Real-time ultrasound: What does image quality mean to genetic evaluations?

M. L. Spangler[1] and D. W. Moser[2]

[1]University of Nebraska, Lincoln; [2]Kansas State University, Manhattan

ABSTRACT: Ultrasonic measurements for the traits of ribeye area (UREA), percentage intramuscular fat (UPFAT) and 12th rib fat (UFAT) and their associated image quality scores (IQ) obtained from Ultrasound Guidelines Council (UGC) certifications held in Ames, Iowa from 2002-2006 were utilized to quantify the relationship between IQ and the absolute value of prediction bias (ABS). Over 5,000 records were used for each trait. Currently, IQ are utilized as a subjective criterion to partially determine the proficiency of an ultrasound technician. IQ represents a 1-7 scale where those images classified as a 1 or 2 are acceptable, 3-5 are marginal, and images scored as a 6 or 7 are rejected and not interpreted in practice. Only IQ scores of 1-5 were used in the current study. The GLM procedure of SAS was used to predict ABS with potential explanatory variables of interpreting technician, IQ, machine, animal, and year. For UPFAT an additional explanatory variable of categorical nature was added (CAT). CAT was a subjectively defined categorical variable used to account for any potential differences in predictive ability by animals that had different levels of ether extract values that are used as the carcass reference for UPFAT. IQ was not statistically significant in the prediction of ABS for the traits of UREA and UPFAT. Although IQ was significant (P=0.0018) for the prediction of ABS for UFAT, IQ did not numerically contribute to the predictive ability of the model. The results from the current investigation suggest that the subjective measure of IQ is not strongly related to ABS and as a consequence bias should not be introduced into genetic evaluations by including multiple IQ scores in a contemporary group for any of the three traits considered here. As a caveat to the objectives of the current study, the effects of machine and CAT, as well as their interaction, were significant in the prediction of ABS for UPFAT. Further investigation is warranted to determine if differences between interpreting software exist with particular regards to the reduction of variation for the trait of UPFAT.

INTRODUCTION

The benefits of real-time ultrasound, particularly the shortened generation interval, have been well described (Wilson, 1992). Real-time ultrasound has become a standard method of collecting phenotypic measurements for the traits of external fat, intramuscular fat, and ribeye area. In order to insure the data quality going into National Cattle Evaluations (NCE) is such that useful Expected Progeny Differences (EPDs) can be estimated guidelines that determine the proficiency of technicians have been established. The governing body that provides guidelines for the collection and interpretation of ultrasound images that are used for NCE is the Ultrasound Guidelines Council (UGC). This council provides criteria that field and lab technicians must meet in order to be certified to collect and/or interpret ultrasound images when the data will be used in NCE. Criterion currently used included the standard error of prediction, standard error of repeatability, correlations, proportion of images accepted (able to be interpreted), and image quality score. The UGC currently utilizes a 1-7 scoring system to subjectively categorize ultrasound images where image quality scores of 1 are the most desirable and those of 7 are the least desirable. Images categorized as either 6 or 7 are not interpreted and thus the data from these images cannot be included in NCE. If this image quality scoring system is able to differentiate images according to the accuracy with which they can be interpreted then it stands to reason that an inherent, and currently unaccounted for, bias is introduced in the case where multiple image quality scores are represented within a contemporary group. Consequently, the objective of the current study was to determine if differences exist in the absolute value of prediction bias between image quality scores of 1-5 for the traits of rib fat, percentage of intramuscular fat, and ribeye area.

MATERIALS AND METHODS

Data

Ultrasound interpretations for the traits of 12th rib fat (UFAT), ribeye area (UREA), and percentage intramuscular fat (UPFAT) and corresponding carcass data were obtained from UGC field certifications held in Ames, Iowa from 2002-2006. Only images scored as 1-5 for image quality were used in the current study, as images quality scores of 6 and 7 are not routinely interpreted during field certifications. All images were interpreted using UGC certified software. Interpretations were performed by UGC certified lab technicians. Trained personnel at a commercial plant collected carcass data. The carcass measure associated with UPFAT is the percentage of intramuscular fat as measured using ether extract. Simple statistics including frequencies can be found in Table 1.

Analysis

The GLM procedure of SAS was used to determine the factors that had the largest impact on the absolute values of the bias between ultrasound predictions and carcass values (ABS). The absolute value of prediction bias was used due to the fact that within an image quality score there are deficiencies that can cause both an overestimation and underestimation. ABS was calculated as follows:

$$ABS = |UTRAIT - CTRAIT| \qquad (1)$$

Where UTRAIT is any trait measured by ultrasound and CTRAIT is the corresponding carcass measurement.

Explanatory variables included animal (ANIM), year (YEAR), ultrasound machine (MACH), interpreting technician (INTERP), image quality score (IQ), and in the case of UPFAT a categorical trait was created to represent differing levels of intramuscular fat in the carcass (CAT). Descriptions of CAT can be found in Table 2.

RESULTS and DISCUSSION

For UREA and UPFAT, IQ was not a significant source of variation contributing to ABS. For the prediction of ABS with regards to ribeye area, explanatory variables of INTERP ($P < 0.0001$) and ANIM nested within YEAR ($P < 0.0001$) were included in a model along with IQ ($P = 0.771$) that explained 32.02% of the variation. The optimal model explained 32.00% of the variation and included effects of INTERP and ANIM nested within YEAR. Although IQ was significant in the prediction of ABS for UFAT (P=0.0018) along with INTERP (P < 0.0001) and the nested effect of ANIM within YEAR (P < 0.0001), a substantial numerical difference in the R^2 value did not exist when IQ was included in the model (R^2 = 53.64%) versus when it was excluded (R^2 = 53.49%). For the prediction of ABS with regards to percentage of intramuscular fat, explanatory variables of MACH (P = 0.0264), CAT (P < 0.0001), INTERP (P < 0.0001), ANIM (P < 0.0001), and IQ (P = .3361) explained 53.73% of the variation. The optimal model included MACH, CAT, INTERP, ANIM and explained 52.43% of the variation in ABS. The fact that CAT was significant in the UPFAT model suggests that extreme ether extract categories impact ABS. Other studies have also identified the range in ether extract data to be a significant source of variation in explaining ABS (Herring et al., 1998). In a reduced model, their existed a significant interaction between CAT and MACH suggesting that differences may exist between interpreting software with regards to their ability to accurately predict percentage of intramuscular fat in extreme categories. The data presented here is insufficient to make strong inferences regarding the implications of this interaction.

Previous studies have shown that the skill level of the image collecting technician as well as the interpreting technician can influence prediction bias (Herring et al., 1994; Perkins et al., 1992). In the case of the current study the field technicians represented various skill levels and the lab technicians represented, presumably, posses a higher skill level than would be expected from the average of those routinely tracing images. With particular regard to UREA, it is reasonable to expect that experienced lab technicians can interpret image quality scores of 1-5 with similar bias given their biological knowledge of the shape of the muscle. Individual animals create unique sources of variation that can create differences in prediction bias including hide thickness (Chambaz et al., 2002) and extreme values for ether extracted fat percentage that may fall outside of the values used to develop UPFAT models. It should also be noted that in the case of 12^{th} rib fat and ribeye area, carcass measures are not without error given differences in hide pulls or potential measurement error.

IMPLICATIONS

Results from the current study suggest that an unaccounted for bias is not introduced into genetic evaluations through the inclusion of multiple image quality scores within a single contemporary group. In the current framework where images are collected then sent to an approved lab for interpretation and then to a breed association for inclusion in National Cattle Evaluations it would seem that image quality is only important from the standpoint of an image being of good enough quality to interpret. The traditional scale of 1-7 best serves as a teaching tool and perhaps as a method of quality control in processing labs. In the event that advancement is made such that auto interpretation is used in the field the impact of image quality would need to be reevaluated. A simultaneous evaluation of current technologies is warranted to directly compare software.

LITERATURE CITED

Chambaz, A., P. A. Dufey, M. Kreuzer, and J. Gresham. 2002. Sources of variation influencing the use of real-time ultrasound to predict intramuscular fat in live beef cattle. Canadian J. Anim. Sci. 133-139.

Herring, W. O., A. L. Kriese, J. K. Bertrand, and J. Crouch. 1998. Comparison of four real-time ultrasound systems that predict intramuscular fat in beef cattle. J. Anim. Sci. 76:364-370.

Herring, W.O., D. C. Miller, J. K. Bertrand, and L. L. Benyshek. 1994. Evaluation machine, technician, and interpreter effects on ultrasonic measures of backfat and longissimus muscle area in beef cattle. J. Anim. Sci. 72:2216-2226.

Perkins, T. L., R. D. Green, K. E. Hamlin, H. H. Shepard, and M. F. Miller. 1992. Ultrasonic prediction of carcass merit in beef cattle: Evaluation of technician effects on ultrasonic estimates of carcass fat thickness and longissimus muscle Area. J. Anim. Sci. 70:2758.

Wilson, D. E. 1992. Application of ultrasound for genetic improvement. J. Anim. Sci. 70: 973:983.

Table 1. Summary statistics including frequencies, means and standard deviations (SD) of the absolute value of bias (ABS) by image quality score (IQ)

Trait	IQ	Frequency	Mean ABS[a]	SD
UREA[b]	1	657	1.05	0.80
	2	1,982	1.03	0.78
	3	1,379	0.99	0.76
	4	865	0.97	0.75
	5	531	1.04	0.76
UFAT[c]	1	657	0.076	0.061
	2	1,982	0.069	0.055
	3	1,379	0.074	0.057
	4	865	0.075	0.057
	5	531	0.080	0.061
UPFAT[d]	1	1,036	1.14	0.96
	2	3,079	1.03	0.87
	3	784	0.99	0.84
	4	543	0.97	0.89
	5	306	0.92	0.75

[a] ABS= |Ultrasound prediction – Carcass value|
[b] Ultrasonically measured ribeye area in square inches
[c] Ultrasonically measured 12th rib fat in inches
[d] Ultrasonically measured percentage of intramuscular fat

Table 2. Frequencies and ranges of categorical classification of ether extract data (CAT)

CAT	Frequency	Range[a]
1	105	0-1.50
2	727	1.50-2.50
3	1,813	2.50-3.50
4	1,110	3.50-4.50
5	1,192	4.50-5.50
6	293	5.50-6.50
7	369	6.50-7.50
8	139	> 7.50

[a] Range is measured in percentage of intramuscular fat as determined by ether extract