AN EVALUATION OF THE PROTOTYPE PORTABLE HUNTERLAB VIDEO IMAGING SYSTEM (BEEFCAM) AS A TOOL TO PREDICT TENDERNESS OF BEEF CARCASSES USING OBJECTIVE MEASURES OF LEAN AND FAT COLOR

A.M. Wyle, R.C. Cannell, K.E. Belk, M. Goldberg, R. Riffle, and G.C. Smith

Summary

This project evaluated the performance of the HunterLab Video Imaging System (BeefCam) for accuracy and precision in predicting the tenderness of beef carcasses. Data were collected on 348 carcasses of Beefmaster steers at two different packing plants. Samples were obtained from the 12th rib portion of the *longissimus* muscle of the left side of each carcass for Warner-Bratzler Shear (WBS) force analysis. The first Trial (plant 1) was conducted to determine the best prediction model for sorting carcasses into "tender" or "tough" groups, and the second Trial (plant 2) was conducted to validate results obtained from Trial I. In Trial I, BeefCam was able to correctly identify 150 of 156 carcasses that actually produced tender (WBS < 4.5 kg) steaks, and in Trial II BeefCam was able to correctly identify 139 of 150 carcasses that actually produced tender steaks. Because there were so few carcasses that had tough (WBS > 4.5 kg) steaks – 6.0 % in Trial I and 5.1% in Trial II—this study may not have provided a meaningful test of the BeefCam. Nevertheless, of the 10 carcasses in Trial I that actually produced tough steaks, BeefCam correctly identified 5 of them, and of the 8 carcasses in Trial II that actually produced tough steaks BeefCam correctly identified 2 of those. BeefCam inappropriately classified very small numbers of carcasses as "likely to produce tough beef", or as "likely to produce tender beef" when the opposite was in fact true. So we concluded that further study is needed to deliver an instrument with higher precision.

Introduction

For youthful carcasses, the current United States Department of Agriculture (USDA) Quality Grading system relies principally on the relationship between the amount of intramuscular fat (marbling) in the ribeye and the palatability of the cooked beef. There is a positive correlation between marbling and tenderness (Smith *et al*., 1987; Morgan *et al*., 1991), but the relationship is not very useful when the range of marbling in a test population is narrow (e.g. "small" versus "slight"). More than 80% of the steer and heifer carcasses presented for quality grading qualify for Select or the lower third of Choice (Smith *et al*., 1995). In this narrow range of marbling scores, the USDA quality grading system does not adequately sort carcasses likely to produce tough vs. tender beef; therefore, some carcasses receive a premium (the difference in price between Choice vs. Select) that is not warranted (Wheeler *et al*., 1994). In order for a value-based marketing system, such as the one envisioned by Cross and Whittaker (1992), to work, both the producer and the packer must be confident that subjectivity has been minimized in the USDA quality grading system.

The National Beef Instrument Assessment Planning (NBIAP) Symposium identified as the most promising technologies for improving quality and consistency of beef, video image analysis (VIA), electromagnetic scanning (ToBEC), carcass ultrasound technology, elastography, Swatland's probe, the Australian Tendertec system and real-time ultrasound of live animals (National Live Stock and Meat Board, 1994). Li *et al.* (1997) determined that software could be written to segment a video image of a ribeye into fat, lean and connective tissue components accurately enough to conduct color analysis on each of those components independently, but the RGB color of lean computed from the video images was not a good predictor of tenderness. Hodgson *et al.* (1992) found a significant ($P < .01$) correlation between the external fat color of cow carcasses and the tenderness of their steaks so it is important to be able to separate the fat from the connective tissue in the video images.

There are many different scales used to measure color, but the CIE L*, a*, and b* color scale has been the most promising for the prediction of beef carcass tenderness. Wulf *et al*. (1997) found that L* (lightness; dark = 0, white = 100), a* (red = + values, green = - values), and b* (yellow = + values, blue = - values) were correlated with tenderness of beef, while Tatum *et al.* (1997) determined that the b* value of CIE color space was valuable in explaining some of the variation in beef tenderness. The color measurement devices used in the Wulf *et al.* (1997) study (Minolta Colorimeter) and in the Tatum *et al.* (1997) study (HunterLab MiniScan) both had small apertures, so color measurement was taken on only a small portion of the exposed *longissimus* muscle.

Cannell *et al.* (1997) investigated a system that used the L*, a*, b* color scale in a video image analysis system to segment a ribeye image into fat, lean and connective tissue components (pixel by pixel) so that an objective color measurement could be recorded for both lean and fat. Cannell *et al.* (1997) reported that the HunterLab Color Vision System could convert the digital RGB (red, green, and blue) color scale of which a video image is composed to the CIE L*, a*, b* color space, while Belk *et al.* (1997) determined that the HunterLab Color Vision System, in conjunction with expert marbling scores, was able to significantly improve the assessment of beef carcass tenderness.

Experimental Procedures

Description of the BeefCam Device

The prototype BeefCam device is a portable, handheld video-imaging unit that utilizes an internal mirror to reflect the image onto a horizontally positioned camera lens, while maintaining the proper focal length. The HunterLab designers managed to create a folded effect in which the camera and the sample are perpendicular to each other which makes the system very easy to align over the exposed *longissimus* muscle. Because the mirror and camera must be 90^o apart, the front of the device has a wedge shape that allows it to slide easily into position over the exposed ribeye muscle.

BeefCam uses a Panasonic camera to acquire images that are stored in a fully functioning, portable computer housed in a separate protective stainless steel case coupled to the BeefCam device by a data/power cord. The protective case can either be worn as a backpack or placed on a cart and moved along with the device.

Before any of the trials were conducted, HunterLab developed software that can be loaded onto the computer, allowing BeefCam to be completely self-contained. The algorithms and conversions allow the device to acquire RGB images and convert them into the CIE L*, a*, and b* color space for lean and fat. In this study, BeefCam algorithms generated, in total, the machine-measured variables shown in Table 2. This software allowed the processing of images to occur without technician input.

Table 2. List of variables measured by BeefCam.

Variable	Variable Name
R, G, and B of muscle	LeanR, LeanG, LeanB
R, G, and B of fat	FatR, FatG, FatB
L*, a*, b* of muscle	Lean L, Lean a, Lean b
L*, a*, b* of fat	Fat L, Fat a, Fat b
Percent intramuscular fat	Marbpct
Average marbling fleck size	fleksize
Standard deviation of fleck size	fleksd
Number of flecks of marbling	flekcnt
Marbling distribution	marbdist
Standard deviation of marbling distribution	mdistsd
Ribeye area	viarea
Principal components[a] of lean L*, a*, b*	pclean1, pclean2, pclean3
Principal components[a] of fat L*, a*, b*	Pcfat1, pcfat2, pcfat3

[a]Determined using a multivariate technique for examining relationships among several quantitative variables (SAS, 1988).

Cattle Selection and Packing Plants

Carcasses for this study were from *Bos indicus* crossbred steers (n = 324) that were selected by representatives of Beefmaster Breeders United. Cattle were slaughtered commercially at packing plants that used high-voltage electrical stimulation and were chilled for 36-48 hours before carcass evaluation occurred.

Data Collection

In order to insure that the color of the *longissimus* muscle was measured accurately, BeefCam was calibrated using a set of color calibration plates. Carcass scanning was conducted by trained personnel from HunterLab and Colorado State University. Proper care was taken to insure that the complete *longissimus* muscle was included in the image and to ensure that the device was properly aligned with the carcass (so that no distortion occurred in the image).

For use in analyses, expert grade data (maturity, marbling, adjusted preliminary yield grade, and kidney/pelvic/heart fat) were collected on each of the carcasses in the study by USDA personnel. In addition, the hot carcass weight and ribeye area of each carcass was recorded. The USDA graders were given as much time as necessary to accurately determine the yield and quality grade factors.

Tenderness Assessment

A one-inch sample was obtained from the *longissimus* muscle at the 12^{th}-13^{th} rib interface on the left side of each carcass and transported to the Colorado State University meat laboratory. The muscle samples were aged for 14 days at 2^{o} C and then frozen. Upon subsequent tempering for 24 hours at 2^{o} C, Warner-Bratzler Shear (WBS) force analysis was performed at an endpoint cooking temperature of 70^{o} C according to American Meat Science Association (1995) guidelines. All of the samples were broiled on a Hobart Char Broiler (Model CB 51, Hobart Corp., Troy, OH). The samples were turned at 4, 8, 12 and, if necessary, 16 minutes until an internal temperature of 70^{o} C was reached. The samples were allowed to equilibrate to room temperature (approximately 20^{o} C) and then, between 6 and 8, 1.27-cm diameter cores were removed from each sample parallel to the muscle fiber orientation. A single, peak shear force measurement was obtained using a Warner-Bratzler Shear force machine, and the mean shear force was calculated and recorded for each steak. Any sample with a mean shear force greater than 4.5 kg was considered to be "tough" and any sample with a mean shear force less than 4.5 kg was considered to be "tender" (National Live Stock and Meat Board, 1994). Results of the WBS force tests were used to determine how accurately the BeefCam device was able to sort the carcasses according to differences in the tenderness of their steaks.

Analysis of Data

Due to changes made in the BeefCam device between the first two Trials, the data had to be analyzed separately, by Trial (n=166 for Trial I and n=158 for Trial II). The purpose of the analysis was to determine if the BeefCam device was capable of sorting beef carcasses into groups that produced steaks that were either "tender" or "tough". It was determined that use of this binomial segmentation of the carcasses would provide the most conservative estimates of BeefCam's performance.

The first step was to determine whether or not any single variable could be used to "pre-certify" carcasses that were tender. This was accomplished using scatter plots of the individual BeefCam-generated data plotted against WBS force at an endpoint cooking temperature of 70^{o} C. Variables were determined to be useful for purposes of pre-certification if a high or low quantitative value corresponded with acceptability in tenderness. The next step was to create multivariate principal components for the fat and lean L*, a*, and b* colors, as well as BeefCam's marbling variables. The principal components for the lean were computed using the SAS statistical software package (SAS, 1988). The lean L*, a*, and b* values were weighted by SAS to produce three principal component variables which modeled the three-dimensional relationship between the three color parameters. The first, second and third principal components of lean L*, a* and b* values accounted for between 63.1 and 73.5%, 19.4 and 36.1%, and 0.07 and 7.2% of the standardized variance across the three Trials, respectively. The first, second and third principal components of fat L*, a* and b* values accounted for between 51.1 and 96.1%, 3.4 and 28.0%, and 0.1 and 20.9% of the standardized variance across the three Trials, respectively. Similarly, the first, second, and third principal components (the fourth, fifth and sixth principal components were not calculated) of the machine marbling measurements accounted for between 49.2 and 78.3%, 18.0 and 26.9%, and 1.7 and 11.1% of the standardized variance across the three Trials, respectively.

After this, a linear regression analysis was performed using the SAS Statistical Software Package (SAS, 1988). Using WBS force as the dependent variable and all of the machine-measured variables and the principal components (excluding those used for pre-certification) as independent variables, a stepwise regression was performed. Only data on those carcasses remaining after pre-certification was completed were included in the regression model development process. Using these procedures, a certification plot of the actual WBS force versus the predicted WBS force for each Trial was developed. By separating each resulting plot into four quadrants (predicted tender-actually tender, predicted tender-actually tough, predicted tough-actually tender, and predicted tough-actually tough), accuracy of the BeefCam device relative to its ability to sort beef carcasses based on tenderness could be determined.

Results and Discussion

Assessment of sample populations. The two sample populations were representative of the kinds of carcasses on which BeefCam would likely be used. Descriptive statistics for the two sample populations are shown in Table 1; the sample populations of carcasses from Trials I and II were very similar in characteristics (Table1). It would have been helpful if the sample populations had differed more in WBS force values; as it turned out there were not enough carcasses that had tough (WBS > 4.5 kg) steaks. Of the 166 carcasses analyzed in Trial I, only 10 had steaks with WBS force values greater than 4.5 kg (6.0%). Similarly, in Trial II only 8 of 158 carcasses produced steaks that had WBS force greater 4.5 kg (5.1%). It was, correspondingly, difficult to determine how effectively BeefCam might sort carcasses with large variations in steak tenderness because the sample population was not comprised of carcasses that exhibited such variability.

Table 1. Descriptive statistics of the sample populations by trial.

Variable	Trial 1 (n = 166) Mean $\pm$ Std Dev	Trial 2 (n = 158) Mean $\pm$ Std Dev
Fat thickness	0.54 $\pm$ 0.19	0.55 $\pm$ 0.18
Hot carcass weight	798.7 $\pm$ 83.0	700.1 $\pm$ 69.8
Ribeye area	13.5 $\pm$ 1.6	12.6 $\pm$ 1.4
KPH fat, %	2.29 $\pm$ 0.62	1.93 $\pm$ 0.35
Final yield grade	3.04 $\pm$ 0.81	2.88 $\pm$ 0.78
Marbling score[a]	391.0 $\pm$ 72.3	373.0 $\pm$ 42.8
Overall maturity score	62.9 $\pm$ 26.9	51.1 $\pm$ 9.3
Final quality grade[b]	269.9 $\pm$ 45.6	266.9 $\pm$ 37.3
WBS force	3.43 $\pm$ 0.65	3.43 $\pm$ 0.60

[a] The marbling scale is as follows: 200 = Traces, 300 = Slight, 400 = Small, 500 = Modest, 600 = Moderate, 700 = Slightly Abundant.
[b] The quality grade scale is as follows: 100 = Standard, 200 = Select, 300 = Choice, 400 = Prime.

Trial I. During Trial I, 166 carcasses were scanned using the BeefCam device. In an attempt to "pre-certify" some of the carcasses, it was determined that any carcass exhibiting a second principal component value of lean L*, a*, and b* color space greater than zero could be certified as "tender". Of the 80 carcasses with a second principal component for lean L*, a*, and b* color space greater than zero, only 4 beef carcasses were "tough". Secondly, any carcasses that possessed a machine-measured percent marbling value of greater than 5.0 also were certified as being "tender". Of the 17 carcasses certified with this method, none were actually found to have steaks that were "tough". After the pre-certification was completed, linear regression was performed on data from all remaining carcasses. The prediction equation developed is shown in Table 3. It was evident that the model did not have a strong linear relationship to beef carcass tenderness ($R^2 = 0.19$, RMSE = 0.657).

Table 3. Beta coefficients for the regression equation predicting WBS force for Trial I.

Variable	Coefficients
Constant	-1.2905
LeanB	0.0249
Fat b	0.0506
Pcfat3	-3.3018
Pcmarb3	0.3231
R^2	0.19
RSME	0.657

A combined logistical model was used to separate the carcasses into "tough" and "tender" groups effectively. It was arbitrarily determined that any carcass with a predicted WBS force value, for its steaks, of greater than 3.8 kg would not be certified as being "tender", and any carcass with a predicted WBS force value for its steaks, of less than 3.8 kg would be certified as being "tender".

The BeefCam device did a very good job of sorting out the "tough" carcasses. In all, the BeefCam device was able to certify 154 carcasses as being "tender" out of 166 (92.8%), while only classifying 4 carcasses as being "tender" that were actually "tough" (2.6%). However, one area that could be improved was the identification of the "tender" carcasses among those carcasses that were classified as being "tough". BeefCam categorized too many carcasses that were actually "tender" as being "tough". While this error was not detrimental to the performance of the device, it would result in lost revenue to anyone selling beef based on BeefCam's determination of carcass palatability. The probability of BeefCam not classifying a carcass as "tender" when it actually was "tender" was 61.6% and is reported in Table 4.

Table 4. Probability of BeefCam incorrectly identifying a carcass as "tough" or "tender" and the mean WBS force of each subset.

Data	Mean	SD	Total Number in Group[a]	Number Correctly Sorted[b]
Trial 1 certified	3.38	0.57	152	146
Trial 1 not certified	4.17	1.12	14	6
Trial 2 certified	3.32	0.57	145	139
Trial 2 not certified	3.70	1.13	13	2

[a] The number of carcasses that are included in that particular group of cattle (i.e., the number of cattle certified versus the number not certified).
[b] Those carcasses in a particular group that were sorted correctly by BeefCam.

Trial II. Before the start of Trial II, there were some unfortunate occurrences that affected results. While the BeefCam device was being transported to the plant, the shipping box broke, the calibration plate that was needed to standardize the color measurement device was lost, and one of the fluorescent light bulbs was broken. In order to complete the trial, the machine was standardized using an alternate method, and the light bulb was replaced with a bulb that created a different light environment as compared to that used in Trial I. As a result, images collected in Trial II yielded slightly different color measurements than did images collected from Trial I. However, in order to try to validate the model developed in Trial I, it was decided to force the same variables into the regression model, but to allow different beta coefficients to be developed for the prediction equation. The prediction model developed for Trial II is shown in Table 5.

Table 5. Beta coefficients for the regression equation predicting WBS force in Trial II.

Variable	Coefficients
Constant	5.5424
LeanB	-0.1379
Fat b	0.0159
Pcfat3	1.3215
Pcmarb3	0.2001
R^2	0.07
RMSE	0.711

The same pre-certification methods were used for Trial II that were used in Trial I. As in Trial I, the arbitrary line used to separate those carcasses that were predicted to be “tender” from those that were predicted to be “tough” was at 3.8 kg of estimated WBS force. Only 1 of the 85 carcasses pre-certified as “tender” by the second principal component for lean L*, a*, and b* color value (see Table 2) were actually “tough” (1.2%). Similarly, none of the 3 carcasses pre-certified as tender using the machine percent marbling value were actually “tough”. The probability that BeefCam incorrectly identified a carcass as “tough” or “tender” is shown in Table 4. In Trial II, BeefCam was able to certify 145 out of 158 carcasses as being “tender” (94.2%), and only 6 carcasses in this category were actually “tough” (4.1%). As evidenced by data presented in Table 4, the model actually did a better job of sorting carcasses in Trial II than it did in Trial I, even though it had a lower R^2 value (0.07 vs. 0.19) and a higher RMSE (0.711 vs. 0.657). The probability that BeefCam identified a carcass as being “tough” when, in fact, it actually was “tender” was 75.9% and is reported in Table 4.

Conclusions

For Trials I and II, BeefCam was able to accurately sort out the carcasses that produced “tough” steaks, such that those carcasses that were certified as being “tender” were “tender” 98-99% of the time. In these Trials, BeefCam was not as accurate in sorting out the carcasses that produced “tender” steaks, as 62-76% of those beef carcasses classified as “tough” were actually “tender”. Trial II was successful in validating the prediction model used in Trial I which suggests that BeefCam could be used as a carcass sorting tool, but it would need to be more accurate in sorting out all of the “tender” carcasses. Again, the lack of carcasses in the two sample populations that produced steaks that were “tough” (WBS force > 4.5 kg) made it difficult to determine the effectiveness of using BeefCam as a sorting tool for identifying carcasses likely to produce “tender” vs. “tough” steaks. It is clear that further validation of the accuracy achieved in these Trials is necessary and should be conducted using a sample population that includes carcasses with steaks representing a wider range of tenderness.

References

American Meat Science Association. 1995.

Belk, K.E., R.C. Cannell, J.D. Tatum, and G.C. Smith. 1997. Video imaging systems for composition and quality. Proceedings of the Meat Industry Research Conference. American Meat Institute, Washington, D.C.

Cannell, R.C., K.E. Belk, C.D. Bunting, H.N. Zerby, J.A. Scanga, M.H. George, J.W. Wise, J.D. Tatum, and G.C. Smith. 1997. An evaluation of the HunterLab Color Vision System (a VIA system) for use in improving grading accuracy and for augmenting application of USDA quality and yield grades and for use in predicting palatability and tenderness attributes of beef. Final Report to the National Cattlemen’s Beef Association. Colorado State University, Ft. Collins.

Cross, H.R. and Whittaker, A.D. 1992. The role of instrument grading in a beef value-based marketing system. J. Anim. Sci. 70:984.

Hodgson, R.R., K.E. Belk, J.W. Savell, H.R. Cross, and F.L. Williams. 1992. Development of a quantitative quality grading system for mature cow carcasses. J. Anim. Sci. 70: 1840.

Li, Jun, J. Tan and F.A. Martz. 1997. Predicting beef tenderness from image texture features. American Society of Agricultural Engineers. Mimeographed Report.

National Cattlemen’s Association. 1994. National Beef Tenderness Conference, Denver, CO. April 22-23, 1994. National Cattlemen’s Association, Englewood, CO.

National Live Stock and Meat Board. 1994. National Beef Instrument Assessment Plan-1994. National Live Stock and Meat Board, Chicago, IL.

Smith, G.C., Z.L. Carpenter, H.R. Cross, C.E. Murphey, H.C. Abraham, J.W. Savell, G.W. Davis, B.W. Berry, and F.C. Parrish, Jr. 1987. Relationship of USDA marbling groups to palatability of cooked beef. J. Food Qual. 7: 289.

Tatum, J.D., M.H. George, K.E. Belk, and G.C. Smith. 1997. Development of a palatability assurance "critical control points" (PACCP) model to reduce the incidence of beef palatability problems. Final report to the National Cattlemen's Beef Association. Colorado State University, Ft. Collins.

Wheeler, T.L., L.V. Cundiff and R.M. Koch. 1994. Effect of marbling degree on beef palatability in *Bos taurus* and *Bos indicus* cattle. J. Anim. Sci. 72: 3145.

Wulf, D.M., S.F. O'Conner, J.D. Tatum, and G.C. Smith. 1997. Using objective measures of color to predict beef *longissimus* tenderness. J. Anim. Sci. 75: 684.