National Library of Canada
Canadian Theses Service
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Services des thèses canadiennes

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

NL-339 (r.86/06)

Canada

THE UNIVERSITY OF ALBERTA

DESCRIPTIVE SENSORY ASSESSMENT OF BEEF STEAKS BY CATEGORY SCALING, LINE SCALING AND MAGNITUDE ESTIMATION

by

PHYLLIS JO-ANNE SHAND

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
FOODS

FACULTY OF HOME ECONOMICS
EDMONTON, ALBERTA
FALL, 1984

ISBN 0-315-30125-2

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Phyllis Jo-Anne Shand

TITLE OF THESIS: Descriptive sensory assessment of beef steaks by category scaling, line scaling and magnitude estimation.

DEGREE FOR WHICH THESIS WAS PRESENTED: Master of Science

YEAR THIS DEGREE GRANTED: 1984

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Phyllis Shand
(signed)

PERMANENT ADDRESS:
115-8604-103 Street
Edmonton, Alberta
T6E 4B6

DATED : September 25, 1984

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "**Descriptive Sensory Assessment of Beef Steaks by Category Scaling, Line Scaling and Magnitude Estimation**", submitted by **Phyllis Jo-Anne Shand** in partial fulfilment of the requirements for the degree of **Master of Science** in **Foods**.

Zenia J. Hawrysh
Supervisor

L.E. Jeremiah

Robert T. Hardin

Dated: September 25, 1984.

ABSTRACT

Eighteen trained panelists evaluated the eating quality of semimembranosus (SM) steaks using an eight-point category scale, a 15 cm line scale with anchors 1.3 cm from each end and modulus free magnitude estimation in different sequences. Comparison of the evaluation techniques focused on the descriptive assessment of a total of 54 delay-chilled (DC) and hot-boned (HB) cooked SM steaks.

When panelists employed category scaling (CS), the HB steaks were rated lower ($P<0.05$) in initial and overall tenderness and higher ($P<0.01$) in overall juiciness than DC steaks. For line scaling (LS), there were no significant differences in quality attributes due to postmortem treatment. With magnitude estimation (ME), the HB steaks were less tender ($P<0.05$) and more juicy ($P<0.01$) than DC steaks. For evaluations by each technique, connective tissue amount and flavor intensity did not differ significantly between the postmortem treatments.

Sensory assessments of tenderness and juiciness were generally supported by objective measurements. Warner Bratzler shear values tended to be higher for HB steaks than for DC steaks. In addition, the HB steaks were visually more ($P<0.01$) rare, had higher ($P<0.001$) Hunter *a* values and greater ($P<0.01$) percentage press fluid than DC steaks.

Treatment F-values showed that CS, with significant F-values for initial tenderness, overall tenderness and overall juiciness, was most sensitive to treatment differences. The LS technique, with no significant F-values, was least sensitive in detecting differences in steak quality attributes. Magnitude estimation, with significant F-values for overall tenderness and overall juiciness, was as sensitive as CS to most treatment differences.

Replication F-values for each of the three evaluation techniques were similar and nonsignificant for most quality attributes. However, for CS, LS and ME, significant (P<0.001) panelist variation was noted for each quality attribute. Except for the ME assessment of overall juiciness, F-values for the treatment by panelist interaction were not significant.

Correlation coefficients between each of the evaluation procedures for initial and overall tenderness were highly significant (P<0.001) and indicated strong linear relationships between scores assigned by judges using CS, LS and ME. The correlations between each of the three techniques for connective tissue amount, juiciness and flavor intensity, while sometimes significant, were moderate to low.

Correlations of tenderness data from each evaluation method with shear values yielded similar significant values ($r<0.63$). Coefficients of determination for linear, power, logarithmic, hyperbolic and parabolic functions between panel tenderness assessments by each technique and Warner Bratzler shear were similar. For initial and overall tenderness, the power functions between ME and shear data yielded exponents of 1.13 and 1.05, respectively.

Panelists found that the CS and LS methods were easier to learn and required less effort for sample evaluation than ME. In addition, CS was most preferred and ME was least preferred by panelists.

In general, findings from this study support the use of category scaling for the descriptive sensory assessment of beef. To optimize the potential of the LS and ME methods for assessments of cooked meat, some procedural modifications may be needed.

ACKNOWLEDGMENTS

The author wishes to express her deepest appreciation to her advisor, Dr. Zenia Hawrysh, for her guidance, encouragement, support and generous contribution of time and energy throughout the project and during the preparation of this manuscript.

Sincere gratitude is also extended to the taste panel members for their dedicated participation. The author also wishes to thank Dr. R.T. Hardin, Chairman of the Department of Animal Science, Mr. R. Weingardt and Mrs. M. Peebles for assistance in statistical treatment of the data; Miss S. Gifford and Miss I. Gordon for technical assistance; and Dr. L.E. Jeremiah, Meat Scientist, Canada Department of Agriculture, for donating the meat for this study.

Financial support provided by the Natural Sciences and Engineering Research Council of Canada and the Agricultural Research Council of Alberta Farming for the Future Program is also gratefully acknowledged.

Finally, the author is indebted to her family and friends for their patience, understanding and support during this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Most studies concerned with factors influencing meat quality (such as breeding, management or processing treatments) utilize sensory measurements to determine the organoleptic quality of cooked meat and to make important decisions regarding all aspects of beef production, processing and marketing. Therefore, it is essential that the most reliable and accurate sensory technique be used to evaluate cooked meat quality.

The American Meat Science Association (AMSA, 1978) has suggested that category scales are the most appropriate system for evaluating the sensory attributes of meat. Semistructured line scales, which form the basis for "Quantitative Descriptive Analysis", have been used by several meat researchers (Harries et al., 1972; Bouton et al., 1980a) and have been reported to be more sensitive to product differences than structured scales (Baten, 1946). However, information on the sensitivity of category and line scales in determining differences in meat quality is limited.

Magnitude estimation, a ratio-scaling technique, has become popular for food research (Moskowitz, 1983) but has not been widely utilized with meat (Larmond, 1976). Magnitude estimation can provide quantitative information on the perception of meat tenderness, which is not possible with either category or line scales. Since instrumental measurements are calibrated in ratios, it may be advantageous to also have sensory measures on a ratio scale (Larmond, 1976). Thus, the feasibility of measuring an attribute of meat such as tenderness using magnitude estimation requires investigation.

Comparative studies examining the use of category scaling, line scaling and magnitude estimation in meat research are lacking. Therefore, this investigation was designed to compare the sensitivity and accuracy of these evaluation techniques for the descriptive sensory assessment of beef. In addition, statistical relationships between the three evaluation procedures and ease of scale usage were examined.

LITERATURE REVIEW

Category Scaling

Category scaling (CS) is routinely employed for the descriptive assessment of meat (AMSA, 1978). For the CS technique, panelists rate the intensity of a given attribute on a graduated scale, usually seven to nine categories in length.

Procedures for the assessment of meat by CS range from an elaborate system (Cover et al., 1962a, 1962b, 1962c) for quantifying meat quality in terms of softness, juiciness, connective tissue and three muscle fiber attributes to an evaluation of the single component of tenderness (Ray et al., 1982). A category scale scorecard, typical of that in current use by meat researchers, includes initial and overall tenderness, initial and overall juiciness, connective tissue amount and flavor intensity (AMSA, 1978).

There is no consensus in the literature as to the optimum scale length. In studies of meat quality, the length of category scales ranges from five (McDowell et al., 1982), to eight (Crouse et al., 1983) to 21 points (Savell et al., 1981).

Psychophysical research (Miller, 1956; Norwich, 1981) has suggested that the optimum number of categories that an individual can manipulate simultaneously is in the order of seven plus or minus two. While scales five to nine categories in length have shown similar reliability (Bendig, 1953, 1954), scales 11 categories in length are less reliable (Bendig, 1953). For an analysis of handwriting legibility, Garner (1960) reported that scale sensitivity increased as the number of categories increased. However, he (Garner, 1960) noted that the optimum number of categories would be a function of the amount of discriminability inherent in the particular stimuli under investigation.

Information on optimal scale length for the sensory evaluation of food products is limited. Hedonic scales ranging in length from five to nine categories were compared by Jones et al. (1955) during a food preference survey of 5,400 soldiers. Scales eight or nine categories in length tended to be more sensitive to differences in food preference (as measured by information transmission) than scales five to seven categories in length.

Category scales for hedonic and intensity evaluations of food are generally anchored with descriptive terms. Verbal anchoring has been shown to increase the reliability and sensitivity of CS (Bendig, 1953, 1954). However, the descriptive terms used as anchors must be carefully selected. Jones and Thurstone (1955) investigated the semantics of 51 descriptive terms used for anchoring hedonic food scales and identified terms with a similar meaning to a wide range of panelists. In a subsequent preference survey

(Jones et al., 1955), 20 food items were rated on hedonic scales, which incorporated suitable descriptive terms identified by Jones and Thurstone (1955). Discrimination of food preferences was improved when the neutral descriptor "neither like nor dislike" was eliminated (Jones et al., 1955).

Raffensperger et al. (1956) developed a toughness-tenderness scale for beef by following techniques similar to those used by Jones and Thurstone (1955) and Jones et al. (1955) for hedonic scales. The degree of ambiguity for each of 41 descriptive terms was identified. Then, a nine-point tenderness-toughness scale was constructed and used in a pilot study to evaluate nine beef roasts. The centre category labelled "neither tough nor tender" was avoided by the panelists. Subsequently, removal of this neutral phrase and a decrease in scale length to eight points resulted in a substantial reduction in systematic error along the scale (Raffensperger et al., 1956).

Results of a recent survey by Cross (1977), described in the American Meat Science Association Guidelines for Cookery and Sensory Evaluation of Meat (AMSA, 1978), and an examination of the current literature indicate that most meat scientists now employ eight-point CS (with no centre neutral category) for the descriptive sensory assessment of beef.

Many parametric and nonparametric statistical procedures have been employed for CS data. The use of parametric statistical procedures for data analysis is considered valid only when certain assumptions are met or approximated. For analyses of variance, these assumptions include: normality, independence and homogeneity of variance (Gaito, 1980).

Although successive intervals along the category scale may not be equal in subjective magnitude (Stevens and Galanter, 1957), integer values from category scales are generally used for data analysis (Chapman and Wigfield, 1970). Cloninger et al. (1976) reported that there was little change to analyses of variance with the normalization of category data from intensity scales. Data from nine sensory studies utilizing scales varying from five to 15 points were normalized by calculating z-values to estimate the position of the categories on the psychological continuum. Analyses of transformed data from a five-point scale resulted in an F-value of 12.73 rather than 12.80 for non-transformed data; and for a nine-point scale of 2.80 versus 3.08.

The popularity and continued use of CS has been attributed to its diversity and apparent simplicity (Amerine et al., 1965). However, CS may have some shortcomings. Data from CS may not be reproducible from one laboratory to another. Panelists may attempt to stretch or contract the scale to fit the number of categories (Kapsalis and Moskowitz, 1977; Moskowitz, 1982). Category scale data also are highly dependent upon the scale length (Kapsalis and Moskowitz, 1977).

Panelists may avoid using the ends of the scale and this "category end effect" could bias the ratings (O'Mahony, 1979). Panelists also may be more sensitive to sample differences at the centre of a scale than at either end (Cloninger et al., 1976). However, Stevens and Galanter (1957) noted that panelists are able to discriminate better at one end of the sensory continuum than at the other.

Finally, descriptive anchors may confuse the judges or may be interpreted differently by panelists (Jones and Thurstone, 1955). However, the careful selection of descriptive terms generally can produce scales with similar meaning for most panelists (Jones et al., 1955; Raffensperger et al., 1956).

The widespread use of CS for the descriptive assessment of meat indicates that this method is considered to be both reliable and sensitive to differences in meat quality. However, few systematic studies of the CS technique have been conducted. In addition, several limitations of the CS procedure have been identified. Since it is important to use the most reliable and accurate sensory technique for determining meat quality, an evaluation of the CS procedure and alternative methods is needed.

Line Scaling

Semistructured line scaling (LS), which forms the basis for "Quantitative Descriptive Analysis" (QDA) (Stone et al., 1974, 1980), has been used by several meat researchers (Harries et al., 1972; Randall and Larmond, 1977; Bouton et al., 1980a; Stanley et al., 1980). Line scales are usually 15 cm in length with verbal descriptive anchors 1.3 cm from each end (Stone et al., 1974). Panelists place a vertical mark across the horizontal line scale at the location which best reflects his/her perception of the intensity of the attribute under evaluation.

Line-scaling procedures used for the descriptive assessment of meat vary considerably. Modifications of the QDA procedure have been used for textural evaluations of beef (Harries et al., 1972), ground beef (Randall and Larmond, 1977) and chicken (Lyon and Klose, 1980). In QDA, trained panelists identify and quantify, in order of occurrence, the quality attributes of a product. Several researchers (Walker et al., 1977; Bouton et al., 1980a) have used unstructured (line) scales for tenderness and juiciness evaluations of beef from various postmortem processing treatments.

Procedures for utilizing LS have not been systematically evaluated. Some workers (Baten, 1946; Randall and Larmond, 1977; Stanley et al., 1980) have allowed direct comparisons of LS ratings between samples.

For each attribute under investigation, several samples were scored either on the same line scale (Randall and Larmond, 1977) or on adjacent line scales (Baten, 1946; Stanley et al., 1981). However, most researchers (Harries et al., 1972; Stone et al., 1974; Lyon et al., 1978; Lyon and Klose, 1980; Vickers, 1983) have used separate scorecards for the evaluation of quality attributes related to each sample.

While most line scales are 15 cm (6 in) in length (Baten, 1946; Randall and Larmond, 1977; Stone et al., 1974, 1980), scales 10 cm long (Harries et al., 1972; Giovanni and Pangborn, 1983) to 18.4 cm long (Vickers, 1983) have been used. In addition, descriptive anchors have been placed 1.5 cm from each end of the scale (Stone et al., 1974, 1980; Lyon and Klose, 1980), 2.2 cm from each end (Vickers, 1983) or at each end of the scale (Baten, 1946; Giovanni and Pangborn, 1983). Occasionally, the line scale has also been anchored at its midpoint (Stone et al., 1974; Lyon et al., 1978). However, Stone et al. (1980) stated that their QDA line scales were improved when the middle anchor point was eliminated.

For data analysis, the location of each response along the line scale is converted to a numerical score which can be analysed by a variety of statistical procedures. To facilitate tabulation of LS data from QDA, Stone et al. (1980) developed a specialized system which incorporated a digitizer with visual display, a storage system, a microprocessor, a printer and interface with a main frame computer. Van de Voort et al. (1981) decribed a variation of the LS technique which allowed the direct transfer of data from the panelist to the computer. Panelists scored attributes on Optical Mark Reader (OMR) cards which consisted of 39 "bubbles" instead of the continuous line scale. Marked cards were then read directly by the computer. Despite the potential advantages of a digitizer or OMR card system, many researchers continue to use rulers or templates to assign numerical values to LS data (Lyon and Kluse, 1980; Stanley et al., 1980; Vickers, 1983).

The LS procedure has some of the same limitations decribed previously for the CS technique. Panelists may avoid positions near the boundaries (the category end effect, O'Mahony, 1979). Although the line scale allows a panelist to discriminate as finely as desired along the scale, the fixed constraints at the ends of the scale can limit the panelist's freedom and generate biased data (Stevens, 1975). In addition, panelists tend to use the upper portion of the line scale more than the lower portion (Moskowitz, 1983).

However, the LS technique possesses advantages over the traditional category scaling procedure. The LS procedure provides a continuous rather than stepwise scale (Hall, 1958) and generally produces equal interval data (Anderson, 1970; Weiss, 1972). In addition, the line scale is less susceptible to numerical preferences (Hall, 1958; Moskowitz, 1983). The reduced verbal anchoring characteristic of the line scale may also minimize the problem of unequal intervals associated with the interpretation of descriptive terms (Jones et al., 1955).

Information on the sensitivity of category and line scales in determining differences in meat quality is limited. Raffensperger et al. (1956) scored the tenderness of beef cuts from different grades using a nine point category scale and a modified line scale (nine divisions were marked) and obtained similar results using each method. However, for assessments of apple quality, Baten (1946) reported that the statistical analysis of data generated from LS resulted in larger t-values than those for data from the CS procedure. Thus, comparative studies of LS and CS procedures are needed to assess the relative merits of these sensory evaluation procedures for the descriptive sensory assessment of meat.

Magnitude Estimation

Magnitude estimation (ME), a ratio-scaling technique, has become popular for food research (Moskowitz, 1983) but has not been widely utilized for meat (Larmond, 1976). In using ME, panelists assign numbers to stimuli in proportion to the perceived intensity of the attribute. The ratios between subjective assessments are assumed to reflect the underlying ratios of perceived magnitude. For example, a stimulus judged "20" is perceived to be twice as intense as one judged "10". For the first sample presented, the experimenter may assign a certain value (fixed modulus ME) or may allow each panelist to assign any value of his/her choice (free modulus ME).

Magnitude estimation has been widely used for intensity and hedonic assessments of taste in model systems (Moskowitz, 1970; Moskowitz and DuBose, 1977; O'Mahony and Heintz, 1981; Williams and Bernhard, 1981; Johnson and Clydesdale, 1982) and in simple food systems (Cardello et al., 1979; DuBose et al., 1980; Johnson et al., 1982; Giovanni and Pangborn, 1983). Textural attributes of foods also have been evaluated using the ME procedure (Cardello et al., 1982a, 1982b). However, few studies have employed ME for assessments of complex, heterogeneous food systems such as meat.

There is limited information on procedures utilizing the ME technique for the sensory evaluation of meat. In an early study, Segars et al. (1975) used 17 experienced judges to score three textural attributes of six beef muscles using free modulus ME. In the ME evaluation of meat samples, panelists were instructed to assign larger numbers to indicate increased difficulty of cutting (one or two chews), increased effort in chewing the meat and increased residue at the end of chewing. Later, Segars et al. (1981) used two slightly different modulus free ME procedures for evaluations of irradiated ham and beef rolls. In one phase of the study, panelists evaluated four attributes of each sample in the series presented. In another phase, panelists evaluated a separate series of samples for each of three attributes (Segars et al., 198[illegible]).

Prior to statistical analysis, a variety of normalization and calibration procedures can be applied to ME data to reduce panel variability due to panelists' use of numbers. In "modulus equalization" (Lane et al., 196[illegible]), the ratings from each panelist are multiplied by a correction factor such that each panelist's geometric mean is equal to the group geometric mean. Moskowitz (1977) described a normalization procedure called "modulus normalization" in which ratings from each panelist for a core stimulus (internal standard) or for a series of stimuli can be used to adjust ratings across panelists and across experiments. In the "external calibration" procedure (Moskowitz, 1977), panelists assign numbers to a series of word descriptors in the same manner that they had used for the evaluation of samples. A correction factor is then determined for each panelist.

In a collaborative study involving 20 laboratories (Powers et al., 1981), the three methods for normalizing magnitude estimates described above were compared. Powers et al. (1981) concluded that the three procedures led to essentially the same results. Moskowitz (1977, 1983) has outlined examples of how these normalization procedures can be implemented.

Magnitude estimation data can be analysed by any appropriate statistical procedure. However, magnitude estimates are distributed log-normally (Marks, 1974) with the majority of the numbers at the lower end of the number range.

Logarithmic transformations are frequently performed to change the distribution of the data from a log-normal to a normal distribution, a requirement of parametric statistical procedures (Moskowitz and Sidel, 1971; Powers et al., 1981; Garcia-Medina, 1981; Giovanni and Pangborn, 1983). For ME data, the geometric mean is the preferred average as it preserves the ratio between the estimates and is not excessively influenced by an occasionally large estimate (Marks, 1974). However, logarithmic transformations of ME data may not always be appropriate. When panelists assign zero ratings, logarithms can not be used and geometric means can not be calculated (Moskowitz, 1977). Moskowitz (1983) reported that similar conclusions were reached from analyses of both nontransformed and transformed data.

Several biases of the ME technique have been identified. Most panelists operate within a limited range of familiar numbers and show aberrant behavior in scaling very large or small stimuli (Moskowitz, 1983). Thus, several researchers (Howard, 1973; Williams and Bernhard,

1981) have suggested that for ME experiments, extremes in sensory magnitude should be avoided.

The "regression effect" is another common bias of ME data (Moskowitz, 1983). When assigning ratings to reflect intensity, panelists show conservative behavior which could lead to poorer, less discriminating data. An additional limitation of ME is that panelists prefer to use round numbers more than odd unusual numbers. O'Mahony and Heintz (1981) used fixed modulus ME to measure the perceived intensity of salt solutions and found that more than 75% of the estimates were simple multiples of the standard 10 or unit values from 0 to 5. A partial remedy of this bias has been the use of modulus free ME (Moskowitz, 1983).

It is unclear whether or not the ME technique produces numbers that have true ratio scale properties or whether this method produces nothing more than a disguised internal type measurement (O'Mahony and Heintz, 1981; Birnbaum, 1982). However, proponents of ME (Stevens and Galanter, 1957; Marks, 1974; Moskowitz, 1982, 1983) have described numerous studies confirming the validity and reliability of this and other ratio-scaling procedures. Concerns about the validity of the ME procedure have not been fully resolved and continue to be expressed in the psychological literature (Weiss, 1981; McBride, 1983; Mellers, 1983).

However, ME has several advantages over the CS and LS procedures. Because the judge is unconstrained as to the range and size of numbers that can be used, the magnitude scale is not truncated at the extremes of sensory magnitude (Moskowitz and Sidel, 1971). In addition, ME data can be expressed as ratios or percentages (Moskowitz, 1975). Thus, magnitude estimates provide quantitative information about how panelists perceive different stimuli while data from CS and LS only convey information about intervals (Moskowitz, 1983).

Workers in psychometrics and psychophysics (Stevens and Galanter, 1957; Marks and Cain, 1972) have found that for many perceptual continua, a power function $S=kI^n$; relates sensory intensity, S, to physical intensity, I. In log-log coordinates, the power function becomes a line

$$\log S = n \log I + \log k$$

with slope 'n' and intercept 'k'. Much attention has been directed to the exponent (slope 'n') of each continuum because the exponent is an index of perceptual sensitivity and governs the rate at which perceived intensity increases with physical magnitude (Moskowitz et al., 1974).

Since instrumental measurements are calibrated in ratios, it may be advantageous to also have sensory measures on a ratio scale. In addition, equations relating the textural properties of meat assessed instrumentally to texture judged by panelists would be a very valuable tool for meat scientists. However, few studies in meat research have employed ME. Segars et al. (1975) reported that exponents relating textural properties of meat to an instrumental measure of tenderness (punch and die shear device) ranged from 1.6 to 2.6. These exponents indicate that the panel was much more sensitive than the machine in detecting textural differences. A two-fold increase in an instrumental property was perceived as a 3.5- to six-fold increase in the sensory attribute. Because ME values could be assigned without contraction of the scale at the extremes of the texture range, sensory data paralleled the instrumental measurements (Segars et al., 1975). Thus, the potential of ME for the descriptive sensory assessment of meat requires further investigation.

In addition, few studies have compared ME to either CS or LS for the descriptive assessment of food attributes. McDaniel and Sawyer (1981a) used nine-point CS and modulus free ME to score the intensity of 19 descriptive profile terms of whiskey sour formulations. The ME technique yielded a similar number of significant differences to CS (McDaniel and Sawyer, 1981a). However, Giovanni and Pangborn (1983) obtained larger treatment F-values for panelists' evaluations of intensities of fat in milk and sucrose in lemonade by the LS technique than by ME. Further comparative studies of CS, LS and ME procedures are needed.

Comparative Sensory Methodology

Many early studies on taste panel methodology have focused on sensory tests such as the paired comparison, ranking, rating and scoring (Baten, 1946; Murphy et al., 1954, 1957; Pilgrim and Wood, 1955; Raffensperger et al., 1956; Filipello, 1957; Gridgeman, 1961). In recent years, there has been a renewed interest in the comparison of different sensory evaluation techniques used for intensity and hedonic assessments of food products (Moskowitz and Sidel, 1971; McDaniel and Sawyer, 1981a, 1981b; Moskowitz, 1982; Giovanni and Pangborn, 1983; Vickers, 1983).

Although the sensory evaluation techniques examined by researchers differ considerably, in general, comparative studies of sensory techniques have used similar procedures and evaluation criteria. In most studies, treatments were selected to provide differences in the attribute/product under investigation. For example, Murphy et al. (1957) used three different strawberry varieties, Gridgeman (1961) stored eggs under four conditions and McDaniel and Sawyer (1981a, 1981b) prepared a number of different whiskey sour formulations.

In several studies (Pilgrim and Wood, 1955; Gridgeman, 1961; Moskowitz and Sidel, 1971), different groups of panelists were used for evaluations by each sensory technique. However, most researchers (Baten, 1946; Murphy et al., 1954, 1957; Carlin et al., 1956; Filipello, 1957; McDaniel and Sawyer, 1981a, 1981b; Giovanni and Pangborn, 1983; Vickers, 1983) employed one group of panelists to compare sensory procedures. Gridgeman (1961) stated that an ideal experimental comparison would involve the same group of panelists for each of the sensory techniques under investigation.

In an early comparative study (Murphy et al., 1957), panelists used three sensory evaluation techniques on the same day (and possibly at the same session). However, to minimize possible memory influences, training and motivational factors, various procedures have been employed. Carlin et al. (1956) presented two scoring techniques on alternate days. McDaniel and Sawyer (1981a, 1981b) employed CS and ME techniques on different days according to randomized designs. Recently, balanced designs were used for a comparison of LS and ME techniques (Giovanni and Pangborn, 1983; Vickers, 1983). In both of these studies, half of the panelists employed LS first, while the remaining panelists began with ME.

Statistical procedures used to quantitatively compare sensory evaluation methods which have different scales of measurement are lacking (Murphy et al., 1957). However, criteria which have been used to compare sensory evaluation techniques include:

(1) the number of significant differences among the treatments (Murphy et al., 1954, 1957; Pilgrim and Wood, 1955; Moskowitz and Sidel, 1971; Moskowitz, 1982; McDaniel and Sawyer, 1981a, 1981b)

(2) the magnitude and significance of the F-value or t-value

from statistical analyses of each evaluation technique (Baten, 1946; Murphy et al., 1954; Gridgeman, 1961; Dransfield et al., 1982; Moskowitz, 1982; Giovanni and Pangborn, 1983)

(3) correlation coefficients (Filipello, 1957; Dransfield et al., 1982; Giovanni and Pangborn, 1983), and

(4) the sensitivity of the judges (Carlin et al., 1956; Murphy et al., 1957; Giovanni and Pangborn, 1983). In addition, several researchers (Baten, 1946; Murphy et al., 1957) have examined panelist preference among the sensory evaluation techniques employed.

Some criteria for examining sensory methods have been specific to comparative studies of CS, LS and ME procedures. For CS and ME data, relationships between hedonic or intensity evaluations and physical concentration have been plotted (Moskowitz and Sidel, 1971) or expressed mathematically (Moskowitz, 1982). In addition, Giovanni and Pangborn (1983) have tested the "goodness of fit" for mathematical models between LS and ME data and between sensory data and physical concentrations by examining the resulting coefficients of determination. Anderson (1977) and Coleman et al. (1981) have suggested that the "goodness of fit" of a mathematical function should be tested by measuring the significance of the deviations from the model by analysis of variance. However, several workers (O'Mahony and Heintz, 1981; Powers et al., 1981) have used correlation coefficients to test mathematical models for ME data.

Comparative studies of CS, LS and ME procedures have focused on either hedonic assessments of food products (Moskowitz and Sidel, 1971; McDaniel and Sawyer, 1981b; Moskowitz, 1982; Vickers, 1983), or on intensity evaluations of food attributes (McDaniel and Sawyer, 1981a; [illegible] (Giovanni and Pangborn, 1983). However, few studies using trained panelists have compared CS, LS and ME for the descriptive assessment of food attributes. McDaniel and Sawyer (1981a) used a trained panel to score the intensity of 19 descriptive profile terms of whiskey sour formulations using CS and ME. In addition, studies comparing sensory procedures for the descriptive assessment of meat are lacking.

Raffensperger et al. (1956) used a nine-point category scale and a modified line scale (nine divisions were marked) to score the tenderness of beef cuts from different grades. The evaluation techniques were compared by examining the discriminability of the scales, agreement among judges and distributions of the ratings. Scale sensitivity and panelist agreement did not differ between scaling techniques; however, the CS procedure had a significantly greater amount of systematic error than the LS procedure (Raffensperger et al., 1956).

The eating quality of beef assessed at each of five European research institutes was compared by utilizing a common eight-point category scale (Dransfield et al., 1982).

Criteria used to compare data from each of the research institutes included: correlation of sensory data, panel discrimination of tenderness differences, panel variation between institutes and an examination of the sensory data. In addition, instrumental measures of tenderness (Warner Bratzler shear and Volodkevich-style jaws) were correlated with panel tenderness scores. Assessments of tenderness by the experienced panels were highly correlated across institutes. Instrumental measures of tenderness also correlated highly with panel tenderness scores. Panelists at one research institute did not discriminate tenderness differences as well as panelists at the other four research institutes (Dransfield et al., 1982).

The procedures and evaluation criteria identified in the preceding discussion may be useful for a comparative study of CS, LS and ME for the descriptive sensory assessment of beef.

Training

Meat scientists routinely employ trained descriptive panels to determine the effects of various production and processing treatments on the quality attributes of cooked meat. However, recommendations regarding the extent of panelist training for CS, LS and ME differ. An "expert" descriptive panel using CS may receive three to four months of intensive training (AMSA, 1978); panelists using LS for sensory assessments of food usually receive about 10 hours of training (Cook and Wessman, [illegible]). With ME, it is felt that a conscientious person can make quantitative estimates of quality attributes of foods after a few introductory sessions (Moskowitz, [illegible]; Williams and Bernhardt, [illegible]). However, use of ME by trained rather than untrained panels may result in improved discrimination of textural attributes of foods (Cardello et al., [illegible]). In addition, practice in the use of ME may result in more stable judgements (Teghtsoonian and Teghtsoonian, [illegible]; Giovanni and Pangborn, 1983).

Cross et al. (1978) and the AMSA (1978) outlined selection and training procedures for the descriptive sensory assessment of beef which have now been adopted by many meat researchers (Kastner et al., 1980; Davis et al., 1981; Berry and Cross, 1982; Bowles Axe et al., 1983; Crouse et al., 1983). Three objectives of training were identified: (1) to familiarize the panelist with the test procedures, (2) to improve a panelist's ability to recognize and identify sensory attributes, and (3) to improve the panelist's sensitivity and memory (Cross et al., 1978; AMSA, 1978). In the study by Cross et al. (1978), training was accomplished through individual and group sessions in which a wide range of beef samples were evaluated and discussed. One or more sessions were devoted toward demonstrating levels of each attribute under study. Although the training procedure described by Cross et al. (1978) was developed for the CS technique, this procedure (with slight modifications) could also be effectively used to train panelists in the use of LS and ME.

Choice of the Meat System

Tenderness is probably the most important single quality attribute of meat (Bailey, 1972). Considerable current research has focused on the effects of various postmortem processing treatments such as hot boning, delay chilling and electrical stimulation on the resulting tenderness and on other quality attributes of cooked beef. Traditionally, carcasses are left intact after slaughter with muscles attached and restrained by the skeleton, chilled for 18 to 48 h at 2 to 3°C and then fabricated and sold as sides, quarters, wholesale cuts, subprimal cuts or boneless cuts (Cross, 1979).

In hot boning (HB), muscles or muscle systems are removed from the carcass prior to chilling (West, 1983). Because HB requires the chilling and storage of only edible meat and not excess fat and bone, potential economic benefits include: savings in energy, cooler space, labor, transportation costs, product shrinkage and in-plant residence time (Kastner et al., 1973; Ray et al., 1982; Kastner, 1983). However, prerigor excision and rapid chilling may cause toughening (Kastner et al., 1973; Cross and Tennent, 1980; Berry and Cross, 1982; Bowles Axe et al., 1983; Lyon et al., 1983). Although flavor is generally not affected by HB (Schmidt and Keman, 1974; Kastner and

Russell, 1975; Cransfield et al., 1976), the connective tissue of hot-boned meat is sometimes perceived as more abundant than that of comparable meat from conventionally boned carcasses (Berry and Cross, 1982; Bowles Ake et al., 1981).

Delay chilling (DC) is an early postmortem treatment in which carcasses are held at elevated temperatures (16°C) for 6 to 24 h postmortem (Cliplef and Strain, 1976; Smith et al., 1979a; Crouse et al., 1983). In several studies (Hostetler et al., 1975; Cliplef and Strain, 1976; Smith et al., 1979a; Lochner et al., 1980), meat from delay chilled carcasses was more tender than that from conventionally processed carcasses. Recent reports (Lochner et al., 1980; Marsh et al., 1980-81) have suggested that the enhanced tenderness of delay-chilled beef was primarily due to the high muscle temperature and pH (Marsh et al., 1980-81) maintained during the first 2 to 4 h postmortem. In addition, DC may also improve meat flavor (Cliplef and Strain, 1976) and reduce detectable connective tissue (Crouse et al., 1983).

Electrical stimulation (ES) of beef involves the application of an electrical current soon after slaughter to accelerate the rate of glycolysis and reduce the onset of rigor mortis (Will et al., 1979). The use of ES immediately post slaughter may prevent toughening of hot-boned meat (Gilbert and Davey, 1976; Walker et al., 1977; Bouton et al., 1981a; Bowles Axe et al., 1983). However, several workers (Taylor et al., 1981-82; Lyon et al., 1983) have found that ES did not improve the tenderness of hot boned meat.

Frequently, ES has been shown to increase the tenderness of conventionally chilled carcasses (Bouton et al., 1978, 1980a, 1981b; Smith et al., 1979b; Cross and Tennent, 1980; George et al., 1980; McKeith et al., 1980, 1981; Riley et al., 1981). However, ES may be more effective in improving the tenderness of tougher meat, such as that from forage fed beef (Davis et al., 1981; Salm et al., 1981; Schroeder et al., 1982). Improvements in flavor and in other meat quality attributes due to ES have occasionally been reported (Smith et al., 1979b, McKeith et al., 1980).

There is limited research on the ES of beef carcasses in conjunction with DC. However, Smith et al. (1979b) and Elgasim et al. (1981) reported that ES contributed more to increased tenderness of the meat cuts than did DC.

Thus, these various postmortem processing treatments (HB, DC, ES) should provide meat with a wide range in organoleptic quality useful for a comparative study of sensory evaluation techniques. In addition, the meat quality differences attributable to postmortem treatment and determined from sensory and objective measurements, should be of a magnitude similar to those found in current research practice.

The longissimus dorsi muscle is a popular muscle for studies on meat quality. However, because muscles vary in their response to different treatments (Hawrysh and Berg, [illegible]), additional information is needed on the effects of various postmortem processing treatments on other major bovine muscles such as the semimembranosus (SM), semitendinosus (ST) and biceps femoris (BF). The SM is a large, easily accessible muscle which could provide a number of similar steaks. Because the SM muscle has displayed a slight end-to-end variation in tenderness (Paul and Bratzler, 1955; Ginger and Weir, 1956; Dransfield and Jones, 1978), sampling from the same anatomical location from each muscle is important. Balancing of steak position from adjacent steaks would further minimize any differences in tenderness due to position.

Choice of the Cooking System

Two dry heat methods described in the literature for cooking SM steaks are roasting (Batcher and Deary, 1975; Berry et al., 1977; McKeith et al., 1981) and broiling (Breidenstein et al., 1968; Dryden et al., 1979; Smith et al., 1982). In roasting, heat is transmitted to the meat by convection, either by normal or forced air, in a closed preheated oven (AMSA, 1978). The meat is not turned during cooking. In broiling, steaks are cooked primarily by radiant energy (Paul and Palmer, 1972). The heat usually radiates from one direction, so the meat must be turned during cooking (AMSA, 1978).

Although broiling more closely simulates current consumer practice, Harrison (1975) recommended roasting instead of broiling for tender meat cuts as precise broiling conditions are difficult to maintain. Radiation intensity varies within and among broiler units (Paul and Palmer, 1972; Batcher and Deary, 1975). In addition, the air temperature of the oven can be influenced by air movement in the laboratory (Cover et al., 1957) and may not reflect the true energy output of the unit (Paul and Palmer, 1972). In broiling, the investigator relies on the final internal temperature of the steaks as his only control (AMSA, 1978).

Several workers (Batcher and Deary, 1975; Cross et al., 1979) have compared the characteristics of beef steaks cooked by roasting and broiling. When cooked to either 60° or 71°, roasted SM steaks were scored more juicy and tender and less mealy than comparable steaks broiled to the same temperature (Batcher and Deary, 1975). In addition, cooking losses were significantly lower for roasted than broiled steaks (Batcher and Deary, 1975), which would result in increased edible portion. Other advantages of roasting over broiling described by Batcher and Deary (1975) included: less splatter, lower energy consumption and minimal effort and attention for the researcher. During preliminary work, it was noted that roasted 3.8 cm SM steaks were also more even in doneness than broiled steaks (Shand and Hawrysh, unpublished).

Cross et al. (1979) found that either roasting or broiling was suitable for 2.5 cm thick longissimus steaks. Tenderness, connective tissue amount and flavor scores for steaks did not differ significantly between cooking methods. However, broiled steaks were less juicy and had greater cooking losses than roasted steaks (Cross et al., 1979).

Cooking methods selected for meat research should maximize the differences due to factors under investigation and minimize variability due to cooking (Mottram, 1981). Thus, roasting of SM steaks appears to be an appropriate and reproducible cooking method to use in a comparative study of sensory evaluation procedures.

EXPERIMENTAL PROCEDURE

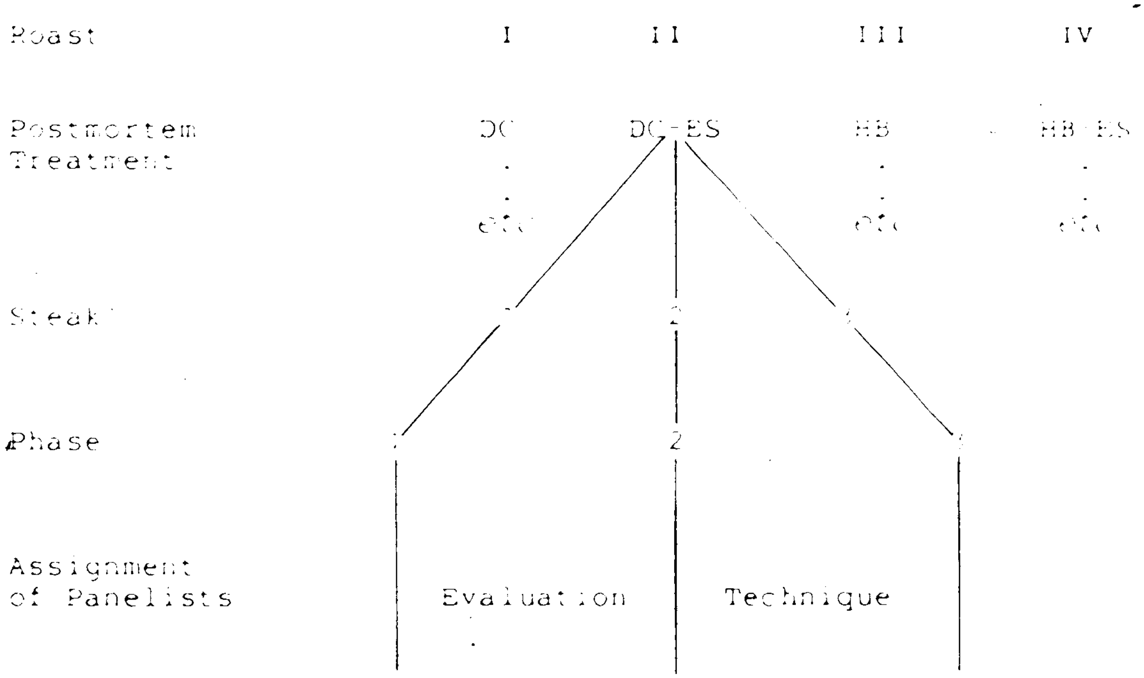
Experimental Design and Statistical Analysis

Eighteen trained panelists evaluated the quality attributes of semimembranosus steaks using three sensory evaluation techniques (category scaling (CS), line scaling (LS) and magnitude estimation (ME)). Semimembranosus (SM) steaks were evaluated according to an experimental design which involved four treatments, nine replications (roasts) per treatment, three steaks per roast, three phases (time periods) and 18 judges. The design for one experimental replication is outlined in **Figure 1**.

The three adjacent steaks, cut from each of nine roasts per treatment, were assigned according to three 3 x 3 Latin square designs for preparation in each of three phases. Each phase involved the evaluation of three steaks taken from each of the three roast positions.

To minimize the effects of scale carry over, learning and motivational factors, panelists utilized each sensory evaluation technique sequentially, in different predetermined orders, in the three phases of the study (**Figure 1**). Within each phase, samples from each steak were evaluated by the 18 trained panelists with six of the judges using the CS procedure, six other judges using the LS procedure and the remaining six judges using ME. At the completion of the study, each judge had evaluated one of

Figure 1. Design for one experimental replication.



1-3	CS	ME	LS
4-6	CS	LS	ME
7-9	LS	CS	ME
10-12	ME	CS	LS
13-15	LS	ME	CS
16-18	ME	LS	CS

DC	Delay Chilling
DC-ES	Delay Chilling with Electrical Stimulation
HB	Hot Boning
HB-ES	Hot Boning with Electrical Stimulation
CS	Category Scaling
LS	Line Scaling
ME	Magnitude Estimation

---

[1] Steak assignment to phase was balanced over the nine replications.

three adjacent steaks from each of the nine roasts per treatment using each evaluation technique.

Sample presentation to panelists was completely balanced for serving order, judge and replication as described by Sidel and Stone (1976). However, for each phase of the study, serving order for each panelist remained the same to allow panelist evaluations with each technique to be similar.

To balance any treatment order effects for each replication, all preparation, sampling and objective measurements made on the steaks were performed according to complete and incomplete Latin square designs.

Arithmetic means were calculated for CS and LS data; geometric means were calculated for ME. Least squares split-plot analyses of variance were computed for the various factors under investigation. Sources of variation were treatment (n=4), replications (n=9), phases (n=3) and steaks (n=3) and for taste panel data, panelists (n=18). For all analyses with more than one observation per cell, means were computed and used in the analyses. Student-Newman-Keuls' multiple range test (Steel and Torrie, 1980) was used to establish significant differences among treatments.

Data for each sensory evaluation technique were analysed separately. The magnitude estimates were transformed to logarithms for data analyses unless otherwise noted. However, it was not necessary to normalize ME data as described by Powers et al. (1981) because panelist variation was removed in the analyses of variance procedure utilized.

Initially, the four postmortem treatments evaluated were included in all analyses of sensory and objective data. However, two treatments had few significant differences and contributed little to a comparison of sensory evaluation techniques, the main interest of this study. Therefore, analyses were performed on data from only two (delay chilling and hot boning) of the four postmortem processing treatments. The sources of variation and valid error terms used in the analysis of data from each sensory evaluation procedure are presented in **Table 1.**

Correlation and regression analyses were performed to assess linear and other mathematical functions for appropriate sensory data between each of the three evaluation techniques. In addition, for tenderness assessments, correlation and regression analyses were performed between each of the evaluation techniques and Warner Bratzler shear data. The functions evaluated included: linear ($Y=a+bX$), power ($Y=aX^n$), logarithmic ($Y=a+b\ln X$), hyperbolic ($Y=a+b(1/X)$) and parabolic ($Y=a+bX+cX^2$). For evaluation of the non-linear functions, appropriate transformations of the data were made and linear

Table 1. Sources of variation, degrees of freedom and valid error terms for the analyses of variance of data from each sensory evaluation procedure.

Source of Variation	Degrees of Freedom[1] Two Treatments	Degrees of Freedom[1] Four Treatments	Valid Error Term
Treatment (T)	1	3	TP
Replication (R)	8	8	TR
TR (T)	8	24	
Phase (P)	2	2	--
Judge/Phase (J/P)[2]	15	15	RP+RJ/P+TRP+TRJ/P
TP	2	6	--
TJ/P[3]	15	45	RP+RJ/P+TRP+TRJ/P
RP	16	16	--
RJ/P	120	120	TRP+TRJ/P
TRP	16	48	--
TRJ/P	120	360	--
Total	323	647	

[1] Actual degrees of freedom for each of the analyses may be reduced because of missing data.

[2] Used as panelist source of variation.

[3] Used as treatment by panelist source of variation.

regressions were computed. All correlations and regressions were computed across treatments rather than within treatments to allow an evaluation of relationships over the entire range of the data.

For the three evaluation techniques, the ranked data for selected criteria were converted to scores according to the method of Fisher and Yates (1949) as described by Larmond (1977) and then submitted to analyses of variance. Mean separation of ranks was done using Student-Newman-Keuls' multiple range test (Steel and Torrie, 1960).

Meat Used for the Study

Thirty-six SM roasts of similar weight, nine from each of four postmortem processing treatments, were obtained from young crossbred steers (average age of 15 months) raised at the Agriculture Canada Research Branch, Lacombe, Alberta. Details of breeding, management and postmortem handling of the animals were described by Jeremiah et al. (1984).

Four postmortem processing treatments were applied to the carcass sides: delay chilling (DC), delay chilling with electrical stimulation (DC-ES), hot boning (HB) and hot boning with electrical stimulation (HB-ES).

Carcass sides assigned to the DC treatments were held for 2 h in a walk-in meat cooler at 10-15° before being chilled at 2°. The SM muscles from the sides assigned to the HB treatments were excised at 40 min postmortem.

Electrical stimulation of delay chilled sides and hot boned SM muscles followed similar procedures. At 45 min postmortem, sides or excised muscles were electrically stimulated (550 V, AC 50-60 cycles per sec, 2 min duration with 1 sec on, 3 sec off) using a Best and Donovan No. 2[illegible] hog stunner attached to a HC8015 Fluke multimeter.

The sides or excised muscles were aged for six days (2°). A roast, taken from the distal end of each SM muscle, was individually vacuum packed in a barrier bag and stored (-30°) for seven months. Frozen roasts were transported to the Meat Laboratory at the Edmonton Research Station and cut into three 3.8 cm steaks, starting from the cut surface (approximate centre of the muscle). Care was taken to cut each of the steaks perpendicular to the muscle fiber direction. Each steak was weighed, placed individually in a polyethylene bag, sealed with a twist tie and drugstore wrapped in waxed freezer paper. The packaged frozen steaks were stored (-25°) for one to two months. The remainder of each roast was wrapped in aluminum foil, placed in a polyethylene bag and stored (-25°) for later chemical analyses.

Prior to cooking, each steak was thawed in its packaging for 3 h at 22° and 20 h at 3°.

Cooking Procedure

Each steak was placed [illegible] side up [illegible] aluminum roasting pan (4[illegible] x [illegible] x [illegible] cm). The steaks were roasted individually to an internal temperature of [illegible] in one of four household ovens (Kenmore, Mark [illegible]) [illegible] at [illegible]°. The internal temperature of each steak was [illegible] by two copper-constantan thermocouples connected to a Honeywell recording potentiometer. After [illegible], each cooked steak was weighed, wrapped in plastic wrap and refrigerated (4°) for approximately one hour to facilitate sampling of the meat.

Sampling Procedure

The sampling pattern for sensory and objective tests (**Figure 2**) was standardized during preliminary work.

Each cooked steak was cut, parallel to muscle fiber direction, into four (steaks 1 and 2) or three (steak 3) 1.3 cm thick slices. Steak slices were cut into rectangular cores (1.3 x 1.3 x 3.2 cm) using an electric meat slicer (Berkel model 1836) and a double-bladed scalpel. After removing the browned exterior surface, two 1.3 cm cubes were cut from the centre of each of 18 cores for each of the 16 panelists (**Figure 2**). The cubes were refrigerated (4°) on a covered plate for one half to three and one half hours before evaluation.

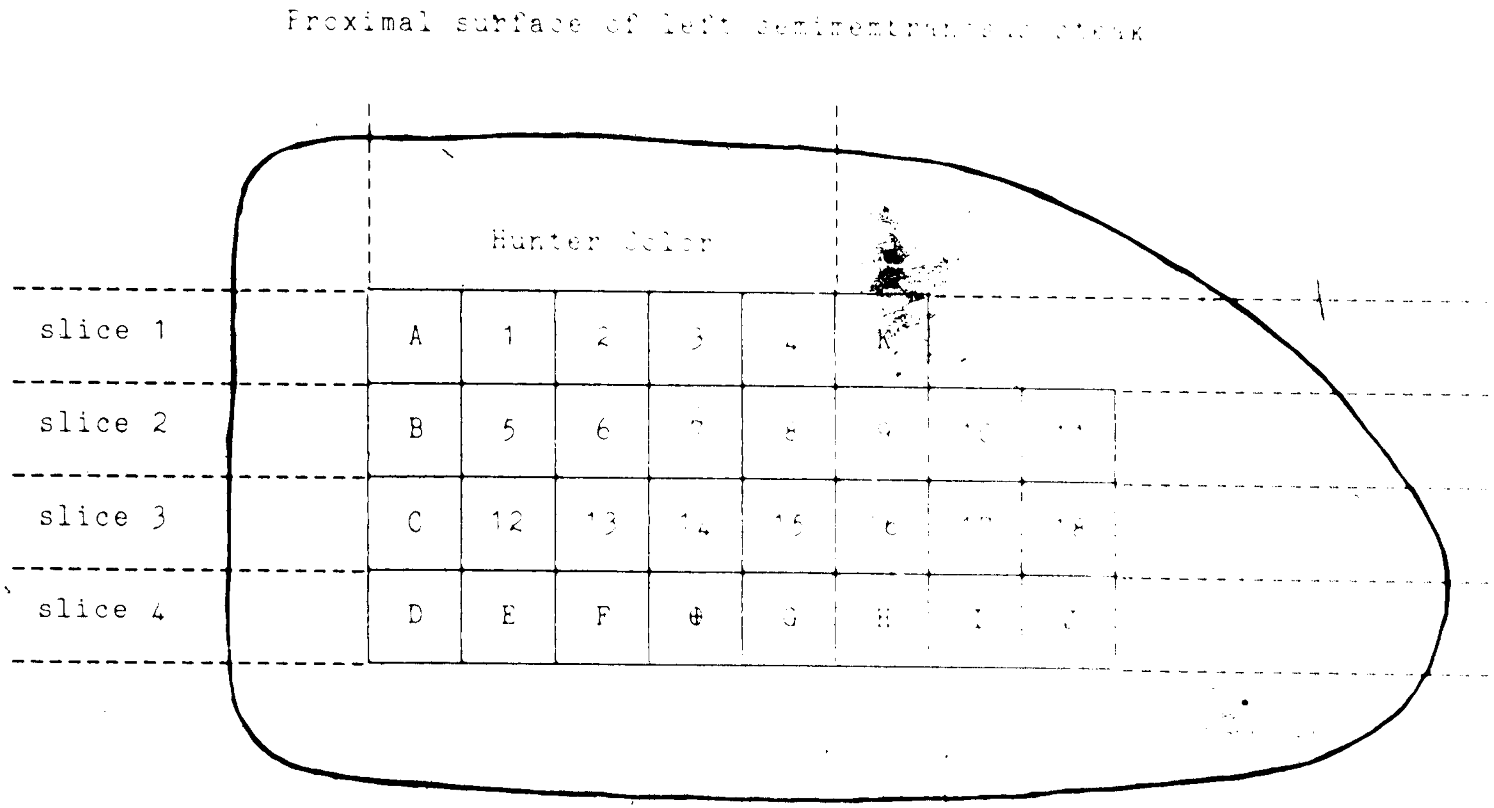


Figure 2. Location of samples for sensory and objective measurements from [illegible] semimembranosus steak.

Positions 1 to 18 - Taste panel
Positions A to K - Warner Bratzler shear
Position ⊕ - Press [illegible]

Cores for Warner Bratzler shear measurements and press fluid determinations were taken from positions [illegible] steaks [illegible] and [illegible]. A portion ([illegible] cm) was removed from each steak, wrapped separately in plastic wrap and aluminum foil and used for Hunter color measurements.

Objective Measurements on Raw Samples

Fat and Moisture

The percentages of fat (ether extract) and moisture of each raw roast were determined by the methods of the Association of Official Analytical Chemists (AOAC, [illegible]). A sample ([illegible] g) from the distal end of each roast was thawed (4°), trimmed of connective tissue and fat, ground (30 sec) to an homogeneous paste, freeze dried (26 h) and reground. For fat determinations, duplicate 2 g portions of each freeze dried sample were placed in a Goldfisch extraction apparatus for 18 h and the resulting ether extract weighed. Moisture was determined by drying duplicate 2 g portions of each freeze-dried sample for an additional 18 h at 105°.

pH

A Fisher Accumet Model 230 pH/ion meter used to determine the raw pH of each roast. A thawed 20 g sample from each roast, free of fat and connective tissue, was blended with 100 ml distilled water (60 sec) and filtered into two beakers to give duplicate readings.

Drip in Thaw

The weight of drip in thaw of each steak was calculated as a percentage based on the weight of the frozen steak.

Objective Measurements on Cooked Samples

Cooking Time

The cooking time required for each steak to reach an internal temperature of [illegible]° from an initial temperature of [illegible]° was determined. The cooking time of each steak in min per 100 g raw weight was also calculated.

Cooking Losses

Percentage total, volatile and drip losses, based on the weight of each thawed trimmed steak were calculated.

Warner Bratzler Shear

Meat cores (1.3 x 1.3 x 3.2 cm) were removed from specified positions on steaks 1 and 2 of each roast, adjacent to those used for sensory evaluation (Figure 2). The cores (21°) were sheared once, perpendicular to the fiber direction, with a Warner Bratzler shear attachment on the Ottawa Texture Measuring System (OTMS) (up speed, 0.25 cm/sec; 5 % range selection). An average of nine shear values were obtained for each steak.

Press Fluid

The percentage of press fluid in the [illegible] steaks was determined using a modification of the method of [illegible] and Swatland (1976). Triplicate [illegible] g samples, taken from steaks 1 and 2, were placed individually between two pieces of preweighed [illegible] aluminum foil. Three pieces of Whatman No. 1 filter paper ([illegible] cm in diam) were placed outside each piece of foil. Each sample was pressed separately between two plexiglass plates (to set; 878.6 kg/cm² force) using a Carver Laboratory Press (Fred S. Carver, Inc., Summit, New Jersey). The pressed meat and foil were weighed immediately after pressing to determine moisture loss. The percentage of press fluid was calculated by dividing the weight of the expressed liquid by the weight of the original meat sample.

Color

The color of each cooked steak was measured using a Hunterlab Color Difference Meter (Model D25A-2) standardized against a white tile with values of L=92.7, a=-1.0 and b=0.3. An exposed interior surface of each steak (**Figure 2**) was placed under the instrument port. Lightness 'L', redness 'a' and yellowness 'b' values were recorded for each sample. Then, each sample was rotated 90° and a second reading was made for each color difference factor.

Degree of Doneness of Cooked Samples

During sample preparation, the internal color of each cooked steak was assessed by three experienced judges. After each steak was cut into slices (Figure 2), the three judges independently scored the degree of doneness of the second slice using a double-pointed category scale (1=well-done, 5=medium, 9=rare) (Appendix, Figure 13, page 132).

Sensory Methodology

Eighteen trained panelists evaluated the organoleptic quality of steaks from each of four postmortem processing treatments using category scaling (CS), line scaling (LS), and magnitude estimation (ME). Panelists assessed five descriptive attributes in each beef sample: initial tenderness, overall tenderness, connective tissue amount, overall juiciness and flavor intensity.

Selection of Panelists

Panelists were screened by a procedure similar to that used by Cross et al. (1978). Twenty-eight students and staff in the Department of Foods and Nutrition, University of Alberta, participated in a series of 16 triangle tests. During each triangle test, panelists were asked to pick the odd sample with respect to tenderness, juiciness or amount of connective tissue and to indicate the degree and direction of difference. Panelists correctly identified the odd sample 73 % of the time, with a range of 56-96 % over the 16 triangles. Twenty-three panelists were selected for training on the basis of their ability to correctly identify the odd sample, interest in the study and availability for the duration of the study.

Training of Panelists

Training sessions were held two to three times per week for 10 weeks. The first session introduced panelists to sensory evaluation and acquainted them with terms used for general texture description. Panelists ranked standard foods used as anchors for softness and juiciness (Forbes, 1973) and then scored the softness or juiciness of these foods on eight-point scales. For example, foods representing different intensities of juiciness included: shortcake biscuits, dehydrated apples, canned potatoes, raisins and canned mushroom caps.

During training, panelists were gradually acquainted with the three evaluation techniques, panel procedures and the quality attributes under investigation. Definitions provided for each of the descriptive attributes are shown in **Figure 3**. The CS method was introduced first. Panelists were asked to rate each attribute in the meat samples using an eight-point descsiptive category scale. A value of 8 represented extremely tender, juicy, meaty and no connective tissue and a value of 1 represented extremely tough, dry, weak and abundant connective tissue. The final CS scorecard is presented in **Figure 4**. Each panelist standardized

INITIAL TENDERNESS -- is the lack of force required to bite through a cube of beef across the grain, between the molar teeth (evaluated after two chews).

OVERALL TENDERNESS -- is the amount of effort and time required to completely masticate a cube of beef.

CONNECTIVE TISSUE -- is the amount of residue felt during chewing and left after complete mastication.

OVERALL JUICINESS -- is the amount of moisture left in the mouth after complete mastication.

FLAVOR INTENSITY -- is the amount of meaty flavor present in the mouth after complete mastication.

Figure 3. Quality attribute definitions used for the sensory evaluation of beef.

SAMPLE #________ DATE____________ NAME________________

INITIAL TENDERNESS
(2 chews)
____

8 extremely tender
7 very tender
6 moderately tender
5 slightly tender
4 slightly tough
3 moderately tough
2 very tough
1 extremely tough

OVERALL TENDERNESS
(# chews   )
____

8 extremely tender
7 very tender
6 moderately tender
5 slightly tender
4 slightly tough
3 moderately tough
2 very tough
1 extremely tough

CONNECTIVE TISSUE AMT.
____

8 none
7 practically none
6 traces
5 slight
4 moderate
3 slightly abundant
2 moderately abundant
1 abundant

OVERALL JUICINESS
____

8 extremely juicy
7 very juicy
6 moderately juicy
5 slightly juicy
4 slightly dry
3 moderately dry
2 very dry
1 extremely dry

FLAVOR INTENSITY
____

8 extremely meaty
7 very meaty
6 moderately meaty
5 slightly meaty
4 slightly weak
3 moderately weak
2 very weak
1 extremely weak

COMMENTS

Figure 4. Category scale scorecard for the sensory evaluation of cooked semimembranosus steaks.

his/her overall tenderness scores by the number of chews required to completely masticate a cube of beef.

During the second week of training, the LS technique was introduced. Each panelist was instructed to mark a vertical line across a 15 cm line at the point which best described his/her impression of each of the atributes in the beef cubes. The line scale was anchored 1.3 cm from each end with the appropriate descriptors (very tender--very tough; small amount connective tissue--large amount connective tissue; very juicy--very dry; very meaty--very weak). A value from 0.0 to 15.0 was assigned to each rating by converting the mark on the line to a numerical score. The final LS scorecard is presented in **Figure 5.**

Magnitude estimation was introduced during the last four weeks of panel training. To acquaint panelists with the method of modulus free ME, panelists evaluated shapes and lines as described by Moskowitz (1977). Each judge then evaluated the quality attributes of the beef samples by assigning any number (greater than 0) of his/her choice to describe each attribute in the first sample presented. For each successive sample, panelists assigned numbers for each attribute in relation to those attributes in the previous sample. Panelists were instructed that in ME, the ratio relationship of the numbers assigned was more important than the actual numbers assigned. Since panelists had difficulty scoring five attributes simultaneously, the final ME

SAMPLE # ________ DATE ____________ NAME ______________

INITIAL TENDERNESS

very tender	very tough

OVERALL TENDERNESS

very tender	very tough

CONNECTIVE TISSUE AMT.

small amount	large amount

OVERALL JUICINESS

very juicy	very dry

FLAVOR INTENSITY

very meaty	very weak

COMMENTS

Figure 5. Line scale scorecard for the sensory evaluation of cooked semimembranosus steaks.

scorecard was divided into two parts (Figure 6).

The introduction and refinement of each sensory procedure followed a similar format. Panelists individually evaluated samples and then participated in a round table discussion. The group leader recorded results using an overhead projector and encouraged participation by all panelists. For CS and LS, actual scores were recorded; however, to avoid influencing panelists' choice of numbers when using ME, only relative differences between meat samples were shown. Discussion of results and comments by panelists during these sessions helped to develop panel consistency and to improve panelists' understanding of the attributes under investigation. Panelists scored a wide variety of beef samples (different muscles from animals of various ages were cooked to varying degrees of doneness) using each evaluation technique (**Table 2**). The number of attributes and samples evaluated per session was increased gradually for each procedure.

The last two weeks of training were arranged to simulate actual testing conditions. The samples evaluated, taste panel booth arrangement and the scorecards used were similar to those in the actual study. Panelists individually evaluated samples in a taste panel room and then participated in a brief discussion after each session.

DATE__________ NAME__________

CUBE ONE

SAMPLE CODE ______ ______ ______

INITIAL TENDERNESS (↑#=↑toughness) ______ ______ ______

FLAVOR INTENSITY (↑#=↑meatiness) ______ ______ ______

OVERALL JUICINESS (↑#=↑juiciness) ______ ______ ______

COMMENTS

DATE__________ NAME__________

CUBE TWO

SAMPLE CODE ______ ______ ______ ______

OVERALL TENDERNESS (↑#=↑toughness) ______ ______ ______ ______

CONNECTIVE TISSUE AMT. (↑#=↑amount) ______ ______ ______ ______

COMMENTS

Figure 6. Magnitude estimation scorecard for the sensory evaluation of cooked semimembranosus steaks.

Table 2. Samples evaluated by panelists during training for the [illegible] procedures

Muscles cooked	CATEGORY SCALE		[illegible] SCALE		[illegible]	
	No. of samples	Internal temp	No. of samples	[illegible]	[illegible]	[illegible]
Semimembranosus	[illegible]	65, [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Biceps femoris	[illegible]	65, 72	[illegible]	[illegible]	[illegible]	[illegible]
Semitendinosus	6	65, 72	[illegible]	[illegible]	[illegible]	[illegible]
Longissimus	[illegible]	65, 75	[illegible]	[illegible]	[illegible]	[illegible]
Biceps femoris (ischiatic head)	[illegible]	65, [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gluteus medius	[illegible]	65	[illegible]	[illegible]	[illegible]	[illegible]
Rectus femoris	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Pectoralis profundis	[illegible]	65	[illegible]	[illegible]	[illegible]	[illegible]

Note: Samples for the evaluation of panelist performance are [illegible]

Evaluation of Panelist Performance

Panelist performance was evaluated after four and eight weeks of training using the procedure described by Cross et al. (1978) and the AMSA (1978). The first evaluation (four replications of six treatments) gave a general indication of the panelists' ability to discriminate differences in the palatability characteristics of the samples presented. The F-value from a one-way analysis of variance for each palatability characteristic was used as an indicator of each panelist's performance. A high F-value indicated that the panelist was consistent in making duplicate judgements and could discriminate differences in the treatments presented for evaluation. Those characteristics (for example, amount of connective tissue) that panelists had difficulty in differentiating were stressed in later training sessions.

The second evaluation of panelist performance (four replications of eight treatments) was used primarily to select panelists for the study. Eighteen panelists (17 females, one male; 18 to 35 years) who were sensitive to sample differences, consistent over time and highly motivated were chosen to participate in the study. Prior to training, 16 of the 18 were naive to sensory evaluation.

Sample Presentation

Panel sessions were held daily between 1030 h and 1430 h in an atmospherically-controlled sensory panel room equipped with eight individual booths and white lights. Each panelist received two 1.3 cm cubes of beef from designated positions of each of two sets of four steaks which represented two replications. The samples in small covered glass jars had been warmed to [illegible] °C in Corningware

Table 3. Assignment of the three evaluation procedures[1] to panelists during each phase[2] of the study.

Panelist	Phase 1	Phase 2	Phase 3
1-3	CS	ME	LS
4-6	CS	LS	ME
7-9	LS	CS	ME
10-12	ME	CS	LS
13-15	LS	ME	CS
16-18	ME	LS	CS

---

[1] Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

[2] Each phase involved the evaluation of nine steaks per treatment.

Individual panelists' scores for each attribute were tabulated for statistical analysis. If data were missing for an attribute, the mean value of that panelist's scores for the other replications of that treatment was utilized for data analysis.

Panelists' impressions regarding each of the sensory evaluation procedures was obtained in two ways. After completing each phase of the study, each panelist was given an open ended questionaire to record his/her impressions of the evaluation technique just employed (Appendix, Figure 17, page 136). For each method, panelists were asked to comment on ease of learning, length of training provided, applicability for the sensory evaluation of beef, effort needed for sample evaluation and on scale accuracy. In addition, panelists were asked to list desirable and undesirable characteristics of each evaluation technique. Upon completion of the study, the panelists individually ranked the three evaluation procedures according to selected criteria (Appendix, Figure 18, page 137).

As a motivational tool and to show appreciation for the taste panelists' participation, small treats were given after each panel session. Judges also received a small honorarium at the end of the study.

RESULTS AND DISCUSSION

Because the electrical stimulation treatments (delay chilling with electrical stimulation (DC ES) and hot boning with electrical stimulation (HB ES)) resulted in few significant differences and contributed little to a discussion of sensory evaluation techniques for the descriptive sensory assessment of meat, the "Results and Discussion" have been limited to a discussion of data for the delay chilling (DC) and hot boning (HB) treatments. In this and subsequent sections, the abbreviations DC and HB are to indicate delay chilled and hot boned, respectively. For most of the tables presented in the text, corresponding tables presenting data for the four treatments appear in the Appendices (Tables 16 to 23, pages 138 to 145).

Objective Measurements

Means and standard errors for chemical measurements and cooking data for steaks from the DC and HB treatments are presented in **Table 4**. Data for pH and for percentages of ether extract and total moisture indicate that the steaks from both treatments were similar. The semimembranosus (SM) muscles used for this study were very lean with an average ether extract of 1.4 %. Hawrysh and Berg (1976) reported a similar ether extract (1.7 %) for the semitendinosus (ST) muscles of eight Grade A1 steers.

Table 4. Means and standard errors for chemical and cooking data for semimembranosus steaks from delay-chilled and hot-boned treatments.

Measurement	Postmortem Treatment Delay-Chilled	 Hot-Boned	SEM[1]
Chemical data[2]			
pH	5.5	5.6	0.07
Ether extract, %	1.3	1.4	0.21
Total moisture, %	73.2	73.7	0.30
Cooking data[3]			
Raw weight, g	457.7	449.6	52.31
Drip in thaw, %	5.6	7.8	0.20***
Initial temp., °C	4.8	6.0	1.39
Final internal temp., °C	65.1	65.1	0.16
Cooking time, min	55.3	58.4	2.75
Cooking time, min/100g	12.2	13.3	0.99
Cooking losses, %			
Total	23.1	22.1	1.29
Volatile	20.1	19.2	1.01
Drip	3.0	2.9	0.72

[1] Standard error of the mean.

[2] Values are the means of 18 determinations, 2 per replicate.

[3] Values are the means of 27 determinations, 1 per steak per replicate.

*** Significant at $P<0.001$.

The raw weight of steaks from the DC and HB postmortem treatments was similar (**Table 4**). However, HB steaks had a significantly greater percentage drip in thaw than comparable DC steaks. Since studies comparing the thaw losses of DC and HB treatments are lacking, the reason for this result obtained in the present study is not readily apparent. Delay-chilled and hot-boned treatments generally have similar thaw losses to their conventionally processed counterparts. Thaw losses were similar for longissimus steaks from DC (16° for 12 h) and control treatments (16° for 1 h, then 2°) (Crouse et al., 1983). Bowles Axe et al. (1983) observed that the percentage drip in thaw of SM muscles, boned 2 h postmortem and aged for either 24 h or 6 days, was similar to that of control SM muscles, boned 48 h postmortem and aged 0 or 6 days. Berry and Cross (1982) also reported that thaw losses of SM roasts boned 2 h postmortem did not differ from comparable roasts boned at 24 h postmortem.

There was no significant difference in the final internal temperature attained by DC and HB steaks (**Table 4**). Steaks were removed from the ovens at an average temperature of 65.1° (range, 64.75° to 65.50°) and did not exhibit a post oven temperature rise. Using a modified oven roasting technique, Moore et al. (1980) also noted that SM steaks cooked to 65° had no post oven temperature rise.

The average cooking times (min or min per 100 g) and percentage cooking losses for steaks showed no significant differences attributable to postmortem treatment. In the present study, steaks took an average total cooking time of 56.8 min to reach 65° with total, volatile and drip losses of 22.6 %, 19.6 % and 3.0 %, respectively. Moore et al. (1980) cooked 3.8 cm SM steaks in a rotary gas oven (177°) to 65° in 55.2 min and obtained total, volatile and drip losses of 20.6 %, 19.0 % and 1.6 %, respectively. Batcher and Deary (1975) roasted 3.8 cm SM steaks to either 60° or 71° and reported total cooking losses of 18.4 % and 26.2 %, respectively.

Few studies have compared the cooking losses of DC and HB treatments. However, Jeremiah et al. (1984), using similar meat and the same postmortem treatments as those in the present study, reported that cooking losses for DC and HB SM muscles were similar (**Table 5**). In addition, Jeremiah et al. (1984) found no significant differences in the cooking losses of longissimus (L) and triceps brachii (TB) muscles attributable to postmortem treatment.

Most researchers (Joseph and Connolly, 1977; Smith et al., 1979a, 1979b; Crouse et al., 1983; Jeremiah et al., 1984) have reported that cooking losses for DC meat cuts are similar to those of their conventionally-chilled counterparts. However, Smith et al. (1979a) observed that delay-chilled SM steaks had significantly greater cooking

Table 5. Means for objective and sensory data for semimembranosus steaks from delay-chilled and hot-boned treatments (adapted from Jeremiah et al., 1984).

Attribute	Postmortem Treatment		
	Delay-Chilled	Hot Boned	
Objective data			
Total cooking losses, %	31.0 ± 0.92[1]	30.2 ± 0.8[1]	
Warner Bratzler shear, kg	6.7 ± 0.24	8.0 ± 0.45	*
Sensory data[2]			
Initial tenderness	4.1 ± 0.21	3.2 ± 0.25	*
Overall tenderness	4.1 ± 0.17	3.2 ± 0.20	*
Connective tissue	4.4 ± 0.15	3.7 ± 0.18	*
Juiciness	4.3 ± 0.17	4.7 ± 0.18	
Flavor intensity	5.2 ± 0.09	5.3 ± 0.08	

[1] Standard error.

[2] Eight-point category scales. Higher values indicate increased tenderness, juiciness, flavor intensity and decreased connective tissue perceptibility.

* Significant at $P<0.05$.

losses than conventionally-chilled SM steaks.

Generally, the cooking losses of hot-boned SM muscles (cooked postrigor) do not differ significantly from muscles boned at 24 h postmortem or later (Kastner et al., 1973; Follett et al., 1974; Berry and Cross, 1982; Bowles Axe et al., 1981; Jeremiah et al., 1984).

Table 6 summarizes means and standard errors for the degree of doneness and for objective measurements of cooked SM steaks. The visual evaluation of steak doneness by a three member panel indicated that the HB steaks were more rare (P<0.01) than comparable DC steaks. However, steaks from both treatments had reached the same average internal temperature of 65.1° (**Table 4, page 56**).

Hunter color and press fluid determinations on the cooked DC and HB steaks supported these panel results which showed a significant difference in steak doneness between the two postmortem treatments. Although Hunter *L* and *b* values for both treatments were similar, the Hunter *a* (redness) value for HB steaks was higher (P<0.001) than that for comparable DC steaks. In addition, a significantly greater percentage of press fluid was released from HB steaks than from DC steaks (**Table 6**).

Although the HB steaks in the present study took slightly longer to cook than comparable DC steaks (**Table 4, page 56**), the difference was not statistically significant.

Table 6. Means and standards errors for the degree of doneness evaluation and objective measurements for semimembranosus steaks from delay-chilled and hot-boned treatments.

Measurement	Postmortem Treatment		
	Delay-Chilled	Hot-Boned	SEM[1]
Degree of doneness[2]	4.9	5.6	.14**
Hunter *L*[3]	44.9	42.6	[illegible]
a	6.8	8.4	0.09***
b	10.4	10.5	[illegible]
Press fluid, %[4]	36.5	40.1	0.6[illegible]**
OTMS-Warner Bratzler shear, kg[5]	7.7	8.7	0.66

[1] Standard error of the mean.

[2] Doneness scale: 1=well-done, 5=medium, 9=rare. Values are the means of 81 judgements, one per steak by each of three panelists.

[3] Values are the means of 54 determinations, 2 on each of three steaks per replication.

[4] Values are the means of 54 determinations, 3 on each of two steaks per replication.

[5] Values are the means of 156 determinations, with an average of 9 shears on each of two steaks per replication.

, * Significant at P<0.01 and P<0.001, respectively.

There is a lack of evidence in the literature to indicate that changes in the rate of heat penetration of this magnitude (approximately 5 %) would produce observable differences in the degree of doneness. Several workers (Shaffer et al., 1973; Batcher and Deary, 1975; Vollmar et al., 1976; McDowell et al., 1982) have shown that when rates of heat penetration between cooking methods vary considerably, differences in degree of doneness are sometimes noted.

The difference in doneness, Hunter *a* values and percentages of press fluid between the DC and HB steaks in the present study (**Table 6**) may have been due to postmortem processing treatment. However, comparisons of data for these particular measurements (degree of doneness, Hunter color and water-holding capacity) in cooked DC and/or HB meat are lacking.

Occasionally, raw HB meat has been found to be darker in color than conventionally-boned beef (Cross and Tennent, 1980; Claus, 1982). Jeremiah et al. (1984), using similar SM steaks and the same postmortem treatments as the present study, reported that HB steaks were darker, as determined instrumentally and subjectively, than comparable DC steaks. The water-holding capacity of raw meat cuts generally is not influenced by either DC (Cliplef and Strain, 1976) or HB (Cross and Tennent, 1980). However, Kastner et al. (1973) observed that the amount of press fluid from raw meat was greater for meat cuts excised at 5 and 6 h postmortem than for control cuts excised at 48 h. These findings for postmortem effects on color and water-holding capacity of raw meat are interesting. However, conclusions about these postmortem effects for cooked meat are not possible since limited information is available on the relationship between cooked and raw meat characteristics.

Although shear force values for HB steaks were higher than for comparable DC steaks, the difference was not statistically significant (**Table 6**). Studies comparing DC and HB treatments are lacking. Using steaks similar to those in the present study, Jeremiah et al. (1984) reported that HB steaks had significantly higher shear force values than the DC steaks (**Table 5, page 59**).

In several studies (Joseph and Connolly, 1977; Smith et al., 1979a, 1979b; Jeremiah et al., 1984), delay-chilled SM, ST and biceps femoris (BF) muscles received similar shear values to comparable conventionally-chilled cuts. The effect of DC on the tenderness of the L muscle has been inconsistent. Several workers (Smith et al., 1979a, 1979b; Crouse et al., 1983) have found no difference in shear values between DC and control L muscles. However, Cliplef and Strain (1976) and Lochner et al. (1980) reported that delay-chilled L muscles were significantly more tender than comparable conventionally-chilled muscles. In the study by Hostetler et al. (1975), a general improvement in tenderness by DC was observed; however, data were not presented to indicate the effects of delay versus conventional chilling on individual muscles.

Generally, the literature indicates that hot-boned muscles of the round (SM, ST, BF) cooked post-rigor are slightly tougher than their conventionally boned counterparts with the differences occasionally reaching statistical significance (Schmidt and Gilbert, 1970; Kastner et al., 1973; Schmidt and Keman, 1974; Falk et al., 1975; Kastner and Russell, 1975; Bowles Axe et al., 1983).

Sensory Assessments

Data for the descriptive sensory evaluation of DC and HB steaks obtained from 18 trained panelists using category scaling (CS), line scaling (LS) and magnitude estimation (ME) are summarized in **Table** 7. For the CS and LS evaluation techniques, higher values indicate increased tenderness, juiciness, flavor intensity and decreased connective tissue amount. For ME, higher values indicate increased toughness, connective tissue amount, juiciness and flavor intensity.

When the trained panel employed CS, the HB steaks were rated lower ($P<0.05$) in initial tenderness than comparable DC steaks (**Table** 7). For panelist evaluations by both LS and ME, HB steaks tended to be less tender than comparable DC steaks, although the differences were not statistically

Table 7. Means and standards errors for category scaling, line scaling and magnitude estimation of semimembranosus steaks from delay chilled and hot boned treatments.

Attribute	Postmortem Treatment		
	Delay Chilled	Hot Boned	SEM[1]
Initial tenderness			
Category scaling[2]	[illegible].4	4.[illegible]	[illegible]**
Line scaling[3]	[illegible]	[illegible]	[illegible]
Magnitude estimation[4]	[illegible]	[illegible]	[illegible]
Overall tenderness			
Category scaling	5.[illegible]	4.[illegible]	[illegible]*
Line scaling	8.9	[illegible].6	[illegible]
Magnitude estimation	16.[illegible]	[illegible]	[illegible]**
Connective tissue amount			
Category scaling	4.9	4.6	[illegible]
Line scaling	9.7	9.2	[illegible].36
Magnitude estimation	10.0	11.5	(0.02[illegible])
Overall juiciness			
Category scaling	4.7	5.0	0.05**
Line scaling	7.7	8.3	0.25
Magnitude estimation	12.3	14.4	(0.014)**
Flavor intensity			
Category scaling	5.0	5.0	0.[illegible]
Line scaling	7.9	8.0	0.23
Magnitude estimation	13.9	14.4	(0.012)

[1] Standard error of the mean.

[2] Maximum score=8. Higher values indicate increased tenderness, juiciness, flavor intensity and decreased connective tissue amount.

[3] Maximum score=15. Higher values indicate increased tenderness, juiciness, flavor intensity and decreased connective tissue amount.

[4] Modulus free scoring. Higher values indicate increased toughness, connective tissue amount, juiciness and flavor intensity.

[5] Standard error expressed as log 10 value.

*, ** Significant at P<0.05 and P<0.01, respectively.

significant. However, the treatment difference for the evaluation of initial tenderness by the LS technique approached significance (P=[illegible]).

Using an eight point category scale similar to that employed in the present study, Jeremiah et al. (1984) found that SM steaks from the HB treatment were significantly lower in initial tenderness than steaks from the DC treatment (Table 5, page 59). In addition, the difference in initial tenderness of 1.9 units between the DC and HB treatments reported by Jeremiah et al. (1984) was the same as that obtained for similar steaks in the present study. However, the actual scores reported by Jeremiah et al. (1984) were lower than the means obtained in the present study. Differences in cooking procedure may have contributed to this result. Jeremiah et al. (1984) roasted 2.5 cm SM steaks in a convection oven (177°) to an internal temperature of 75 ± 3°; in the present study, 3.8 cm steaks were roasted in conventional ovens (176°) to 65°.

Data for overall tenderness (**Table** 7) show that for CS and ME, HB steaks were less tender ($P<0.05$) than comparable DC steaks. When panelists employed LS, the overall tenderness of HB steaks was also lower than that of comparable DC steaks, but the difference did not reach statistical significance ($P=0.054$).

In this study, sensory assessments of tenderness tend to be supported by shear force data (**Table** 6). Although the difference was not statistically significant, the average shear force value for HB steaks was higher than that of comparable DC steaks.

Data from Jeremiah et al. (1984) for overall tenderness (Table 5, page 59) are consistent with the DC and HB results of the present study. Using an eight-point category scale, Jeremiah et al. (1984) obtained significantly lower overall tenderness scores for HB steaks than for comparable DC steaks. The magnitude of the tenderness difference that exists between DC and HB steaks in the study by Jeremiah et al. (1984) is the same as that obtained for similar SM steaks in the present study. As noted earlier for initial tenderness, variations in the cooking procedures employed by Jeremiah et al. (1984) and in the present study probably account for differences in the actual tenderness data obtained.

In several studies (Joseph and Connolly, 1977; Smith et al., 1979b; Jeremiah et al., 1984), trained panelists, using eight- or nine-point category scales, detected no significant differences in overall tenderness between delay and conventionally-chilled SM muscles.

Schmidt and Keman (1974) obtained similar tenderness scores (nine-point hedonic scale) for SM muscles excised at 1 and 24 h postmortem. Results of duo trio tests (Will et al., 1976) indicated that SM muscles boned at 3, 5 or 7 h postmortem did not differ significantly from control muscles cured at 48 h. Using eight point category scales, Berry and Cross (1982) found that SM roasts cured 1 h postmortem were similar in overall tenderness to cuts cured at 24 h. However, braised SM steaks from the HB treatment were significantly tougher than comparable SM steaks from the conventionally cured treatment (Berry and Cross, 1982).

In a recent study (Bowles Axe et al., 1983), panelists rated hot-boned SM steaks (aged 24 h) significantly lower on an eight point tenderness scale than comparable control cuts (aged 48 h). However, Bowles Axe et al. (1983) found that with six days of aging, HB and control cuts were similar in tenderness. In contrast, Jeremiah et al. (1984) reported that L, TB, and SM cuts excised 40 min postmortem and aged 6 days were tougher than their conventionally-boned counterparts.

Using CS, LS and ME, panelists did not detect significant differences in the amount of connective tissue between comparable steaks from the DC and HB postmortem treatments (**Table 7**). However, trends for CS and ME data (P=0.07 and P=0.08, respectively) indicate that more connective tissue was detected in the HB steaks than in comparable DC steaks. Using procedures similar to the present study, Jeremiah et al. (1984) observed that hot-boned SM steaks had significantly more detectable connective tissue than either control or delay-chilled SM steaks. Semimembranosus steaks from either a DC treatment

Smith et al. (1980) or a HB treatment. [illegible] et al., ([illegible]) reported connective tissue scores (eight point scale) similar to their conventionally processed counterparts. Berry and [illegible] ([illegible]) also used eight point [illegible] scale and reported that SM roasts boned 1 h postmortem were similar in amount of detectable connective tissue to those boned at 48 h. However, [illegible] SM steaks from the HB treatment had significantly more connective tissue than comparable conventionally boned SM steaks (Berry and [illegible], [illegible]).

When panelists employed CS and ME, they found that juiciness of HB steaks significantly higher (P<0.[illegible]) than that of comparable DC steaks (**Table 7**). In addition, with the LS technique, HB steaks tended (P<0.12) to be juicier than comparable DC steaks. Since steaks from the DC and HB treatments attained the same average internal temperature of 65.1° (range, 64.5° to 65.5°) (**Table 4, page 56**), juiciness differences were not expected. However, the significant differences for juiciness (**Table 7**) are supported by objective measurements (**Table 6**). Visual evaluation of steak doneness indicated that the HB steaks were more rare (P<0.01) than comparable DC steaks. Although Hunter *L* and *b* values for both treatments were similar, the Hunter *a* (redness) value for HB steaks was higher (P<0.001) than that of comparable DC steaks. In addition, a greater (P<0.01) percentage of press fluid was released from HB steaks than from comparable DC steaks.

Few studies have compared the juiciness of meat from the DC and HB postmortem treatments. Jeremiah et al. (1984), using an eight-point category scale, found that the juiciness of comparable DC and HB steaks was similar (**Table 5, page 59**). Delay-chilled and HB meat cuts generally have similar juiciness scores to their conventionally-processed counterparts (Smith et al., 1979b; Bowles Axe et al., 1983; Crouse et al., 1983; Crouse and Seideman, 1984; Jeremiah et al., 1984). However, in the study by Schmidt and Keman (1974), SM steaks from the HB treatment tended to be more juicy than comparable steaks from the control treatment. Although the difference was not statistically significant, juiciness means for the HB and control treatments were 5.1 and 4.1, respectively, on a nine-point scale (Schmidt and Keman, 1974).

In the present study, the difference in juiciness between the HB and DC steaks was very small and may not be of practical importance. However, it is interesting to note that for both the CS and ME procedures, the panelists were sensitive $(P<0.01)$ to these small differences in juiciness.

Employing each of the three evaluation techniques, trained panelists detected no differences in flavor intensity attributable to postmortem processing treatment (**Table 7**). Jeremiah et al. (1984) also reported that flavor scores for HB and DC steaks were not significantly different (**Table 5, page 59**). Flavor intensity and/or flavor desirability of either DC or HB SM muscles is usually similar to that of conventionally processed SM cuts (Schmidt and Keman, 1974; Smith et al., 1979b; Bowles Axe et al., 1983; Jeremiah et al.; 1984).

Results (Table 7) for the descriptive sensory assessment of the quality attributes of DC and HB steaks, show that evaluations by the 18 trained panelists were consistent among the sensory evaluation techniques. Although there were differences in statistical significance, the direction and extent of the difference between mean scores, assigned to each of the quality attributes in the DC and HB steaks, were generally similar for CS, LS and ME. Panelist assessments by each of the three sensory evaluation procedures indicated that HB steaks were tougher, more juicy and tended to have more detectable connective tissue than comparable DC steaks. Using CS, LS and ME, panelists assigned scores indicating similar flavor intensity to steaks from both postmortem treatments. In addition, for each evaluation technique, differences between steak treatment means tended to be large for initial tenderness and overall tenderness, intermediate for amount of connective tissue and juiciness, and slight for flavor intensity.

Distribution of Tenderness Data

Distributions of the overall tenderness assessments by the 18 panelists for the DC and HB postmortem treatments are illustrated in **Figures** 7 and 8. For evaluations of overall tenderness with CS, the frequency of usage of each category was plotted for each of the two postmortem treatments. Frequency of usage of appropriate scale intervals for assessments with LS and ME was also plotted. For each postmortem treatment, eight intervals were plotted for CS, 16 for LS, 20 for ME (real numbers) and 17 for ME (logarithms).

For the CS technique (**Figure 7a**), the entire eight-point scale was used for overall tenderness assessments with no apparent avoidance of the scale extremes. Steaks from the HB treatment tended to receive lower tenderness scores than comparable samples from the DC treatment. This finding supports results from the analysis of variance of these data which showed a significant difference in overall tenderness attributable to postmortem treatment.

For continuous distributions, the Kolmogorov-Smirnov goodness-of-fit test (Steel and Torrie, 1980) has been used to test whether or not the data follow a normal distribution. However, because CS data is from eight discrete categories, this test was not considered appropriate (Keeping, 1962).

Overall tenderness responses for LS for the two postmortem treatments were evenly distributed along most of the scale length (**Figure 7b**). However, for each postmortem

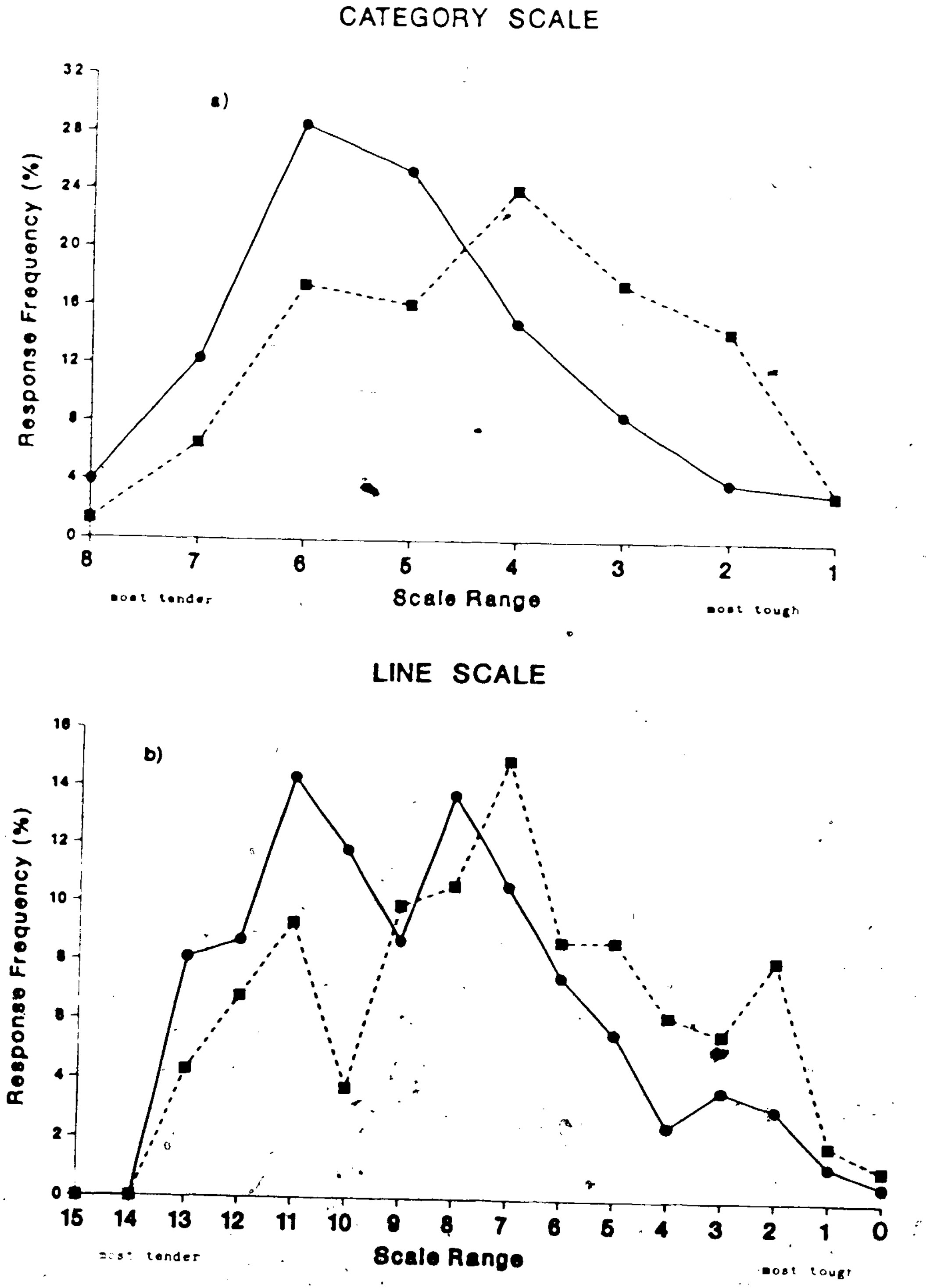
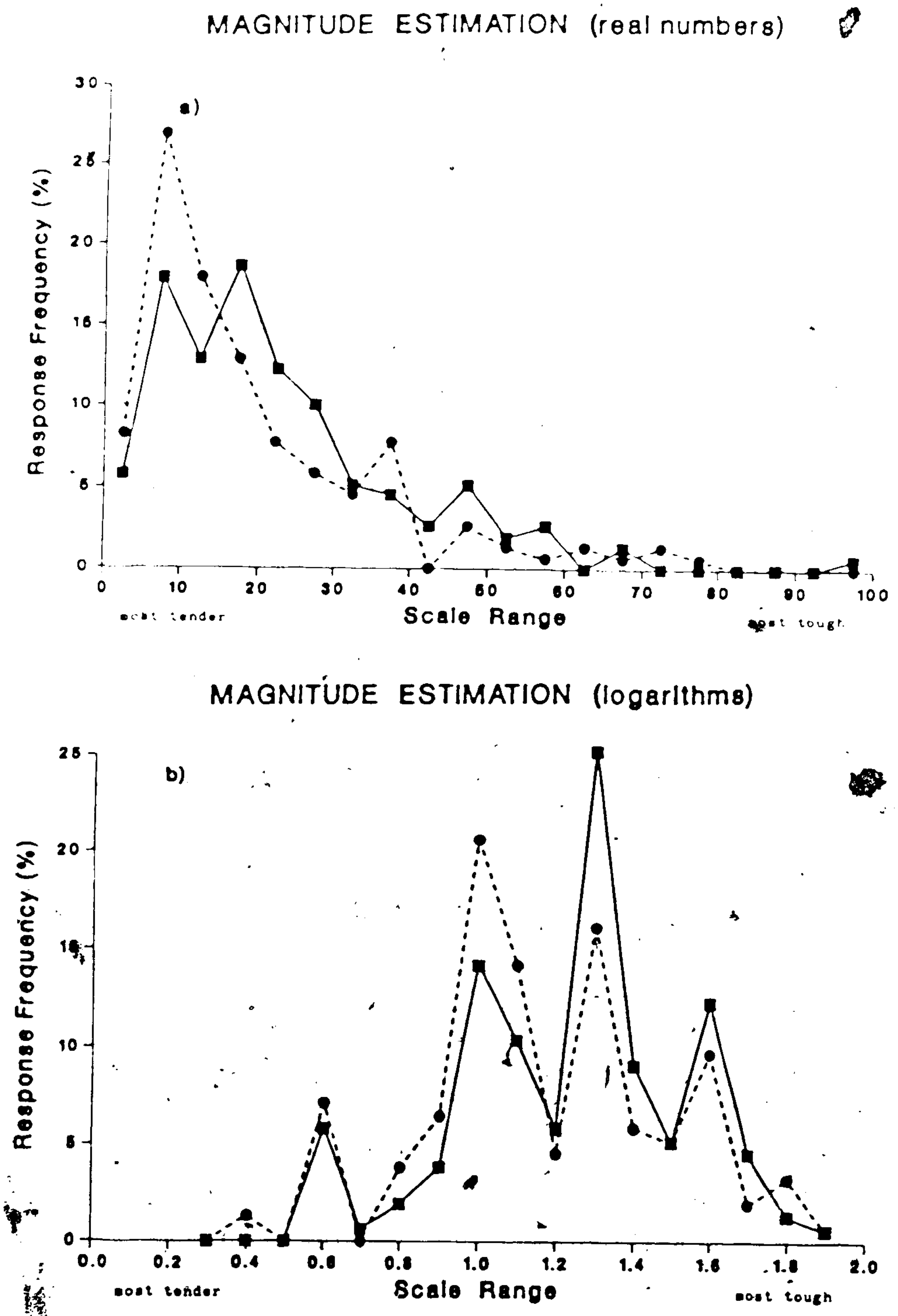


Figure 7. Frequency of panelists' usage of (a) the eight-point category scale (N=310) and (b) the 15 cm line scale (N=322) for the assessment of overall tenderness in delay-chilled (•) and hot-boned (■) steaks.


MAGNITUDE ESTIMATION (real numbers)

MAGNITUDE ESTIMATION (logarithms)

Figure 8. Frequency of panelists' usage of values for magnitude estimation (N=312) (a) before and (b) after logarithmic transformation of the assessment of overall tenderness in delay-chilled (●) and hot-boned (■) steaks.

treatment, few panelists provided ratings near the ends of the line scale. The descriptive anchors (located 1.3 cm from each end of the scale) (See **Figure 5, page 47**) may have constrained panelists' use of the scale extremes. As determined for CS, HB steaks tended to receive lower tenderness scores than comparable DC steaks. However, in the analysis of variance of overall tenderness data for LS, this difference in overall tenderness between HB and DC steaks did not reach statistical significance.

The frequency of panelists' usage of numerical values for the ME assessment of overall tenderness in steaks from DC and HB is depicted in **Figure 8a**. For each postmortem treatment, the ME responses are very skewed to the left. A logarithmic transformation of the tenderness data (**Figure 8b**) changed the distributions of ME responses from log-normal to approximately normal. Some points along the ME scale were more frequently used due the panelists' choice of certain favored numbers (10, 15, 20). Thus the transformed ME distributions for the two postmortem treatments remained somewhat jagged. Overall tenderness distributions for the ME responses (**Figure 8a** and **8b**) show that steaks from the HB treatment tended to be less tender than comparable steaks from the DC treatment. These data support results of the analysis of variance of ME overall tenderness data.

In general, panelist responses (**Figures 7 and 8**) by each evaluation technique show that the HB steaks tended to be less tender than comparable DC steaks. For each postmortem treatment, panelist assessments of overall tenderness with CS and LS were more evenly distributed than corresponding overall tenderness assessments with ME.

Sources of Variation in the Statistical Analyses

The magnitude and significance of F-values for selected sources of variation from the split-plot analyses of variance of category scale, line scale and magnitude estimation data were examined. The valid error terms used for testing the significance of these sources of variation were shown in **Figure 1, page 31.**

Treatment F-Values

Treatment F-values (**Table 8**) were used as a measure of the sensitivity of each evaluation technique to differences in palatability due to postmortem treatment. For initial tenderness, the treatment F-value for CS was significant and larger than comparable F-values for the LS and ME methods. The LS F-value approached statistical significance (P=0.056). For ME, the treatment F-value was very low and statistically nonsignificant. Thus, for initial tenderness, the CS technique was most sensitive to treatment differences, followed by the LS technique. When panelists used the ME method, differences in initial tenderness

Table 8. Treatment F-values for quality attributes from the analyses of variance of category scale, line scale and magnitude estimation data.

Attribute	Treatment F-values		
	Category Scaling	Line Scaling	Magnitude Estimation
Initial tenderness	6.62*	4.97	[illegible]
Overall tenderness	7.33*	5.[illegible]	[illegible]*
Connective tissue amount	4.33	1.21	[illegible]
Overall juiciness	13.55**	3.03	12.26**
Flavor intensity	0.00	0.10	[illegible]

*, ** Significant at P<0.05 and P<0.01, respectively.

between the DC and HB treatments were poorly discriminated.

Difficulties with the ME procedure may have contributed to panelists' inability to discriminate differences in initial tenderness between the two postmortem treatments. Because of the effort and time required for sample mastication, panelists noted difficulty in remembering their impressions of one sample relative to another. However, Moskowitz (1983) observed that panelists develop their own internal frame of reference for assigning scores to each sample and thus may not need to remember attributes in successive samples.

The F-values for overall tenderness for CS and ME were significant (P<0.05) and larger than the F-value from the LS analysis of variance (Table 8). However, the LS F-value was close to statistical significance (P=0.054). Thus, for overall tenderness, the CS and ME procedures were equally sensitive to treatment differences and superior to the LS technique.

Initial tenderness F-values for the CS and LS techniques, respectively, were similar to those obtained for overall tenderness. However, the ME treatment F-value for overall tenderness was larger than the corresponding F-value obtained for initial tenderness, indicating that panelists were better able to detect differences in overall tenderness than in initial tenderness between the two postmortem treatments.

Although none of the treatment F-values for connective tissue amount (Table 8) was statistically significant, the F-values for the CS and ME techniques were slightly larger than that determined for the LS method.

For overall juiciness, treatment F-values for CS and ME were similar and highly significant (P<.[illegible]) (Table 8). Analysis of juiciness data from the LS technique for the MR and DC treatments resulted in a low nonsignificant F-value. Thus, the CS and ME evaluation procedures were more sensitive than the LS method for the detection of treatment differences related to meat juiciness.

Flavor intensity, a trait not expected to be affected by any of the postmortem treatments, had very low treatment F-values for each sensory evaluation technique (Table 8).

In general, the data (Table 8) indicate that the evaluation techniques differed in their sensitivity to treatment differences. The CS technique was most sensitive to treatment differences as indicated by the three significant F-values obtained for initial tenderness, overall tenderness and juiciness. For LS, F-values for each of the five attributes were not statistically significant. For ME, treatment F-values for overall tenderness and juiciness were significant and similar to those obtained for CS; however, the F-value for initial tenderness was not statistically significant.

In the present study, the CS technique was considerably more sensitive to differences in meat quality than the LS technique. The panelists' inability to distinguish treatment differences by the LS procedure was unexpected. Stone et al. (1980) have reported success with line scales for "Quantitative Descriptive Analysis" (QDA). However, few studies have compared the sensitivity of the LS technique to other evaluation methods such as CS.

Symonds (1924) observed that although graphic (line) scaling allows a panelist to discriminate as finely as desired, this scaling technique does not force such a judgement. However, for assessments of apple quality, judges distinguished a greater contrast between the two apple varieties using LS than CS (Baten, 1946). In addition, statistical analyses of the LS data resulted in larger t-values than those for the CS data (Baten, 1946). Variation in LS procedures between the study of Baten (1946) and the present study may have contributed to the differences in the results obtained in the two studies. Baten (1946) allowed and encouraged a direct comparison of LS ratings between samples by having panelists score the apple quality of two samples on adjacent line scales. However, in the present study, a separate line scale scorecard was used for each sample under investigation.

The sensitivity of a nine-point category scale and of a modified line scale (nine divisions were marked) to differences in meat quality were compared by Raffensperger et al. (1956). For both CS and LS, analyses of variance showed similar significance (P<.01) for treatment effects, grade and the three beef cuts evaluated.

In the present study, the ME technique was as sensitive as CS to most treatment differences. Comparisons of CS and ME procedures have frequently been made by untrained panels for hedonic evaluations of food products (Moskowitz and Sidel, 1971; McDaniel and Sawyer, 1981a; Moskowitz, 1977). However, few studies using trained panelists have compared CS and ME for the descriptive assessment of food attributes. McDaniel and Sawyer (1981a) used nine point CS and modulus free ME to score the intensity of 19 descriptive profile terms of whiskey sour formulations. The ME tehnique yielded a similar number of significant differences to CS (13 significant differences noted for ME; 14 for CS) (McDaniel and Sawyer, 1981a).

Results (Table 8) indicate that the ME technique was superior to the LS technique in sensitivity to differences between the two postmortem meat treatments. However, in a recent study, Giovanni and Pangborn (1983) noted that the LS procedure was more sensitive than the ME procedure to differences in intensity among six levels of fat in milk (0, 1, 2, 4, 8, 16 %) and seven concentrations of sucrose in lemonade (4, 6, 8, 10, 14, 20, 30 %). For LS and ME data, statistically significant differences (P<0.001) were found among the levels of fat in milk and among the sucrose concentrations in lemonade. However, for the LS technique, treatment F-values for fat level and sucrose concentration (331.9 and [illegible], respectively) were larger than the corresponding F-values for ME (152.7 and 28.2, respectively) (McDaniel and Pangborn, 1983).

Replication F-Values

Replication F-values (Table 9) for each of the three evaluation techniques were similar and nonsignificant for most of the quality attributes under investigation. However, the CS replication F value for overall juiciness was significant (P<0.01). Significant (P<0.05) replication F-values for overall juiciness and flavor intensity were also determined for ME data. For all attributes, replication F-values for LS were lower than corresponding F-values for CS and ME.

The magnitude of the replication F-value reflects the variation in scoring among the nine replications evaluated. For CS and ME, there was a greater tendency for the panel to assign high scores at one session and low scores at another than for the LS technique. However, replication variation only reached statistical significance once for CS and twice for ME.

In a comparative evaluation of whiskey sours by CS and ME, McDaniel and Sawyer (1981a) observed that replication

Table 9. F values for selected sources of variation from the analyses of variance of category scale, line scale and magnitude estimation data.

	Source of Variation		
Attribute	Replication	Panelist	Treatment by Panelist
Initial tenderness			
Category scaling	[illegible]	[illegible]***	[illegible]
Line scaling	[illegible]	[illegible]***	[illegible]
Magnitude estimation	[illegible]	[illegible]***	[illegible]
Overall tenderness			
Category scaling	[illegible]	[illegible]***	[illegible]
Line scaling	[illegible]	[illegible]***	[illegible]
Magnitude estimation	[illegible]	[illegible]***	[illegible]
Connective tissue amount			
Category scaling	2.68	5.55***	[illegible]
Line scaling	[illegible]	[illegible]***	[illegible]
Magnitude estimation	[illegible]	[illegible]***	[illegible]
Overall juiciness			
Category scaling	[illegible]**	4.22***	[illegible]
Line scaling	[illegible]	[illegible]***	[illegible]
Magnitude estimation	[illegible]*	24.20***	[illegible]*
Flavor intensity			
Category scaling	1.68	8.38***	[illegible]
Line scaling	0.88	9.81***	[illegible]
Magnitude estimation	[illegible]*	25.27***	[illegible]

*, **, *** Significant at P≤0.05 and P<0.01, and P<0.001, respectively.

F-values for CS were significant 14 times; however, for ME, replications were not significant in any of the 114 comparisons made.

Panelist F-Values

Panelist F-values (**Table 9**) were significant ($P<0.001$) for each quality attribute assessed with each of the three evaluation methods. For each attribute, the CS technique had the lowest panelist variation; LS was intermediate and ME had the highest variation. The magnitude of the panelist F-values for each of the evaluation procedures probably reflects the scale range used for each technique. For CS, panelists worked within a confined scale range of 1 to 8 units; for LS, the scale range was 0 to 15 units; and for ME, panelists used any numbers they wanted to score the samples.

Significant panel variability in sensory tests is considered common and to be expected (McDaniel, 1974). Cross et al. (1978) determined significant panelist F-values for tenderness, juiciness and connective tissue (eight-point category scales) for each of four trained descriptive attribute panels under investigation.

In an evaluation of whiskey sour formulations, McDaniel and Sawyer (1981a) used trained panelists to compare nine-point CS nd modulus free ME. One significant panelist F-value was observed for ME; and 83 for CS. However, in their study (McDaniel and Sawyer, 1981a), the ME data had been normalized prior to statistical analysis. In the present study, panelist variability (due to scale range) was removed in the analysis of variance procedure as the panelist source of variation.

Treatment by Panelist F-Values

F-values for the treatment by panelist interaction are shown in **Table 9**. Except for the ME assessment of overall juiciness, all treatment by panelist F-values were not significant indicating that individual panelists agreed with one another when rating the treatments by CS, LS or ME. Examination of the data for overall juciness assessments by ME revealed that two panelists had scored the DC and HB treatments opposite to the scores assigned by the other 16 panelists. Since these panelists had performed satisfactorily in evaluating all quality attributes during training and in judging the other attributes during the study, no reason for their method of assigning juiciness scores is readily apparent.

In general, the data (**Tables** 8 and 9) show that the magnitude and significance of F-values, especially treatment F-values, were useful criteria for a comparison of the three evaluation techniques. Several workers (Moskowitz, 1982; Giovanni and Pangborn, 1983) have also used treatment F-values as criteria for comparing sensory evaluation techniques.

Relationships Between Evaluation Techniques

In addition to subjecting the data to analyses of variance, correlation and regression analyses were performed to assess relationships between each of the evaluation techniques. Data for these analyses were averaged over all panelist responses for each of the 18 roasts (nine per treatment).

Pearson correlation coefficients and coefficients of determination between each of the three evaluation procedures are presented in **Table 10**. For initial tenderness, correlation coefficients for CS, LS and ME data were highly significant ($P<0.001$) and indicated strong linear relationships between the scores assigned by judges using each of the evaluation methods (**Table 10**). Shindell (1964) suggested that correlation coefficients were low, if $r\leq 0.39$; moderate, if r is between 0.40 and 0.79; and high, if $r\geq 0.80$, irrespective of the sign. In the present study, the assignment of higher numbers indicated greater tenderness for both CS and LS. High numbers signified toughness with ME. Thus, the positive and negative relationships (**Table 10**) are as expected. Coefficients of determination show that 74 to 86 % of the variation in the scoring of initial tenderness may be explained by the relationships between each of the three sensory evaluation procedures.

Table 10. Pearson correlation coefficients (r) and coefficients of determination (r^2) between panel assessments from each evaluation technique[1].

Attribute	r	r^2
Initial tenderness		
CS vs LS	0.93***	0.86
CS vs ME	-0.86***	0.74
LS vs ME	-0.91***	0.82
Overall tenderness		
CS vs LS	0.93***	0.86
CS vs ME	-0.92***	0.84
LS vs ME	-0.92***	0.84
Connective tissue amount		
CS vs LS	0.54*	0.29
CS vs ME	-0.34	0.12
LS vs ME	-0.53*	0.28
Overall juiciness		
CS vs LS	0.44*	0.20
CS vs ME	0.74***	0.54
LS vs ME	0.65**	0.42
Flavor intensity		
CS vs LS	0.37	0.14
CS vs ME	0.59**	0.35
LS vs ME	0.36	0.13

[1] Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

*, **, *** Significant at $P<0.05$, $P<0.01$ and $P<0.001$, respectively.

Correlation coefficients for overall tenderness between each of the evaluation techniques were similar in magnitude to those determined for initial tenderness (**Table 10**). All correlations were highly significant (P<0.001) and accounted for 84 to 86 % of the variation in these data. The coefficient of determination between CS and ME data for overall tenderness was 10 % higher than that determined for initial tenderness. **Figures 9 to 11** illustrate the linear nature of the relationships between overall tenderness assessments for each of the three evaluation procedures.

In a study conducted by Cross et al. (1978), four panels, given similar training, evaluated similar L steaks using eight-point category scales. Tenderness assessments were highly correlated across panels (r-values ranged from 0.88 to 0.94) (Cross et al., 1978).

Recently, Dransfield et al. (1982) compared the eating quality of beef assessed at each of five European research institutes. For a common eight-point tenderness scale, correlation coefficients between the trained panel scores from each of the five institutes ranged from 0.69 to 0.94. When panel assessments using each institute's usual tenderness scale (six to 11 categories in length) were correlated, r-values of 0.73 to 0.94 were determined (Dransfield et al., 1982).

For connective tissue amount, correlations between the three evaluation methods were moderate to low and accounted

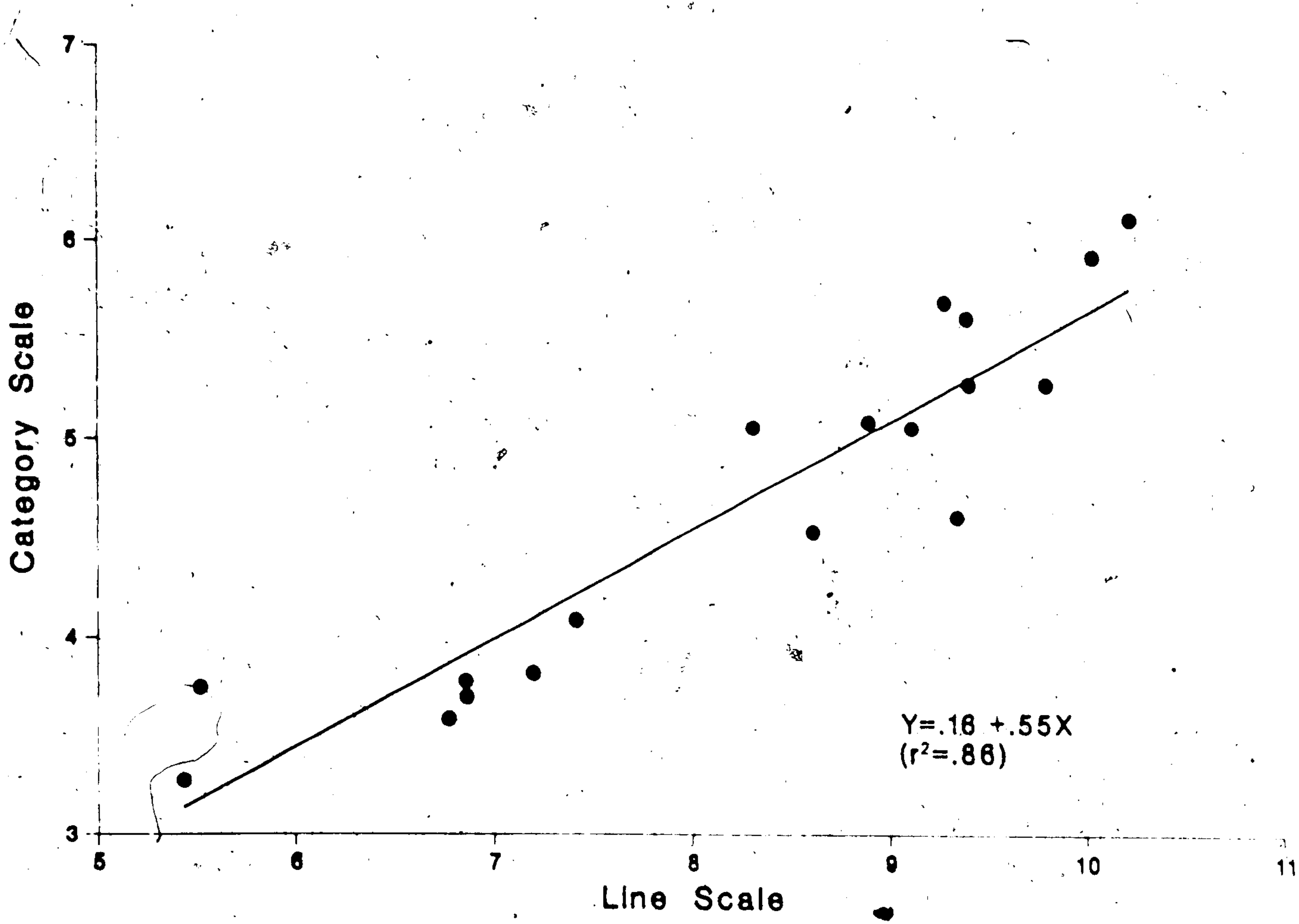


Figure 9. Relationship between panel overall tenderness assessments by category scaling and line scaling (N=18).

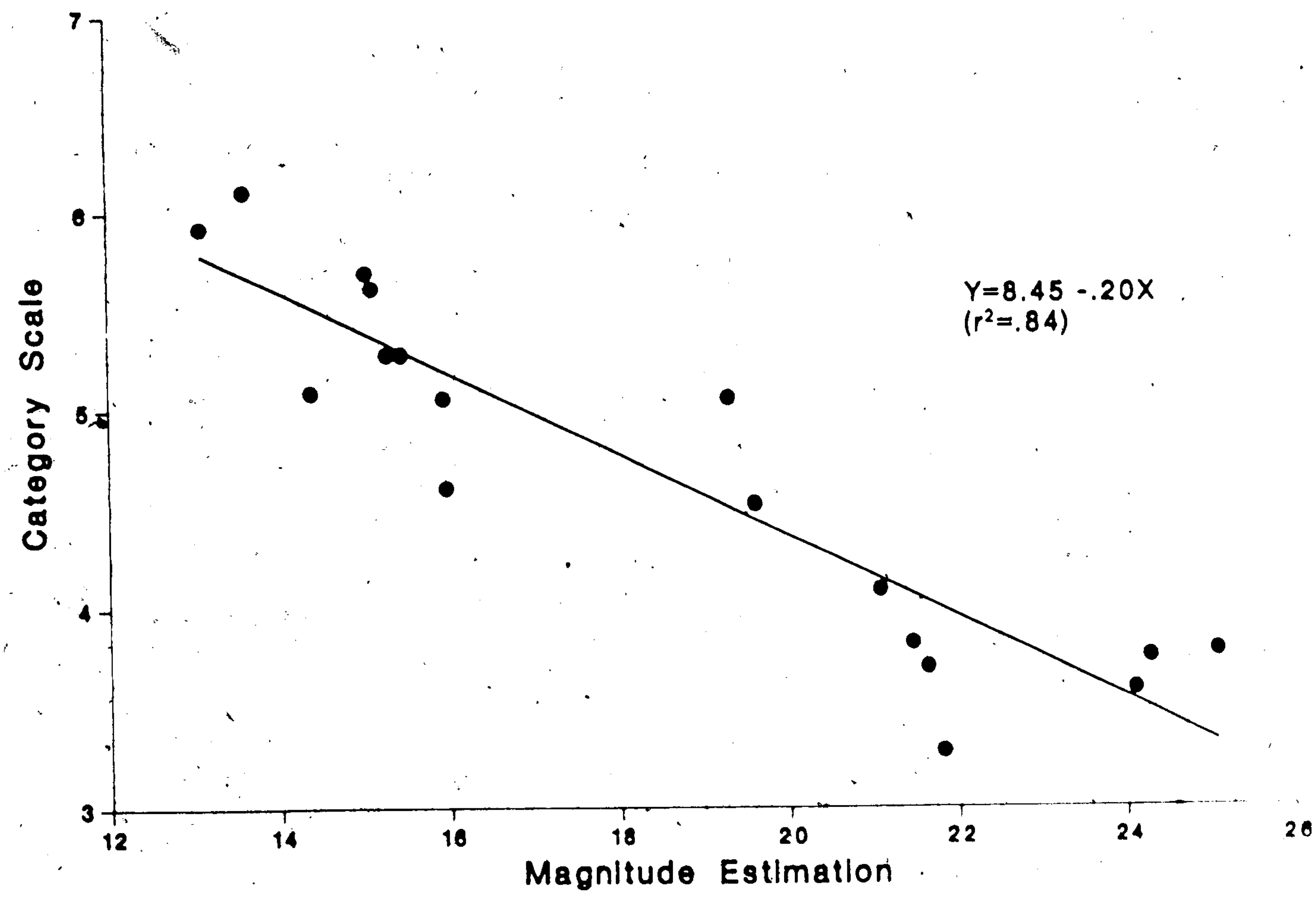


Figure 10. Relationship between panel overall tenderness assessments by category scaling and magnitude estimation (N=18).

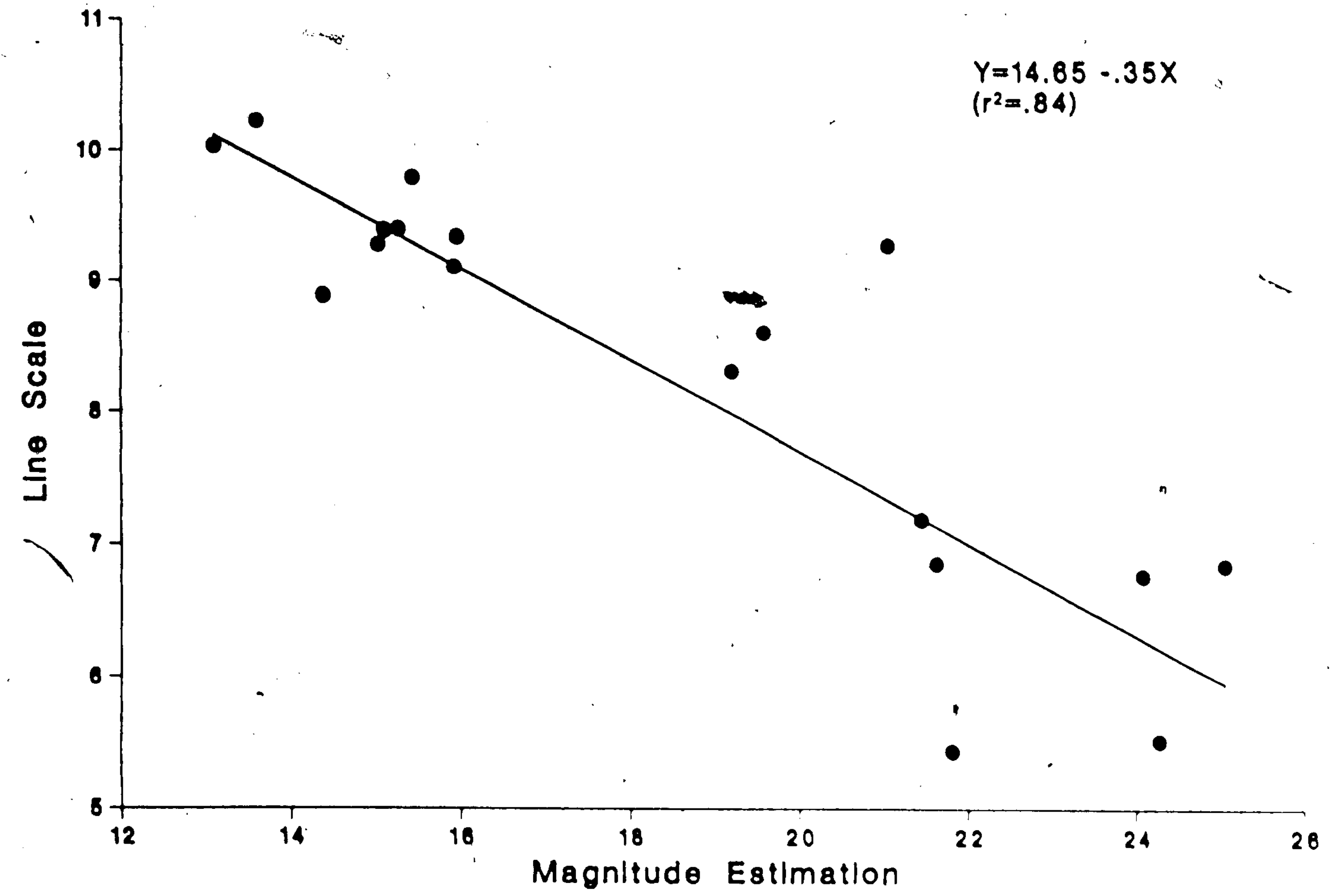


Figure 11. Relationship between panel overall tenderness assessments by line scaling and magnitude estimation (N=18).

for only 12 to 29 % of the variation in the data (**Table 10**). In the present study, sample heterogeneity may have contributed to these low correlations for connective tissue amount between the evaluation procedures. In addition, for each evaluation technique, mean connective tissue values for the 18 roasts were from a smaller stimulus range than the corresponding range in means for overall tenderness for each of the roasts (**Appendix, Table 24, page 146**). Thus, for the category scale, the range in means for the 18 roasts was 1.8 units for connective tissue amount and 2.8 units for overall tenderness. Several researchers (Szczesniak, 1968; Cross et al., 1978) have noted that correlation of values from a small stimulus range may result in low correlation coefficients.

Cross et al. (1978) obtained r-values of 0.79 to 0.86 for assessments of connective tissue amount between each of four trained panels. Similar L steaks from 11 maturity/marbling groups were evaluated by each panel. In the present study, the SM steaks were obtained from young animals (15 mo) of similar grade (A1) and thus might not be expected to contain wide ranges of connective tissue.

Correlation coefficients for overall juiciness between each of the sensory procedures were moderate (**Table 10**). Since all steaks were cooked to the same final internal temperature, the range of juiciness scores for each evaluation technique was small (**Appendix, Table 24, page**

146). For steaks cooked to one degree of doneness, several workers (Cross et al., 1978; Dransfield et al., 1982) have also reported that correlations of juiciness between different panels were moderate to low.

Flavor intensity data yielded low to moderately low correlation coefficients between each of the three evaluation techniques (**Table 10**). Thirteen to 35 % of the variation in the data could be explained by the relationships between data from each procedure. A small stimulus range for flavor intensity was noted since steaks from the two postmortem treatments received similar flavor scores (**Appendix, Table 24, page 146**).

Dransfield et al. (1982) correlated flavor intensity scores between each of three European research institutes and obtained r-values (0.41 to 0.51) similar to those of the present study (**Table 10**).

In general, correlation coefficients betweeen each of the evaluation techniques (**Table 10**) for initial and overall tenderness were highly significant and indicated strong linear relationships between scores assigned by judges using CS, LS and ME. For the remaining attributes, the correlation between each of the three methods, while sometimes significant, were moderate to low. In the present study, sample heterogeneity, especially for connective tissue amount, and a narrow stimulus range, particularly for juiciness and flavor intensity, may have contributed to the low correlations obtained (Szczesniak, 1968; Cross et al., 1978).

Table 11 summarizes coefficients of determination for tenderness assessments from each of the three evaluation techniques fitted to five mathematical functions. The coefficients of determination obtained between each of the evaluation techniques were similar. None of the mathematical functions provided a consistently superior fit of the data.

For the parabolic functions between CS and ME tenderness data, the quadratic (X^2) term did not make a significant contribution to the function. Thus, for initial and overall tenderness, the relationships between CS and ME data are not curvilinear. The linear nature of the relationship betweem CS and ME data for overall tenderness was shown in **Figure 10, page 90**. In contrast, Cardello et al. (1982b) rescaled six standard texture scales using ME and found that the CS data was concave downward relative to the ME data. This indicated that at higher intensities of each attribute evaluated, CS was less sensitive than ME to differences in these attributes (Cardello et al., 1982b). In the present study, however, data for the evaluation of beef samples using CS did not cover the entire stimulus range of the scale and thus may have influenced the shape of the resulting functions.

Table 11. Coefficients of determination (R^2) for initial and overall tenderness between each of the three evaluation techniques[2] fitted to five mathematical functions

Comparison Y vs X	Linear (y=a+bx)	Power (y=ax^b)	Logarithmic (y=a+b ln x)	Hyperbolic (y=a+b(1/x))	Parabolic (y=a+bx+cx²)
Initial tenderness					
CS vs LS	0.86	0.89	[illegible]	[illegible]	[illegible]
LS vs CS	0.86	0.89	[illegible]	[illegible]	[illegible]
CS vs ME	0.74	0.76	[illegible]	[illegible]	[illegible]
ME vs CS	0.74	0.76	[illegible]	[illegible]	[illegible]
LS vs ME	0.82	0.79	[illegible]	[illegible]	[illegible]
ME vs LS	0.82	0.79	[illegible]	[illegible]	[illegible]
Overall tenderness					
CS vs LS	0.86	0.86	[illegible]	[illegible]	[illegible]
LS vs CS	0.86	0.86	0.81	[illegible]	[illegible]
CS vs ME	0.84	0.85	0.85	[illegible]	[illegible]
ME vs CS	0.84	0.85	0.81	[illegible]	0.81
LS vs ME	0.84	0.79	0.81	[illegible]	0.81
ME vs LS	0.84	0.79	[illegible]	[illegible]	[illegible]

[1] All coefficients of determination were significant at P<0.001

[2] Category scaling (CS), line scaling (LS) and magnitude estimation (ME)

Relationships Between Sensory and Instrumental Measures of Tenderness

Correlation coefficients and coefficients of determination between panel tenderness assessments from each evaluation procedure and Warner Bratzler shear data are presented in **Table 12**. For each evaluation technique, correlations of initial and overall tenderness with Warner Bratzler shear values yielded similar significant r values (**Table 12**). The positive and negative relationships are as would be expected since for CS and LS data, high numbers indicate tenderness; and for ME and shear force data, high numbers signify toughness.

For CS and LS data, r-values between initial tenderness and shear were similar to corresponding r-values between overall tenderness and shear. However, the data for the ME technique (**Table 12**) suggest that the relationship between overall tenderness and shear was slightly better than that for initial tenderness and shear.

Even though all correlations of subjective tenderness data with shear force measurements were statistically significant, the correlations only accounted for 40 to 51 % of the variation in the data. Sample heterogeneity may have contributed to the unexplained data variation (Kap[illegible]lis and Moskowitz, 1977). In the present study, the high unexplained variation in the data may be a reflection of the lack of sensitivity of the Warner Bratzler shear to

Table 12. Pearson correlation coefficients (r) and coefficients of determination (r^2) between panel tenderness assessments from each evaluation technique[1] and Warner Bratzler shear data.

Attribute	r	r^2
Initial tenderness		
CS	-0.71***	0.51
LS	-0.68***	0.47
ME	0.63**	0.40
Overall tenderness		
CS	-0.68***	0.46
LS	-0.69***	0.48
ME	0.70***	0.49

[1] Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

, * Significant at P<0.01 and P<0.001, respectively.

structural components that influence taste panel assessments. Gullett et al. (1983) reported that shear values may not be related to panelists' perceptions of meat tenderness.

Correlations between Warner Bratzler shear values and sensory tenderness assessments vary considerably in the literature (Szczesniak and Torgeson, 1965). In addition, correlation coefficients are difficult to compare between studies. The magnitude of the correlation coefficient obtained in any study is highly dependent upon the range of values covered, the number of samples involved and the amount of variation within the block of samples (Szczesniak, 1968).

For the SM muscle, correlation analysis of tenderness assessments by the CS technique with shear values have resulted in r-values of -0.81 (Breidenstein et al., 1968) to -0.53 (Smith et al., 1978) or lower (McCurdy et al., 1981). However, studies reporting correlations of panelist assessments of tenderness using the LS method with Warner Bratzler shear values are lacking. In addition, no reports have correlated ME tenderness values with Warner Bratzler shear data. However, Segars et al. (1975) reported correlation coefficients ranging from 0.92 to 0.98 for texture evaluations (difficulty of cutting, chewiness and residue) by the ME procedure and an instrumental measure of tenderness (punch and die test cell). Since only six muscles representing extremes in tenderness were used, the results of this study must be interpreted with caution. Recently, Segars et al. (1981) obtained a correlation of 0.76 between magnitude estimates of hardness in irradiated restructured meat products and a punch and die shear cell.

Correlations of sensory scores and mechanical measurements have been used as an indicator of panel accuracy (Hovenden et al., 1979). Results of the present research (**Table 12**) show that each of the evaluation techniques yeilded similar significant r-values.

Summarized in **Table 13** are coefficients of determination between panel tenderness assessments from each evaluation technique and Warner Bratzler shear data fitted to five mathematical functions. Similar coefficients of determination between sensory assessments of tenderness and shear data were determined for each of the five mathematical functions. Thus, the CS, LS and ME data can be explained equally well by each of these functions. Furthermore, linear, logarithmic, hyperbolic and parabolic functions described ME data as well as the more traditionally used power function. This finding agrees with a recent report by Giovanni and Pangborn (1983).

The power functions between panel tenderness assessments using ME and Warner Bratzler shear data are illustrated in **Figure 12**. The exponent or slope of the power function is thought to be an index of perceptual

Table 13. Coefficients of determination (R²) between panel tenderness assessments from each evaluation technique¹ and Warner Bratzler shear data fitted to five mathematical functions

	Mathematical function				
Attribute	Linear ($y=a+bx$)	Power ($y=ax^b$)	Logarithmic ($y=a+b\ln x$)	Hyperbolic ($y=a+b(1/x)$)	Parabolic ($y=a+bx+cx^2$)
Initial tenderness					
CS	0.51	0.53	0.51	0.52	0.51
LS	0.47	0.47	0.36	[illegible]	0.47
ME	0.40	0.3[illegible]	0.39	0.3[illegible]	0.4[illegible]
Overall tenderness					
CS	0.46	0.47	0.36	0.4[illegible]	0.46
LS	0.48	0.48	0.4[illegible]	[illegible]	0.49
ME	0.49	0.4[illegible]	0.48	0.4[illegible]	0.4[illegible]

Coefficients of determination ≥ 0.46 were significant at P<0.001.
Coefficients of determination ≥ 0.31 were significant at P<0.01.

¹Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

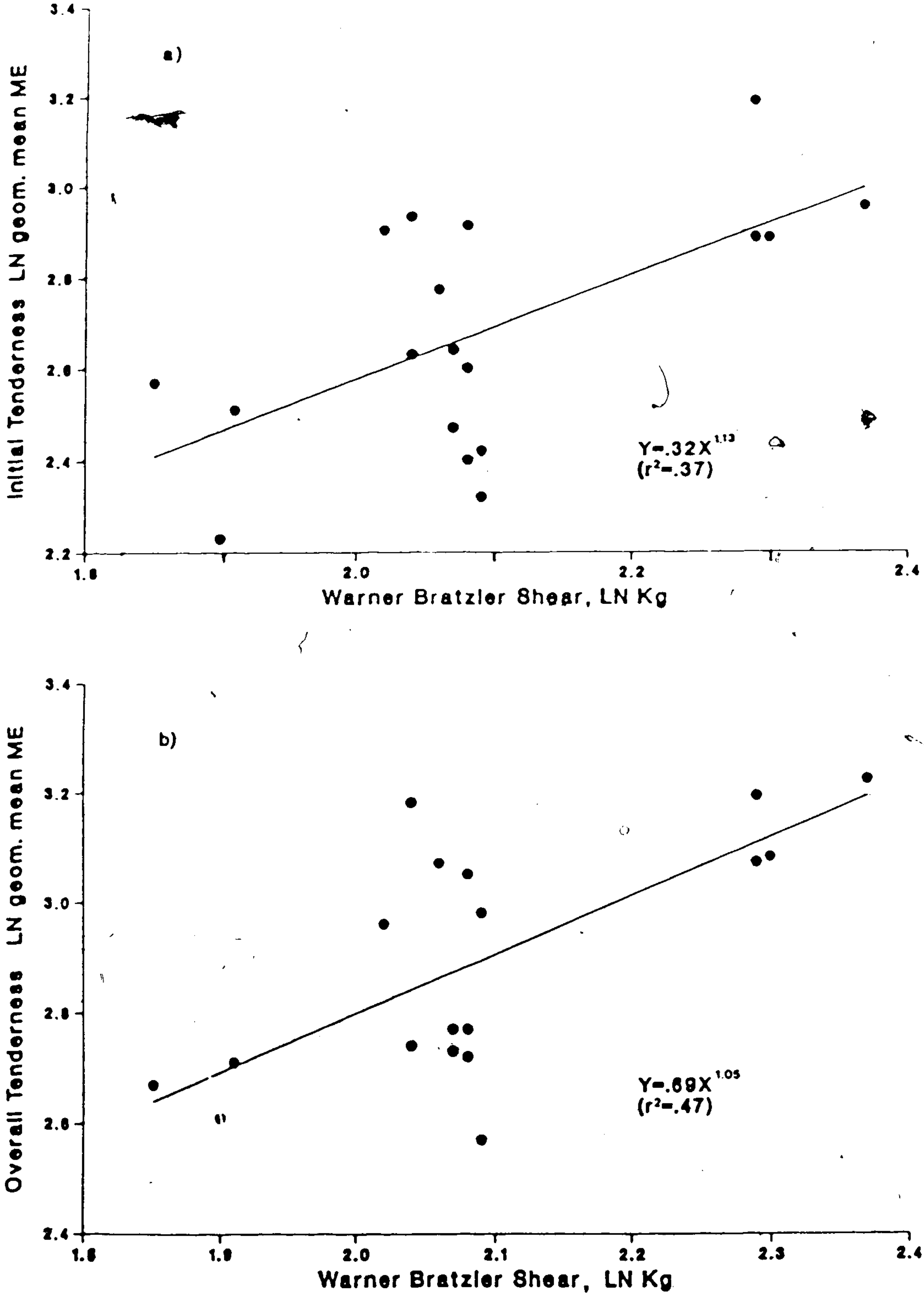


Figure 12. Power functions between panel (a) initial tenderness and (b) overall tenderness assessments by magnitude estimation (ME) and Warner Bratzler shear (N=18).

sensitivity . For initial tenderness and overall tenderness, the power functions relating ME and shear data yielded exponents of 1.13 and 1.05, respectively. These exponents indicate that panel tenderness assessments using ME and the Warner Bratzler shear determinations on the Ottawa Texture Measuring System were similar in sensitivity to tenderness differences. However, because of data variability shown in **Figure 12**, the standard errors for initial tenderness (0.37) and overall tenderness (0.28) are large. Thus the exponents determined in the present study may not be reliable indicators of sensory response. Segars et al. (1975) reported that exponents relating textural properties of meat to an instrumental measure (punch and die shear device) ranged from 1.8 to 2.6. However, comparison of exponents between studies is difficult because exponents may be dependent on stimulus range and response range (Birnbaum, 1982).

Scale Preference

Mean values for panelist rankings of selected criteria for the three evaluation techniques are presented in **Table 14.** In addition, **Table 15** provides data regarding panelist response by rank for each evaluation procedure. Comments provided by panelists using the open-ended questionaire (**Appendix, Figure 17, page 136**) have been incorporated into the discussion of each criterion.

Table 14. Mean values[1] for panelist ranking of selected criteria for category scaling, line scaling and magnitude estimation.

Criteria[2]	Evaluation Technique		
	Category Scaling	Line Scaling	Magnitude Estimation
Ease of learning	1.4b	1.6b	3.0a
Applicability	1.8	1.8	2.2
Scale accuracy	1.3c	2.1b	2.6a
Effort for Sample evaluation	2.4a	2.4a	1.2b
Scale preference	1.4c	2.1b	2.7a

[1] Values are the mean of ranks assigned by the 18 panelists. For applicability, values are the mean of ranks assigned by 17 panelists.

[2] Ranking based on:
1=easiest, most applicable, most accurate, most effort and most preferred;
3=most difficult, least applicable, least accurate, least effort and least preferred.

a,b,c Means within the same row sharing a common superscript are not significantly different at $P<0.05$.

Table 15. Frequency of panelist ranks[1] assigned to each of the evaluation techniques for selected criteria.

Criteria	Rank	Frequency of Ranks[1] Category Scaling	 Line Scaling	 Magnitude Estimation
Ease of learning	1 (easiest)	56	44	0
	2	44	56	0
	3 (most difficult)	0	0	100
Applicability for evaluation of beef	1 (most)	35	41	24
	2	47	35	18
	3 (least)	18	24	58
Scale accuracy	1 (most)	78	22	0
	2	17	44	39
	3 (least)	6	33	61
Effort needed to evaluate each sample	1 (most)	11	0	89
	2	39	56	6
	3 (least)	50	44	6
Scale preference	1 (most)	72	17	11
	2	17	61	22
	3 (least)	11	22	67

---

[1] Frequencies of ranks are expressed as a percentage of panelists. For most criteria, ranks were assigned by 18 panelists; however, for applicability, only 17 panelists provided ranks.

The CS and LS techniques were ranked easier (P<0.001) to learn than the ME technique. Fifty-six percent (10) of the panelists ranked CS as easiest to learn; the remaining 44 % (8) ranked the LS procedure as the easiest to learn (**Table 15**). However, all panelists ranked ME as the most difficult of the three evaluation techniques to learn. One panelist stated that the principle of ME was easy to understand but that its application to meat was difficult.

Each evaluation method received similar mean ranks for scale applicability (**Table 14**). However, panelist responses (**Tables 14** and 15) suggest that the CS and LS procedures may have been considered slightly more applicable for the sensory evaluation of beef than the ME technique.

For scale accuracy, mean ranks for each of the three sensory evaluation methods differed significantly (P<0.001). The CS method was ranked as the most accurate; LS was ranked intermediate and ME was ranked as the least accurate sensory procedure.

Panelists commented that they could be more consistent in scoring the five descriptive attributes of beef with CS than with either LS or ME. One panelist observed that it was more difficult to score along the centre portion of the line scale than at positions near each end. Many panelists commented that the magnitude estimates they assigned to the quality attributes varied considerably from day to day. While this day to day variability was removed in the statistical procedures employed for data analysis, concern about the variability of the ME responses is probably reflected in the low ME rankings assigned by judges for scale accuracy.

Both the CS and LS techniques required similar and less effort (P<0.001) for sample evaluation than ME (Table 14). Many panelists commented that evaluation with ME required intense concentration in order to assess quality attributes of one sample relative to another.

Usually, the descriptive sensory assessment of beef involves the simultaneous evaluation of several characteristics. This type of assessment may be more difficult with ME than with either CS or LS, and may result in decreased sensitivity to treatment differences. Segars et al. (1981) modified their ME procedure for irradiated beef, ham and poultry rolls in order to reduce memory interference and to make ME evaluations easier for the panelists (Cardello, personal communication). Originally, panelists were instructed to evaluate four attributes in each sample of the series presented; in a later phase of the investigation, panelists evaluated a separate series of samples for each of three attributes (Segars et al., 1981).

Scale preference differed significantly (P<0.001) among the evaluation techniques (Table 14). Seventy-two percent of the panelists preferred CS to either LS or ME (Table 15). LS was ranked intermediate in preference and ME was least preferred.

The effect of scale preference on panelist motivation is unknown. If a panelist is not comfortable with the task, sensory performance may be affected. However, Baten (1946) reported that even though panelists preferred the CS technique to the LS technique, their discrimination of differences between apple varieties was greater with LS than with CS.

In general, the data (**Tables 14** and **15**) indicate that the CS method was easy to learn, accurate, required little effort for sample evaluation and was preferred by panelists for the descriptive sensory assessment of beef. The LS procedure was also considered easy to learn, required little effort for sample evaluation and was ranked intermediate in scale accuracy and preference. However, panelists found that the ME technique was more difficult to learn, less accurate and required more effort for sample evaluation than the other techniques. In addition, the ME procedure was least preferred by panelists for sensory assessments of cooked meat quality.

SUMMARY AND CONCLUSIONS

Three sensory evaluation techniques employed for the descriptive sensory assessment of beef were compared: (1) category scaling (CS) which is most commonly used by meat scientists, (2) line or semistructured scaling (LS) which forms the basis for "Quantitative Descriptive Analysis" and (3) magnitude estimation (ME) which is becoming popular for food research but has not been widely utilized with meat.

Eighteen trained panelists evaluated the initial tenderness, overall tenderness, connective tissue amount, overall juiciness and flavor intensity of semimembranosus (SM) steaks using an eight-point category scale, a 15 cm line scale with anchors 1.3 cm from each end and modulus free magnitude estimation.

Panelists were screened and then intensively trained in the use of each evaluation technique, panel procedures and the quality attributes under investigation.

Semimembranosus roasts used for the study were from four postmortem meat treatments (delay chilling (DC), delay chilling with electrical stimulation (DC-ES), hot boning (HB) and hot boning with electrical stimulation (HB-ES)) selected to provide differences in organoleptic quality.

A total of 108 steaks, three from each of 36 roasts, were individually roasted at 176° to an internal temperature of 65°. To minimize the effects of scale carry over, learning and motivational factors, panelists utilized each evaluation technique in sequence in different predetermined orders. Each judge evaluated one of three adjacent steaks from each of nine roasts per treatment using each evaluation technique.

In addition to sensory evaluation of the samples, visual evaluation of steak doneness and objective measurements of tenderness (Warner Bratzler shear on the Ottawa Texture Measuring System), juiciness (percentage press fluid) and color (Hunter Color Difference Meter) were also made.

Few treatment differences attributable to ES were found. Therefore, the comparison of the three sensory evaluation techniques focused on differences obtained between two of the postmortem treatments (DC and HB). Criteria used to compare the three evaluation techniques included an examination of: the objective and sensory data for the two postmortem treatments, the sensitivity of the evaluation methods to differences in quality attributes between the two postmortem treatments, the replication and panelist variation, the significance of the treatment by panelist interaction, the relationships between the three evaluation procedures, the relationships between tenderness assessments and Warner Brazler shear data, and scale preference.

Chemical data, raw weight, cooking time, final internal temperature and cooking losses for SM steaks showed no differences attributable to DC and HB. However, HB steaks had a significantly greater percentage drip in thaw than comparable DC steaks. In addition, HB steaks were more rare than comparable DC steaks as determined subjectively (degree of doneness) and objectively (Hunter a values, percentage press fluid).

When the 18 trained panelists employed CS, HB steaks were rated lower ($P<0.05$) in initial tenderness and overall tenderness and higher ($P<0.01$) in overall juiciness than comparable DC steaks. Using LS, there were no significant differences in the quality attributes of the SM steaks due to postmortem treatment. With the ME technique, the HB steaks were less tender ($P<0.05$) and more juicy ($P<0.01$) than comparable DC steaks.

Although there were differences in statistical significance, the direction and extent of the difference between mean scores assigned to each of the descriptive attributes in the DC and HB steaks were consistent among the three evaluation techniques. For each evaluation method, differences between steak treatment means tended to be large for initial tenderness and overall tenderness, intermediate for amount of connective tissue and juicines, and slight for flavor intensity. This consistency in scoring between each of the three evaluation techniques provides an internal validation for the results of the present study.

Sensory assessments of tenderness tend to be supported by Warner Bratzler shear data. Although the difference was not statistically significant, the average shear force value for HB steaks was higher than that of comparable DC steaks. Since steaks from the DC and HB treatments attained the same average internal temperature of 65.1° (range, 64.75° to 65.50°), juiciness was not expected to differ between the two postmortem treatments. However, the significant differences for juiciness observed by panelists using CS and ME are supported by objective measurements. The HB steaks were visually more rare ($P<0.01$), had higher ($P<0.001$) Hunter *a* (redness) values and greater ($P<0.01$) percentage press fluid than their DC counterparts.

An examination of the distributions of panelist responses for overall tenderness for the two postmortem meat treatments revealed that for each evaluation technique, HB steaks tended to be less tender than comparable DC steaks. These distributions support results from the analysis of variance of these data for CS and ME which showed significant differences in overall tenderness attributable to postmortem treatment.

The magnitude and significance of F-values for selected sources of variation from the split-plot analyses of CS, LS and ME data were examined. Treatment F-values were used as a measure of the sensitivity of each evaluation technique to differences in palatability due to postmortem treatment. The CS procedure was most sensitive to treatment differences as indicated by the three significant F-values obtained for initial tenderness, overall tenderness and juiciness. The LS technique was least sensitive in detecting differences in steak quality attributes. No significant treatment F-values for LS were determined. Magnitude estimation was as sensitive as CS to most treatment differences. For ME, treatment F-values for overall tenderness and juiciness were significant and similar to those obtained by CS; however, the F-value for initial tenderness was not statistically significant.

Replication F-values for each of the three evaluation techniques were similar and nonsignificant for most of the quality attributes under investigation. However, replication variation reached statistical significance once for CS (overall juiciness) and twice for ME (overall juiciness and flavor intensity) For all attributes, replication F-values were lower for the LS method than corresponding F-values for CS and ME.

Panelist F-values were significant ($P<0.001$) for each quality attribute assessed with each of the three evaluation methods. The magnitude of the panelist F-values appeared to reflect the scale range used for each procedure. The CS technique, with a confined scale range of eight categories, had the lowest panelist variation. The LS method, with a scale range of 15 cm, was intermediate in panelist variation. The ME technique, in which panelists were free to assign any value they wanted, had the highest panelist variation.

Except for the ME assessment of overall juiciness, F-values for the treatment by panelist interaction were not significant, indicating that individual panelists agreed with one another when rating the treatments by CS, LS or ME.

Correlation and regression analyses were performed to assess relationships between each of the evaluation methods and between sensory and instrumental measures of tenderness. Data for these analyses were averaged over all panelist responses for each of the 18 roasts (nine per treatment).

Correlation coefficients between each of the evaluation procedures for initial and overall tenderness were highly significant ($P<0.001$) and indicated strong linear relationships between scores assigned by judges using CS, LS and ME. For the remaining attributes, the correlations between each of the three techniques, while sometimes significant, were moderate to low. Sample heterogeneity, especially for connective tissue amount, and a narrow stimulus range, particularly for juiciness and flavor intensity, may have contributed to the low correlations obtained.

Coefficients of determination obtained for tenderness assessments from CS, LS and ME fitted to each of five mathematical functions were similar. None of the mathematical functions provided a consistently superior fit of the data.

Correlations of tenderness data from each evaluation method with Warner Bratzler shear values yielded similar significant r-values. However, these correlations only accounted for 40 to 51 % of the variation in the data. Warner Bratzler shear determinations may not be sensitive to the same structural components which influence taste panel assessments.

Similar coefficients of determination between sensory assessments of tenderness and Warner Bratzler shear data were determined for each of the five mathematical functions. Thus, the CS, LS and ME data can be explained equally well by each of these functions. Furthermore, linear, logarithmic, hyperbolic and parabolic functions described ME data as well as the more traditionally used power function.

The power functions relating the texture judged by panelists using ME to the textural properties of meat assessed instrumentally can provide additional information about the perception of meat tenderness, which is not possible with either the CS or LS techniques. For initial and overall tenderness, the power functions between ME and shear data yielded exponents of 1.13 and 1.05, respectively.

These exponents (which are close to 1) indicate that panel assessments using ME and Warner Bratzler shear determinations were similar in sensitivity to tenderness differences in the meat samples. However, because of data variability, these exponents may not be reliable indicators of sensory response.

The evaluation techniques were ranked by the 18 trained panelists for selected criteria. The CS and LS methods were easier to learn and required less effort for sample evaluation than ME. Each evaluation procedure received similar mean ranks for scale applicability. However, CS was ranked as most accurate for the sensory evaluation of beef; LS was intermediate; and ME was judged the least accurate sensory procedure. In addition, panelists preferred CS over both LS and ME. The LS technique was ranked intermediate in preference and ME was least preferred. The effect of scale preference on motivation is unknown. If a panelist is not comfortable with the task, sensory performance may be affected.

Results of this work indicate that for the descriptive sensory assessment of SM steaks, the CS evaluation technique, most frequently used to determine meat quality, was easy to learn, required little effort for sample evaluation, was preferred by panelists and was most sensitive to treatment differences. The LS method was easy to learn, required little effort for sample evaluation and was ranked intermediate in preference. However, panelist evaluations by the LS procedure did not result in treatment differences for any of the attributes scored. Although panelists found that the ME technique was more difficult to learn, required more effort for sample evaluation and was least preferred, ME was as sensitive as the CS method to most treatment differences. Therefore, findings from this research generally support the use of category scaling for the descriptive sensory assessment of beef by trained panelists.

The heterogeneous meat system employed in the present study is typical of that usually studied by meat researchers. Further comparative studies, using a less heterogeneous meat system such as a restructured meat product, may provide additional information on the accuracy and precision of panelist assessments from each of these evaluation techniques.

The lack of significant treatment differences by LS was unexpected. The line scale has been successfully used for the "Quantitative Descriptive Analysis" of a number of food products. Further work could be done to systematically evaluate the LS procedure for the descriptive sensory assessment of meat. An evaluation of the direct comparison of LS ratings between samples on the same scorecard versus the use of a separate scorecard for each sample would be useful.

Panelists stated that the principle of magnitude estimation was easy to understand but that its application to meat was difficult. The difficulties encountered in the present study with the use of ME for meat have not previously been described in the literature. Usually, the descriptive sensory assessment of beef involves the simultaneous evaluation of several characteristics. This type of assessment may be more difficult with ME than with either CS or LS, and may result in decreased sensitivity to treatment differences. To optimize the potential of the ME procedure for the descriptive assessment of cooked beef, some procedural modifications may be needed. The ME method may be most useful for an evaluation of just a few attributes at one time. However, the evaluation of products requiring mastication such as meat may always require greater effort.

BIBLIOGRAPHY

AMERICAN MEAT SCIENCE ASSOCIATION (AMSA). 1978. "Guidelines for Cookery and Sensory Evaluation of Meat." Am. Meat Sci. Assn. and Natl. Live Stock and Meat Board, Chicago, IL.

AMERINE, M.A., PANGBORN, R.M. and ROESSLER, E.B. 1965. "Principles of Sensory Evaluation of Food." Academic Press, New York, NY.

ANDERSON, N.H. 1970. Functional measurement and psychophysical judgement. Psychol. Rev. 77:153.

ANDERSON, N.H. 1977. Note on functional measurement and data analysis. Percept. Psychophy. 21:201.

ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS (AOAC). 1980. "Official Methods of Analysis." 13th ed., AOAC, Washington, DC.

BAILEY, A.J. 1972. The basis of meat texture. J. Sci. Food Agric. 23:995.

BATEN, W.D. 1946. Organoleptic tests pertaining to apples and pears. Food Res. 11:84.

BATCHER, O.M. and DEARY, P.A. 1975. Quality characteristics of broiled and roasted beef steaks. J. Food Sci. 40:745.

BENDIG, A.W. 1953. The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. J. Appl. Psychol. 37:38.

BENDIG, A.W. 1954. Reliability and the number of rating scale categories. J. Appl. Psychol. 38:38.

BERRY, B.W. and CROSS, H.R. 1982. Processing systems for hot- and cold-boned primals from mature cow carcasses. J. Food Sci. 47:875.

BERRY, B.W., WHEELING, M.R. and CARPENTER, J.A.,Jr. 1977. Effects of muscle and cookery method on palatability of beef from several breeds and breed crosses. J. Food Sci. 42:1322.

BIRNBAUM, M.H. 1982. Problems with so-called "direct" scaling. In "Selected Sensory Methods: Problems and Approaches to Hedonics." J.T. Kuznicki, R.A. Johnson and A.F. Rutkiewic, Eds., ASTM STP 773, Philadephia, PA, p. 34.

BOUTON, P.E., FORD, A.L., HARRIS, P.V. and SHAW, F.D. 1978. Effect of low voltage stimulation of beef carcasses on muscle tenderness and pH. J. Food Sci. 43:1392.

BOUTON, P.E., FORD, A.L., HARRIS, P.V. and SHAW, F.D. 1980a. Electrical stimulation of beef sides. Meat Sci. 4:145.

BOUTON, P.E., SHAW, F.D. and HARRIS, P.V. 1980b. Electrical stimulation of beef carcasses in Australia. Proc. 26th Eur. Meet. Meat Res. Workers, Colorado Springs, CO., Vol II. p. 23.

BOWLES AXE, J.E., KASTNER, C.L., DIKEMAN, M.E., HUNT, M.C., KROPF, D.H. AND MILLIKEN, G.A. 1983. Effects of beef carcass electrical stimulation, hot boning, and aging on unfrozen and frozen longissimus dorsi and semimembranosus steaks. J. Food Sci. 48:332.

BREIDENSTEIN, B.B., COOPER, C.C., CASSENS, R.G., EVANS, G. and BRAY, R.W. 1968. Influence of marbling and maturity on the palatability of beef muscle. I. Chemical and organoleptic considerations. J. Anim. Sci. 27:1532.

CARDELLO, A.V. 1981. Personal communication.

CARDELLO, A.V., HUNT, D. and MANN, B. 1979. Relative sweetness of fructose and sucrose in model solutions, lemon beverages and white cake. J. Food Sci. 44:748.

CARDELLO, A.V., MALLER, O., KAPSALIS, J.G., SEGARS, R.A., SAWYER, F.M., MURPHY, C. and MOSKOWITZ, H.R. 1982a. Perception of texture by trained and consumer panelists. J. Food Sci. 47:1186.

CARDELLO, A.V., MATAS, A. and SWEENEY, J. 1982b. A research note. The standard scales of texture: rescaling by magnitude estimation. J. Food Sci. 47:1738.

CARLIN, A.F., KEMPTHORNE, O. and GORDON, J. 1956. Some aspects of numerical scoring in subjective evaluation of foods. Food Res. 21:273.

CHAPMAN, L.D. and WIGFIELD, R. 1970. Rating scales in consumer research. Food Manuf. 45(8):59.

CLAUS, J.R. 1982. Effects of beef carcass electrical stimulation and hot boning on muscle display color of unfrozen and frozen steaks. M.S. Thesis, Kansas State University, Manhattan, KS.

CLIPLEF, R.L. and STRAIN, J.H. 1976. Tenderness and related organoleptic characteristics of beef carcass as affected by postmortem chilling temperature. Can. J. Anim. Sci. 56:417.

CLONINGER, M.R., BALDWIN, R.E. and KRAUSE, G.F. 1976. Analysis of sensory rating scales. J. Food Sci. 41:1225.

COLEMAN, B.J., GRAF, R.G. and ALF, E.F. 1981. Assessing power function relationships in magnitude estimation. Percept. Psychophy. 29:178.

COVER, S., BANNISTER, J.A. and KEHLENBRINK, E. 1957. Effect of four conditions of cooking on the eating quality of two cuts of beef. Food Res. 22:635.

COVER, S., RITCHEY, S.J. and HOSTETLER, R.L. 1962a. Tenderness of Beef. I. The connective tissue component of tenderness. J. Food Sci. 27:469.

COVER, S., RITCHEY, S.J. and HOSTETLER, R.L. 1962b. Tenderness of Beef. II. Juiciness and the softness component of tenderness. J. Food Sci. 27:476.

COVER, S., RITCHEY, S.J. and HOSTETLER, R.L. 1962c. Tenderness of Beef. III. The muscle-fiber component of tenderness. J. Food Sci. 27:483.

CROSS, H.R. 1977. Meat cookery and sensory evaluation. Proc. 30th Annu. Recipro. Meat Conf. Am. Meat Sci. Assoc. p. 124.

CROSS, H.R. 1979. Pre-rigor meats in fresh sausage. Proc. 21st Annu. Meat Sci. Inst. National Independent Meat Packers Assoc., Athens, Georgia. Described by Claus, J.R. 1982. Effects of beef carcass electrical stimulation and hot boning on muscle display color of unfrozen and frozen steaks. M.S. Thesis, Kansas State University, Manhattan, KS.

CROSS, H.R. and TENNENT, I. 1980. Accelerated processing systems for USDA choice and good beef carcasses. J. Food Sci. 45:765.

CROSS, H.R., MOEN, R. and STANFIELD, M.S. 1978. Training and testing of judges for sensory analysis of meat quality. Food Technol. 32(7):48.

CROSS, H.R., STANFIELD, M.S., ELDER, R.S. and SMITH, G.C. 1979. A research note. A comparison of roasting versus broiling on the sensory characteristics of beef longissimus steaks. J. Food Sci. 44:310.

CROUSE, J.D. and SEIDEMAN, S.C. 1984. Effect of high temperature conditioning on beef from grass or grain fed cattle. J. Food Sci. 49:157.

CROUSE, J.D., SEIDEMAN, S.C. and CROSS, H.R. 1983. The effects of carcass electrical stimulation and cooler temperature on the quality and palatability of bull and steer beef. J. Anim. Sci. 56:81.

DAVIS, G.W., COLE, A.B., JR., BACKUS, W.R. and MELTON, S.L. 1981. Effect of electrical stimulation on carcass quality and meat palatability of beef from forage and grain-finished steers. J. Anim. Sci. 53:651.

DRANSFIELD, E. and JONES, R.C.D. 1978. Effect of rate of chilling on the variability in texture of the beef round. J. Sci. Food Agric. 29:607.

DRANSFIELD, E., BROWN, A.J. and RHODES, D.N. 1976. Eating quality of hot deboned beef. J. Fodd Technol. 11:401.

DRANSFIELD, E., RHODES, D.N., NUTE, G.R., ROBERTS, T.A., BOCCARD, R., TOURAILLE, C., BUCHTER, L., HOOD, D.E., JOSEPH, R.L., SCHON, I., CASTEELS, M., COSENTINO, E. and TINBERGEN, B.J. 1982. Eating quality of European beef assessed at five research institutes. Meat Sci. 6:163.

DRYDEN, F.D., MARCHELLO, J.A., TINSLEY, A., MARTINS, C.B., WOOTEN, R.A., ROUBICEK, C.B. and SWINGLE, R.S. 1979. Acceptability of selected muscles from poor condition and realimented cull range cows. J. Food Sci. 44:1058.

DuBOSE, C.N., CARDELLO, A.V. and MALLER, O. 1980. Effects of colorants and flavorants on identification, perceived flavor intensity, and hedonic quality of fruit-flavored beverages and cake. J. Food Sci. 45:1393.

ELGASIM, E.A., KENNICK, W.H., McGILL, L.A., ROCK, D.F. and SOELDNER, A. 1981. Effects of electrical stimulation and delayed chilling of beef carcasses on carcass and meat characteristics. J. Food Sci. 46:340.

FALK, S.N., HENRICKSON, R.L. and MORRISON, R.D. 1975. Effect of boning beef carcasses prior to chilling on meat tenderness. J. Food Sci. 40:1075.

FILIPELLO, F, 1957. Organoleptic wine-quality evaluation. I. Standards of quality and scoring vs. rating scales. Food Technol. 11(1):47.

FISHER, R.A. and YATES, F. 1949. "Statistical Tables for Biological, Agricultural and Medical Research." 3rd ed., Oliver and Boyd Ltd., London, England.

FOLLETT, M.J., NORMAN, G.A. and RATCLIFF, P.W. 1974. The ante-rigor excision and air cooling of beef semimembranosus muscles at temperatures between [illegible] and +15°C. J. Food Technol. 9:509.

FORBES, S.M.C. 1973. The relationship between consumer criteria for choosing beef and beef quality. M. S. Thesis, University of Manitoba.

GAITO, J. 1980. Measurement scales and statistics: resurgence of an old misconception. Psychol. Bull. 87:564.

GARCIA MEDINA, M.R. 1981. Flavor odor taste interactions in solutions of acetic acid and coffee. Chem. Senses 6:1[illegible].

GARNER, W.R. 1960. Rating scales, discriminability, and information transmission. Psychol. Rev. 67:343.

GEORGE, A.R., BENDALL, J.R. and JONES, R.C.D. 1980. The tenderising effect of electrical stimulation of beef carcasses. Meat Sci. 4:51.

GILBERT, K.V. and DAVEY, C.L. 1976. Carcass electrical stimulation and early boning of beef. 1976. N. Z. J. Agric. Res. 19:429.

GINGER, B. and WEIR, C.E. 1958. Variations in tenderness within three muscles from beef round. Food Res. 23:662.

GIOVANNI, M.E. and PANGBORN, R.M. 1983. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. J. Food Sci. 48:1175.

GRIDGEMAN, N.T. 1961. A comparison of some taste-test methods. Food Res. 26:171.

GULLETT, E.A., ROWE, D.L. and HINES, R.J. 1983. Tenderness parameters in sensory evaluation of meat. Can. Inst. Food Sci. Technol. J. 16(3):xii (abstract).

HALL, R.L. 1958. Flavor study approaches at McCormick & Company, Inc. In "Flavor Research and Food Acceptance." Reinhold Publishing Corp., New York, NY. p. 224.

HARRIES, J.M., RHODES, D.N. and CHRYSTALL, B.B. 1972. Meat texture. I. Subjective assessment of the texture of cooked beef. J. Texture Stud. 3:101.

HARRISON, D.L. 1975. Selection of cooking method based on research objectives. Proc. 28th Annu. Recipro. Meat Conf. Am. Meat Sci. Assoc. p. 340.

HOWARD, A. 1973. Taste panel techniques II. A validating technique. CSIRO Food Res. Q. 33:8.

HAWRYSH, Z.J. and BERG, R.T. 1975. Eating quality of beef from young bulls as influenced by breed. Can. Inst. Food Sci. Technol. J. 8(3):149.

HAWRYSH, Z.J. and BERG, R.T. 1976. Studies on beef eating quality in relation to the current Canada grade classifications. Can. J. Anim. Sci. 56:383.

HOSTETLER, R.L., CARPENTER, Z.L., SMITH, G.C. and DUTSON, T.R. 1975. Comparison of postmortem treatments for improving tenderness of beef. J. Food Sci. 40:223.

HOVENDEN, J.E., DUTSON, T.R., HOSTETLER, R.L. and CARPENTER, Z.L. 1979. Variation and repeatability of an untrained beef sensory panel. J. Food Sci. 44:1598.

JEREMIAH, L.E., MARTIN, A.H. and MURRAY, A.C. 1984. The effects of various postmortem treatments on certain physical and sensory properties of three different bovine muscles. Meat Sci. In press.

JOHNSON, J. and CLYDESDALE, F.M. 1982. Perceived sweetness and redness in colored surcose solutions. J. Food Sci. 47:747.

JOHNSON, J.L., DZENDOLET, E., DAMON, R., SAWYER, M. and CLYDESDALE, F.M. 1982. Psychophysical relationships between perceived sweetness and color in cherry-flavored beverages. J. Food Prot. 45:601.

JONES, L.V. and THURSTONE, L.L. 1955. The psychophysics of semantics: an experimental investigation. J. Appl. Psychol. 39:31.

JONES, L.V., PERYAM, D.R. and THURSTONE, L.L. 1955. Development of a scale for measuring soldiers' food preferences. Food Res. 20:512.

JOSEPH, R.L. and CONNOLLY, J. 1977. The effects of suspension method, chilling rates and post mortem ageing period on beef quality. J. Food Technol. 12:231.

KAPSALIS; J.G. and MOSKOWITZ, H.R. 1977. The psychophysics and physics of food texture. Food Technol. 31(4):91.

KASTNER, C.L. 1983. Optimal hot-processing systems for beef. Food Technol. 37(5):96.

KASTNER, C.L. and RUSSELL, T.S. 1975. Characteristics of conventionally and hot-boned bovine muscle excised at various conditioning periods. J. Food Sci. 40:747.

KASTNER, C.L., HENRICKSON, R.L. and MORRISON, R.D. 1973. Characteristics of hot-boned bovine muscle. J. Anim. Sci. 36:484.

KASTNER, C.L., DIKEMAN, M.E., NAGELE, K.N., LYON, M., HUNT, M.C. and KROPF, D.H. 1980. Effects of carcass electrical stimulation and hot boning on selected beef muscles. Proc. 26th Eur. Meet. Meat Res. Workers, Colorado Springs, CO., Vol II p. 40.

KEEPING, E.S. 1962. "Introduction to Statistical Inference." D. Van Nostrand Co., Inc., Toronto.

LANE, H.L., CATANIA, A.C. and STEVENS, S.S. 1961. Voice level: autophonic scale, perceived loudness, and effects of sidetone. J. Acoust. Soc. Am. 33:160.

LARMOND, E. 1976. Texture measurement in meat by sensory evaluation. J. Texture Stud. 7:87.

LARMOND, E. 1977. "Laboratory Methods for Sensory Evaluation of Food." Research Branch, Canada Department of Agriculture, Publication 1637.

LOCHNER, J.V., KAUFFMAN, R.G. and MARSH, B.B. 1980. Early-postmortem cooling rate and beef tenderness. Meat Sci. 4:227.

LYON, B.G. and KLOSE, A.A. 1980. Sensory profiling of canned boned chicken: comparisons of retail, school lunch, and military canned boned chicken. J. Food Sci. 45:1336.

LYON, C.E., LYON, B.G., TOWNSEND, W.E. and WILSON, R.L. 1978. Effect of level of structured protein fiber on quality of mechanically deboned chicken meat patties. J. Food Sci. 43:1524.

LYON, M., KASTNER, C.L., DIKEMAN, M.E., HUNT, M.C. KROPF, D.H. and SCHWENKE, J.R. 1983. Effects of electrical stimulation, aging, and blade tenderization on hot-boned beef psoas major and triceps brachii muscles. J. Food Sci. 48:131.

MARKS, L.E. 1974. "Sensory processes: the new psychophysics." Academic Press, New York, NY.

MARKS, L.E. and CAIN, W.S. 1972. Perception of intervals and magnitudes for three prothetic continua. J. Exp. Psychol. 94:6.

MARSH, B.B., LOCHNER, J.V., TAKAHASHI, G. and KRAGNESS, D.D. 1980-81. Research note. Effects of early postmortem pH and temperature on beef tenderness. Meat Sci. 5:479.

McBRIDE, R.L. 1983. Pychophysics: could Fechner's assumption be correct? Aust. J. Psychol. 35:85.

McCURDY, S.M., HARD, M.M. and MARTIN, E.L. 1981. Sensory properties of rib and round muscle roasts from two beef breed-types on two feeding regimes. J. Food Sci. 46:991.

McDANIEL, M.R. 1974. Magnitude estimation versus category scaling in food preference testing and food quality rating. Ph.D. Dissertation, University of Massachusetts, Amherst, MA.

McDANIEL, M.R. and SAWYER, F.M. 1981a. Descriptive analysis of whiskey sour formulations: magnitude estimation versus a 9-point category scale. J. Food Sci. 46:178.

McDANIEL, M.R. and SAWYER, F.M. 1981b. Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. J. Food Sci. 46:182.

McDOWELL, M.D., HARRISON, D.L., PACEY, C. and STONE, M.B. 1982. Differences between conventionally cooked top round roasts and semimembranosus muscle strips cooked in a model system. J. Food Sci. 47:1603.

McKEITH, F.K., SMITH, C.G., DUTSON, T.R., SAVELL, J.W., HOSTETLER, R.L. and CARPENTER, Z.L. 1980. Electrical stimulation of intact or split steer or cow carcasses. J. Food Prot. 43:795.

McKEITH, F.K., SAVELL, J.W. and SMITH, G.C. 1981. Tenderness improvement of the major muscles of the beef carcass by electrical stimulation. J. Food Sci. 46:1774.

MELLERS, B.A. 1983. Evidence against "absolute" scaling. Percept. Psychophy. 33:523.

MILLER, G.A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63:81.

MOORE, L.J., HARRISON, D.L. and DAYTON, A.D. 1980. Differences among top round steaks cooked by dry or moist heat in a conventional or a microwave oven. J. Food Sci. 45:777.

MOSKOWITZ, H.R. 1970. Ratio scales of sugar sweetness. Percept. Psychophy. 7:315.

MOSKOWITZ, H.R. 1975. Applications of sensory measurement to food evaluations. II. Methods of ratio scaling. Lebensm. Wiss. +Technol. 8:249.

MOSKOWITZ, H.R. 1977. Magnitude estimation: notes on what, how, when, and why to use it. J. Food Qual. 1:195.

MOSKOWITZ, H.R. 1982. Utilitarian benefits of magnitude estimation scaling for testing product acceptability. In "Selected Sensory Methods: Problems and Approaches to Measuring Hedonics." J.T. Kuznicki, R.A. Johnson and A.F. Rutkiewic, Eds., ASTM STP 773, Philadephia, PA. p. 11.

MOSKOWITZ, H.R. 1983. "Product Testing and Sensory Evaluation of Foods. Marketing and R & D Approaches." Food & Nutrition Press Inc., Westport, CN.

MOSKOWITZ, H.R. and SIDEL, J.L. 1971. Magnitude and hedonic scales of food acceptability. J. Food Sci. 36:677.

MOSKOWITZ, H.R. and DUBOSE, C. 1977. Taste intensity, pleasantness and quality of aspartame, sugars, and their mixtures. Can. Inst. Food Sci. Technol. J. 10(2):126.

MOSKOWITZ, H.R., SEGARS, R.A., KAPSALIS, J.G. and KLUTER, R.A. 1974. Sensory ratio scales relating hardness and crunchiness to mechanical properties of space cubes. J. Food Sci. 39:200.

MOTTRAM, D.S. 1981. Cooking meat for sensory and instrumental assessment. J. Sci. Food Agri. 32:523 (summary).

MURPHY, E.F., BAILEY, R.M. and COVELL, M.R. 1954. Observations on methods to determine food palatability and comparative freezing quality of certain new strawberry varieties. Food Technol. 8:113.

MURPHY, E.F., COVELL, M.R. and DINSMORE, J.S.,JR. 1957. An examination of three methods for testing palatability as illustrated by strawberry flavor differences. Food Res. 22:423.

NORWICH, K.H. 1981. The magical number seven: making a "bit" of "sense". Percept. Psychophy. 29:409.

O'MAHONY, M. 1979. Psychophysical aspects of sensory analysis of dairy products: a critique. J. Dairy Sci. 62:1954.

O'MAHONY, M. and HEINTZ, C. 1981. Direct magnitude estimation of salt taste intensity with continuous correction for salivary adaptation. Chem. Senses 6:101.

PAUL, P. and BRATZLER, L.J. 1955. Studies on tenderness of beef. III. Size of shear cores: end to end variation in the semimembranosus and adductor. Food Res. 20:635.

PAUL, P.C. and PALMER, H.H. 1972. "Food Theory and Applications." John Wiley and Sons, Inc., Toronto.

PILGRIM, F.J. and WOOD, K.R. 1955. Comparative sensitivity of rating scale and paired comparison methods for measuring consumer preference. Food Technol. 9:385.

POWERS, J.J., WARREN, C.B. and MASURAT, T. 1981. Collaborative trials involving three methods of normalizing magnitude estimations. Lebensm. Wiss. Technol. 14:86.

RAFFENSPERGER, E.L., PERYAM, D.R. and WOOD, K.R. 1956. Development of a scale for grading toughness-tenderness in beef. Food Technol. 10:627.

RANDALL, C.J. and LARMOND, E. 1977. Effect of method of comminution (flake-cutting and grinding) on the acceptability and quality of hamburger patties. J. Food Sci. 42:728.

RAY, E.E., STIFFLER, D.M. and BERRY, B.W. 1982. Effects of electrical stimulation and hot-boning on physical changes, cooking time and losses, and tenderness of beef roasts. J. Food Sci. 47:210.

RILEY, R.R., SAVELL, J.W., MURPHEY, C.E., SMITH, G.C., STIFFLER, D.M. and CROSS, H.R. 1983. Palatability of beef from steer and young bull carcasses as influenced by electrical stimulation, subcutaneous fat thickness and marbling. J. Anim. Sci. 56:592.

SALM, C.P., MILLS, E.W., REEVES, E.S., JUDGE, M.D. and ABERLE, E.D. 1981. A research note. Effect of electrical stimulation on muscle characteristics of beef cattle fed a high energy diet for varying lengths of time. J. Food Sci. 46:1284.

SAVELL, J.W., McKEITH, F.K. and SMITH, G.C. 1981. Reducing postmortem aging time of beef with electrical stimulation. J. Food Sci. 46:1777.

SCHMIDT, G.R. and GILBERT, K.V. 1970. The effect of muscle excision before the onset of rigor mortis on the palatability of beef. J. Food Technol. 5:331.

SCHMIDT, G.R. and KEMAN, S. 1974. Hot boning and vacuum packaging of eight major bovine muscles. J. Food Sci. 39:140.

SCHROEDER, J.W., CRAMER, D.A. and BOWLING, R.A. 1982. Postmortem muscle alterations in beef carcass temperature, pH and palatability from electrical stimulation. J. Anim. Sci. 54:549.

SEGARS, R.A., HAMEL, R.G. KAPSALIS, J.G. and KLUTER, R.A. 1975. A punch and die test cell for determining the textural qualities of meat. J. Texture Stud. 6:2[illegible].

SEGARS, R.A., CARDELLO, A.V. and COHEN, J.S. 1981. Objective and subjective texture evaluation of irradiation sterilized meat products. J. Food Sci. 46:999.

SHAFFER, T.A., HARRISON, D.L. and ANDERSON, L.L. 1973. Effects of end point and oven temperatures on beef roasts cooked in oven film bags and open pans. J. Food Sci. 38:1205.

SHINDELL, S. 1964. "Statistics, Science, and Sense." University of Pittsburgh Press, Pittsburgh, PA.

SIDEL, J.L. and STONE, H. 1976. Experimental design and analysis of sensory tests. Food Technol. 30(11):32.

SMITH, G.C., CULP, G.R. and CARPENTER, Z.L. 1978. Postmortem aging of beef carcasses. J. Food Sci. 43:823.

SMITH, G.C., JAMBERS, T.G., CARPENTER, Z.L., DUSTON, T.R., HOSTETLER, R.L. and OLIVER, W.M. 1979b. Increasing the tenderness of forage-fed beef. J. Anim. Sci. 49:1207.

SMITH, G.C., CROSS, H.R., CARPENTER, Z.L., MURPHEY, C.E., SAVELL, J.W., ABRAHAM, H.C. and DAVIS, G.W. 1982. Relationship of USDA maturity groups to palatability of cooked beef. J. Food Sci. 47:1100.

SMITH, M.E., KASTNER, C.L., HUNT, M.C., KROPF, D.H. and ALLEN, D.M. 1979a. Elevated conditioning temperature effects on beef carcasses from four nutritional regimes. J. Food Sci. 44:158.

STANLEY, D.W. and SWATLAND, H.J. 1976. The microstructure of muscle tissue- a basis for meat texture measurement. J. Texture Stud. 7:65.

STANLEY, D.W., VOISEY, P.W. and SWATLAND, H.J. 1980. Texture-structure relationships in bacon lean. J. Texture Stud. 11:217.

STEEL, R.G.D. and TORRIE, J.H. 1980. "Principles and Procedures of Statistics. A Biometrical Approach." 2nd ed., McGraw-Hill Book Co., Toronto.

STEVENS, S.S. 1975. "Psychophysics: An Introduction to its Perceptual, Neural and Social Prospects." John Wiley and Sons, Inc., Toronto.

STEVENS, S.S. and GALANTER, E.H. 1957. Ratio scales and category scales for a dozen perceptual continua. J. Exp. Psychol. 54:377.

STONE, H., SIDEL, J., OLIVER, S., WOOLSEY, A. and SINGLETON, R.C. 1974. Sensory evaluation by quantitative descriptive analysis. Food Technol. 28(11):24.

STONE, H., SIDEL, J.L. and BLOOMQUIST, J. 1980. Quantitative descriptive analysis. Cereal Foods World. 25:642.

SYMONDS, P.M. 1924. On the loss of reliability in ratings due to coarseness of the scale. J. Exp. Psychol. 7:456.

SZCZESNIAK, A.S. 1968. Correlations between objective & sensory texture measurements. Food Technol. 22:981.

SZCZESNIAK, A.S., and TORGESON, K.W. 1965. Methods of meat texture measurement viewed from the background of factors affecting tenderness. Adv. Food Res. 14:33.

TAYLOR, A.A., SHAW, B.G. and MacDOUGALL, D.B. 1980-81. Hot deboning beef with and without electrical stimulation. Meat Sci. 5:109.

TEGHTSOONIAN, M. and TEGHTSOONIAN, R. 1971. How repeatable are Steven's power law exponents for individual subjects? Percept. Psychophy. 10:147.

VAN DE VOORT, F.R., MILLER, H. and ASHTON, G.C. 1981. Research note. Automated reading and recording of sensory data for semi-structured scales. Can. Inst. Food Sci. Technol. J. 14:220.

VICKERS, Z.M. 1983. Magnitude estimation vs category scaling of the hedonic quality of food sounds. J. Food Sci. 48:1183.

VOLLMAR, E.K., HARRISON, D.L. and HOGG, M.G. 1976. Bovine muscle cooked from the frozen state at low temperature. J. Food Sci. 41:411.

WALKER, D.J., HARRIS, P.V. and SHAW, F.D. 1977. Accelerated processing of beef. Food Technol. Aust. 29:504.

WEISS, D.J. 1972. Averaging: an empirical validity criterion for magnitude estimation. Percept. Psychophy. 12:385.

WEISS, D.J. 1981. The impossible dream of Fechner and Stevens. Perception. 10:431.

WEST, R.L. 1983. Functional characteristics of hot-boned meat. Food Technol. 38(5):57.

WILL, P.A., HENRICKSON, R.L. and MORRISON, R.D. 1976. The influence of delay chilling and hot boning on tenderness of bovine muscle. J. Food Sci. 41:1102.

WILL, P.A., HENRICKSON, R.L., MORRISON, R.D. and ODELL, G.V. 1979. Effect of electrical stimulation on ATP depletion and sarcomere length in delay-chilled bovine muscle. J. Food Sci. 44:1646.

WILLIAMS, J.G. and BERNHARD, R.A. 1981. Amino acid lactose interactions and their sensory consequences. J. Food Sci. 46:1245.

ZOOK, K. and WESSMAN, C. 1977. The selection and use of judges for descriptive panels. Food Technol. 31(11):56.

APPENDIX

SCORECARD FOR BEEF

JUDGE ____________

DATE ____________

SAMPLE NO. ____________

	5	4	3	2	
Internal Color	Medium				Well Done
					Rare

COMMENTS:

Figure 13. Scorecard for the degree of doneness assessment of cooked semimembranosus steaks.

DEFINITIONS AND TASTING PROCEDURE FOR THE SENSORY EVALUATION OF BEEF

Evaluate the palatability characteristics of each of the beef samples. Circle a number along the 8-point scale that best describes your impression of each of the characteristics. Evaluate each sample individually, trying not to make comparisons between samples.

Please rinse your mouth with water before beginning and between samples. Take a short break between the first and second set of samples.

INITIAL TENDERNESS- is the lack of force required to bite through a cube of beef across the grain, between the molar teeth. Evaluate initial tenderness after two chews.

OVERALL TENDERNESS- is the amount of effort and time required to completely masticate a cube of beef. Record the number of chews required to completely masticate the cube on the scorecard. Then refer to your chew range card to rate the beef cube for overall tenderness.

CONNECTIVE TISSUE-is the amount of residue felt during chewing and left after complete mastication.

OVERALL JUICINESS- is the amount of moisture left in the mouth after complete mastication.

FLAVOR INTENSITY- is the amount of meaty flavors present in the mouth after complete mastication. Please comment if other flavors mask the meatiness of the sample.

COMMENTS- Your comments about each sample are welcome and would be very helpful.

Before leaving your booth, please check to insure that you have completed the entire scorecard.

Thank You!

Figure 14. Definitions and tasting procedure for the sensory evaluation of beef using the category scale.

DEFINITIONS AND TASTING PROCEDURE
FOR THE
SENSORY EVALUATION OF BEEF

Evaluate the palatability characteristics of each of the beef samples. Place a vertical line on the horizontal line at the point that best describes your impression of each of the characteristics. Evaluate each sample individually, trying <u>not</u> to make comparisons between samples.

Please rinse your mouth with water before beginning and between samples. Take a short break between the first and second set of samples.

INITIAL TENDERNESS- is the lack of force required to bite through a cube of beef across the grain, between the molar teeth. Evaluate initial tenderness after two chews.

OVERALL TENDERNESS- is the amount of effort and time required to completely masticate a cube of beef.

CONNECTIVE TISSUE- is the amount of residue felt during chewing and left after complete mastication.

OVERALL JUICINESS- is the amount of moisture left in the mouth after complete mastication.

FLAVOR INTENSITY- is the amount of meaty flavors present in the mouth after complete mastication. Please comment if other flavors mask the meatiness of the sample.

COMMENTS- Your comments about each sample are welcome and would be very helpful.

Before leaving your booth, please
check to insure that you have
completed the entire scorecard.

Thank You!

Figure 15. Definitions and tasting procedure for the sensory evaluation of beef using the line scales.

DEFINITIONS AND TASTING PROCEDURE
FOR THE
SENSORY EVALUATION OF BEEF

Evaluate the palatabilty characteristics of each of the beef samples. Assign any numbers you wish to describe each characteristic of the first sample in each set. For each successive sample, assign numbers for each characteristic in relation to the sample you had just evaluated.

Please rinse your mouth with water before beginning and between samples. Take a short break between the first and second set of samples.

INITIAL TENDERNESS- is the lack of force required to bite through a cube of beef across the grain, between the molar teeth. Evaluate initial tenderness after two chews.

FLAVOR INTENSITY- is the amount of meaty flavors present in the mouth after complete mastication. Please comment if other flavors mask the meatiness of the sample.

OVERALL JUICINESS- is the amount of moisture left in the mouth after complete mastication.

OVERALL TENDERNESS- is the amount of effort and time required to completely masticate a cube of beef.

CONNECTIVE TISSUE- is the amount of residue felt during chewing and left after complete mastication.

COMMENTS- Your comments about each sample are welcome and would be very helpful.

Before leaving your booth, please
check to insure that you have
completed the entire scorecard.

Thank you!

Figure 16. Definitions and tasting procedure for the sensory evaluation of beef using magnitude estimation.

NAME______________ DATE______________

RE______________

I would like some information on the scaling technique that you have been using for the past six days. Please answer each question as fully as possible.

1. Please comment on the following:

any difficulties you had in learning this scale

the length of training given for learning this scale

the applicability of this scale for the sensory evaluation of beef

the effort needed to evaluate each sample

your impression of the accuracy provided by this scale

2. List several desirable and undesirable features of this scale.

desirable	undesirable

Figure 17. Scorecard for ranking the three sensory evaluation techniques.

NAME ______________________ DATE ________________

RE: MAGNITUDE ESTIMATION, 8-POINT SCALE, LINE SCALE

Please compare the scaling techniques used in this study. Answer the following questions by ranking the three scales. Consider each question separately.

1. ease of learning

	RANK	SCALE
easiest to learn	1	____________
	2	____________
most difficult to learn	3	____________

2. applicability of the scale for the sensory evaluation of beef

	RANK	SCALE
most applicable	1	____________
	2	____________
least applicable	3	____________

3. your impression of scale accuracy

	RANK	SCALE
most accurate	1	____________
	2	____________
least accurate	3	____________

4. effort needed to evaluate each sample

	RANK	SCALE
most effort	1	____________
	2	____________
least effort	3	____________

5. scale preference

	RANK	SCALE
most preferred	1	____________
	2	____________
least preferred	3	____________

COMMENTS

Thank you for your co-operation.

Figure 18. Questionaire for panelists'comments regarding the three sensory evaluation techniques.

Table 16. Means and standards errors for chemical and cooking data of semimembranosus steaks from the four postmortem meat treatments.

	Postmortem Treatment				
Measurement[1]	DC	DC-ES	HB	HB-ES	SEM[2]
Chemical data[3]					
pH	5.5	5.6	5.6	[illegible].6	[illegible]
Ether extract, %	[illegible].3	[illegible].0	[illegible].4	[illegible]	[illegible]
Total moisture, %	[illegible]3.2	[illegible]3.8	72.[illegible]	73.[illegible]	[illegible].4
Cooking data[4]					
Raw weight, g	457.[illegible]	432.4	449.6	443.[illegible]	[illegible].4
Drip in thaw, %	5.6b	6.[illegible]b	7.8a	7.1ab	[illegible]**
Initial temp., °C	4.8	4.6	6.0	5.3	0.6[illegible]
Final temp., °C	65.1	65.1	65.1	65.[illegible]	0.[illegible]
Cooking time, min	55.3	57.0	58.4	57.0	[illegible].94
Cooking time, min/100g	12.2	13.7	13.3	13.[illegible]	0.96
Cooking losses, %					
Total	23.[illegible]	23.7	22.1	21.5	1.22
Volatile	20.[illegible]	20.7	19.2	18.9	0.92
Drip	3.0	3.1	2.9	2.6	0.40

---

[1] Delay chilling (DC), delay chilling with electrical stimulation (DC-ES), hot boning (HB), hot boning with electrical stimulation (HB-ES).

[2] Standard error of the mean.

[3] Values are the means of 18 determinations, 2 per replicate.

[4] Values are the means of 27 determinations, 1 per steak per replicate.

a,b Means within the same row sharing a common superscript are not significantly different at P<0.05.

** Significant at P<0.01.

Table 17. Means and standards errors for the degree of doneness evaluation and objective measurements f semimembranosus steaks from the four postmortem meat treatments.

	Postmortem Treatment[1]				
Measurement	DC	DC-ES	HB	HB-ES	SEM[2]
Degree of doneness[3]	4.9b	5.1ab	5.6a	5.5ab	0.30*
Hunter L[4]	44.9	44.1	42.6	43.0	1.11
a	6.8	7.8	8.4	7.5	0.45
b	10.4	10.5	10.5	10.3	0.19
Press fluid, %[5]	36.5b	38.2ab	40.1a	39.2ab	1.08*
OTMS-Warner Bratzler shear, kg[6]	7.7	7.5	8.7	8.6	0.45

[1] Delay chilling (DC), delay chilling with electrical stimulation (DC-ES), hot boning (HB), hot boning with electrical stimulation (HB-ES).

[2] Standard error of the mean.

[3] Doneness scale: 1=well-done, 5=medium, 9=rare. Values are the means of 81 judgements, one per steak by each of three panelists.

[4] Values are the means of 54 determinations, 2 on each of three steaks per replication.

[5] Values are the means of 54 determinations, 3 on each of two steaks per replication.

[6] Values are the means of 156 determinations, with an average of 9 shears on each of two steaks per replication

a,b Means within the same row sharing a common superscript are not significantly different at P<0.05.

* Significant at P<0.05.

Table 18. Means and standard errors for category scaling, line scaling and magnitude estimation of semimembranosus steaks from the four postmortem meat treatments.

	Postmortem Treatment[1]				
Attribute[2]	DC	DC-ES	HB	HB-ES	SEM[3]
Initial tenderness					
Category scale	5.4	4.9	4.5	5.0	0.33
Line scale	9.8	8.5	8.2	8.7	0.58
Magnitude estimation	13.7	16.1	15.5	14.6	(0.041)*
Overall tenderness					
Category scale	5.3	4.5	4.2	4.6	0.30
Line scale	8.9	7.8	7.6	8.1	0.53
Magnitude estimation	16.3	18.5	20.0	17.3	(0.028)
Connective tissue					
Category scale	4.9	4.9	4.6	5.0	0.16
Line scale	9.7	9.8	9.2	9.7	0.37
Magnitude estimation	10.0	10.2	11.5	9.8	(0.024)
Overall juiciness					
Category scale	4.7b	5.1a	5.0a	5.1a	0.09*
Line scale	7.7	8.1	8.3	8.3	0.25
Magnitude estimation	12.3b	12.9b	14.4a	14.7a	(0.015)**
Flavor intensity					
Category scale	5.0	5.2	5.0	5.2	0.11
Line scale	7.9	8.0	8.0	8.2	0.23
Magnitude estimation	13.9	14.4	14.4	15.7	(0.015)

---

[1] Delay chilling (DC), delay chilling with electrical stimulation (DC-ES), hot boning (HB), hot boning with electrical stimulation (DC-ES).

[2] For a description of each evaluation technique, see **Table 7, page 65.**

[3] Standard error of the mean.

[4] Standard error expressed as log 10 value.

*, ** Significant at P<0.05 and P<0.01, respectively.

Table 19. F-values for selected sources of variation from the analyses of variance of category scale, line scale and magnitude estimation data for the four postmortem treatments.

	Source of Variation'			
Attribute	T	R	P	TxP
Initial tenderness				
Category scaling	1.40	0.47	9.34***	0.63
Line scaling	1.34	0.50	18.63***	0.85
Magnitude estimation	0.55	0.43	32.29***	0.81
Overall tenderness				
Category scaling	1.54	1.07	10.63***	0.68
Line scaling	1.25	0.81	17.98***	0.51
Magnitude estimation	1.74	0.99	67.96***	1.06
Connective tissue amt.				
Category scaling	1.32	1.19	10.03***	0.76
Line scaling	0.57	0.27	20.91***	0.57
Magnitude estimation	1.69	2.47*	34.81***	0.88
Overall juiciness				
Category scaling	4.33*	3.61**	10.31***	0.96
Line scaling	1.40	1.02	14.14***	0.87
Magnitude estimation	5.79**	4.50**	46.47***	1.21
Flavor intensity				
Category scaling	1.10	2.45*	15.82***	0.99
Line scaling	0.23	2.29	16.64***	1.07
Magnitude estimation	2.27	3.17*	61.42***	0.85

'Treatment (T), replication (R), panelist (P) and treatment by panelist (TxP).

*, **, *** Significant at $P<0.05$ and $P<0.01$, and $P<0.001$, respectively.

Table 20. Pearson correlation coefficients (r) and cofficients of determination (r^2) between panel assessments from each evaluation technique' (N=36).

Attribute	r	r^2
Initial tenderness		
CS vs LS	0.91***	0.82
CS vs ME	0.87***	0.76
LS vs ME	0.91***	0.82
Overall tenderness		
CS vs LS	0.91***	0.83
CS vs ME	0.89***	0.79
LS vs ME	-0.87***	0.76
Connective tissue amount		
CS vs LS	0.61***	0.37
CS vs ME	-0.52***	0.27
LS vs ME	-0.58***	0.34
Overall juiciness		
CS vs LS	0.48***	0.23
CS vs ME	0.59***	0.34
LS vs ME	0.53***	0.29
Flavor intensity		
CS vs LS	0.54***	0.29
CS vs ME	0.53***	0.28
LS vs ME	0.44**	0.20

'Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

, * Significant at P<0.01 and P<0.001, respectvely.

Table 21. Coefficients of determination (R^2) for initial and overall tenderness between each of the three evaluation techniques[2] fitted to five mathematical functions (N=36)[1]

Comparison Y vs X	Mathematical Function				
	Linear ($Y=a+bX$)	Power ($Y=aX^b$)	Logarithmic ($Y=a+b\ln X$)	Hyperbolic ($Y=a+b/X$)	Parabolic ($Y=a+bX+cX^2$)
Initial tenderness					
CS vs LS	0.82	0.84	0.82	0.8[illegible]	0.82
LS vs CS	0.82	0.84	0.82	[illegible]	0.83
CS vs ME	0.76	0.75	[illegible]	0.74	[illegible]
ME vs CS	0.76	0.75	[illegible]	0.74	[illegible]
LS vs ME	0.82	0.81	0.83	0.74	0.83
ME vs LS	0.82	0.81	0.83	0.83	0.83
Overall tenderness					
CS vs LS	0.83	0.80	0.8[illegible]	0.72	0.81
LS vs CS	0.83	0.80	0.81	[illegible]	0.83
CS vs ME	0.79	0.72	0.76	0.72	0.79
ME vs CS	0.79	0.72	[illegible]	0.68	0.79
LS vs ME	0.76	0.70	0.74	[illegible]	0.78
ME vs LS	0.76	0.70	0.74	[illegible]	[illegible]

[1] All coefficients of determination were significant at P<0.001

[2] Category scaling (CS), line scaling (LS) and magnitude estimation (ME)

Table 22. Pearson correlation coefficients (r)[1] and coefficients of determination (r^2) between panel tenderness assessments from each evaluation technique[1] and Warner Bratzler shear data (N=36).

Attribute	r	r^2
Initial tenderness		
CS	-0.68***	0.47
LS	-0.70***	0.49
ME	0.66***	0.44
Overall tenderness		
CS	-0.70***	0.49
LS	-0.73***	0.54
ME	0.68***	0.47

[1] Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

*** Significant at P<0.001.

Table 23. Coefficients of determination (R²)[1] between panel tenderness assessments from each evaluation technique[2] and Warner Bratzler shear data fitted to five mathematical functions (N=30)

Attribute	Mathematical function				
	Linear (y=a+bx)	Power (y=ax^b)	Logarithmic [illegible]	Hyperbolic [illegible]	Parabolic [illegible]
Initial tenderness					
CS	0.47	0.48	[illegible]	[illegible]	[illegible]
LS	0.49	0.50	[illegible]	[illegible]	[illegible]
ME	0.44	0.36	[illegible]	[illegible]	[illegible]
Overall tenderness					
CS	0.49	0.50	[illegible]	[illegible]	[illegible]
LS	0.54	0.51	[illegible]	[illegible]	[illegible]
ME	0.47	[illegible]	[illegible]	[illegible]	[illegible]

[1] All coefficients of determination were significant at P<0.001.

[2] Category scaling (CS), line scaling (LS) and magnitude estimation (ME).

Table 24. Range of mean values for the delay-chilled and hot boned postmortem treatments [illegible]

	Evaluation Technique		
Attribute	Category Scale	Line Scale	Magnitude Estimation
Initial tenderness	3.4 to 6.8 (3.4)	[illegible]	[illegible]
Overall tenderness	3.3 to 6.1 (2.8)	[illegible]	[illegible]
Connective tissue amount	3.7 to 5.5 (1.8)	[illegible]	[illegible]
Overall juiciness	4.3 to 5.4 (1.1)	[illegible]	[illegible]
Flavor intensity	4.3 to 5.4 (1.1)	[illegible]	[illegible]