do you take annually and for what length of time? _____

_____

HEALTH

Please note, if you have any of the physical conditions listed below, you will be unable to participate in this sensory training program due to the nature of the project.

Dentures	Hypertension	Take any medications which affect your senses, especially taste and smell
Diabetes	Dietary restrictions or food intolerances	
Oral or gum disease	Food allergies	Have lost all or part of your sense of taste or smell
Hypoglycemia		

Do you have **any** of the physical conditions listed above? _____Yes _____ No

If selected to participate in this sensory training program, you may be tasting a variety products, including products with high spicy/heat intensity. Do you have any problem tasting these or any other types of products? Please indicate any problems below. NOTE: Panelists are encouraged not to swallow most samples but some may be ingested. _____

_____

_____

Figure 12. Example of prescreening questionnaire for selection of panelists for a meat texture and flavor descriptive attribute sensory panel (modified from Meilgaard et al., 2007).

PRESCREENING QUESTIONNAIRE Part 2

FOOD HABITS

Are you currently on a restricted diet? If yes, please explain. _____

_____

_____

How often do you eat fast foods out in a month? _____

What is (are) your favorite foods? _____

_____

_____

What is (are) your least favorite foods? _____

_____

_____

What foods do you ***not like*** to eat? _____

_____

_____

Is your ability to distinguish smell, taste and textural characteristics…

	SMELL	TASTE	TEXTURE
Better than average	☐	☐	☐
Average	☐	☐	☐
Worse than average	☐	☐	☐

Think about what you had for dinner last night. Please describe the meal, including how the food and beverage tasted, and what you liked and disliked about them in as much detail as possible.

_____

_____

_____

_____

2

Figure 12. (continued).

PRESCREENING QUESTIONNAIRE Part 3

FLAVOR AND TEXTURE QUIZ

If a recipe calls for thyme and none is available, what would you substitute?

_____

What are some foods that taste like yogurt? _____

_____

How would you describe the difference between flavor and aroma? _____

_____

How would you describe the difference between flavor and texture? _____

_____

What is the best one- or two-word description of grated Italian cheese (Parmesan or Romano)?

_____

Describe some of the noticeable flavors in pork chops. _____

_____

Describe some of the noticeable flavors in hot dogs. _____

_____

Describe some of the textural properties of foods in general. _____

_____

For what type of products is texture important? _____

_____

3

Figure 12. (continued)

c. Screening

Candidates who have passed the prescreening stage are then invited to participate in a series of screening exercises. According to ASTM STP758 (1981), the purpose of screening is to select candidates that have the following qualities:

- Normal sensory acuity
- Interest in sensory evaluation
- Ability to discriminate and reproduce results
- Appropriate behavior, such as cooperation, motivation, and promptness

It is often necessary to screen as many as four times the number of panelists that will be needed in the actual panel. Sensory protocols and procedures during screening should be similar to those used later in actual studies. It is important to use the same meat products for screening that will be evaluated by the panel after training. A higher level of discrimination is required for descriptive analysis panelists than for discriminative tests; therefore, the screening procedures should be more rigorous for descriptive analysis panelists and will require that more panelists be screened in order to achieve the final number of panelists needed. Screening panelists can include more than one of the strategies discussed below.

(1) LOGIC TESTS

Logic tests are conducted to determine a candidate's ability to follow directions and make decisions. Figure 13 provides an example of logic tests; additional logic tests can be found in Meilgaard et al. (2007). Candidates should mark the line in the approximate area for a correct response. Candidates who provide incorrect answers by marking the line in the opposite direction as requested should be immediately eliminated. These candidates will not follow instructions and the panel leader will have to continually work with them to assure their understanding.

Scaling Tests

Directions: Mark on the line following each figure the proportion of the area that is shaded.

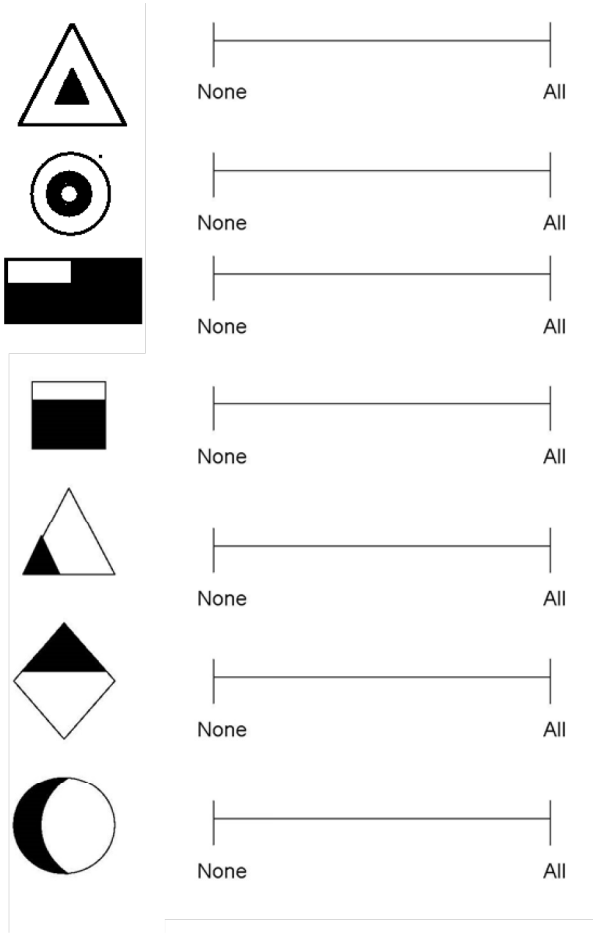


Figure 13. Examples of logic tests used for screening potential sensory panelists.

(2) MATCHING TESTS

These tests are designed to determine a potential panelist's ability to describe or identify descriptive attributes. Meat products contain multiple flavors and textures, especially further-processed products. Even if the original objective of the trained panel is to evaluate whole-muscle beef steaks, panelists need to have the ability to describe, identify, and rate flavor attributes. If the panelists have the ability to describe texture attributes of a further-processed meat product, such as a frankfurter or sausage product, the panel can be expanded or can be further trained to evaluate these products. Meilgaard et al. (2007) describe these tests in detail. It is useful to present matching tests for taste and aroma in order to better assess a candidate's ability to discriminate.

(A) MATCHING TESTS—TASTE

In an example of a matching test for taste provided in Meilgaard et al. (2007), the candidate is served a set of cups labeled sweet, sour, salty, bitter, and water, as well as a set of five cups

coded with three-digit random numbers. The labeled cups contain the appropriate stimulus at just more than threshold levels, and the coded cups contain a corresponding stimulus. Examples of these are listed in Table 3. The candidate is instructed to taste the labeled cups first to familiarize themselves with the basic taste. They then are asked to taste the coded samples and indicate on a score sheet the matching code number for each stimulus.

Table 3. Suggested samples for matching taste test. Adapted from Meilgaard et al. (2007).

BASIC TASTE	STIMULUS	CONCENTRATION
Sweet	Sucrose	2.0%
Salty	Sodium chloride	0.2%
Sour	Citric acid	0.05%
Bitter	Caffeine	0.05%
Water	Filtered water	---

(B) MATCHING TESTS—AROMA

Sniff tests are conducted for matching tests on aroma. To conduct sniff tests, cut 0.75 cm-wide strips of Whatman filter paper at least 2.54 cm long. Cotton balls can be used as an alternative to Whatman filter paper. Obtain essential oils representing a variety of aromas from a business that uses essential oils to formulate food or personal care products (Table 4). Dip the filter paper strip in the oil sufficiently to concentrate the oil aromatic. Allow the strip to dry under a hood for 30 minutes, then place it in a glass jar with a lid (i.e., a baby food jar) or in a test tube with a lid. Label the containers with random three-digit codes. Ask the candidate to remove the lid and sniff the contents of the container without touching the contents. Candidates then are asked to match the aroma they detect with the appropriate aroma descriptor selected from a list at the bottom of the score sheet; list two to four more descriptors than the number of stimuli presented. Accept candidates that correctly identify approximately 70% of the attributes presented.

Table 4. Suggested samples for sniff tests. Adapted from Meilgaard et al. (2007).

AROMA DESCRIPTORS	STIMULUS
Anise, licorice	Anise oil
Almond, cherry	Amaretto, benzaldehyde, oil of bitter almond
Cinnamon	Cinnamaldehyde, cassia oil
Clove, dentist's office	Eugenol, oil of clove
Ginger	Ginger oil
Green, freshly mown lawn	*cis*-3-Hexenol
Lemon, lime or orange	Lemon, lime, or orange oil
Peppermint, minty	Peppermint oil
Vanilla	Vanilla extract
Wintergreen	Methyl salicylate, oil of wintergreen

(3) RANKING TESTS

Ranking tests are used to determine a candidate's ability to discriminate graded levels of intensity of a given attribute. As described in Meilgaard et al. (2007), candidates are presented with a series of samples in random order. The samples are coded with three-digit random numbers and cover a range of a specific attribute. Ask the candidates to rank the samples in order of increasing intensity of the stated attribute. Examples of sample sets are listed in Table 5.

Table 5. Suggested sample sets for ranking tests.

ATTRIBUTE	SENSORY STIMULI	CONCENTRATION			
Taste					
Sweet	Sucrose / water, g/L	10	20	50	100
Salty	Sodium chloride, g/L	1.0	2.0	5.0	10.0
Texture					
Hardness	Cream cheese,[b] American cheese,[b] peanuts, carrot[c]				
Juiciness	Banana,[c] carrot,[c] mushroom,[c] apple[c]				

[a]Attributes adapted from Meilgaard et al. (2007).

[b]½-inch cubes

[c]½-inch slices

(4) IDENTIFICATION TESTS

Sniff tests also can be used as identification tests to understand a panelist's ability to recognize, describe, or identify descriptors. These tests are used mainly in selection of descriptive panelists. Sniff tests as described above can be used. Jars with varying flavor aromatics are presented individually to panelists using three-digit random codes. The candidate is asked to describe the attribute that they detect on a sheet of paper. When evaluating their score sheet, a correct answer is when the candidate uses descriptors similar to the attribute. For example, if cedar oil is used in the sniff test, correct answers could be cedar, wood, a forest, sweaters, stored sweaters, or wooden chest.

(5) DISCRIMINATIVE TESTS

Triangle tests are recommended, but note that Duo-Trio tests also can be used to determine a panelist's ability to discriminate differences in sensory attributes. For discriminative panel selection, fewer tests need to be run (from six to eight); for descriptive panelists, a larger number of tests should be conducted to provide a greater range of attributes.

A sequential analysis procedure is used to minimize the number of tests needed for screening (Bradley, 1953). That procedure facilitates an early decision on very good or very poor candidates.

The sequential procedure makes one of the following decisions after each triangle test:

- Accept the candidate as a potential panelist
- Reject candidate
- Continue testing

The decisions are based on specifications of four parameters:

- P_0 = Maximum proportion of correct decisions ruled as an unacceptable candidate
- P_1 = Minimum proportion of correct decisions ruled as an acceptable candidate
- a = Probability of selecting an unacceptable candidate
- b = Probability of rejecting an acceptable candidate

By plotting test numbers against the accumulated number of correct test results, a decision is made on the basis of the region in which the point is plotted (Figure 14). The regional boundaries are described in greater detail by Cross, Moen, and Stanfield (1978).

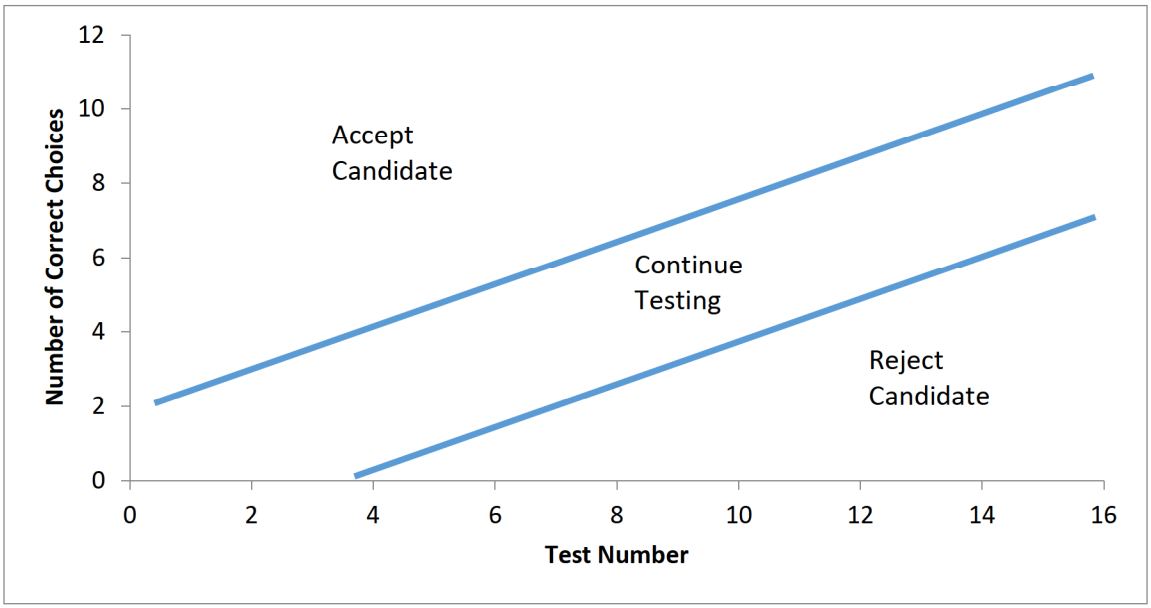


Figure 14. Sequential analysis chart used for screening potential panelists. Adapted from Cross et al. (1978).

The following is an example of the screening procedure. Test samples for Triangle tests are prepared to give a two-unit difference (eight-point scale, Table 6) in the attribute being tested. In this example, the attributes are tenderness, juiciness, and connective tissue amount. More than five triangles are recommended. These are five examples of how to create differences. Other examples can be created. Differences should be of a magnitude that could easily be detected by an experienced panel leader. The values $P_0 = 0.45$, $P_1 = 0.70$, a = 0.10, and b = 0.10 are used. Figure 14 then is applied to accept, continue testing, or accept panelists.

Table 6. Examples of Triangle tests that can be used and the sensory attribute of interest.

Triangle Tests	Explanation of attribute differences targeted for testing
1.	Strip loin steak cooked to 70°C and a strip loin steak cooked to 75°C with juices pressed out to get differences in juiciness
2.	Eye of round steak cooked to 70°C and a strip loin steak cooked to 70°C to get differences in tenderness
3.	Top sirloin steak and a strip loin steak both cooked to 70°C to get differences in flavor intensity (or ST muscle "leached" in water for a few hours before cooking)
4.	Ground beef patty and a ground beef patty containing 1.0% added ground liver evenly mixed and distributed to get differences in liver flavor
5.	Brisket or bottom round steak grilled and a strip loin steak grilled to get differences in connective tissue

At the end of the screening period, the candidates in the "accept" region can be selected for training. If time is a factor, screening can be stopped after 15 sessions. It is desirable to select for training only those in the "accept" region, but if most of the candidates are in the "continue testing" region, they also could be selected. Another option would be to recruit more candidates and reinitiate the screening process. In panel selection, screening should not be considered a part of training, but rather a test to quickly eliminate those individuals who cannot detect large differences in attributes. At least twice as many individuals should be screened as are needed on the final panel.

Meilgaard et al. (2007) suggest, when using Triangle tests, to reject candidates scoring less than 60% correct on the easy tests or less than 40% on the moderately difficult tests. When using Duo-Trio tests, reject candidates scoring less than 75% on the easy tests or less than 60% on the moderately difficult tests.

After completion of the three stages of testing for potential panelists, accept panelists for training that pass all stages of candidate screening. During candidate testing, observe panelist behavior, timeliness, ability to interact with others, confidence level, and dependability. Do not

accept a panelist who misses an appointment, is difficult to work with, or is disruptive or domineering in a group setting. Panelists who are not dependable during a "job interview" will not be dependable after they have the job. If the panel leader determines that an individual is disruptive, the panelist's behavior pattern will be difficult to alter and might take more effort than desired to change. Final panelists selected for training should be acceptable on all criteria discussed above.

3. Training

The objectives of training are the following:

- Familiarize an individual with test procedures
- Improve an individual's ability to recognize and identify sensory attributes
- Improve an individual's sensitivity and memory, permitting precise and consistent sensory judgments

Panelists should understand the importance of the study to insure their cooperation and motivation. Let them know that you are pleased to have them participate and that their cooperation is appreciated. Without influencing the panelists' future responses, give them as much specific information as possible on the purpose of the test. The importance of concentration should be stressed.

Panelists need to learn to be objective early in training. While all panelists are consumers, their opinions or preferences should not be expressed in their evaluations; nor should they influence others through discussion. A number of decisions have to be reached early in training regarding protocols. The amount of sample that a judge places in his/her mouth must always be standardized. The decision of whether or not the panelist should swallow a sample should be standardized. To further standardize the evaluation methods, palate cleansing procedures should be standardized. After sample evaluation, taste bud refreshers and mouth rinsing should follow to minimize taste bud fatigue. Each panel member should rinse their mouth between samples. Room temperature water—bottled, filtered, or distilled—is the most common rinse. When there is a great deal of aftertaste, taste bud refreshers such as unsalted crackers, seltzer water, apple slices (as long as a water rinse is used afterward to avoid flavor carryover), or ricotta cheese are useful. Taste bud refreshers, however, should be used with caution. Even though they can eliminate lingering aromatics, mouthfeels, and aftertaste after evaluation of a product, taste bud refreshers can contribute to taste bud fatigue.

The interval between samples should be standardized and is dependent upon the product under study. Enough time should be allowed between samples to permit recovery from flavor buildup, yet not so much time that the taster loses his/her ability to discriminate.

Depending on the descriptive method used and level of experience with the product, the panel leader can provide panelists with predetermined lexicons (Table 1) and with definitions and procedures for use in evaluation. Examples of training sessions for meat descriptive palatability attributes are listed in Table 7 and for beef flavor in Table 8. In some instances, the panelists will develop their own descriptors and scaling techniques, such as with QDA. Training is best accomplished through individual and group sessions in which various samples of the product types usually involved in the tests are evaluated and discussed. Multiple sessions should be

devoted to demonstrating levels of each attribute under study as shown in Tables 7 and 8. All attributes defined in Tables 7 and 8 may not need to be included in a study. The sensory professionals should determine the attributes to include and train for based on the objectives and treatments in the study.

With ground beef patties, the following processes can provide excellent variations in sensory properties for training. Potential sensory training sessions for ground beef would follow similar logic as presented in Tables 7 and 8.

- Hot processing; variation in grind size
- Fast vs. slow freezing
- Use of gels, gums, TVP
- Different fat levels
- Cookery method (charbroiling vs. microwave)

During the early stages of training, the panel leader should try to identify the extremes and middle of the rating scale. During training, it is necessary to refer to some standards, such as the *psoas major* muscle for extremely tender and hot-boned *longissimus* or *semitendinosus* from an old cow carcass chilled in ice water for extremely tough samples. As training progresses, the panelists should be able to identify other points along the rating scale. Once panelists have begun to scale for individual attributes, training sessions that include the evaluation of a combination of two attributes should be conducted. Successive additions of other attributes should occur as panelists illustrate knowledge and confidence in their abilities as the complexity of the evaluation increases.

The panel leader will need to provide very specific instructions to panelists regarding procedures to follow in measuring the various sensory attributes during chewing. With the complexity of added ingredients now being used in such products as ground beef, an overall tenderness score might not suffice. Use of gels and gums might create softness in the product, but also a tacky or cohesiveness property. For tenderness, stating which teeth should be used as well as the number of chews should be standardized. Positioning of the sample on the teeth, such as orientation of the muscle fibers, should be addressed. Molars can be used to detect properties of firmness, compression, and rate of breakdown, while incisors can provide information relative to shear. Incisors can prove helpful in detecting connective tissue, and they can detect the rubberiness or elasticity of a small piece of connective tissue. The tongue also can be helpful in detecting connective tissue by pushing it through the chewed mass. Panelists must be instructed to perform considerable sample breakdown, often employing a high number of chews (ready to swallow) before assigning a score for the volume of residual connective tissue present. The presence of surface crust on cooked samples might present unique problems relative to sample breakdown during chewing and uniformity of chewed pieces. In studies where designed differences exist in fat content, juiciness should be evaluated early (initial) and later (sustained) in the chewing process.

Table 7. Examples of training session for meat descriptive palatability attributes.

Day 1

Goals:

- To introduce the basic meat descriptive attribute method to the panel
- To familiarize the panelists with the ballot
- To provide some initial descriptor for juiciness and muscle fiber (MF) tenderness

Meat samples and order of presentation:

Juiciness

1. Hand out the AMSA guidelines and a ballot.
   Explain what meat descriptive attribute sensory evaluation is.
   Concentrate on understanding what juiciness is: juicy versus dry.
2. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
   Use this sample to ask panelists to just give an initial evaluation for juiciness with panel leader scoring as well.
3. **Strip loin steak**—setting lower scale for juiciness; cooked to 75°C and pressed
4. **Eye of round steak**—cooked to 90°C; scaling between 1 & 2 for juiciness
5. **Strip loin steak**—standard or baseline steak; cooked to 70°C

Muscle Fiber Tenderness

1. Concentrate on understanding muscle fiber tenderness: Tough versus tender
2. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
   Use this sample to anchor on 5 or 6 of scale.
3. **Tenderloin steaks**—for a very tender sample; cooked to 65°C
   Use this sample for a 7 or 8 on MF tenderness.
4. **Old cow steak**—for a tough sample; cooked to 70°C
   Use this to give a 2 or 3 for MF tenderness.

Day 2

Goals:

- To develop a base line for juiciness and tenderness
- To begin scaling for each of the two attributes

Meat samples and order of presentation:

Juiciness

1. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
   Use this sample to ask panelists to just give an initial evaluation for juiciness with panel leader scoring as anchor.
2. **Strip loin stea**k—cooked to 75°C and pressed; setting lower scale for juiciness
3. **Eye of round steak**—cooked to 85°C; scaling between 1 and 2 for juiciness
4. **Strip loin steak**—standard or baseline steak; cooked to 70°C
5. **Strip loin steak**—cooked to 60°C; scaling for higher juiciness

Table 7. Examples of training session for meat descriptive palatability attributes. (continued)

Muscle Fiber Tenderness: Tough versus Tender; also rate for juiciness to start combining

1. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
   Use this sample to anchor on 5 or 6 of scale.
2. **Tenderloin steaks**—for a very tender sample; cooked to 65°C
   Use this sample for a 7 or 8 on MF tenderness.
3. **Old cow steak**—for a tough sample; cooked to 70°C
   Use this to give a 2 or 3 for MF.
4. **Eye of round steak**—broiled to 85°C for lower part of scale
5. **Strip loin steak**—standard or warm-up sample; cooked to 70°C

Day 3

Goals:

- To develop a baseline for juiciness and tenderness
- To begin scaling for each of the two attributes
- To begin differentiating between muscle fiber tenderness and connective tissue amount
- To begin scaling for both attributes of tenderness

Meat samples and order of presentation:

1. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
   Use this sample to ask panelists to just give an initial evaluation for juiciness and muscle fiber tenderness.
2. **Strip loin steak**—cooked to 75°C and pressed; setting lower scale for juiciness and toughness
3. **Eye of round steak**—cooked to 85°C; scaling between 1 and 2 for juiciness and toughness
4. **Tenderloin steaks**—cooked to 65°C; for a very tender sample
5. **Old cow steak**—for a tough sample; cooked to 70°C
6. **Strip loin steak**—cooked to 60°C; for higher juiciness

Introduce the concept of CT: Need to separate tough versus tender for muscle fiber tenderness and amount of connective tissue

7. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
8. **Tenderloin steaks**—for a very tender sample; cooked to 65°C
9. **Bottom round steak—**to show CT
10. **Old cow steak**—for a tough sample; cooked to 70°C
11. **Broiled brisket steak**—broiled to 75°C
12. **Strip loin steak**—standard or warm-up sample; cooked to 70°C

Day 4

Goals:

- To begin differentiating between muscle fiber tenderness and connective tissue amount
- To begin scaling for both attributes of tenderness

Table 7. Examples of training session for meat descriptive palatability attributes. (continued)

Meat samples and order of presentation:

1. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
2. **Cold shortened steak**—setting lower scale for tenderness; cooked to 70°C
   Describe muscle fiber tenderness.
   Describe connective tissue amount.
3. **Top butt steak**—scaling for tenderness; cooked to 70°C
4. **Tenderloin steak**—scaling for tenderness; cooked to 70°C
5. **Break**
6. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
7. **Brisket steak**—scaling for tenderness/connective tissue; cooked to 70°C
8. **Cold shortened steak**—scaling for tenderness; cooked to 70°C
9. **Strip loin steak**—scaling for tenderness and baseline determinations; cooked to 70°C

Day 5

Goals:

- To fine tune evaluation for juiciness, muscle fiber tenderness, and connective tissue amount
- To scale for all attributes

Meat samples and order of presentation:

1. **Strip loin steak**—standard or warm-up sample; cooked to 70°C
2. **Tenderloin steak**—differences in flavor, connective tissue, and juiciness
3. **Top butt steak**—differences in flavor, connective tissue, and juiciness
4. **Cold shortened steak**—differences in flavor, connective tissue, tenderness, and juiciness
5. **Strip loin steak**—baseline evaluation
6. **Water soaked strip steak or eye of round**—differences in all attributes
7. **Strip steak**—cooked to 75°C and pressed

Day 6

Goal: To evaluate samples in the booths and begin independent evaluations

Meat samples and order of presentation:

Session number 1: Four steaks that vary in juiciness, muscle fiber tenderness, or connective tissue amount

Break

Session number 2: Four steaks that vary in juiciness, muscle fiber tenderness, or connective tissue amount

Continue with training by increasing number of steaks per session to six. Each day evaluate and review scores. Add exercises needed to increase understanding and repeatability of attributes. This example is for an experienced panel that is being refreshed. For a new panel, these exercises would be expanded and presented in more sessions depending on panelist response.

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011).

Session 1 - Introduce scaling
Present universal scale for flavor intensity (Meilgaard et al., 2007): • 2.0—soda flavor in saltless saltine cracker • 5.0—apple flavor in Mott's® Applesauce • 7.0—orange flavor in Minute Maid® Orange Juice • 10.0—grape flavor in Welch's® Grape Juice • 12.0—cinnamon flavor in Big Red chewing gum Introduce basic tastes and recognize intensity levels across attributes using solutions from Meilgaard et al. (2007) for salt, sweet, bitter, and sour. Taste nonmeat items for flavor intensity and basic tastes from Meilgaard et al. (2007). • Lay's® Classic Potato Chips: Rate potato flavor using universal scale—salt = 12.0; sweet = 4.5; sour = 1.5; and bitter = 2.0 • Minute Maid® Orange Juice Frozen Concentrate Reconstituted: Rate orange flavor using universal scale—sweet = 8.0; sour = 3.5; and bitter = 1.5. • Haagen-Dazs® Vanilla Ice Cream: Rate vanilla flavor using universal scale—sweet = 12.0; salt = 2.0; sour = 2.0; and bitter = 1.0.
Session 2 - Review universal scale and basic tastes, train on attribute 1—beef flavor and aroma identity (ID)
Repeat exercises from session 1. Provide definition and references for beef flavor and aroma ID. Panelists will evaluate individually and come to consensus: • High beef flavor/aroma ID—prime top loin steak cooked on grill to 70°C • Low beef flavor/aroma ID—standard top loin steak cooked on grill to 70°C

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011). (continued)

Session 3 - Review beef flavor/aroma ID and have references available; continue to anchor using universal scale and basic taste references; introduce warm-up sample concept; introduce brown/roasted flavor/aroma attribute
Have references for universal scale, basic tastes, and beef flavor/aroma ID available; panelists work through references at their own pace to anchor on attributes. Present warm-up sample—choice top loin steak, aged 14 days, grilled to 70°C Panelists evaluate beef flavor/aroma ID and basic tastes; discuss and come to consensus. Introduce definition and references for brown/roasted flavor and aroma. Present three samples and evaluate for all attributes introduced to this point: • Choice top loin steak cooked to 57°C for low brown/roasted flavor • Choice top loin steak cooked to 80°C for high brown/roasted flavor • Choice top loin steak cooked to 70°C—unknown level; panelists will determine
Session 4 – Retrain using previous exercises if panelists are having difficulty; introduce bloody/serumy and metallic flavor/aroma attributes.
Present any previous references as needed to anchor panelists *(Note: do this at the beginning of all sessions as needed; this will not be repeated for future sessions, but should be considered as an exercise for any session where panelists are showing inconsistencies in scaling for an attribute).* Present definition and references from lexicon for bloody/serumy flavor/aroma attribute. Present definition and references from lexicon for metallic flavor/aroma attribute. Present two samples for panelist evaluation of all attributes introduced to this point: • Select tenderloin steak grilled to 70°C—high bloody/serumy and metallic attributes • Select top sirloin steak grilled to 70°C—unknown; panelists determine and come to consensus
Session 5 – Introduce fat-like flavor/aroma and liver-like flavor/aroma
Present definition and references from lexicon for fat-like flavor/aroma attribute. Present definition and references from lexicon for liver-like flavor/aroma attribute. Present two samples for panelist evaluation of all attributes introduced to this point: • Prime top loin steak broiled to 70°C—should have strong fat-like flavor and low liver • Cow top loin steak broiled to 57°C—strong liver notes

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011). (continued)

Session 6 –Introduce green-haylike flavor/ aroma and umami flavor

Present definition and references from lexicon for green-haylike flavor/aroma attribute.

Present definition and references from lexicon for umami flavor attribute.

Present one sample for panelist evaluation of all attributes introduced to this point:

- Prime grass-fed steak broiled to 57°C—green and umami notes

Session 7 - Introduce overall sweet flavor, sweet aroma, and sour aroma

Present definition and references from lexicon for overall sweet flavor/aroma attribute.

Present definition and references from lexicon for sweet and sour aroma attributes.

Present two samples for panelist evaluation of all attributes introduced to this point:

- Choice top loin steak grilled to 70°C

Session 8 – Calibration day

Panelists will individually evaluate three muscles for the MAJOR NOTES.

Warm-up using a USDA Select strip steak grilled to 70°C – Panelists will come to consensus.

- Sample 1—knuckle roast; no consensus, individual evaluation
- Sample 2—top sirloin steak; no consensus, individual evaluation
- Sample 3—eye of round roast; no consensus, individual evaluation

Determine panel proficiency on MAJOR NOTES.

Session 9 - Overview major notes; introduce animal hair aroma and barnyard aroma/flavor

Present definition and references from lexicon for animal hair aroma attribute.

Present definition and references from lexicon for barnyard aroma/flavor attribute.

Present three samples for panelist evaluation of all attributes introduced to this point:

- Tenderloin from a bull broiled to 74°C for animal hair aroma and barnyard flavor/aroma
- Select top loin steak grilled to 70°C
- Choice knuckle roast roasted to 70°C

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011). (continued)

Session 10 – Review major notes, introduce burnt aroma/flavor and rancid aroma/flavor

Present definition and references from lexicon for burnt aroma/flavor attribute.

Present definition and references from lexicon for rancid aroma/flavor attribute.

Present four samples for panelist evaluation of all attributes introduced to this point:

- Standard top loin steak grilled to greater than 80°C or until visibly burnt
- Low Choice top loin steak stewed to 68°C
- Select inside round roast (70°C)
- Choice flat iron steak (70°C)

Session 11 – Review major notes; introduce heated oil aroma/flavor and chemical aroma/flavor

Present definition and references from lexicon for heated oil aroma attribute.

Present definition and references from lexicon for chemical aroma/flavor attribute.

Present three samples for panelist evaluation all attributes introduced to this point:

- Top Choice top sirloin broiled to 68°C—chemical aromas and flavors
- Choice eye of round roast (70°C)
- Choice tenderloin steak (70°C)

Session 12 – Review major notes; introduce leather (old) aroma and apricot flavor

Present definition and references from lexicon for leather aroma attribute.

Present definition and references from lexicon for apricot flavor attribute.

Present three samples for panelist evaluation of all attributes introduced to this point:

- Cow top round roast stewed to 74°C—leather aromas
- Select tenderloin steak (70°C)
- Choice bottom round roast (70°C)

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011). (continued)

Session 13 - Review major notes; introduce green aroma/flavor and asparagus flavor
Present definition and references from lexicon for green aroma/flavor attribute. Present definition and references from lexicon for asparagus flavor attribute. Present four samples for panelist evaluation of all attributes introduced to this point: • Cow top loin steak broiled to 57°C—green flavors • Cow top loin stewed to 63°C • Select knuckle roast (70°C) • Select flat iron steak (70°C)
Session 14 – Review major notes; introduce musty-earthy/humus aroma and cumin aroma/flavor
Present definition and references from lexicon for musty/earthy/humus attribute. Present definition and references from lexicon for cumin attribute. Present four samples for panelist evaluation of all attributes introduced to this point: • Cow tenderloin roasted to 63°C • Choice inside round roast (70°C) • Select top sirloin steak (70°C)
Session 15 – Review major notes; introduce floral aroma/flavor rand beet aroma/flavor
Present definition and references from lexicon for floral aroma/flavor attribute. Present definition and references from lexicon for beet aroma/flavor attribute. Present four samples for panelist evaluation of all attributes introduced to this point: • Cow tenderloin stewed to 80°C—beet flavors • Select bottom round roast (70°C) • Choice top loin steak (70°C)

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011). (continued)

Session 16 – Review major notes; introduce chocolate/cocoa aroma/flavor and medicinal aroma

Present definition and references from lexicon for chocolate/cocoa attribute.

Present definition and references from lexicon for medicinal attribute.

Present four samples for panelist evaluation of all attributes introduced to this point:

- Bull tenderloin grilled to 63°C—chocolate aroma
- Choice top sirloin steak (70°C)
- Select eye of round roast (70°C)

Session 17 – Review major notes; Introduce charcoal aroma/flavor, wood aroma, and spoiled-putrid aroma

Present definition and references from lexicon for charcoal aroma/flavor attribute.

Present definition and references from lexicon for wood aroma/flavor attribute.

Present definition and references from lexicon for spoiled-putrid aroma attribute.

Present four samples for panelist evaluation of all attributes introduced to this point:

- Smell spoiled standard tenderloin grilled to 68°C for spoiled-putrid aroma
- Prime top loin steak food service gas grilled to 68°C—smoky charcoal aroma
- Select top loin steak (70°C)
- Choice top loin steak (70°C)

Session 18 – Review major notes; introduce dairy aroma/flavor and buttery aroma/flavor

Present definition and references from lexicon for dairy aroma/flavor attribute.

Present definition and references from lexicon for buttery aroma/flavor attribute.

Present four samples for panelist evaluation of all attributes introduced to this point:

- Top Choice top sirloin roasted to 63°C—dairy aroma
- Bull tenderloin grilled to 63°C—buttery flavor
- Select eye of round roast (70°C)
- Choice eye of round roast (70°C)

Table 8. Examples of training sessions for a beef flavor descriptive attribute panel using the Beef Flavor Lexicon (Adhikari et al., 2011). (continued)

Session 19 – Review major notes; Introduce milk aroma/flavor and sour milk/sour dairy aroma/flavor

Present definition and references from lexicon for milk aroma/flavor attribute.

Present definition and references from lexicon for sour milk/sour dairy aroma/flavor.

Present three samples for panelist evaluation of all attributes introduced to this point:

- Milk-fed veal top loin steak—cooked milk
- Select top sirloin steak (70°C)
- Choice top sirloin steak (70°C)

Session 20 – Review major notes; introduce stale aroma/flavor, soapy aroma, and warmed-over aroma/flavor

Present definition and references from lexicon for stale aroma/flavor attribute.

Present definition and references from lexicon for soapy aroma attribute.

Present definition and references from lexicon for warmed-over aroma/flavor attribute.

Present four samples for panelist evaluation of all attributes introduced to this point:

- Cow top loin steak broiled to 57°C—refrigerator stale notes and warmed-over flavor
- Select tenderloin steak broiled to 78°C—warmed-over aroma and flavor
- Choice flat iron steak (70°C)
- Select flat iron steak (70°C)

Session 21 – Calibration; evaluate all attributes

Warm-Up: Choice top loin steak

- Sample 1—select eye of round roast
- Sample 2—choice top sirloin steak
- Sample 3—choice knuckle roast
- Sample 4—select flat iron steak

Session 22 – Calibration; evaluate all attributes

Warm-Up: Choice eye of round roast

- Sample 1—select inside round roast
- Sample 2—choice bottom round roast
- Sample 3—choice flat iron steak
- Sample 4—select tenderloin steak

***Training schedule is adjustable and may be changed to optimize panel performance as level of previous training may influence ability of panelists to successfully complete each exercise.*

Training time is a function of product testing method and procedural variables, and it increases with the number of attributes to be studied. The availability of panelists and the responses of panelists to training influence training time. Individuals should be evaluated during training and while the study is in progress. Delays or interruptions of more than two weeks in a series of tests should be followed by refresher training sessions and perhaps an evaluation. Training is never completed, and day-to-day variation among panelists is an issue that must continually be monitored.

Trained descriptive attribute panels should not be asked to evaluate any attribute in terms of like/dislike or acceptability. Those responses should be obtained only from a consumer panel (see Consumer Panels section).

After the panel has been trained, ballot development sessions begin with the standard lexicon and are enhanced with attributes unique to a given study. References for these attributes can be found in ASTM DS72 (2011) as well as the scientific literature. Munoz and Civille (1998) discussed attribute scaling and development of common lexicons for use in descriptive analysis in ballot development sessions.

4. Performance evaluation for descriptive attribute panels

Performance evaluation can begin soon after training starts. The initial and subsequent evaluations help the panel leader identify problems among individual panelists. Nine samples are selected to cover the full range of test attributes. Panel evaluation is spread over four days with three sessions per day and three samples per session. The design is outlined in Table 9.

Table 9. Example of how to randomize nine samples to three sessions over four sensory days for performance evaluation using a random numbers table.

Day1			Day 2			Day 3			Day 4		
Session			Session			Session			Session		
1	2	3	1	2	3	1	2	3	1	2	3
S9	S5	S6	S8	S6	S3	S2	S6	S7	S6	S8	S4
S8	S3	S1	S1	S9	S7	S4	S3	S1	S7	S3	S2
S2	S4	S7	S5	S2	S4	S9	S5	S8	S1	S5	S9

S=Sample number

Samples also can be arranged in a William's square arrangement to permit each treatment to appear in every session as well as every serving position within the sessions. An example of a 6 × 6 William's square is shown in Table 10.

Table 10. Example of a 6 X 6 Williams square.

	Serving Order					
Sensory Session	1	2	3	4	5	6
1	A	B	F	C	E	D
2	B	C	A	D	F	E
3	C	D	B	E	A	F
4	D	E	C	F	B	A
5	E	F	D	A	C	B
6	F	A	E	B	D	C

Adapted from Williams (1949).

Data analysis treats the data for each candidate as a one-way ANOVA with nine treatments and four observations per cell. The data also could be evaluated as two-way ANOVA with one observation per cell. The design can be treated as a balanced-lattice design (Cochran and Cox, 1957). With this design, day and session effects can be studied. The layout for data is given in Table 9. From the ANOVA table, calculate the F-ratio defined as F = MS treatments/MS error.

Used in this context, the F-ratio is a measure of a panelist's ability to award different scores to different samples while being able to repeat them on the same sample on different days. The degree to which a person discriminates among samples and is consistent in replicate judgments is reflected in the F-ratio (ASTM STP434, 1968; Cross et al., 1978). A larger F-ratio indicates a better-performing panelist. Candidates can be ranked on the basis of their F-ratios, which are product, attribute, and study dependent.

The data can be analyzed by ANOVA, and the effect of panelist, treatment, and panelist by treatment interaction can be tested. Order effects can be defined as random effects in the model. A discussion is provided in the data analysis section. By using ANOVA, the residual error will include sensory day and unexplained differences between samples.

The number of panelists selected should be based on the performance evaluation results. A person whose evaluation is less than satisfactory should not be included just to achieve a predetermined panel size. ASTM STP434 (1968) requires a minimum of five panelists because fewer would represent too much dependence upon any one individual's response. A minimum of eight panelists for a trained descriptive attribute panel, with an optimum starting number of twelve to allow for panel attrition, however, is recommended. ASTM MNL26-2nd (1996) states that it is important that panelists show the ability to discriminate using the product and attributes for the study. It is more important to have panelists who can discriminate and have fewer numbers than to have higher numbers of panelists who are not adequately trained.

The decision of who should or should not be on the panel should not be based solely on the initial evaluation. Subsequent evaluations of panelists' performance are necessary throughout the study. Training for new panels could last two to three months, depending on the individuals involved and the descriptive analysis method being used. Refer to Cross et al. (1978) for review and validation of this technique or see discussion of panelist effects under the data analysis section of these guidelines.

5. Monitoring panelist performance

Performance records should be maintained for each panelist and should be periodically reviewed by the panel leader. Performance evaluations (ANOVA that includes panelist effects and panelist by treatment interactions) alert the panel leader to any problems a panelist might have. Consistently poor performance may indicate that the panelist was incorrectly selected for the study or that physical or psychological distractions are preventing them from making good evaluations. For panelists with poor performance identified during or between studies, additional training sessions should be conducted. We recommend their data should only be used once they have passed the subsequent performance evaluations.

6. Panel maintenance

The importance of cooperation and performance as it relates to the entire project should be stressed. Feedback to the panelists as to their importance to the total research program is an integral part of keeping the group willing to participate. Some institutions have successfully used paid panelists. A system of rewards also can be instituted. For example, cookies, candy, cake, or ice cream can be used. Those who have used paid panelists have found them to be more highly motivated and more consistently available than "in-house" personnel. As a result of being on the sensory panel, panelists may develop other outside activities together. Holding periodic social events among management and panelists may help create the stimulus to stay with the group. Care should be taken, however, to ensure outside activities and discussions are not brought into the sensory evaluation environment. It also is important that attributes of test products are not discussed during outside social activities until after an experiment is completed.

Panelists should be continually trained to maintain relevant attributes fresh in their memory. Reference samples should be available if requested. When necessary, screen other potential panel members in order to maintain or enlarge the panel. In any particular study or test, substitution of panel members should be avoided. Refresher training should be conducted before starting a new experiment and after extended interruption of testing.

D. Consumer Sensory Panel

1. Scope

While use of a trained panel to assess treatment differences in a set of products is critical in determining if the treatment differences are detectable, it is also critical to know if these differences impact consumer acceptance. Are these differences large enough for consumers to detect? Are they important enough to affect overall acceptance? Consumer testing often is a key part of a study as the fate of a food product always has rested on the acceptance by the consuming public. It is only through careful planning of the consumer test, including choosing the most appropriate test method, that the researcher can get reliable, repeatable results that can be projectable to the real consumer population.

Trained and consumer panels can both be useful to determine if differences exist in and between products. Consumer tests should be used if it is important to determine if consumers can detect the differences in the product due to the addition of some treatment. In this situation, Triangle tests can be conducted with consumers to determine if the consumers can find a difference in the product. An example of this type of test would be the addition of a microbiological intervention to ground beef and its impact on the sensory properties of the product. As consumers are less likely to detect differences compared with trained panelists, it is advisable to use consumer Triangle test data in conjunction with trained panel data when making decisions on the products.

Additional points to consider when designing and conducting a consumer sensory test are experimental design, protocol design, test method, ballot development, moderator selection, and interaction with the sensory subjects. Meilgaard et al. (2007), Lawless and Heymann (2010), ASTM MNL26-2nd (1996), and Stone and Sidel (2004) are excellent references for information on the design of consumer panels.

2. Determining the test type

The first step in setting up a consumer study is determining the most appropriate test method to use. The test method should be selected based on the study objectives and on how the results will be used. Consumer panels can be defined as either qualitative or quantitative tests.

a. Qualitative tests

Qualitative tests provide a subjective response of a sample from consumers by having those consumers talk about their feelings regarding the sensory properties of a set of products (Meilgaard et al., 2007). Examples of qualitative tests include focus groups or panels; mini groups; diads and triads; and one-on-one interviews.

b. Quantitative tests

Quantitative tests are used to determine consumers' sensory perception of products by using a set of questions to measure preference, liking, and impressions of various sensory attributes. Quantitative tests can be conducted using central location testing (CLT) with pre recruited participants; nonprerecruited, such as a mall intercept test; or using home use testing (HUT).

If the intent of the consumer evaluation is to determine which product is preferred, preference testing should be performed. How well a product is liked by consumers can be ascertained through acceptance tests. Hedonic scales (like/dislike) are used to indicate the degree of acceptability. Just-about-right (JAR) scales and intensity scales can be used to determine when an attribute is too high/strong or too low/weak.

3. Protocol design

The number and type of panelists, as well as sample preparation and presentation, can significantly influence consumer responses and, therefore, significantly affect or limit interpretation of the results from the study.

a. Panel size

Determining the optimal panel size is influenced by a number of factors, such as the test objectives, test method (CLT, HUT, etc.), and the expected difference to be detected. ASTM MNL26-2nd (1996) states that 100 consumers usually are adequate for most small consumer tests, but the exact number depends on the experimental design. According to Hough et al. (2006), the optimum panel size is dictated by the following:

- d = size of difference to be detected, i.e., 0.4 units vs. 0.8 units on a 9-point hedonic scale.
- s = standard error of the experiment. This value is influenced by several factors such as the amount of product variability, both within treatment and among treatments, as well as product acceptability differences driven by age, gender, and geographical location.
- α level = probability of Type I error.
- β level = probability of Type II error.

The standard error of the experiment is not known beforehand but can be estimated from previous studies. For consumer studies, data from past studies can be analyzed by analysis of variance and standard error of the experiment determined by the square root of the mean square error (Hough et al., 2006).

Type I and Type II errors are key measures based on the test objectives and have a large impact in determining sample size. If the objective of the study is to determine whether or not the test product is significantly better in flavor than the control, then Type I error is the critical measure. In this case, Type I error is the risk in declaring that the test product has a better flavor, although, in reality, the test product does not. Conversely, if the objective is to determine whether or not the test product is liked as well as the control after an extended storage period, Type II error is the critical measure. In this situation, the risk would be in declaring the test product equally acceptable to control, although, in reality, the test product is inferior. Larger

sample sizes are typically required for parity tests (Type II error focus) than for superiority tests (Type I error focus).

In addition to sample variability, researchers also must consider the testing conditions. Under controlled conditions of sensory panel booths, there are fewer distractions, so fewer panelists are needed compared with a less controlled setting, such as an HUT. Sample size also will need to be increased if testing in multiple locations because regional differences in acceptability might exist. A minimum of 50 consumers per location is generally recommended in order to test for regional differences. In situations where consumer segments exist, such as preferences for spice level, sample size also will need to be increased in order to allow for data segmentation.

Review and evaluation of previous studies with a statistician is beneficial for establishing the appropriate number of consumers because consumer testing can be quite expensive. Additional references for sample size determination can be found in ASTM E1958-07e1 (2012) and Lawless and Heymann (2010).

b. Panelist selection

The application of results will depend extensively on the criteria used in selecting test participants. Therefore, selecting the proper panelists is key to obtaining test results that are projectable to the target population.

When selecting panelists for a consumer test, the following factors should be considered:

- Who is the target population? Testing with target consumers is recommended as the best way to get an accurate measure of consumer acceptance. Is it sufficient to test with company employees or is it necessary to recruit local area residents? Or, in some cases, it is sufficient to test from the general population by recruiting from the targeted user group in either one or multiple national markets in order to more accurately reflect the consumer base. A word of caution when using employees as consumer panelists—limit the number of times they can participate so that they don't become overly sensitive to product differences. Additionally, do not use employees who are knowledgeable about the study. It is also best to avoid Research & Development and Quality Assurance employees because of their technical backgrounds.
- What are the product usage patterns for your product? Is the interest in heavy users of the product or are light users acceptable? Preferences for the degree of doneness also are critical because serving a medium-rare meat sample to a panelist who typically eats their meat well done would not give a true read on consumer acceptance.
- What are the demographics of the target population? Factors such as age, gender, income, education, and household size are often used to help identify the target consumer.
- In how many markets does one need to test? Which cities have the largest number of the target consumers? Are there regional preferences in the product?

4. Ballot development

The consumer sensory ballot, which is defined as the sensory instrument for consumer testing, is critical to ensure that the sensory professional is accurately testing consumer responses and

not biasing the responses with a poorly designed ballot. In designing a ballot, some general rules are as follows:

- Appearance, flavor, then texture. Within each session, ask for responses in the order in which they are normally encountered while eating.
- Use a combination of close-ended (i.e. hedonic scales) and open-ended (no defined response) questions.
- Do not change scales in the middle of the ballot; if a nine-point hedonic scale is being used, do not change to a five- or seven-point hedonic scale. It is appropriate, however, to include intensity (five-, seven-, or nine-point) and/or JAR (typically five-point) scales for impressions of specific attributes.
- When using JAR scales, always use penalty analysis to determine the impact of an attribute not being JAR.
- Do not reverse the end-anchors on the scales within one ballot.
- Provide clear, concise directions and unambiguous questions to assure that the test is measuring consumers' sensory perceptions accurately.
- Keep questions to a minimum.
- The most important consumer question is almost always the consumer's overall like/dislike of the product. The position of this question can influence the answer. When asked after appearance and taking the first bite of the product, the answer reflects the consumer's first impression. When asked at the end of the ballot after questions concerning the appearance, juiciness, tenderness, and flavor, the answer reflects the consumer's first impression and any influence of questions asked. The researchers must consider this when designing and interpreting the results.

a. Hedonic scales

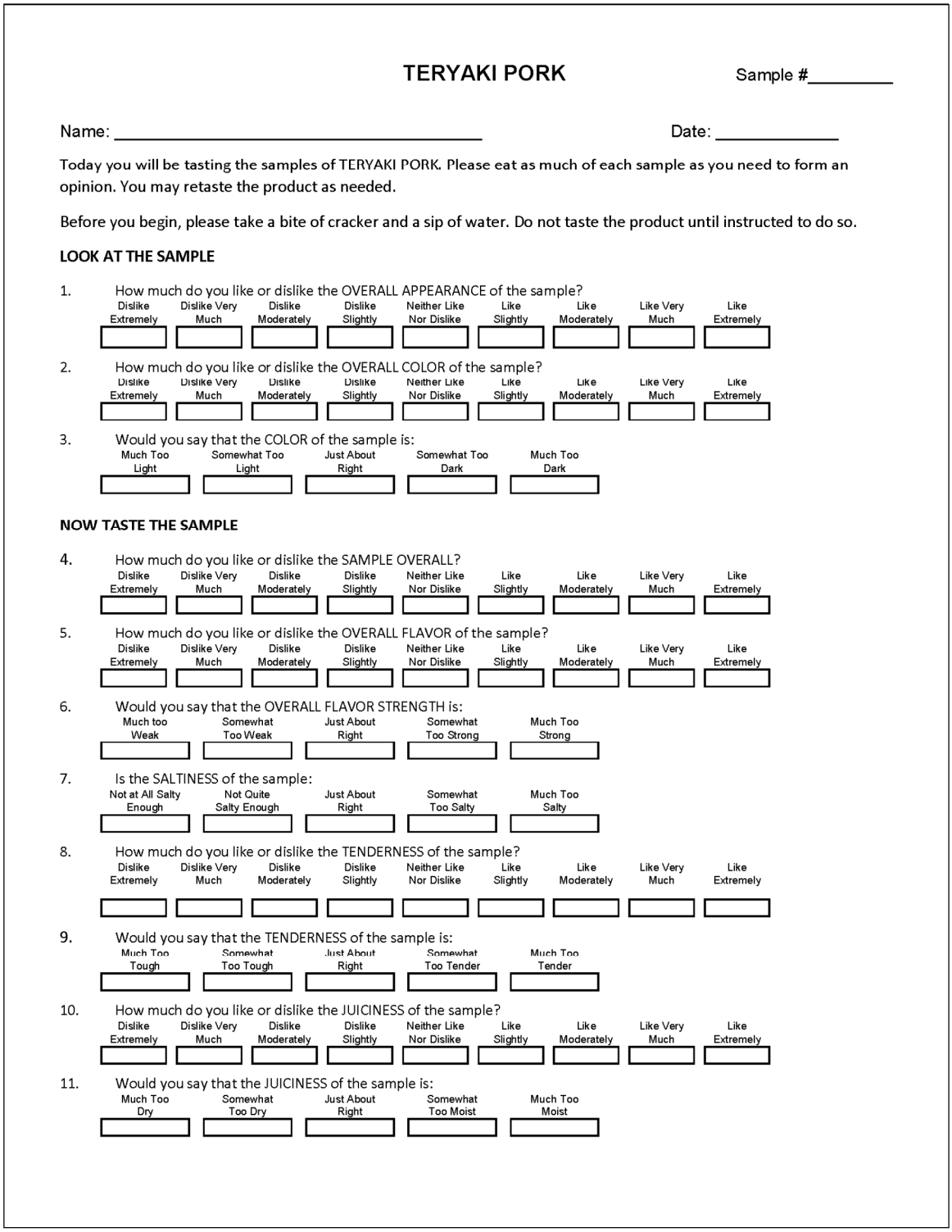
The hedonic scale can be shown horizontally or vertically, and scales can be verbally anchored at each point along the scale with nine categories as follows:

- Like extremely
- Like very much
- Like moderately
- Like slightly
- Neither like nor dislike
- Dislike slightly
- Dislike moderately
- Dislike very much
- Dislike extremely

When using the hedonic scale horizontally, it should be anchored with Dislike Extremely on the left and Like Extremely on the far right, as shown in the sample consumer ballot in Figure 15.

The purpose of anchoring the scale at each point is to encourage a continuum with equal spaces between each successive increase in like/dislike or preference. Meilgaard et al. (2007) and Lawless and Heymann (2010) give examples of hedonic scales that can be used in consumer testing. These scales range from the anchored, nine-point hedonic; to the end-anchored, nine-

point hedonic; to the end- and neutral-anchored, nine-point hedonic; to the non-balanced scale (more or less categories of like in relation to dislike). Many different forms of hedonic scales can be used without major effects on the value of the results, as long as the essential feature of verbal anchoring of clearly successive categories is retained. There should be at least five categories. Replacement of the verbal categories with caricatures representing degree of pleasure and displeasure (smiley scale) have been used, but studies on this type of scale have indicated potential issues with how children interpret the caricatures, so these scale types are not recommended (ASTM E2299-11, 2011). If testing with children, a nine-point “super good/super bad” scale developed by Kroll (1990) has been used successfully. It can be truncated to seven points for younger children if needed.



TERYAKI PORK Sample #_____

Name: _____ Date: _____

Today you will be tasting the samples of TERYAKI PORK. Please eat as much of each sample as you need to form an opinion. You may retaste the product as needed.

Before you begin, please take a bite of cracker and a sip of water. Do not taste the product until instructed to do so.

LOOK AT THE SAMPLE

1. How much do you like or dislike the OVERALL APPEARANCE of the sample?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

2. How much do you like or dislike the OVERALL COLOR of the sample?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

3. Would you say that the COLOR of the sample is:

Much Too Light	Somewhat Too Light	Just About Right	Somewhat Too Dark	Much Too Dark
☐	☐	☐	☐	☐

NOW TASTE THE SAMPLE

4. How much do you like or dislike the SAMPLE OVERALL?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

5. How much do you like or dislike the OVERALL FLAVOR of the sample?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

6. Would you say that the OVERALL FLAVOR STRENGTH is:

Much too Weak	Somewhat Too Weak	Just About Right	Somewhat Too Strong	Much Too Strong
☐	☐	☐	☐	☐

7. Is the SALTINESS of the sample:

Not at All Salty Enough	Not Quite Salty Enough	Just About Right	Somewhat Too Salty	Much Too Salty
☐	☐	☐	☐	☐

8. How much do you like or dislike the TENDERNESS of the sample?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

9. Would you say that the TENDERNESS of the sample is:

Much Too Tough	Somewhat Too Tough	Just About Right	Somewhat Too Tender	Much Too Tender
☐	☐	☐	☐	☐

10. How much do you like or dislike the JUICINESS of the sample?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

11. Would you say that the JUICINESS of the sample is:

Much Too Dry	Somewhat Too Dry	Just About Right	Somewhat Too Moist	Much Too Moist
☐	☐	☐	☐	☐

Figure 15. Sample Ballot for consumer testing.

b. JAR and intensity scales

Intensity scales and Just About Right (JAR) scales can be used to determine if products differ significantly in the levels of specific attributes. These scales can be bipolar or unipolar. Intensity scales can be line scales or category scales, and they can be end anchored or fully anchored. When used as category scales, intensity scales are often five-point, seven-point, or nine-point. If using a mixture of JAR and intensity scales on the ballot, it is good practice to use different scale lengths for the two scale types in order to minimize confusion on what is being asked.

Just-about-right scales are typically category scales and are most often five-point, with seven-point and nine-point scales used less often (Figure 15). Because an attribute can vary without negatively impacting consumer liking, it is critical when using these scales that penalty analysis be used to assess the impact or penalty of an attribute not being JAR. Penalty analysis links the drop in overall liking to the proportion of respondents rating an attribute either too high or too low. It, therefore, provides the ability to prioritize product optimization opportunities by determining the critical product attributes that penalize product acceptance the most. It is important to note that penalty analysis should only be run for JAR attributes with 20% above or below JAR. Running the analysis with less than the 20% cutoff results in too much dependency on any given respondent. For studies with smaller sample sizes or when analyzing subgroups within a study, the cutoff should be increased to at least 25%.

Several statistical analysis software packages, such as XLSTAT, offer an automated penalty analysis feature. If unavailable, however, the penalty for an attribute not being at JAR is calculated in the following manner:

- Collapse JAR scales to three points.
    - Too Little (TL)
    - JAR
    - Too Much (TM)
- For each attribute with $\geq$ 20% of respondents rating it either too high or too low, calculate the mean drop in overall liking for not being JAR.
    - $\text{Mean Drop}_{TL} = \text{Average}_{JAR} - \text{Average}_{TL}$
    - $\text{Mean Drop}_{TM} = \text{Average}_{JAR} - \text{Average}_{TM}$
- Calculate total penalty for each attribute not JAR.
    - Total Penalty = (Mean Drop)*(% Not JAR)

Penalties can be shown graphically in two ways, as shown below in Figures 16 and 17.

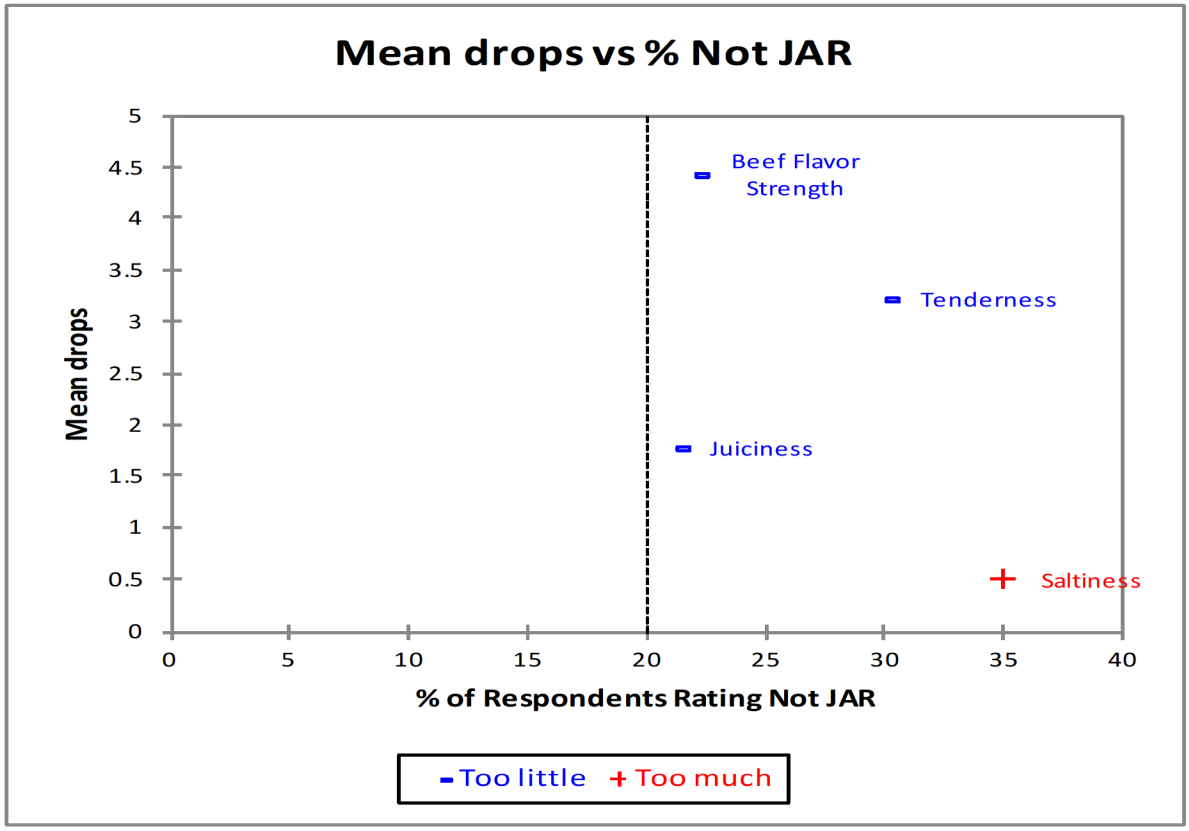


Figure 16. Mean drop in overall liking as a function of the percent not at JAR ratings for the various attributes.

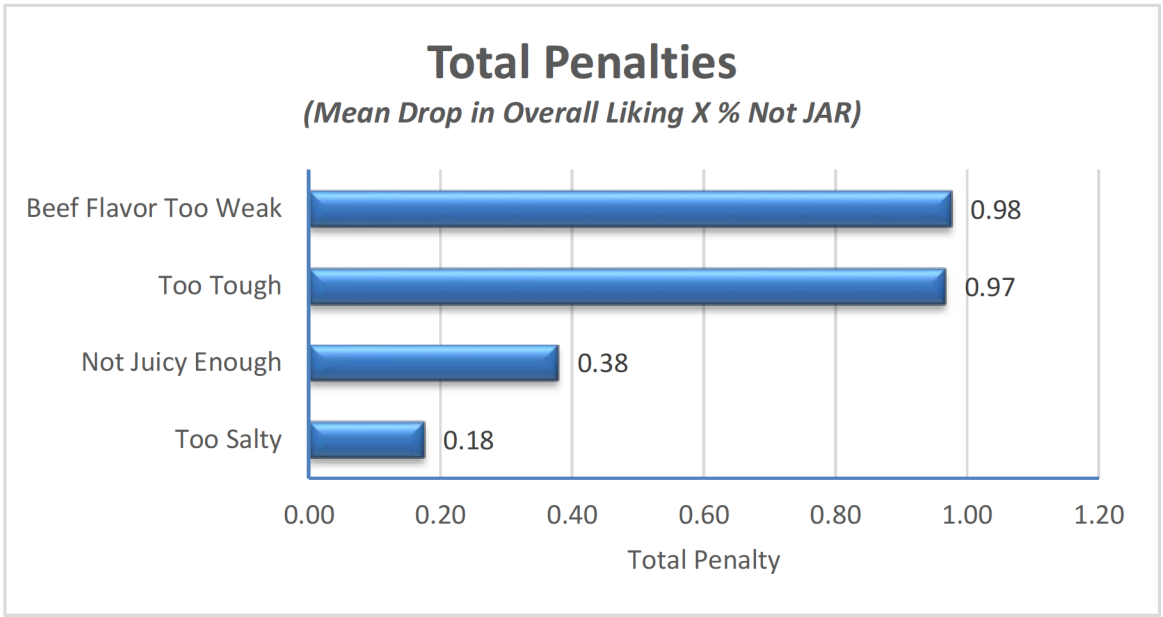


Figure 17. Total penalties of the various attributes.

Total penalties then can be categorized into different tiers based on potential opportunity to improve overall liking if the offending attribute is brought to JAR:

- < 0.25 = low tier penalty: no changes needed
- 0.25 – < 0.50 = mid-tier penalty: consider changing
- ≥ 0.50 = high tier penalty: must change

In the above example, results showed that 35% of the panelists rated the sample “too salty,” but the drop in mean overall liking was 0.5 units, which resulted in a total penalty of 0.18 units. This penalty is considered low tier, indicating the salt level does not need to be adjusted.

Additional information on penalty analysis and the uses and abuses of JAR scales can be found in ASTM MNL63 (2009).

VIII. INSTRUMENTAL MEASURES OF TENDERNESS AND TEXTURAL PROPERTIES

Instrumental methods of measuring meat tenderness are—like trained descriptive attribute panels—objective measures of meat tenderness. Conversely, a consumer panel can provide a subjective measure of tenderness. Historically, the trained descriptive attribute sensory panel tenderness rating has been considered the gold standard to which all other measures are compared. It should be clear, however, that both trained sensory panel and instrumental measures give only measures of relative differences in tenderness. They give little indication of the acceptability of a given tenderness measure, other than that derived from associating objective measures with consumer acceptance data. Acceptability of a given level of tenderness can only be determined by the ultimate users, consumers. Historically, the lack of consumer data on meat tenderness has led many researchers to over-interpret their objective measures as indicating whether meat would be considered tough, tender, or acceptable. Although much progress has been made in collecting meaningful consumer acceptance data, more data on the relationship between consumer impressions of meat tenderness and objective measures of tenderness are needed. A critical problem has been the diversity in shear force measurement protocols that made it nearly impossible to appropriately compare shear force values among data published by different institutions. The large variability in mean shear force and repeatability of shear force from different institutions and different protocols has been demonstrated by comparing Warner-Bratzler and slice shear force (SSF) values on matched steaks from the same animals (Wheeler et al., 1997; Wheeler, Shackelford, & Koohmaraie, 2007). Those data highlight not only the benefits of standard protocols, but also the problem with thresholds such as those published by Shackelford, Morgan, Cross, and Savell (1991) relating WBSF to trained sensory panel tenderness rating in that, depending on the protocol, the threshold may only be applicable to data collected at that same institution or in that single study.

It is often very useful to be able to not only compare objective measures among institutions, but also make useful comparisons of consumer tenderness data with objective measures of tenderness. Toward this goal, much progress has been made on use of more standardized protocols. When there is not one methodology that is superior to others, it seems obvious that the advantages of everyone having comparable data outweigh the disadvantages of restrictive protocol. A superior method, however, should not be discarded for the sake of uniformity of methodology. It also is recognized that the recommendation of certain procedures in these guidelines should not discourage research to evaluate alternative methods or instruments.

A. Whole-Muscle Steaks/Roasts/Chops

1. Instruments/measurements

Numerous devices have been tested for their ability to measure meat tenderness. The measurement most often used has been WBSF, which is still frequently used today. Slice shear force, however, now is commonly used in many laboratories, and its use continues to grow. In fact, the National Cattlemen's Beef Association (NCBA) National Beef Instrument Assessment Plan II—Tenderness Committee recommended the industry use SSF to begin collecting baseline

tenderness data (NCBA, 2002), which at least one commercial company has been doing for several years.

Beef longissimus WBSF of 1.27 cm diameter cores is highly repeatable when measurement protocols are executed properly (Wheeler, Koohmaraie, Cundiff, & Dikeman, 1994; Wheeler et al., 1996, 1997). Several sources of error have been identified, however, that contribute to errors in shear force assessment within and among institutions (Wheeler et al., 1994, 1996, 1997). While developing a method for online assessment of meat tenderness, Shackelford, Wheeler, and Koohmaraie (1999a) developed a simplified technique for measuring beef longissimus shear force, which is referred to as SSF, which appeared to be more accurate than WBSF. The repeatability of SSF (0.89) exceeded repeatability estimates (0.53 to 0.86) that have been reported for longissimus WBSF (Wheeler et al., 1996, 1997). Because of time constraints associated with online assessment of meat tenderness, some aspects of the SSF protocol that Shackelford et al. (1999a) developed for online assessment of beef longissimus tenderness may not be necessary or desirable for routine collection of shear force data in a laboratory setting. Thus, Shackelford, Wheeler, and Koohmaraie (1999b) conducted a series of experiments to develop an optimal protocol for routine SSF measurement in research and to evaluate SSF as an objective method of assessing beef longissimus tenderness. Subsequent experiments were conducted to adapt this technique for use on pork longissimus, lamb longissimus, and many other beef muscles (Shackelford, Wheeler, & Koohmaraie, 2004a,b; Shackelford, King, & Wheeler, 2014).

Because the SSF technique allows for a substantial increase in laboratory throughput, it has allowed for the development of novel tenderness management systems. To date, by far, the most common use of the SSF technique has been for the evaluation of beef longissimus tenderness. This has included routine testing of commercial product lines as well as in research.

a. Warner-Bratzler shear force

This measure can be obtained either with the original WBSF machine or with WBS blade attachments to an automated testing machine (e.g., Instron, United, Texture Technologies, etc.). In addition to peak load (maximum shear force), other traits that might be useful also can be obtained with an automated testing machine. V-notch blades used for shear force should be either the blades made for WBS machines by G-R Manufacturing (Manhattan, Kansas) or blades sold by the testing machine manufacturer. These blades are milled to exact specifications, including the bevel on the cutting edge. Unless in-house manufactured blades meet these exact specifications, they should not be used.

Warner-Bratzler shear blade specifications include: (1) blade thickness of 1.1684 mm (0.046 inches); (2) V-notched (60° angle) cutting blade; (3) cutting edge beveled to a half-round; (4) corner of V rounded to a quarter-round of a 2.363 mm diameter circle; (5) spacers providing gap for cutting blade to slide through of 2.0828 mm thickness. After cooking and recording final cooked temperature and weight, steaks should be chilled overnight at 2 to 5°C before coring. Chilling firms the steak and makes it easier to obtain uniform diameter cores. If chilling is not used, some protocol to obtain consistent steak temperature before coring should be followed, such as allowing steaks to reach room temperature (23°C). Round cores should be uniformly 1.27 cm (0.5 inches) in diameter and removed parallel to the longitudinal orientation of the

muscle fibers so that the shearing action is perpendicular to the longitudinal orientation of the muscle fibers.

Cores can be obtained using a handheld coring device (cork borer) or an automated coring device (drill press with cork borer attached). Coring devices must be in good condition and sharp or the core diameters will not be consistent and will result in spurious increased variation in shear values. Hand coring requires very careful attention to the amount of pressure used as the corer is turned. A minimum of six cores should be obtained from each sample (this might require two pork chops or three lamb chops). Cores that are not uniform in diameter, have obvious connective tissue defects, or otherwise would not be representative of the sample should be discarded. If cooked steaks/chops were chilled, cores should be kept refrigerated until sheared to maintain consistent temperature. Obtaining good uniform cores is the most critical variable in WBSF measurements.

All values obtained should be used for mean calculation, unless visual observation indicates some reason a value should be discarded (e.g., a piece of connective tissue in the shear plane). Each core should be sheared once in the center to avoid the hardening that occurs toward the outside cooked edge of the sample. Warner-Bratzler shear force tests using automated testing machines should be conducted with a crosshead speed of 200 to 250 mm/minute. Shear tests that do not follow these equipment or sample specifications should not be referred to as WBSF (such as square holes in the shear blade, square meat samples, straight-edged shear blade, or blade not properly beveled, etc.). For a demonstration of WBSF, watch the video at http://www.meatscience.org/sensory.

b. Slice shear force

Immediately after cooking, a 1-cm thick, 5-cm long slice is removed from each steak parallel to the muscle fibers. The slice is acquired by first cutting across the width of the longissimus at a point approximately 2 cm from the lateral end of the muscle. Using a sample sizer, a cut is made across the longissimus parallel to the first cut at a distance 5 cm from the first cut. Using a knife that consists of two parallel blades spaced 1 cm apart, two parallel cuts are simultaneously made through the length of the 5-cm long steak portion at a 45° angle to the long axis of the longissimus and parallel with the muscle fibers.

The 5-cm long, 1-cm thick slice is sheared perpendicular to the muscle fibers using a universal testing machine equipped with a flat, blunt-end blade. The SSF blade is designed to replace the WBSF blade on an automated testing machine. The SSF blade has the same thickness (1.1684 mm) and degree of bevel (half-round) on the shearing edge as WBSF blades and should be used with the same amount of gap (2.0828 mm) for the blade to pass through during shearing. The crosshead speed is set at 500 mm/min to minimize the time required for measurement of shear force. Optionally, SSF could be measured using a WBSF machine equipped with an SSF blade, as described in the “small volume” protocol listed below. In that case, the crosshead speed is dictated by the WBSF machine.

For a demonstration of SSF, watch the video at http://www.meatscience.org/sensory. More details of the optimal protocol for longissimus SSF measurement, including detailed pictures and equipment sources are also posted at http://www.meatscience.org/sensory.

(1) Chilled vs. hot

Available data have not demonstrated that more accurate or repeatable data are obtained when WBSF is conducted on chilled samples compared to hot or room-temperature samples (Wheeler et al., 1994). Thus, although empirical observation indicates it is easier to obtain more uniform diameter cores from chilled steaks than warm or room-temperature steaks, steak temperature at coring may not be critical as long as it is consistent within an experiment. The limited data available comparing hot vs. cold SSF, however, indicates hot SSF resulted in higher correlations to WBSF and trained sensory tenderness ratings than did cold SSF (Shackelford et al., 1999b). Thus, the implication of more accurate data and the increased convenience of cooking and shearing in the same day leads to a recommendation for hot SSF.

2. When NOT to use shear force

It has been clearly shown that shear force (both WBSF and SSF) does not properly reflect tenderness differences among muscles (Bouton et al., 1978; Harris & Shorthose, 1988; Shackelford, Wheeler, & Koohmaraie, 1995; Rhee, Wheeler, Shackelford, & Koohmaraie, 2004; King, Wheeler, Shackelford, & Koohmaraie, 2009). For example, beef biceps femoris and longissimus have similar WBSF values, but longissimus is more tender as assessed by a trained descriptive attribute panel (Shackelford et al., 1995; Rhee et al., 2004). Therefore, it is inappropriate to use shear force to compare tenderness differences among muscles. Unfortunately, there are numerous cases in the literature in which this has been done. This likely will lead to misinterpretation of data and false conclusions. Sensory evaluation should be used to compare muscles. When it is necessary to measure shear force in multiple muscles in a given experiment, we urge the investigators to not compare (statistically or otherwise) shear force among muscles and to report the data in such a manner that does not encourage the reader to compare shear force among muscles. Furthermore, we urge the investigators to use a qualifying statement either as a footnote (legend entry) or as a part of the statistical analysis section of the document such as the following sentence from King et al. (2009):

Slice shear force data were analyzed independently for each muscle because measures of shear force do not accurately reflect tenderness differences between muscles (Shackelford et al., 1995; Rhee et al., 2004), because shear force does not accurately represent the contribution of connective tissue to muscle tenderness (Bouton et al., 1978; Harris and Shorthose, 1988).

3. Instrument calibration

Calibration is essential with any instrument, and verification is recommended at 12- to 18-month intervals. Calibration is the daily spot-check of the instrument accuracy. It is performed by placing a known weight on the transducer or by applying a known voltage to the cell via a shunt. Adjustments can be made so that the instrument output matches the known weight or voltage input. The calibration procedure is usually performed using a single value or with weights at approximately 20% and 80% of force values expected in the test (see ASTM E4, 2010, for other approaches). The shear blade attachment must be in place during calibration so that its weight is tared from the machine.

Automated testing machines should be verified according to the manufacturers' instructions. When crosshead speed is critical to the results, it is recommended to have it verified as well.

When force values form the basis of value, for example premiums for tenderness, instruments will need to be validated daily (ASTM F2343, 2006). This process is referred to as validation; a procedure that documents, confirms and provides assurances that force measurements will consistently meet predetermined specifications and force attributes. Further, validation also provides assurances that instruments are installed and operated within the manufacturers' specifications (ASTM F2341, 2005).

The United States Department of Agriculture's (USDA's) Livestock, Poultry and Seed Program approves third-party laboratories for conducting SSF/WBSF testing to ensure that the requirements of an ASTM International tenderness standard have been met. Third-party laboratories are approved to conduct SSF and/or WBSF through proficiency testing conducted in conjunction with an AMS-designated SSF/WBSF reference laboratory (USDA Livestock, Poultry and Seed Program, 2012).

B. Ground Beef/Patties

Evaluation of ground beef almost always requires more than just an evaluation of overall tenderness. Thus, some type of automated testing machine is essential. Potentially useful measurements include the following:

- Shear force (by Allo-Kramer)
- Hardness
- Springiness
- Cohesiveness
- Gumminess
- Chewiness

Additional measures also can provide useful information depending on the objectives.

a. Shear force measures

One strip (2.5 cm wide) should be cut from the center of each of 10 patties per formulation. Each strip should have one of the following qualities:

- Sheared once with a multi-bladed Allo-Kramer shearing device
- Sheared three times with a single-blade, straightedge Allo-Kramer shearing device
- Sheared three times with a straightedge SSF blade attachment

A crosshead speed of 200 to 250 mm/min should be used if using an automated testing machine.

b. Compression measures

Hardness, springiness, cohesiveness, gumminess, and chewiness should be determined according to Bourne (1978). One core (2.54 cm diameter) should be removed from the center of 10 cooked patties and compressed twice to 70% its original height. Hardness is the peak force during the first compression cycle ("first bite"). Cohesiveness is the ratio of the peak force area during the second compression to the peak force area during the first compression (Area1/Area2). Springiness (originally called elasticity) is the height that the food recovers during the time elapsed between the end of the first compression and the start of the second

compression. Gumminess is the product of hardness and cohesiveness. Chewiness is the product of gumminess and springiness. A crosshead speed of 100 mm/min should be used.

IX. DATA ANALYSES

Data analysis is a critical component of sensory evaluation. Sensory data are, in the true sense, multivariate. When a human, either trained or consumer, evaluates a meat sample, they utilize multiple senses during the evaluation process. Trained, descriptive attribute sensory panelists are educated to incur sensory input, segment the information into single attributes, and then rate the intensity of each attribute using an anchored, defined scale. Consumers incur sensory input and may or may not segment that information into individual attributes unless specifically asked to rate a specific trait. Therefore, overall like/dislike consumer ratings are really ratings of multivariate information, and consumer ratings of specific attributes—such as overall flavor like/dislike, overall texture like/dislike, overall juiciness like/dislike, intensity of beef flavor, or level of juiciness—are univariate attributes.

There are multiple statistical analyses tools for sensory data. Researchers should focus on tools that most effectively address the hypothesis of the study. This section will provide information on the most commonly used data analyses tools. Univariate and multivariate tools will be discussed for descriptive and consumer data that are currently available. Additional information can be obtained in Meilgaard et al. (2007), O'Mahony (1986), and ASTM Committee E18 documents on statistical analyses.

A statistician or sensory professional with extensive experience in analyzing data should be consulted before conducting any sensory experiment. The experimental design, randomization, and hypothesis should be clearly defined before initiation of data collection. Initial statistical models and level of significance should be defined so that the experiment is properly executed and predetermined biases do not inadvertently affect the experimental results.

Most descriptive sensory data generated using trained panels are analyzed using ANOVA, where each sensory attribute defined in the ballot is evaluated for effects of treatment independently. This type of analysis provides the opportunity to understand if specific attributes differ among treatment levels. The analysis will include three basic steps: (1) data preparation to include data entry, data verification, summary statistics, and tests for normality and homogeneity; (2) determining the efficacy of the panel; and (3) data analyses that might include univariate and/or multivariate techniques.

A. Data Preparation

Sensory data are traditionally either collected by using a handwritten ballot filled out by each panelist or entered into a computer software program. Regardless of the collection procedure for the data, it needs to be organized so that one row represents the responses of one panelist for one sample. Variables for sensory day, session within sensory day, order, three-digit sensory codes, panelists, treatment codes, and sensory variables for each response should be included as columns. Variables that might be potential covariates or blocks, such as cook time and cook yield, should be included.

If data have been entered into a spreadsheet with panelists' results for each sample, then reorganization is not necessary and the next step of analysis can begin. If data have been collected using a computer program, the data might need to be unscrambled from the randomized sample sheet to assign the observed values to the actual treatments using a

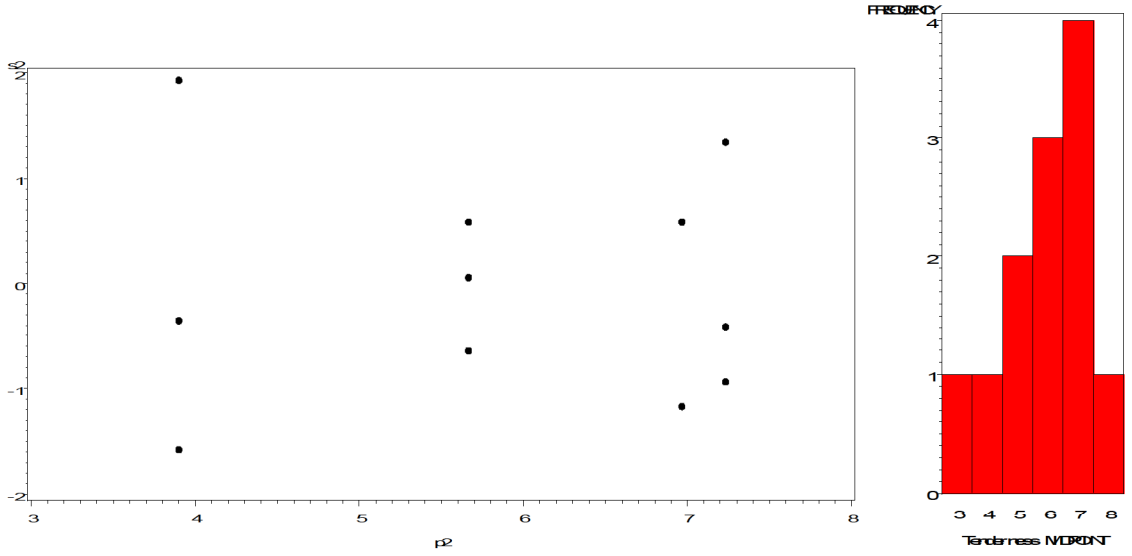
program like Unscrambler from SAS[1] (v. 9.2). If handwritten ballots are used, data should be entered into a spreadsheet and then verified by independent personnel using the original data sheets. Computerized data sheets should be checked for layout and format. If a panelist is not present for a sensory day or sample, a "." should be entered into the cell, indicative of missing data. If a panelist evaluated an attribute and does not find it present, a numerical value of "0" should be entered into the cell if 0 = none or absence of the attribute. Summary statistics for mean, standard deviation or standard error, minimum, and maximum values should be calculated either within the spreadsheet or by a data analysis program. Summary statistics can be useful in conducting diagnostics to ensure that data are within range and that gross errors in data entry are found.

It is assumed that researchers have experience using the statistical package for data analyses of their choice. Many statistical packages are available for use, such as SAS (v 9.3), SPSS, R, XLSTAT, and others. Resources to understand programming language should be used to implement the data analyses and will not be covered here.

After assurance that the data are correctly entered, tests for normality are conducted. Consumer data are commonly abnormal in distribution, but all data should be tested for normality because normality is an assumption when conducting ANOVA. Data can be analyzed with nonnormal assumption procedures like PROC GLIMMIX of SAS (v 9.3) that generalize the MIXED procedure. The GLIMMIX procedure generalizes the MIXED and GENNOD procedures in two important ways. First, the response can have a nonnormal distribution. The MIXED procedure assumes that the response is normally (Gaussian) distributed. Second, the GLIMMIX procedure incorporates random effects in the model and thus allows for subject-specific (conditional) and population-averaged (marginal) inference. If data are not normally distributed and the data are analyzed using a model that does not account for nonnormal distributions, the data should be tested before analysis. Normality of data can be tested in many ways. Some statisticians recommend plotting frequency distributions of the data and examining the shape of the curve. If the curve is bell shaped, the data are normally distributed. This method, however, is not highly accurate. Most statistical programs have methods to test for normality. In SAS (v 9.3), the Proc Univariate function performs many descriptive statistics for normality such as Q-Q plot, stem-and-leaf plot, box plot, and normal-probability plot. This procedure also conducts Kolmogorov-Smirnov, Shapiro-Wilk, Anderson-Darling, and Cramer-von Misers tests to evaluate normality. The Box-Cox transformations in SAS (v. 9.3) can be used to determine logarithmic transformation constants that can be used to convert the data to a normal distribution. Other transformations also might be appropriate. If data are transformed—for ease of interpretation—least squares or unadjusted means should be transformed back to the original scale but LSD values from the original scale cannot be used. The researcher also should check to assure that the back-transformed means are not substantially lower than the means in the original data. This can be an issue when the transformation is logarithmic.

[1] All SAS versions refer to the SAS Institute in Cary, North Carolina.

A second way to test for normality is to plot the residuals of the data. It is suggested, however, that if there are missing observations, a "means" statement not be used and the researcher should use least squares means to compare treatment means. In sensory data, it is not uncommon to have a panelist missing from a day of evaluation; in that case, least squares means should be used. Diagnostics are run to ensure that data are in compliance with the assumptions of an ANOVA; i.e., model errors are normally and independently distributed random variables with mean = 0 and variance = σ^2, variance constant and homogeneously dispersed, and the measurements performed in a random order in a completely randomized design (Devore & Peck, 2005; Huntsberger & Billingsley, 1979; Steel & Torrie, 1980). To run diagnostics, the predicted and studentized residual values of the data need to be calculated and obtained from the ANOVA procedure. The predicted and studentized residual values can be plotted as in Figure 18. For example, the "gplot" procedure in SAS (v. 9.2) will plot the studentized residuals by the predicted observation values to determine whether or not the variance is constant and homogeneously dispersed. If the variance is constant and homogeneously dispersed, then the plot should look similar to the scatterplot in Figure 18 where there is no pattern to the residuals. Normality of the data can be assessed using the "gchart" procedure, and if the data are normally distributed, then the histogram should somewhat resemble a bell-shaped curve. It should be noted that the ANOVA is not very sensitive to departures from normality or unequal variance and is a relatively robust statistical test (Ott & Longnecker, 2001). Slight deviations from normality will most likely not affect analyses and interpretation of results. If there is concern, a statistician should be consulted.



(a) (b)

Figure 18. Example plot of residuals (a) to evaluate whether or not variance is constant and homogeneously dispersed, and histogram (b) to evaluate normality of the data.

Significance levels for analysis should be predetermined. While an α-error of < 0.05 is commonly used, in a small data set, an α of < 0.10 could be justified, whereas in a large data set, an α of < 0.01 might be more appropriate. It is important to predetermine significance levels based on experimental design and observation values before initiation of the study to prevent potential biases when interpreting sensory data. In addition, α-error should always reflect the level of confidence that the researchers will require to make inferences before any data are collected or analyzed.

B. Panelist Effects

Trained sensory data are generated using multiple panelists who evaluate a subsample from an experimental unit. Extensive training, performance evaluation, and validation of trained panelists are conducted prior to managing a study. If the procedures defined in Section VII are followed, the sensory professional has assurance that panelists are consistently and uniformly evaluating samples. During testing, however, factors can influence sensory verdicts and data should be evaluated for panelist effects before final analyses. It should be understood that the more training and experience a panel has with a product, the more sensitive the panel. To test panelist effects and panelist interactions with treatments, an experimental unit is defined as an individual panelist's response to a sample. For example, eight panelists independently, in separate booths, evaluated a steak from an animal and 20 animals are tested during a four-day sensory test, resulting in 160 observations. An ANOVA is conducted with sensory day, treatment, panelist, and panelist by treatment as effects. In a larger study, order would be included as a random effect, or if order is balanced for the study (each sample is included in each order equally across sensory days), it is included as a block. In the ANOVA table, the Type III sums of squares for panelist and panelist by treatment(s) interaction are examined. A significant panelist by treatment interaction, using the predetermined level of α-error, indicates that panelists did not evaluate treatments the same. Examination of the interaction is warranted, and interpretation by the panel leader is needed prior to averaging across panelists. The panel leader determines if the data are presented as an interaction or as treatment effects.

If panelist and panelist by treatment interactions are included in a model and evaluated, there can be significant differences that are meaningless or meaningful. It is up to the panel leader and researchers to understand these effects and their interpretation. When an interaction is significant, the first step is to reevaluate data accuracy. If a decimal point is improperly placed or a variable incorrectly entered in a small study, an interaction could result. It should be noted that it is not unusual for panelist effects to be significant. This does not necessarily mean that the panelists or some of the panelists are not doing a good job in evaluating the products. If the panelists have been trained to detect a one-point difference and the root mean square error is 0.50, then 0.5 differences might be significant.

When examining the least squares means between panelists, there might be only a 0.75 difference between the panelists scoring the lowest and highest within an attribute. In this case, the sensory panel leader would not be concerned about panelist effect. On the other hand, if, when examining the least squares means, one panelist is scoring two to three points different from the remainder of the panelists, additional training is warranted. If a significant panelist by treatment interaction is determined to be meaningful, the data should be presented as an

interaction and not averaged across panelists. Theoretically, if the sensory panel leader has adequately tested the panelists' readiness through performance evaluation and panelists conduct warm-up samples daily prior to testing for calibration, panelist by treatment interactions should be minimized. If, during testing, a panelist rates samples within a treatment differently from other panelists, averaging across panelists might result in misinterpretation of data. This also might be a result of variation within subsamples, and this information might be important in the interpretation of the data.

If a panelist is erratically evaluating treatments and is obviously different than other panelists, there might be sufficient justification for removing their data from the data set. Careful consideration should be given to the decision to remove panelists because if the panel leader has sufficiently trained panelists, panelists passed performance evaluation, and appropriate and consistent warm-up samples are used to calibrate panelists on a daily basis, it should not be automatically assumed that the panelist is rating samples inadequately. A panelist that is not evaluating the samples within the range or levels of other panelists, however, should be targeted for additional training. The panel leader needs to carefully evaluate this panelist's performance and determine whether or not to keep their data.

C. Analysis Options

1. Univariate Analyses

For trained and consumer data, an initial model that includes blocking effects such as sensory day, session, order and treatment effects, and subsequent interactions should be developed to reduce risk of encountering β-error. Effects that are random should be identified and potential covariates included in the model if they also will help to explain error. If researchers desire to include a random variable in the model to control for additional variation, and to make inference to the population from which the experimental units were derived, it is wise to consult with a qualified statistician to ensure that the analysis is being conducted correctly.

Examination of the full model ANOVA should be conducted. Interaction effects that are not significant can be removed. There is not a uniform consensus on when to remove or pool interaction effects into the error term. If the p-value for the interaction effect is large (greater than 0.50), it is obvious that the interaction is not partitioning variation from the model and removing this interaction will not affect the outcome of the analysis. Depending on the sample size and the p-value for effects of interest (treatment and treatment interactions that are close to being significant), however, removing interactions that have p-values between 0.05 and 0.25 can affect the level of significance of the effects of interest. When that occurs, the ANOVA tables with and without the interaction should be conducted and compared. If removing interactions with p-values greater than 0.5 affects the significance so that they are now significant, and they were not significant in the original model, the researchers must decide what to include in the final model. The conservative approach is to include interactions with p-values less than 0.25 and exclude interactions with p-values greater than 0.25. Researchers must understand how significance of effects of interest is affected in both models for correct interpretation of results. In studies with large numbers of observations, it is likely that removal of interactions that have p-values greater than 0.25 will not appreciably affect the significance level of tests of hypothesis for effects of interest. Covariates that are not significant can be excluded from the full model;

however, covariates that account for some variation (p-values greater than 0.05 and less than 0.25) can be included in the model to explain error, but they are most likely not affecting the experimental outcome. After consideration of p-values for all effects defined in the full model, a final model is defined and the ANOVA is calculated. Least squares means are usually calculated for consumer and trained sensory data because it is not uncommon to have missing or unbalanced subcell numbers.

If the final model includes the effects of panelist and panelist interactions as discussed above, treatment effects that might be included in the model but are not part of the panelists by treatment interaction may need to be tested using a different error term than the model residual error. If you are unsure of the proper error term for testing specific effects, consult a statistician. Many statistical packages allow for the definition of error terms for individual effects included in the model and may provide methods to determine the correct error term to use.

In the final model, significant main effects and interactions should be identified. Effects that are not significant can be as important as effects that are significant with regard to conclusions. If a significant interaction occurs upon analysis of multiple factors in a model (excluding block effects), then only the interaction effects should be presented and interpreted. For confirmation that the proper analysis was used, consult a statistician.

When analyzing consumer data, it is also important to look at the percentage of "Like Extremely" and "Like Very Much" ratings (top two boxes in a 9-point hedonic scale) as well as "Dislike Extremely" and "Dislike Very Much" ratings (bottom two boxes) to check for shifts in rating patterns among treatments. It is possible for mean scores to not be significantly different but have shifts in top two/bottom two box ratings, indicating consumer segmentation or polarized ratings.

2. Multivariate Analyses

As discussed, sensory data are by nature multivariate. When a human evaluates a sample, activated sensory receptors send information to the brain. Trained sensory panelists are proficient in either ignoring information (i.e., ignore information that would be considered on visual appearance and its relationship to flavor), segment or pull out a specific attribute from the sensory stimulus (i.e., juiciness is evaluated independently of muscle fiber tenderness or beef flavor identity is measured independently of cooked beef fat), and quantify it. Consumers might rate their overall liking of a sample and then be asked to evaluate specific attributes, such as their overall liking or intensity of juiciness and/or tenderness. Univariate analysis of independent attributes provides for interpretation of treatment effects. Key attributes that differ by treatment provide researchers the ability to determine if treatments differ or not. Whereas, multivariate analyses provide the opportunity for understanding how multiple variables are impacted by treatments and how consumer and trained sensory data might be related. These can be very powerful tools in understanding if treatments affect the sensory properties of meat samples. Multivariate analyses tools provide additional information. If the data set does not have relevant information, such as differences in ANOVA and variation is very low, however, multivariate techniques are not a way to gain information.

In using multivariate techniques, the number of observations is not as important as the strength of the relationship. For example, if a quantitative attribute used in trained panel analysis, such as cooked beef flavor identity, does not have a quantitative relationship in univariate analysis, then these data will not be useful in multivariate analyses. There are useful resources that discuss multivariate analysis—Meilgaard et al. (2007) and Lawless and Heymann (2010). Many of the statistical packages have training that can also assist in understanding how to conduct these analyses. It is assumed that individuals using this guideline are familiar with these or comparable resources. A general description of common multivariate tools that can be used in sensory data, how they are used, what type of information they provide, and general interpretation of data are discussed below. A data set was extracted from Glascock (2014). These data include beef samples (16 treatments) that were selected to differ in flavor due to Quality grade, cut, cooking method, and cooked internal cook temperature endpoint. These samples were evaluated by a trained descriptive flavor attribute panel using the Beef Lexicon, by consumer sensory evaluation in four different cities (n=80 per city), and for volatile flavor chemical analyses. Not all of the data or attributes were included in these analyses, but the data were used to illustrate how to use multivariate techniques to understand relationships within these data.

a. Principal component analysis

Principal component analysis (PCA) assists in understanding relationships among variables. Like correlation analysis, PCA studies the orthogonal relationships among variables in a data set, but it allows the visualization of how variables move in relation to one another. In PCA, orthogonal variation in the data set is explained by combining variables that are related into a set of new variables, called factors, which are uncorrelated and orthogonal to one another. Factors also are often called principal components. The first factor (Factor 1) is derived to account for the greatest amount of variation; subsequent factors account for progressively less variation. The number of factors that can be derived is equal to n - 1, where n equals the number of variables in the data set. Generally, only the first few factors explain significant portions of the variation, and, thus, only the first few are scrutinized. The percent of the variation that is accounted for by each factor is provided in the analysis. A bi-plot is generated that shows the relationships between factors, usually Factors 1 and 2, and the sensory variables and treatments.

Figure 19 is a bi-plot generated from the aforementioned data (Glascock, 2014). The objective is to understand relationships between beef flavor attributes and the 16 treatments. The original data (n=640) was averaged across panelist and treatment so that there was16 lines of data used for PCA analysis. The first factor or Principal Component 1 accounted for 44.53% of the variation and the second factor or Principal Component 2 accounted for 20.96% of the variation. The third component and subsequent components were not significant and were not presented. To interpret the bi-plot, attributes and treatments are plotted. If attributes are in the center of the plot, they are not contributing to variation in the two principal components. Attributes or treatments that are in a similar area are highly related. Attributes or treatments near the 0 on the vertical axis, but located away, either negatively or positively, from the horizontal axis are accounting for more of the variation in Principal Component 2. Conversely, attributes or treatments near the 0 on the horizontal axis, but located away, either negatively or positively, from the vertical axis are accounting for more of the variation in Principal Component 1.

Variables in the four quadrants between axes are accounting for some of the variation, either negatively or positively, for both principal components. Variables used in the PCA should be measured using the same scale; these data were collected using a 16-point Spectrum Universal scale. To interpret the bi-plot, treatments clustered with variables are related. Therefore, Choice and Select bottom round roasts cooked in a crockpot to 176°F and top sirloin steaks grilled to 176°F had similar umami and beef identity. Liver-like flavor was related to Choice top loin steaks cooked on a George Foremen grill to an internal cook temperature of 176°F, whereas, metallic flavor was most closely related to top sirloin steaks cooked to 137°F on either a George Foreman or grill. Other relationships are apparent but will not be discussed. Data where treatment may be represented by different products across companies could be analyzed for either trained or consumer sensory attributes to understand how these products differ in either trained descriptive sensory attributes or consumer attributes.

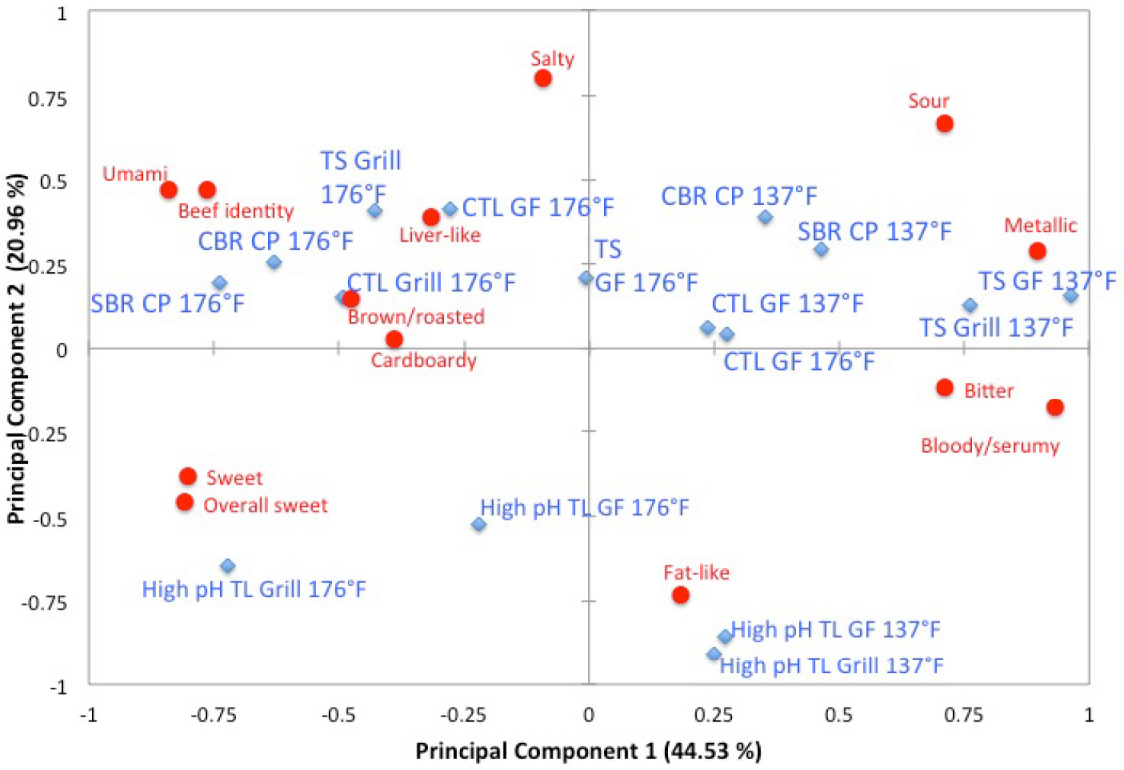


Figure 19. Principal component bi-plot of trained, beef descriptive flavor attributes from the Beef Flavor Lexicon (red) and 16 treatments (blue) where TL = top loin steaks, GF = George Foreman grill, CTL = Choice top loin steaks; SBR = Select bottom round roasts, CP = crockpot cooked, CBR = Choice bottom round roasts, and TS = top sirloin steaks.

b. Partial least squares regression

Partial least squares regression is a one-step iterative process that computes latent vectors to explain both independent and dependent variables. This tool can be used when more than one dependent variable is being predicted or when independent variables are measured using different scales. For example, if trained and consumer sensory responses were obtained on a set of samples, the trained sensory responses could be defined as the independent variables and the consumer responses defined as the dependent variables (variable being predicted so that a trained panel could be used to measure a product and understand how consumer responses would be affected). A bi-plot is shown in Figure 20 that was calculated from consumer and trained descriptive attribute sensory data for the same data presented for PCA analysis. Note that this bi-plot presents the relationships between consumer responses and trained descriptive sensory flavor attributes adapted from Glascock (2014).

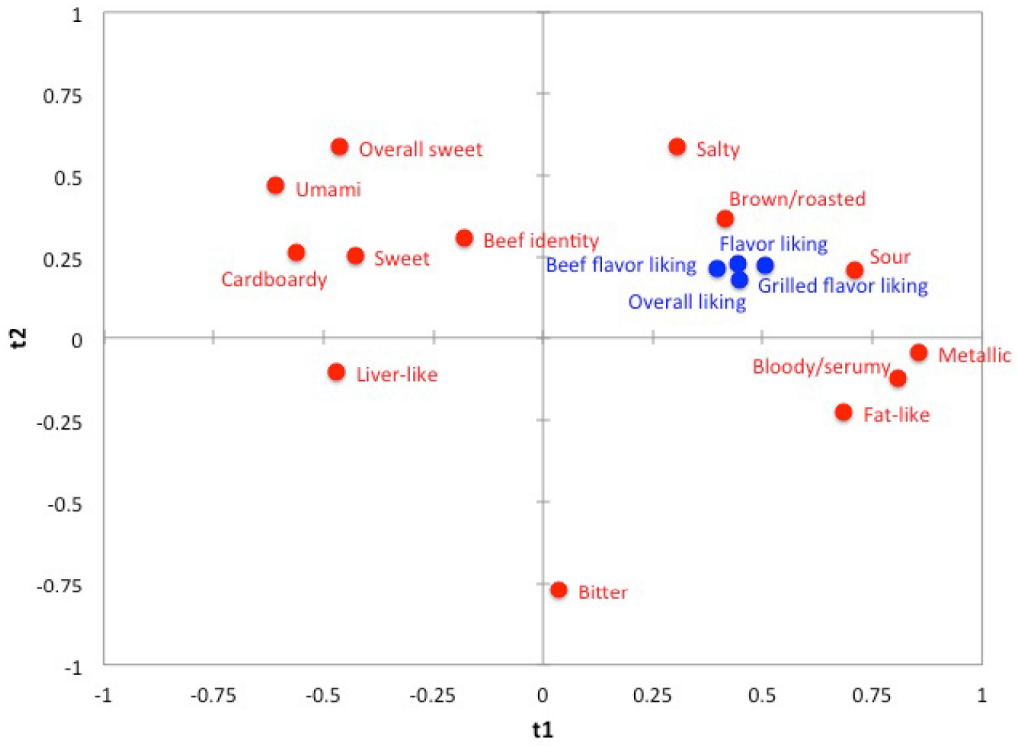


Figure 20. Partial least squares regression bi-plot for trained and consumer sensory data for beef adapted from Glascock (2014) where trained descriptive panel attributes are defined in red and consumer sensory attributes are defined in blue.

These data assist the researcher in understanding what trained panel attributes to target for evaluation or that are most likely related to overall consumer liking. Once a relationship between consumer responses and trained panel responses is established, trained panel evaluation can be used to evaluate additional products at a lower cost. In this analysis, the experimental unit is the sample evaluated by each consumer and the trained descriptive

panelist; a higher n is used to understand these relationships (n=451 for the same data presented in Fig. 19). The variation defined by the model is defined as 20.2% of X explains 45.5% of the variability in Y for the first two dimensions. The type of information obtained from PLS bi-plots examine variables that cluster. For example, from the bi-plot presented in Figure 20, it is apparent that consumer attributes of grilled flavor, flavor, and beef flavor liking were highly related to overall liking. Additionally, liver-like and bitter were not related to overall consumer liking. Brown/roasted tended to be most closely related to overall consumer liking and sour basic taste while closely related to the consumer liking attributes was most closely related to grilled flavor liking. The grouping of attributes provides an understanding of interrelationships within trained and consumer attributes and between these attributes. Different data can be used in PLS analyses, such as understanding chemical measures and either consumer or trained panel descriptive attributes.

c. Preference mapping

This statistical tool links consumer acceptance to trained sensory attributes and has become extensively used in industry. It is a powerful statistical tool to answer the question of how trained panel results are related to consumer overall like ratings. Preference mapping provides a rapid, objective, and reproducible tool for making product decisions. There are two main preference mapping tools: internal preference mapping and external preference mapping. Each tool will be discussed to provide information on how to use these tools and how to interpret these analyses.

(1) INTERNAL PREFERENCE MAPPING

Internal preference mapping uses hedonic data and uses a variation of principal component analysis. The data is converted so that it is orthogonal. On the bi-plot, instead of variables from trained panelists, consumer data and treatments are plotted. This method assists the researcher in understanding the relationship between consumer clusters or groups and overall like and/or treatments. If consumers are clustered with a treatment, they have a preference for that treatment. An example of an IPM bi-plot is presented in Figure 21. Note that Principal Component 1 accounted for 88.79% of variation and Principal Component 2 accounted for 6.70% of variation. These data are for the same data set used for Figures 19 and 20, except the data is presented where consumer response is an experimental unit. Consumer clusters show that not all consumers have the same preferences. In Figure 21, there are four general consumer clusters—one in each quadrant. Consumers in the upper right quadrant have a stronger preference for grilled flavor liking whereas consumers in the left quadrants where furthest away from overall liking. The sensory professional could use these data to further understand why consumers segmented as defined by pulling these consumers from the data set and further examining their responses and/or other information obtained from these consumers such as demographic data, one-on-one interview data, etc.

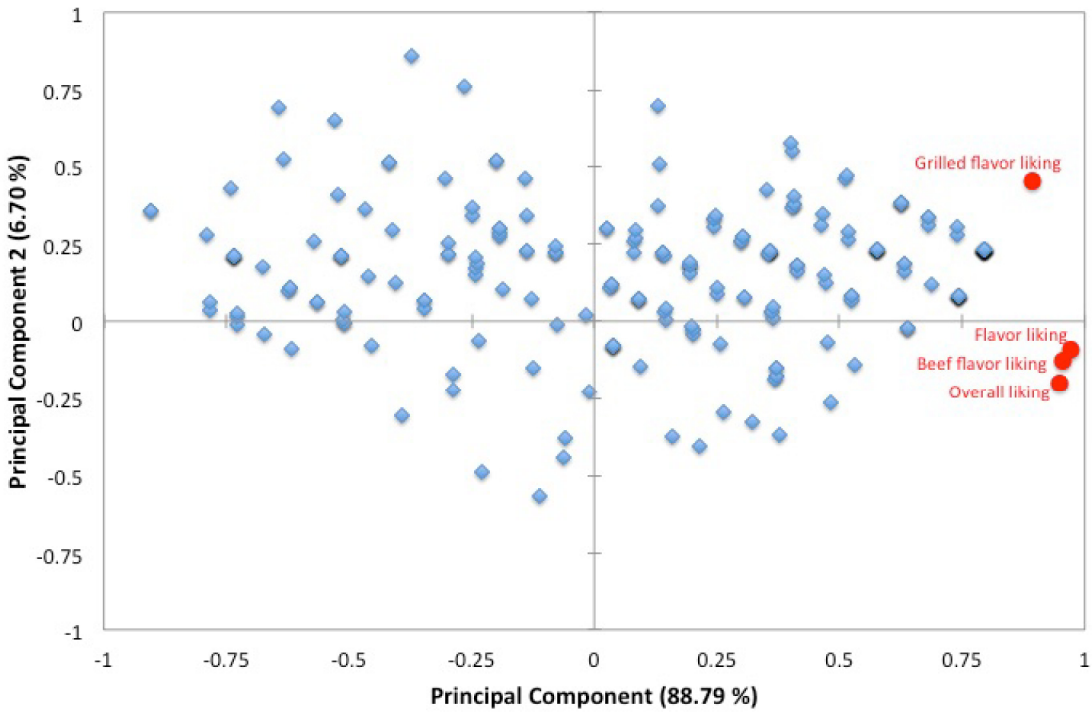


Figure 21. Internal preference mapping bi-plot from consumer overall liking attributes in red and individual consumers in blue adapted from Glascock (2014).

(2) EXTERNAL PREFERENCE MAPPING

External preference mapping (EPM) is used for product optimization and in product development. This tool correlates consumer preference data and trained sensory attribute and/or chemical data. Information from EPM assists the researcher in understanding what trained sensory attributes or chemical characteristics drive or are highly related to consumer likes and dislikes. The bi-plot is generated using PLS as the data was measured using different scales. Also note that the data defined consumer is an experimental unit. While the same data set was used as discussed in Figures 19, 20 and 21, fewer data points were used for simplicity. Attributes are mapped in relationship with consumer overall like ratings (Figure 22). These data show that bloody/serumy was most closely related to consumer overall liking. However, pentanal, hexanal, bitter and umami were negatively related to consumer overall liking. Consumers from two different quadrants could be identified and further information on what was driving overall like positively and negatively could be studies. These data provide direction for products that are higher in the positive and lower in the negative flavor attributes, and help in understanding volatile chemical compounds that may be driving flavor.

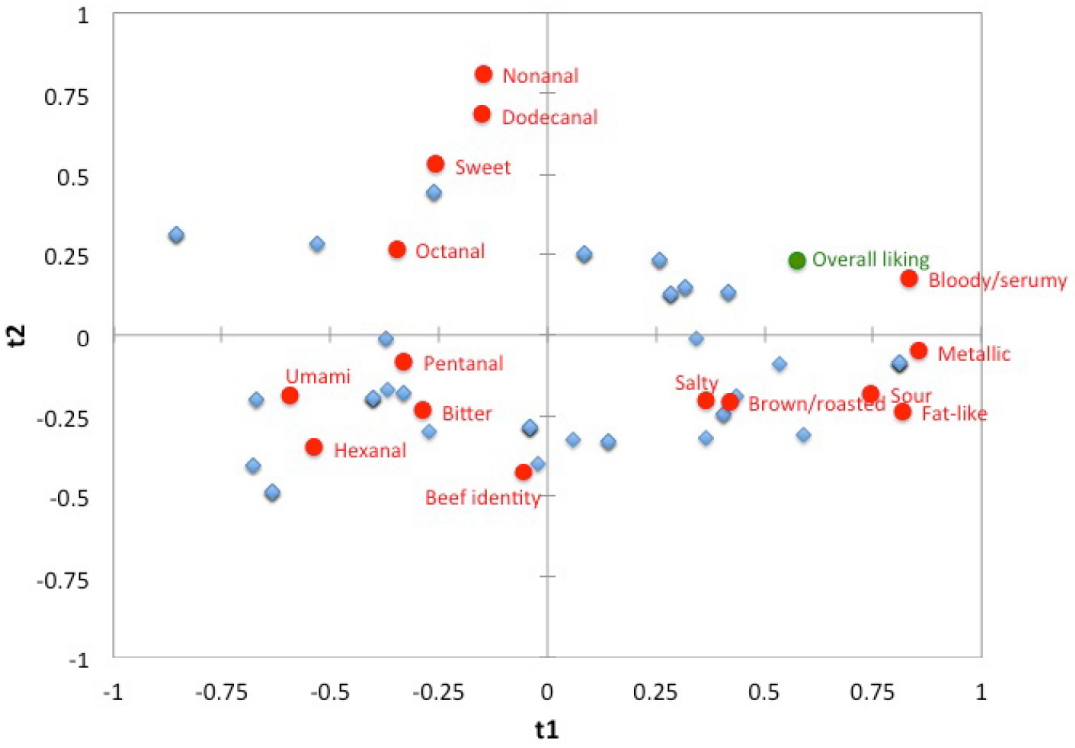


Figure 22. External preference mapping bi-plot from consumer overall liking in green, trained descriptive flavor attributes and volatile flavor chemicals in red, and consumers in blue adapted from Glascock (2014).

X. LITERATURE CITED

Adhikari, K., Chambers IV, E., Miller, R., Vazquez-Araujo, L., Bhumiratana, N., and Philip, C. (2011). Development of a lexicon for beef flavor in intact muscle. *Journal of Sensory Studies*, *26*, 413-420.

AMSA (1978). *Guidelines for cookery and sensory evaluation of meat*. Chicago, IL: American Meat Science Association.

AMSA (1983). *Guidelines for sensory, physical and chemical measurements in ground beef*. Chicago, IL: American Meat Science Association and National Live Stock and Meat Board.

AMSA (1995). *Research guidelines for cookery, sensory evaluation and instrumental tenderness measurements of fresh meat*. Chicago, IL: American Meat Science Association.

ASTM DS72 (2011). *Lexicon for sensory evaluation: aroma, flavor, texture and appearance*. West Conshohocken, PA: ASTM International.

ASTM E4 (2010). *Standard practice for force verification of testing machines*. West Conshohocken, PA: ASTM International.

ASTM E1871 (2010). *Standard guide for serving protocol for sensory evaluation of foods and beverages*. West Conshohocken, PA: ASTM International.

ASTM E1885-04 (2004). *Standard test method for sensory analysis—triangle test*. West Conshohocken, PA: ASTM International.

ASTM E1958-07e1 (2012). *Standard guide for sensory claim substantiation*. West Conshohocken, PA: ASTM International.

ASTM E2299-11 (2011). *Standard guide for sensory evaluation of projects by children*. West Conshohocken, PA: ASTM International.

ASTM E2610-08 (2008). *Standard test method for sensory analysis—Duo-Trio test*. West Conshohocken, PA: ASTM International.

ASTM F2341 (2005). *Standard practice for user requirements for livestock, meat, and poultry evaluation devices or systems*. West Conshohocken, PA: ASTM International.

ASTM F2343 (2006). *Standard test method for livestock, meat, and poultry evaluation devices*. West Conshohocken, PA: ASTM International.

ASTM MNL13-EB (1992). *Manual on descriptive analysis testing for sensory evaluation*. In R. C. Hootman (Ed.). West Conshohocken, PA: ASTM International.

ASTM MNL26-2nd (1996). *Sensory testing methods*.2nd Edition. In E. Chambers & M. B. Wolf (Eds.). West Conshohocken, PA: ASTM International.

ASTM MNL60-2nd (2008). *Physical requirement guidelines for sensory testing laboratories.* 2nd Edition. West Conshohocken, PA: ASTM International.

ASTM MNL63 (2009). *Just-about-right (JAR) scales: design, usage, benefits, and risks*. L. Rothman & M. Parker (Eds.). West Conshohocken, PA: ASTM International.

ASTM STP434 (1968). *Manual on sensory testing methods*. Committee E18. West Conshohocken, PA: ASTM International.

ASTM STP758 (1981). *Guidelines for the selection and training of sensory panel members*. West Conshohocken, PA: ASTM International.

Berry, B. W., & Dikeman, M. E. (1994). AMSA cookery and sensory guidelines. *Proceedings of the Reciprocal Meat Conference, 47*, 51-52.

Berry, B. W., & Leddy, K. F. (1990). Influence of steak temperature at the beginning of broiling on palatability, shear and cooking properties of beef loin steaks differing in marbling. *Journal of Foodservice Systems, 5*, 287.

Bohnenkamp, J. S., & Berry, B. W. (1987). Effect of sample numbers on sensory activity of a trained ground beef texture panel. *Journal of Sensory Studies*, *2*, 23.

Bourne, M. C. (1978). Texture profile analysis. *Food Technology*, *7*, 62.

Bouton, P. E., Ford, A. L., Harris, P. V., Shorthose, W. R., Ratcliff, D., & Morgan, J. H. L. (1978). Influence of animal age on the tenderness of beef: muscle differences. *Meat Science*, *2*, 301-311.

Bradley, R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics*, *9*, 22.

Chu, S.K. 2015. Development of an intact muscle pork flavor lexicon. MS Thesis. Texas A&M University, College Station.

Cochran, W. G., & Cox, G. M. (1957). *Experimental design*. 2nd Edition. New York, NY: John Wiley and Sons.

Cross, H. R., Moen, R., & Stanfield, M. S. (1978). Training and testing of judges for sensory analysis of meat quality. *Food Technology*, *32*, 48.

Cross, H. R., Stanfield, M. S., Elder, R. S., & Smith, G. C. (1979). A comparison of roasting versus broiling on the sensory characteristics of beef longissimus steaks. *Journal of Food Science, 44*, 310.

Devore, J., & Peck, R. (2005). *Statistics: the exploration and analysis of data*. 5th Edition. In C. Crockett (Eds.), *Summarizing Bivariate Data* (pp. 190-191). Belmont, CA: C. Crockett, ed. Brooks/Cole–Thompson Learning, Inc.

FSIS (2013). Food safety education. Retrieved from http://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/appliances-and-thermometers/kitchen-thermometers/ct_index/!ut/p/a1/jZBNboNADIXP0gOMxpQUNcsIqQqkgFCUIswmssDAqDCDGCtpc_rSdtUof_bK9vf0rCeVLKQyuNcNsrYGu59ZBTvIIfDmIcTZ3HuBKH3Ls1UYwvP6aQK2V4DUv1N_oRZwSx_fYfA4JmHSSDUgt0Kb2sqiIRZo3IFGJ4va2ko4rIm_RI0IC9cS8XTAYeg0mpLcBFeCWxp72xP_qj40Iy2Zk23JO20q-pTvUv1_Dbypo9Rfz5Zx6kM2OwXOZPcHXA5n6DfF8XWxBB09fAMZVa3b/#12

Gardner, G. A., Nelson, J. L., Dolezal, H. G., Morgan, J. B., & Novotny, K. K. (1996). Effects of cooking method, endpoint temperature, and quality grade on tenderness and cooking traits of Holstein beef strip steaks. *Journal of Animal Science*, *74*(Suppl. 1), 6.

Glascock, R. A. (2014). *Beef flavor attributes and consumer perception*. (Master's thesis, Texas A&M University, College Station).

Grayson, A. L., King, D. A., Shackelford, S. D., Koohmaraie, M., and Wheeler, T. L. (2014). The effects of freezing and thawing or freezing, thawing, and aging on beef tenderness. *Journal of Animal Science* (submitted January 2014).

Harris, P. V., & Shorthose, W. R. (1988). Meat texture. In R. A. Lawrie (Ed.), *Developments in meat science—4* (pp. 245-286). London, England: Elsevier Applied Science Publishers.

Hiner, R. L., Madsen, L. L., & Hankins, O. G. (1945). Histological characteristics, tenderness and drip losses of beef in relation to temperature of freezing. *Journal of Food Science*, *10*, 312-324.

Hough, G., Wakeling, I., Mucci, A., Chambers IV, E., Gallardo, I. M., & Alves, L. R. (2006). Number of consumers necessary for sensory acceptability tests. *Food Quality and Preferences*, *17*, 522-526.

Huntsberger, D. V., & Billingsley, P. (1979). *Elements of statistical inference*. 4th Edition. In *Analysis of variance* (pp. 314-316). Boston, MA: Allyn and Bacon.

IFT—Sensory Evaluation Division (1995). Guidelines for the preparation and review of papers reporting sensory evaluation data. *Journal of Food Science*, *60*, 211.

King, D. A., Wheeler, T. L., Shackelford, S. D., & Koohmaraie, M. (2009). Comparison of palatability characteristics of beef gluteus medius and triceps brachii muscles. *Journal of Animal Science*, *87*, 275-284.

Kroll, B. J. (1990). Evaluating rating scales for sensory testing with children. *Food Technology 44*(11), 78-86.

Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: principles and practices*. 2nd Edition. New York, NY: Springer Science+Business Media.

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (2007). *Sensory evaluation techniques*. 4th Edition. Boca Raton, FL: CRC Press.

Miller, R. K. (2013). Personal communication. College Station, TX: Meat Science Section, Department of Animal Science, Texas A&M University.

Munoz, A .M., & Civille, G. V. (1998.) Universal, product and attribute scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies*, *13*(1), 57-75.

Munoz, A. M, Civille, G. V., & Carr, B. T. (1992). *Sensory evaluation in quality control*. New York, NY: Van Nostrand Reinhold.

NAMI (2015). *Meat buyers guide*. Washington, DC. North American Meat Institute.

NCBA (2002). *National Cattlemen's Beef Association meeting summary: National Beef Instrument Assessment Plan II: focus on tenderness*. Centennial, CO: Funded by The Beef Checkoff.

NCBA (2009). Fact sheet. Slice shear force. Retrieved from http://www.beefresearch.org/CMDocs/BeefResearch/Slice%20Shear%20Force.pdf

NIH (1979). *The Belmont report. Ethical principles and guidelines for the protection of human subjects of research*. Retrieved from http://ohsr.od.nih.gov/guidelines/belmont.html.

O'Mahony, M. (1986). *Sensory evaluation of food: statistical methods and procedures*. New York, NY: Marcel Dekker, Inc.

Ott, R. L., & Longnecker, M. (2001). An introduction to statistical methods and data analysis. 5th Edition. In C. Crockett (Ed.), *Inferences about more than two population central values*. Pacific Grove, CA: Duxbury—Thompson Learning, Inc.

Rhee, M. S., Wheeler, T. L., Shackelford, S. D., & Koohmaraie, M. (2004). Variation in palatability and biochemical traits within and among eleven beef muscles. Journal of Animal Science, 82, 534-550.

SAS Inst. (2004). SAS Online Doc® 9.1.3, Cary, NC:SAS Institute, Inc.

Shackelford, S. D., King, D. A., & Wheeler, T. L. (2014). Evaluation of multiple options for prediction of tenderness of several beef muscles based on visible and near infrared spectroscopy. Journal of Animal Science, 92,

Shackelford, S. D., Morgan, J. B., Cross, H. R., & Savell, J. W. (1991). Identification of threshold levels for Warner-Bratzler shear force in beef top loin steaks. *Journal of Muscle Foods*, *2*, 289-296.

Shackelford, S. D., Wheeler, T. L., & Koohmaraie, M. (1995). Relationship between shear force and trained sensory panel tenderness ratings of 10 major muscles from *Bos indicus* and *Bos taurus* cattle. *Journal of Animal Science, 73*, 3333-3340.

Shackelford, S. D., Wheeler, T. L., & Koohmaraie, M. (1999a). Tenderness classification of beef. II. Design and analysis of a system to measure beef longissimus shear force under commercial processing conditions. Journal of Animal Science, 77, 1474-1481.

Shackelford, S. D., Wheeler, T. L., & Koohmaraie, M. (1999b). Evaluation of slice shear force as an objective method of assessing beef longissimus tenderness. Journal of Animal Science, 77, 2693-2699.

Shackelford, S. D., Wheeler, T. L., & Koohmaraie, M. (2004a). Technical note: use of belt grill cookery and slice shear force for assessment of pork longissimus tenderness. Journal of Animal Science, 82, 238-241.

Shackelford, S. D., Wheeler, T. L., & Koohmaraie, M. (2004b). Evaluation of sampling, cookery, and shear force protocols for objective evaluation of lamb longissimus tenderness. Journal of Animal Science, 82, 802-807.

Smith, A. M. (2012). Dry versus wet aging of beef: retail cutting yields and palatability evaluations of steaks using alternative cutting styles. (Master's thesis, Texas A&M University, College Station). Retrieved from http : / /hdl .handle .net /1969 .1 /148309

Steel, R. G. D., & Torrie, J. H. (1980). Principles and procedures of statistics: a biometrical approach. 2nd Edition. In C. Napier & J. W. Maisel (Eds.), *Analysis of variance I: the one-way classification*. New York, NY: McGraw-Hill Book Company.

Stone, H., & Sidel, J. L. (2004). *Sensory evaluation practices*. 3rd Edition. San Diego, CA: Elsevier Academic Press.

USDA Livestock, Poultry and Seed Program (2012). Laboratory proficiency testing for shear force measurements. Washington, DC: USDA.

USDA (2010). IMPS 100 fresh beef. USDA Agricultural Marketing Service. Retrieved from http://www.ams.usda.gov/AMSv1.0/getfile?dDocName=STELDEV3003281

Wheeler, T. L., Koohmaraie, M., Cundiff, L. V., & Dikeman, M. E. (1994). Effects of cooking and shearing methodology on variation in Warner-Bratzler shear force values in beef. *Journal of Animal Science*, *72*, 2325-2330.

Wheeler, T. L., Shackelford, S. D., & Koohmaraie, M. (1996). Sampling, cooking, and coring effects on Warner-Bratzler shear force values in beef. *Journal of Animal Science*, *74*, 1553-1562.

Wheeler, T. L., Shackelford, S. D., & Koohmaraie, M. (1998). Cooking and palatability traits of beef longissimus steaks cooked with a belt grill or an open hearth electric broiler. *Journal of Animal Science*, *76*, 2805-2810.

Wheeler, T. L., Shackelford, S. D., & Koohmaraie, M. (2007). Beef longissimus slice shear force measurement among steak locations and institutions. *Journal of Animal Science*, *85*, 2283-2289.

Wheeler, T. L., Shackelford, S. D., Johnson, L. P., Miller, M. F., Miller, R. K., & Koohmaraie, M. (1997). A comparison of Warner-Bratzler shear force assessment within and among institutions. Journal of Animal Science, 75, 2423-2432.

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, *Ser. A 2*, 149-168.