Techniques to Identify Palatable Beef Carcasses: Hunterlab Beefcam™

Keith E. Belk
Colorado State University

Follow this and additional works at: http://digitalcommons.unl.edu/rangebeefcowsymp

Part of the Animal Sciences Commons

http://digitalcommons.unl.edu/rangebeefcowsymp/112

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Range Beef Cow Symposium by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
TECHNIQUES TO IDENTIFY PALATABLE BEEF CARCASSES:
HUNTERLAB BEEFCAM™

Keith E. Belk
College of Agricultural Sciences
Department of Animal Sciences
Colorado State University

Introduction

Most cattlemen agree that instrument technology, combined with mechanisms to trace livestock through the processing chain, would assist in developing a true "value- or quality-based" marketing system--where economic signals are transmitted across the entire production chain so that customer preferences are communicated to producers. Tatum et al. (1999) demonstrated the importance of being able to sort beef carcasses, based on an accurate measurement of their subsequent eating quality, if true quality management practices are to ever be implemented to reduce variation and inconsistency in beef palatability. Because Video Image Analysis (VIA) technology has been a priority for commercial testing by the beef industry (NLSMB, 1994), researchers have focused a great deal of time and effort evaluating such technology since 1994. As part of this effort, Colorado State University and Hunter Associates Laboratories, Inc. (Reston, VA) have collaborated to develop the HunterLab BeefCam™ System for sorting beef carcasses on the basis of projected eating quality.

An Overview of Beef Carcass Instrument Grading Issues

The role that USDA quality grades should play in beef marketing has been hotly contested over the years. A primary source of contention is whether the relationship between quality grade and subsequent palatability is strong enough to warrant large differences in price between carcasses differing in quality grade. Although linear quantitative relationships between quality grade and subsequent eating quality generally have been low ($R^2 < .30$), the purpose of quality grades has never been to predict, with pinpoint accuracy, eating quality per se; but, rather, to provide information to retailers and consumers concerning the risk of a specific cut or carcass being unacceptable in eating quality. In a study of 1,005 heterogeneous beef carcasses spanning the entire range of biodiversity experienced in the U.S. beef industry, Smith et al. (1987) demonstrated how USDA quality grades could be utilized effectively to determine the risk of encountering an unacceptable eating experience.

Today, however, over 80 percent of all U.S. fed beef grades either USDA Select or low Choice (Smith et al., 1995). Thus, the concept of "sorting" beef carcasses based on the risk of generating undesirable eating quality is less effective today using only USDA quality grades because normal production does not span a wide range of diversity in marbling scores. In other words, customer dissatisfaction with beef eating quality today relates to that percentage of unacceptable carcasses that fall within a narrow window of quality grade variation. Ideally, a beef carcass instrument grading system installed to estimate carcass yields also should generate
information that would enhance sortation of beef carcasses based on their expected palatability, within a narrow range of USDA quality grades.

For two decades, the beef industry has explored how various instrument technologies could best be utilized to improve the characterization, sorting and pricing of cattle and beef carcasses. Instrument assessment research was initiated in the U.S. during the 1970s (Cross and Whittaker, 1992), but progress has been slow in developing technology specifically useful for predicting both beef carcass composition and eating quality.

The National Live Stock and Meat Board convened a National Beef Instrument Assessment Planning (NBIAP) Symposium in 1994 to assess state-of-the-art capabilities in carcass/cut evaluation tools and to recommend industry research focus. The NBIAP Symposium resulted in the following conclusions: (1) "reliable, accurate tools for instrument assessment hold the promise of more accurately measuring the factors that contribute to consumer satisfaction with beef, while reducing production costs and waste," (2) "testing experimental technology under real-world conditions is critical to achieving commercial success," and (3) VIA was ready for commercial testing and was the most promising technology for short-term implementation (NLSMB, 1994).

Discussion concerning how beef grading should occur in the future and how instruments should be used encompasses several divergent positions. At one extreme are those that would eliminate federal USDA grading altogether and replace it with grading services provided by a private company (Helming, 1996) that may or may not utilize instrument technology—much like the system implemented by Canada during the '90s (although for different reasons). Advocates claim that privatization of grading would force beef packers to take responsibility for their own quality and not rely on USDA grades as an "excuse" for diminishing beef demand. On the opposite end of the spectrum are those that would maintain the status quo, making no adjustments to the current USDA grading system. In the middle are two additional positions: (1) replace USDA graders with instruments, but hold USDA accountable for operating and maintaining such equipment, or (2) use instrument technology to augment USDA field application of grade standards; i.e., allow USDA graders to provide input into the grading process that is not currently reproducible with an instrument (or not provided at all), while allowing an instrument to provide information that cannot be provided accurately by graders and to make the time-sensitive computations required at commercial packing plant chain speeds (Belk et al., 1996).

Those supporting augmentation of USDA grades believe that accuracy and repeatability of grade placement could be improved using instrument assessment technologies, but that (1) privatization of the grading system would not prove to be a credible, third-party conformity assessment system, (2) the current USDA system is voluntary and, hence, if grading were not desired by the customers of beef packing companies, it could already have been eliminated, (3) eliminating USDA grades would most likely require a change in the Agricultural Marketing Act of 1946, (4) current USDA grades are extremely important merchandising tools in the international market, (5) privatization of the grading system would have adverse peripheral effects on other marketing services offered by USDA, such as certification and Process Verification programs, and (6) USDA cannot be held accountable for low beef prices because
USDA, during this century, has made every effort possible to assist cattle producers as they market their products.

History of Beefcam™ Development

Because beef carcass composition is a common denominator in determining value, irrespective of where carcasses are produced, most of the world's instrument technology research has focused on predicting carcass yields. However, for the U.S. beef industry in particular, sortation of carcasses based on expected beef eating quality also is of significant importance, but most instruments that have been evaluated scientifically for their ability to specifically predict eating quality have not been found to be effective.

Koohmaraie et al. (1994) stated that tenderness is the major determinant of beef eating quality and proposed that tenderness be estimated (and carcasses classified) by determining shear force values for carcasses at approximately one day postmortem to predict tenderness following an appropriate aging period. Although shear force measures of carcass samples were capable of sorting carcasses based on tenderness following aging, logistical concerns, as well as strong industry resistance to invasive methodologies, continue to hamper acceptance of this system.

Studies evaluating other objective technologies for predicting beef palatability, such as ultrasound and penetration resistance measurements, have provided mixed results. Swatland (1991) developed an optical fiber probe to detect connective tissue. That instrument is currently being tested in Canada; but applied implementation has not advanced as quickly as was hoped. George et al. (1997) and Belk et al. (1996) demonstrated that the Australian Mark II Beef Grading Probe (Tendertec) was not effective in estimating subsequent palatability traits of beef carcasses. Phillips (1992) provided data from a device developed in New Zealand (MIRENZ) to predict meat texture and tenderness. However, the MIRENZ device has not yet been demonstrated to be effective in the U.S., and pilot data collected with the instrument on U.S. beef carcasses were not encouraging (Unpublished CSU data, 1997).

By 1996, several experiments suggested that lean and fat color may be related to subsequent cooked beef palatability (Hodgson et al., 1992; Hilton et al., 1997; Tatum et al., 1997; Wulf et al., 1997). Muscle and fat color can be used to measure several traits related to palatability, including: (1) presence/absence of marbling, (2) physiological maturity of the lean, (3) muscle pH, (4) production and feeding management history, and (5) ultrastructural status of sarcomeres and connective tissue within the muscle. Recent work (Tatum et al., 1997) also suggests that color may be related to calpastatin activity of postmortem muscle. Thus, by 1996, it was thought that color parameters derived from the surfaces of ribeyes may provide an additional tool for sorting beef carcasses on the basis of expected eating quality, and Colorado State University initiated pilot work with Hunter Associates Laboratory, Inc. (Reston, VA) to develop carcass palatability assessment techniques using VIA technology to evaluate color in beef carcass ribeyes the HunterLab BeefCam™ system.

BeefCam™ operates based on measurements of lean and fat color reflectance that are captured using VIA images containing up to 250,000 data points (pixels) per measurement. It
can separate out different colors from irregularly shaped surfaces and be used to calculate relative areas that each color represents within the video image, as well as provide feedback on Commission Internationale de l'Eclairage (International Commission on Illumination; CIE) values for L* (psychometric lightness; dark = 0, white = 100), a* (red = + values; green = - values) and b* (yellow = + values; blue = - values) color parameters--color measurements that reflect how the human eye perceives color.

As an example of how such systems are used by other industries, HunterLab provided the following scenario: A manufacturer of blue denim apparel is interested in monitoring (within a range of tolerances) the degree to which "stone-washed" jeans are faded or abraded before shipment and, thus, in measuring (and hence controlling) color change within the denim during washing. However, stone-washed denim is not a solid or uniform color. The HunterLab equipment allows the manufacturer to monitor, quantitatively, not only the overall color, but also the degree of uniformity in color which determines the "character" of the finished garment. This is a similar application to that expected of a grading instrument that would sort beef carcasses based on palatability characteristics of lean and fat (i.e., marbling, lean and fat color, lean and fat texture, lean and fat firmness, etc.), and could allow enhancements to those characteristics currently assessed in beef because of the capability to quantify L*, a* and b* color parameters of fat and lean—which are known to be correlated with eating quality of beef (Hodgson et al., 1992; Hilton et al., 1997; Tatum et al., 1997; Wulf et al., 1997).

To initiate BeefCam™ development, pilot data were collected using a prototype HunterLab benchtop video imaging system located in Reston, VA. Pilot data were encouraging (Belk et al., 1996), particularly if one considers that data were collected with generic software (not developed to assess meat) and a benchtop unit that required beef samples to be transported to Virginia (from Colorado) for evaluation in the laboratory.

Because the benchtop pilot data were encouraging, NCBA and Beefmaster Breeders United next funded a study that allowed development of the prototype HunterLab BeefCam™ which contained both hardware and software specifically designed for the purpose of evaluating beef carcasses. Researchers at Colorado State University evaluated the prototype system for its ability to sort beef carcasses based on expected eating satisfaction of the subsequent cooked product. One of the limitations of the study was the lack of tough (WBS > 4.5 kg) cattle in the population; but carcasses selected by BeefCam™ as being tender were, in fact, tender 95-97% of the time. The major error encountered related to those carcasses not selected by BeefCam™ as being tender, where 60-75% of such carcasses were actually acceptable in tenderness (Wyle et al., 1998). Nevertheless, the study allowed evolution and refinement of the system to generate the commercial HunterLab BeefCam™ now available.

Commercial Beefcam™ System

During 1998-99, researchers at Colorado State University conducted a study for the National Cattlemen's Beef Association (NCBA) to evaluate BeefCam's™ ability to accurately sort carcasses, on the basis of expected eating quality, by making comparisons with both Warner-Bratzler shear force values and untrained, consumer taste panel ratings for steaks from sample carcasses (Wyle et al., 1999). The study was conducted in four commercial beef packing
plants (N = 500) and was based on the premise that beef marketing impetus (those mechanisms that determine beef value) may be shifting from a system based on "commodity" pricing to a system focused on "branding" (via cooperatives, alliances, etc.) and consumer loyalty. It was thought that a beef carcass instrument assessment system that is capable of effectively sorting beef carcasses based on expected eating quality, within a narrow range of marbling scores, would add substantial value to those beef carcasses determined to yield steaks of acceptable (or higher) eating quality, and therefore, those steaks would be qualified to be marketed under "branded" beef labels.

At each packing plant, 60 carcasses were selected to represent each product line that was to be sampled at that plant. Two product lines were chosen to represent the upper two-thirds of Choice: Certified Angus Beef (CAB) and Premium Top Choice (PTC). Two products were selected that comprised the lower 1/3 of USDA Choice carcasses: Premium Low Choice (PLC) and Commodity-Trimmed Choice (CH). Both the CAB and PLC product lines also were restricted to "breed" criteria for eligibility. The product line referred to as High Select/Low Choice (HSLC) was comprised of carcasses with Slight or Small marbling. Finally, ¼-inch trim USDA Select (SE) product line samples were obtained from any carcass with Slight marbling. Carcasses were selected after application of USDA yield and quality grades, and after plant personnel had assigned carcasses to their respective product lines. It was not possible to obtain samples of all six products at each of the four plants, since not all of the plants produced all six product lines. Samples from three product lines (CAB, CH, and SE) were obtained from all four packing plants, while samples from the HSLC product line were obtained from two packing plants, and samples from the PTC and PLC products only were obtained from a single packing plant.

Two BeefCam™ models were developed to sort beef carcasses into classification groupings of either "certified as palatable" or "rejected" (Table 1). The first model (Model 1) utilized only BeefCam™ generated output to predict palatability. To test BeefCam's ability to augment USDA quality grade application, a second model (Model 2) was developed that utilized the same BeefCam™ generated output plus USDA quality grade information (where: Select = 1, Low Choice = 2, and upper two-thirds Choice = 3) to certify carcasses as palatable.

Both Model 1 and Model 2 regression equations were developed to predict a single value reflecting Warner-Bratzler shear force plus consumer-determined overall palatability--a BeefCam™ Quality Score. To determine whether or not BeefCam™ successfully sorted carcasses into differing palatability groups based on the Quality Score, results were compared against both Warner-Bratzler shear force values and untrained consumer taste panel ratings (Table 1). Any carcass that generated a steak having a shear force value of 4.5 kg or above was considered to be tough. Also, consumers rated whether or not (yes or no) they would have been pleased with the overall palatability of the sample had they purchased, prepared and eaten it at home. Consumer satisfaction rates (% like/dislike) for certified carcasses were used as a second means of measuring BeefCam™ accuracy.

With respect to both shear force values and overall palatability ratings (Table 1), using Model 1, the percentage of carcasses producing unpalatable steaks was lower--both for all carcasses and within a product line--for carcasses certified using BeefCam™ than for the
unsorted sample population as a whole. Steaks from Model 1 BeefCam™ certified carcasses had a lower mean shear force value, as well as higher mean consumer taste panel ratings for overall palatability, tenderness and flavor than did the non-certified carcasses (data not presented). Using Model 1, BeefCam™ was most accurate in identifying Choice (CAB, PTC, PLC, CH) carcasses that produced tender steaks, and was least accurate in identifying Select (SE, HSLC) carcasses that produced tender steaks. Of all carcasses that were certified by BeefCam™ using Model 1, 7.1% produced steaks that were considered tough based on shear force values (92.9% accurate), but in the certified group were SE carcasses for which 13.0% produced steaks that were considered tough. Sorting carcasses using BeefCam™ Model 1 decreased the percentage of steaks that consumers disliked, but the rate of dissatisfaction was still higher (20.7%) than desired--probably because the steaks were prepared for consumers to a constant degree of doneness and with no seasoning. The percentage of carcasses certified by BeefCam™ using Model 1 was as low as 35.0% (PLC) and as high as 65.7% (PTC). Use of Model I certified a lower percentage of carcasses with lower quality grades, and the accuracy in predicting overall palatability of steaks from these lower quality carcasses was lower than desired. Use of Model I also did not identify all of the palatable carcasses inasmuch as the percentage of carcasses in the non-certified group (those rejected for certification using BeefCam™) that generated steaks that were unpalatable only was 21.1% (based on shear force values) and 30.6% (based on taste panel ratings).

When BeefCam™ was used to sort carcasses in a USDA grading augmentation system (Model 2; Table 1), only 5.7% of the carcasses that were certified generated steaks that would be considered too tough based on shear force, while 23.4% of all carcasses rejected were too tough. BeefCam™ certified carcasses generated fewer unacceptable steaks--to consumers--than the corresponding sample population as a whole (except within the PTC product line). The BeefCam™ certified group of carcasses (using Model 2) had steaks with lower mean shear force values and higher mean taste panel ratings for overall palatability, tenderness, juiciness, and flavor than the non-certified group of carcasses (data not presented). The percentage of those carcasses certified by BeefCam™ using Model 2 that actually produced tough steaks was similar for all six products. The percentage of carcasses certified within each product ranged from 18.5% (SE) to 83.3% (CAB).

Contrasting the effectiveness of sorting carcasses using BeefCam™ Model 1 versus Model 2, it appeared that Model 2 currently provides the most accuracy. Augmentation of USDA quality grade application (Model 2) resulted in certification of fewer carcasses that would generate unacceptably tough steaks (5.7% vs 7.1%) than Model 1. Both models were more effective at segregating carcasses into differing product lines than was use of current USDA quality grade criteria (plus any other criteria that were required to be met for eligibility in the specific product line; e.g., breed characteristics). A closer look at the two models revealed that BeefCam™ Model 2 certified a higher percentage of carcasses from the CAB product line than Model 1 (83.3% vs 55.3%), while the percentage of certified CAB carcasses that generated tough steaks was the same for both models (about 3-4%). Furthermore, BeefCam™ Model II certified a lower percentage of SE carcasses (18.5% vs 35.4%), but a much lower percentage of those carcasses produced tough steaks (0.0% vs 13.0%).
In a study conducted on 292 beef carcasses selected from a commercial Colorado packing plant (Cannell et al., 1999; unpublished data) a different plant from those sampled in Wyle et al. (1999) researchers at Colorado State University sought to validate whether or not use of the BeefCam™ Model 1 algorithm (Wyle et al., 1999) was effective for sorting beef carcasses into groups differing in expected eating quality using a completely different population of carcasses to that with which the original sorting algorithms were developed. The sample population evaluated contained carcasses that were assigned USDA quality grades ranging from U.S. Standard to U.S. Prime, with the greatest proportion of carcasses falling into the U.S. Select and U.S. Choice grades (similar to the actual U.S. beef population). Sample carcasses were assigned USDA yield grades ranging between 1 and 5, and all carcasses were selected to reflect the normal variability in composition, dressing defects and quality attributes encountered by the facility on a daily basis.

Of all 292 carcasses evaluated, 47.3% (138 carcasses) were certified as palatable using the Model 1 BeefCam™ system. Of those carcasses that were certified as palatable using BeefCam™ Model 1, only 2 carcasses (1.4%) generated steaks that actually had Warner Bratzler shear force values of greater than 4.5 kg, 9 carcasses (6.5%) generated steaks that had Warner Bratzler shear force values of greater than 4.0 kg, and only 28 carcasses (20.3%) generated steaks that were assigned trained taste panel ratings of less than 5.0 for overall tenderness (on an 8-point scale where: 1 = extremely tough). BeefCam™ certified carcasses generated steaks with a mean shear force value and overall tenderness panel rating of 3.2 kg and 5.6, respectively, while those carcasses rejected for certification generated steaks with a mean shear force value and overall tenderness panel rating of 3.6 kg and 5.0, respectively. Of those carcasses rejected for certification, 21 carcasses (13.6%) generated steaks that had Warner Bratzler shear force values of greater than 4.5 kg, 48 carcasses (31.2%) generated steaks with Warner Bratzler shear force values of greater than 4.0 kg, and 70 carcasses (45.5%) generated steaks that were assigned trained taste panel ratings of less than 5.0 for overall tenderness. Thus, when tested on a separate and unique beef carcass sample population, and relative to shear force and trained taste panel responses (consumer taste panel data were not collected in the validation trial), BeefCam™ performed similarly (if not better) in accuracy to how it performed on the initial population from which the sorting algorithms were originally developed.

Although the BeefCam™ system cannot be portrayed as the "silver bullet" for sorting beef carcasses with 100% accuracy, it appears from these studies that BeefCam™ could be used effectively to further sort beef carcasses on the basis of projected eating quality, particularly in branded beef marketing systems designed to improve consumer acceptability with--and loyalty for--U.S. beef products.

References

Belk, K. E., M. H. George, H. N. Zerby, J. A. Scanga, J. D. Tatum, J. W. Wise, and G. C.


Helming, B. 1996. The time has come to shift the responsibility of beef quality grading from the government to the private sector (Parts 1 and 2). Report No. 442, 443. Bill Helming Consulting Services, Olathe, KS.


Table 1. BeefCam™ certification accuracy (for all carcasses and by product line) when system output variables were used alone (Model 1), or in combination with USDA Quality Grade factors (Model 2), to sort beef carcasses on the basis of eating quality (four packing plants; N=500).

<table>
<thead>
<tr>
<th>Mode/Product Line</th>
<th>Carcasses BeefCam™ certified, %</th>
<th>WBS values that were tough, %&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Consumer panelists that disliked the product, %&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Certified</td>
<td>Rejected</td>
</tr>
<tr>
<td><strong>Model 1:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Carcasses</td>
<td>50.8</td>
<td>14.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Certified Angus Beef</td>
<td>55.3</td>
<td>4.4</td>
<td>3.2</td>
</tr>
<tr>
<td>Premium Top Choice</td>
<td>65.7</td>
<td>17.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Premium Low Choice</td>
<td>35.0</td>
<td>10.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Commodity Choice</td>
<td>63.4</td>
<td>10.7</td>
<td>4.8</td>
</tr>
<tr>
<td>High Select/Low Choice</td>
<td>50.0</td>
<td>20.0</td>
<td>16.0</td>
</tr>
<tr>
<td>Commodity Select</td>
<td>35.4</td>
<td>23.8</td>
<td>13.0</td>
</tr>
<tr>
<td><strong>Model 2:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Carcasses</td>
<td>53.0</td>
<td>14.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Certified Angus Beef</td>
<td>83.3</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>Premium Top Choice</td>
<td>74.3</td>
<td>17.1</td>
<td>7.7</td>
</tr>
<tr>
<td>Premium Low Choice</td>
<td>32.5</td>
<td>10.0</td>
<td>7.7</td>
</tr>
<tr>
<td>Commodity Choice</td>
<td>63.4</td>
<td>10.7</td>
<td>4.8</td>
</tr>
<tr>
<td>High Select/Low Choice</td>
<td>48.0</td>
<td>20.0</td>
<td>16.7</td>
</tr>
<tr>
<td>Commodity Select</td>
<td>18.5</td>
<td>23.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<sup>a</sup>Tough = percent of carcasses producing strip loin steaks with a Warner-Bratzler shear force value > 4.5 kg (cooked to 70°C).

<sup>b</sup>Disliked = percent of carcasses for which consumers indicated (yes or no) that they would be displeased with the overall like/dislike of the sample had they purchased the product and prepared it at home (cooked to 70°C).