Development of the Meat Standards Australia (MSA) prediction model for beef palatability

R. Watson[A,D], *R. Polkinghorne*[B] *and J. M. Thompson*[C]

[A]Department of Mathematics and Statistics, University of Melbourne, Parkville, Vic. 3010, Australia.
[B]Marrinya Agricultural Enterprises, 70 Vigilantis Road, Wuk Wuk, Vic. 3875, Australia.
[C]Cooperative Research Centre for Beef Genetic Technologies, School of Environmental and Rural Sciences,
  University of New England, Armidale, NSW 2351, Australia.
[D]Corresponding author. Email: rayw@ms.unimelb.edu.au

Abstract.   In this paper, the statistical aspects of the methodology that led to the Meat Standards Australia (MSA) prediction model for beef palatability are explained and described. The model proposed here is descriptive: its intention is to describe the large amounts of data collected by MSA. The model is constrained to accord with accepted meat science principles. The combined dataset used in development of the prediction model reported is around 32 000 rows × 140 columns. Each row represents a sample tasted by 10 consumers; each column specifies a variable relating to the sample tested. The developed model represents the interface between experimental data, scientific evaluation and commercial application. The model is used commercially to predict consumer satisfaction, in the form of a score out of 100, which in turn determines a grade outcome. An important improvement of the MSA model relative to other beef grading systems is that it assigns an individual consumer-based grade result to specific muscle portions cooked by designated methods; it does not assign a single grade to a carcass.

Additional keywords: *Bos indicus* content, carcass suspension and carcass weight cooking methods, consumer sensory testing, hormonal growth implants, ossification and marbling scores.

Introduction

Previous papers (Ferguson *et al*. 1999; Polkinghorne *et al*. 1999; Thompson *et al*. 1999; Thompson 2002) and papers in this special issue (Polkinghorne *et al*. 2008*a*; Watson *et al*. 2008*a*) have described the development of the Meat Standards Australia (MSA) beef grading system as a useful tool for the assessment of beef palatability in a commercial environment. These papers have described the gradual development of the MSA system from an idea to a practical tool, currently being used in Australia to grade the eating quality of beef. In this paper, some of the statistical aspects employed in the development of the current MSA grading model are described. The MSA grading model is dynamic and further corrections, modifications and improvements are anticipated as more data becomes available, or as better ways of describing the present dataset are found.

The model aims to use animal and carcass processing factors to describe the large amounts of palatability data collected by the MSA consumer panels. The data represents consumer test results obtained under standard test conditions (see Watson *et al*. 2008*a* for more detail) related to a large number of variables pertinent to the source muscle, animal, production and processing practices and the cooking methods employed. The model was constrained to accord with accepted meat science principles.

The dataset was for the most part collected from unrelated experiments. The combined dataset used in the development of the prediction model was around 32 000 rows × 140 columns. Each row represented a sample tasted by 10 consumers; each column specified a variable relating to the sample tested. The regression procedures employed identified several available and practically useable input variables for use in the model and discarded those which did not assist.

Early MSA research trials sought to establish consumer benchmarks to use in evaluating the impact of known or proposed critical control points from breeding to consumption and to determine the predictive value of both traditional and potential grading inputs (Polkinghorne *et al*. 2008*b*). In this paper, the principal indicators of palatability used in the MSA grading model are reported and their inclusion in the model is discussed.

An important distinction in relation to previous and alternative commercial beef grading systems is that the MSA grading scheme assigns a grade to a specific piece of beef cooked by a designated method; it does not assign a single grade to an entire carcass. This is in contrast to the United States Department of Agriculture (USDA 1989), Japanese (JMGA 1988), Canadian (Canadian Beef Grading Agency 1997) and Korean (Kim and Lee 2003) quality grading systems, which assign a quality grade to the carcass after considering a limited number of traits available at the time of grading the chilled carcass. Polkinghorne (2005) used the MSA prediction model to conclude that a single carcass grade

was not capable of accurately describing palatability when carcasses were produced from a variety of production systems.

Significance testing was not the criteria upon which the model was built. Rather, significance testing was a separate exercise based on the individual controlled experiments, which are reported elsewhere. The results of the model were of course in accordance with these results, in the sense that the parameter values used generally conformed to confidence intervals based on the experimental results. The full dataset, is for the most part collected from unrelated standard experiments, and so must be regarded as observational data. Thus, the model building process was a large meta-analysis of data from a series of small experiments. As these different experiments were conducted using different abattoirs this obviously increased the variance but it also provided an indication of transportability, which was important for a model that was intended to be used throughout the Australian industry. The objective was to derive a plausible and smooth model that described the data well. This is not to say that the data could not be described at least as well by some other model. But what is true, is that the MSA grading model does provide a reasonable description of the data, which does appear to be useful in prediction.

Available data – muscle samples and cooking methods

Numbers of muscle samples for each cooking method in the MSA model development database are shown in Table 1. The database, collection history and methodology were described in detail by Polkinghorne *et al*. (2008*a*). Not all cooking methods were tested on all muscles. Some high connective tissue cuts were not grilled; and most low connective tissue cuts were not slow cooked. The protocols used in testing for each cooking method are described in Gee *et al*. (2005) and summarised in Watson *et al*. (2008*a*). Most of the data was collected on the striploin (*M. longissimus dorsi et lumborum*) comprising 68% of all grilled muscles and 34% of all muscles tested by all cooking methods. This reflects the early stages of the MSA consumer testing program where the focus was on striploin testing. Also, to a lesser degree, the commercial importance of the striploin has meant that it was often the common link between experiments which involved a combination of cuts × cooking techniques. Of the five different cooking methods, grilling represented 50% of the samples tested. Stir-frying and roasting were the next most common cooking methods tested, with the thin slice method having the least number of samples tested.

Developing a model

The aim of the modelling procedure was to try to find an overall smooth pattern which aligned with known meat science. To use a diagrammatic analogy: the final model was like the smooth, dotted curve shown in Fig. 1; the jagged curve relates to the effects found in the various analyses.

The model production process was a robust procedure in the sense that a common model was established. Effects were retained if they commonly appeared in a range of analyses at a similar level. Analyses that did not fit the common pattern were investigated to ascertain why, and included or discounted. In many cases, the questions raised by the model-fitting led to further experimentation and investigation.

Table 1. Number of consumer tested samples for each muscle and cooking method used in developing the Meat Standards Australia prediction model (version SP2004)

Muscle type	Cooking method					Total
	Grill	Roast	Stirfry	Thin slice	Slow cook	
M. spinalis dorsi	116	9	289	36	0	450
M. iliacus	12	0	34	0	0	46
M. psoas major	797	334	120	4	0	1255
M. longissimus dorsi thoracis	557	462	66	43	0	1128
M. longissimus dorsi et lumborum (ant.) post)	10 839	626	273	12	18	11 768
M. infraspinatus	361	126	382	24	1	894
M. triceps brachii caput laterale	0	0	110	2	0	112
M. triceps brachii caput longum	793	320	95	37	193	1438
M. supraspinatus	0	29	65	30	22	146
M. biceps femoris	127	0	286	39	0	452
M. gluteus medius	60	60	0	0	0	120
M. gluteus profundis	0	0	254	4	0	258
M. tensor fasciae latae	0	11	171	13	4	199
M. gluteus medius (Eye)	805	405	239	34	28	1511
M. gluteus medius (D portion)	696	232	126	36	0	1090
M. rectus femoris	358	330	377	39	231	1335
M. vastus intermedius	0	0	129	17	19	165
M. vastus lateralis	64	75	288	45	53	525
M. vastus medialis	0	0	183	9	19	211
M. biceps femoris (Syn. *gluteobiceps*)	0	717	285	70	367	1439
M. gastrocnemius	0	0	175	22	87	284
M. semitendinosus	54	582	259	33	255	1183
M. adductor femoris	82	0	277	29	78	466
M. gracilis	2	0	257	27	59	345
M. semimembranosus	193	951	522	35	416	2117
M. rhomboideus	0	0	191	22	5	218
M. semispinalis capitis	26	61	227	32	7	353
M. serratus ventralis cervicis	59	173	153	27	55	467
M. spinalis dorsi	0	0	190	5	15	210
M. splenius	0	0	71	5	5	81
M. obliquus externus abdominis	0	0	200	0	46	246
M. obliquus internus abdominis	0	0	39	8	12	59
M. rectus abdominis	0	0	150	12	18	180
M. latissimus dorsi	0	0	227	0	0	227
M. pectoralis profundus	0	0	211	72	278	561
M. pectoralis superficialis	0	0	157	35	45	237
All muscles (forequarter shin)	0	0	0	0	124	124
All muscles (hindquarter shin)	0	0	0	0	125	125
Mm. intercostales externus and *internus*	0	0	208	4	0	212
Total no. of muscle samples	16 001	5503	7286	862	2585	32 237

Datasets analysed included the following:

(1) Individual experiments;
(2) Meta-analysis by combining results from individual experiments;
(3) Animal effect analyses;
(4) Cut-by-cut analyses; and analyses by cut groups (e.g. very high connective tissue cuts, high connective tissue cuts, low connective tissue cuts, and very low connective tissue cuts); and
(5) Subsets of all of the above – both random subsets and non-random subsets (for example, categorised by variable).

Methods used included (Polkinghorne *et al*. 2008*b*):

(1) Descriptive statistics – dot plots, cross-tabulations, scatter plots, etc.;
(2) Regression models and general linear model fits; and
(3) Validation – both statistical validation and meat scientific validation.

A commonly used tool in the model development was regression modelling, i.e. a model of the form:

Meat quality or 'MQ' score = (formula involving animal variables and treatment variables) + (consumer variation)

The consumer variation can be averaged out (by averaging over lots of consumers). The initial intention was to try to divide the MQ formula into separate components corresponding to the meat production process:

$$\underset{(\text{farm})}{\text{Animal}}\ \underset{(\text{abattoir})}{+\ \text{Processing}}\ \underset{(\text{consumer})}{+\ \text{Preparation}}$$

characteristics + variables + method
(farm) (abattoir) (consumer)

Variables were needed to describe each of these components; and then a way found to put them together. However, these independent parts were found to be anything but independent. Furthermore, even within the parts, many of the quantities involved were interdependent.

This finding has influenced the commercial implementation of the MSA program, requiring education and reinforcement that all industry sectors interact in determining the eating quality of beef delivered to the ultimate consumer. It has also been important in dispelling various 'silver bullet' solutions to beef quality variously advanced and including particular breeds, attributes
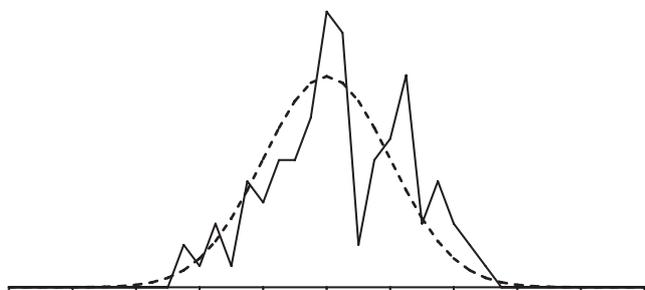
such as marbling or processes such as aging (Polkinghorne *et al*. 2008*a*).

Selection of variables

The 140 database columns record a large number of variables in addition to identification relating to animal groups, ear tag and slaughter body numbers and unique sample codes. These variables include breed, age and feed regime, processing inputs (including carcass suspension and electrical stimulation), conventional chiller assessment measures such as marbling, fat depth and pH, individual muscle detail (including position of the sample within muscle) and days aged. The consumer data comprised cooking method and tenderness (tn), juiciness (ju), like flavour (fl) and overall liking (ov) scores recorded in consumer testing. The MQ score was formed by weighting the four sensory scores, $MQ = 0.4 \times tn + 0.1 \times ju + 0.2 \times fl + 0.3 \times ov$, obtained using optimum linear discriminating function (Watson *et al*. 2008*a*). The model development process, as discussed in the following sections, sought to identify and quantify variables, which assisted in describing the consumer tested outcomes.

Variables which consistently added to predictive ability were retained in the model whereas those which demonstrated no relationship, or were inconsistent, were discarded. An example of this were texture and firmness scores, which despite repeated analysis, failed to predict eating quality either as single variables, or in combination with others. In some instances, a selection needed to be made between alternative variables which essentially described different aspects of the same characteristic in the carcass. Examples of this were the different measures for fatness which generally were positively correlated within any of the datasets, e.g. P8 and rib fat depths and USDA and AUS-MEAT marbling scores were all positively correlated. In cases like this, trade-offs between reliability, availability, sensibility and transparency had to be made. After considerable discussion, statistical input and debate, the variables presented in Table 2 were considered as candidates for modelling.

The primary variables were included as they all had significant predictive value for MQ; whereas the secondary variables were

Fig. 1. A diagrammatic representation of the developed model (dashed curve) in relation to individual experimental data (solid line).

Table 2. Primary and secondary variables considered for modelling meat quality score

Variables class	Primary	Secondary
Animal and carcass characteristics	Percentage *Bos indicus* content, carcass weight, sex, ossification and marbling scores and ultimate pH	Meat colour score, rib fat depth, muscle texture and firmness scores, weight adjusted for maturity score
Treatment effects	Muscle portion, cooking method, carcass suspension, days aged	Feed (feedlot or pasture-finished and days on feed for feedlot animals) and whether the calves were unweaned (i.e. milk-fed veal)

subject to further investigation. Ultimately, the feed variables (such as feed, days on feed) were deleted as the effect of feedlotting on palatability was found to be adequately covered by carcass weight, ossification and marbling score effects. The further addition of feed or days on feed was found to have little consistent additional effect.

The weight adjusted for maturity score variable was calculated as a ratio of estimated liveweight (i.e. carcass weight/0.56) relative to age (estimated by ossification score) in an attempt to describe the lifetime growth rate of the animal. However, as reported by Thompson *et al.* (1999), the use of this ratio to predict the MQ score had some statistical complications. Examination of the interaction between ossification score and carcass weight was found to provide a better model to describe the relationships with the MQ score. A gradual change in the slope of carcass weight to the MQ score, within ossification score provided a model which allowed a meat science interpretation to be proposed.

Although the effect of ultimate pH (pH_u) below 5.7 on the MQ score was found to be minimal, pH_u was retained as a predictor variable as it was the basis of a cut-off criterion. In the presence of pH_u both AUS-MEAT and USDA meat colour scores added little to the prediction of the MQ score. Rib fat was also used as a censoring variable, but since it was found to have some impact for several muscle portions it was retained in the model in preference to other fat measures.

To indicate some of the effects of the model, an ANOVA is presented in Table 3. The model fitted for the common ANOVA is as follows:

$$MQ = cook + hang|dagd + epbi + osswt|sex + umb$$
$$+ \ rbf + pH_u + hgp$$

where cook is cooking method, hang is tenderstretch or Achilles hung, dagd is days aged post-mortem, osswt is ossification carcass weight score, sex is heifer or steer, umb is USDA marbling score, rbf is rib fat depth and hgp is hormonal growth implant indicator. In this model, 'hang|dagd' represents

hang effect + aging effect + hang × aging interaction, which indicates the difference in aging effect for different methods of hanging. Similarly, for 'osswt|sex' represents osswt + sex + osswt × sex. The latter term indicates the difference in the carcass weight/ossification effects between males and females. The estimated percentage *Bos indicus* content was represented by epbi and treatment using hormone growth promotant implants by HGP.

This ANOVA is too simple in the sense that it does not allow for a range of effects. Essentially, everything interacts with cook × muscle; and there are interactions between several of the other variables, which are also ignored in this analysis. However, it provides an indication that each of these variables had some effect.

This model also ignored animal effects; as do most of the model analyses: the model was intended to replace the animal effects by observable variables such as epbi, ossification score, carcass weight, sex, rib fat depth, marbling and so on, which can be used for prediction. In most instances, when the model is fitted, it is fitted for each cut separately, so that there is generally only one cut from each animal included in such an analysis.

The ANOVA obtained for the striploin (*M. longissimus dorsi et lumborum*) data is given in Table 4.

Again, each of the parameters is shown to have some effect indicating the significant variation found within this muscle. The striploin has been used extensively in beef research and there are many reports of variation and sensitivity to animal and processing effects (e.g. Dransfield 1977; Shorthose and Harris 1990; Shackelford *et al.* 1995). The appropriateness of using the striploin as an indicator cut for other carcass muscles has also been discussed and challenged by Shorthose (1996) who suggested the eye round (*M. semitendinosus*) as an alternative less subject to processing variation. Koohmaraie *et al.* (1998) and Rhee *et al.* (2004) also found the correlation between meat quality

Table 3.   ANOVA table for meat quality score (all muscles)
The term for hormone growth promotant (HGP) was on 2 d.f., corresponding to three levels: no implant, implant and unknown

Independent variables	d.f.	*F*-ratio	*P*-value
Cook × muscle	180	76.64	0.000
Hang	1	138.55	0.000
Days aged	1	38.67	0.000
Hang × days aged	1	27.27	0.000
Estimated percentage *Bos indicus* content	1	1016.34	0.000
Ossification/carcass weight	1	521.10	0.000
Sex	1	22.46	0.000
Ossification/carcass weight × sex	1	35.07	0.000
Rib fat depth	1	3.32	0.069
Marbling	1	561.77	0.000
Ultimate pH	1	10.77	0.001
HGP implant	2	70.75	0.000
Error	32 962		
Total	33 154		

Table 4.   ANOVA for meat quality score of the striploin (*M. longissimus dorsi et lumborum*)
As for the analysis in Table 3 hormone growth promotant (HGP) implant has three levels: no implant, implant and unknown. The position within muscle also has three levels: anterior, centre, posterior

Independent variables	d.f.	*F*-ratio	*P*-value
Cook	5	12.35	0.000
Position within muscle	2	97.16	0.000
Hang	1	55.32	0.000
Days aged	1	57.21	0.000
Hang × days aged	1	4.49	0.034
Estimated percentage *Bos indicus* content	1	1026.76	0.000
Ossification/carcass weight	1	181.8	0.000
Sex	1	45.56	0.000
Ossification/carcass weight × sex	1	38.99	0.000
Rib fat depth	1	6.42	0.011
Marbling	1	660.81	0.000
Ultimate pH	1	25.56	0.000
HGP implant	2	149.24	0.000
Error	11 890	–	–
Total	11 909		–

measurements (both sensory and objective) for the striploin and other muscles in the carcass to be weak.

ANOVA for each individual muscle (data not shown) produced varying rankings of importance, as measured by *F*-ratios, for the principal variables further indicating that the palatability of each muscle would be best estimated individually rather than from a common indicator muscle. This presented a challenge to the traditional approach of grading a carcass as a unit *v.* its individual muscles. Polkinghorne (2005) used the MSA model prediction of palatability for a range of muscles from carcasses from different production systems subjected to a range of processing conditions. Polkinghorne concluded that it is not possible to provide meaningful estimates of palatability for a range of cuts from a simple striploin relationship.

Input variable analysis

Epbi

The epbi is among the most consistent animal-based indicators identified in the modelling process, appearing as significant in most analyses. Table 5 gives the raw mean MQ score by epbi split into five categories; for striploin data, and for all data. This table is simply an indicator of the negative relationship that exists between consumer scores and the epbi. Clearly, many other factors are involved. Here it is assumed that with the large numbers of samples involved the other factors are averaging out. Negative eating quality effects with *Bos indicus* cattle have also been widely reported by others including Koch *et al.* (1976), Sherbeck *et al.* (1996) and Koohmaraie *et al.* (1998).

Epbi relationship to carcass hump height

The epbi of research cattle groups was known, but for commercial cattle, the epbi was based on the producer's report. As grading results and producer payment were related to the epbi declared it was also felt there could be some uncertainty in its veracity. Furthermore, in some cases, an average value for a group of cattle was given, and there was clear variation within the group. A check was devised to provide an objective assessment by considering the carcass hump height in relation to carcass weight. The relationship between carcass hump height and epbi is indicated in Fig. 2.

A discriminant analysis for the epbi (categorised into five groups) on hump height and carcass weight yielded the best discriminant function: hump (mm) – $0.1 \times$ carcass weight (kg).

With this discriminator the results presented in Table 6 were obtained.

The percentage *Bos indicus* group cut-offs used in the analysis above were 37, 52, 60 and 84. Using hwd <40 as the epbi = 0 indicator (to be conservative and reduce the risk of misclassifying too many *Bos taurus* animals), the results presented in Table 7 were produced from a comparison of the proposed
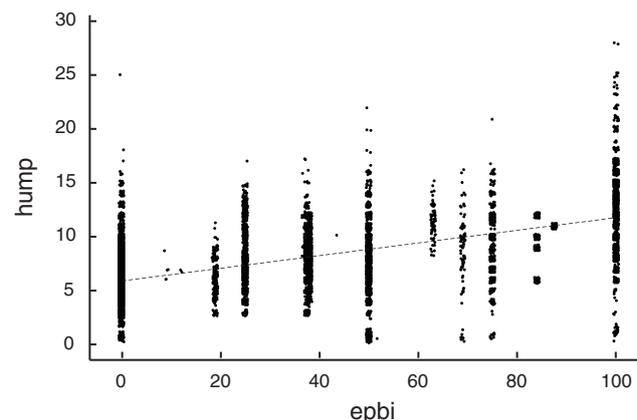

Fig. 2. Scatter plot of carcass hump height (hump, cm) *v.* percentage *Bos indicus* content (epbi). The dashed line represents the line of best fit.

Table 6. Classification accuracy for percentage *Bos indicus* content (epbi) as estimated from hump height and carcass weight
Overall: $n = 23\,493$; n correct = 17\,671; proportion correct = 0.752

Predicted group	True group				
	Z	L	M	H	F
Z (zero epbi)	14 474	628	359	118	13
L (low epbi)	742	1067	636	55	24
M (moderate epbi)	191	264	329	234	26
H (high epbi)	441	604	477	361	477
F (full epbi)	86	103	185	159	1440
Total *n*	15 934	2666	1986	927	1980
n correct	14 474	1067	329	361	1440
Proportion correct	0.908	0.400	0.166	0.389	0.727

Table 7. Classification accuracy for percentage *Bos indicus* (BI) estimated from hump height and carcass weight by breed type

	Classification			Percentage BI
	Non BI	BI	All	
British	12 387	342	12 729	3
British × Euro	2239	2	2241	0
Euro	91	1	92	1
British × *Bos indicus*	1405	4029	5434	74
Euro × *Bos indicus*	2	16	18	89
Belmont Red	32	702	734	96
Belmont Red × *Bos indicus*	11	208	219	95
100% *Bos indicus*	18	1884	1902	99
All	16 185	7184	23 369	–

Table 5. Average meat quality (MQ) score by percentage *Bos indicus* category

Bos indicus percentage	0	0–50	50	50–100	100	All
Striploin samples MQ score	61.4	53.3	52.0	49.9	45.7	55.6
n	5206	2637	2133	627	1307	11 910
All samples MQ score	59.1	54.6	53.2	52.1	50.1	56.4
n	19 042	6031	3084	1396	3602	33 155

hump-based *Bos indicus* estimate in relation to breed type as recorded in the dataset.

Apart from the Belmont Reds, very few zero (reported) epbi were misclassified and most of any part *Bos indicus* animals were correctly classified. The hump-estimated percentage *Bos indicus* adjustment upgraded epbi to a higher level if hump was significantly greater than the equivalent for the reported epbi.

As a result of this analysis, a hump and carcass weight-based adjustment was incorporated into the model. Carcasses which fell outside the established relationship were reassigned to a more appropriate percentage *Bos indicus* level. In addition to providing a cross check against the declared epbi of carcasses presented for grading, this also provided a means to grade carcasses on-line without a producer declaration. While there is some risk that the calculated epbi may be greater than the true percentage, this procedure can be commercially useful where slaughter groups vary widely in epbi. The alternative was to draft animals into epbi categories before slaughter, which would have generally subjected the animals to stress before trucking or lairage and, therefore, potentially reduced glycogen content.

The above adjustment based on hump height and carcass weight, rated Belmont Reds, a tropically adapted *Bos taurus* and Africander-derived breed, as if they were part *Bos indicus*. An analysis of the Belmont Red animals indicated that their meat palatability rated close to that of animals with epbi = 50. In other words, they were similar to 50% epbi animals, in palatability as well as in their hump height. This was generally consistent with relationships of tenderness in Brahman-based cattle to phenotype reported by Sherbeck *et al.* (1996). Thus, a further adjustment to the epbi conversion for Belmont Reds was not required.

Ossification score and carcass weight

Ossification is an assessment of the calcification of the cartilage in the sacral and dorsal vertebrae. Other workers have examined the relationship between tenderness and skeletal maturity scores with some studies showing a negative relationship between maturity and tenderness scores (Smith *et al.* 1982, 1988; Hilton *et al.* 1998; Park *et al.* 2008), whereas others have failed to find any relationship (Romans *et al.* 1965; Carroll *et al.* 1976; Field *et al.* 1997). The USDA grading system utilises ossification score in conjunction with meat colour to create a maturity score, which is in turn associated with marbling to assign a quality grade. It was identified as a potential modelling input due to its USDA system inclusion, so carcass ossification was recorded from the first MSA trials. The score assigned by MSA graders is a numeric presentation of the USDA scale in increments of 10 where A0 = 100, B0 = 200, C0 = 300 out to E90 = 590 (Romans *et al.* 1965).

Figure 3 presents the distribution of ossification values in the model development dataset. The majority of the data are in the 100–200 ossification range. The early MSA pathways and initial models limited ossification to a maximum of 200, reflecting opinion from the USDA B0 cut-off, and there was a lack of data beyond this threshold: the early testing was primarily of young cattle destined for the Australian domestic market. Although this constraint was subsequently removed, the skewed distribution is a legacy of the earlier grading constraints.
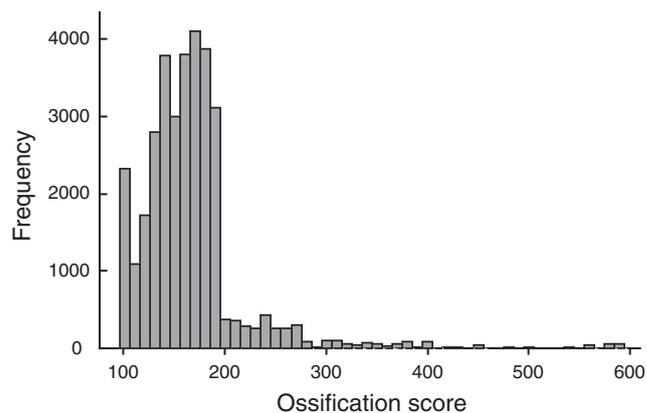

Fig. 3. Frequency distribution of ossification score for all cuts in the Meat Standards Australia dataset.

A majority of the 100 ossification scores relate to cuts derived from milk-fed vealers (i.e. they were unweaned calves still suckling the cow when transported for slaughter and coded MFV in the database). While the data encompassed a considerable range in ossification score, this number varied by muscle. There were far fewer samples with high ossification scores and for some muscles there were very few ossification scores at the high end of the range. MSA is testing further samples with high ossification scores in order to better support the model for ossification scores greater than 300.

The dataset shows a general downward effect on the MQ score with increasing ossification score, and this is especially evident with lower ossification scores. As expected, younger animals tend to be associated with higher MQ scores.

Modelling carcass weight is not so straightforward. Ignoring other factors, the regression of MQ score on weight is curvilinear. The MQ score is high when it is small and when the weight is large, with a minimum at around 300 kg.

The following ossification/carcass weight model allowed a differential effect of carcass weight within each ossification score category subject to the criteria that: 'MQ increased with carcass weight within each ossification category' and 'for any given carcass weight, the lower ossification score (younger animal) was more palatable'. The relationships between carcass weight and palatability within ossification score category were assumed to be linear and are shown in Fig. 4. At low ossification scores the slope between carcass weight and palatability was close to zero, but it increased gradually as the ossification scores increased.

This model was fitted by using a weighted average of the separate fits for each of the muscles. For many of the muscles in which there were reasonable numbers of observations and a wide range in ossification score (e.g. the striploin) the result needed some adjustment to ensure that the slopes of the lines were not negative and did not intersect within the range of likely carcass weights. Although all slopes were positive, the fitted line for ossification score 100 had a slightly negative gradient. This was retained in the model as it was felt that, for this ossification category, smaller weight may reflect lesser age and, therefore, for the MFV, may be associated with increased palatability.

For muscles in which the datasets were more sparse, the intercept and slope values for carcass weight within
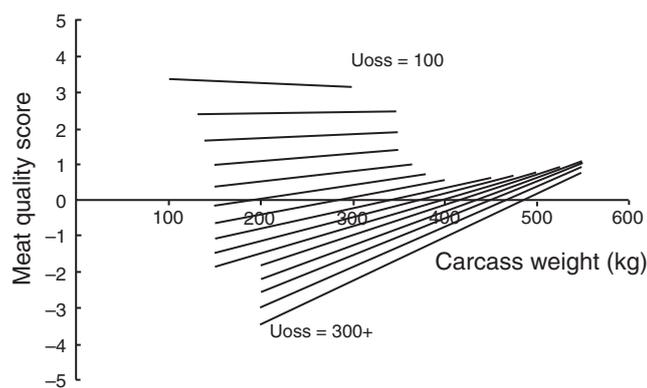
Fig. 4. Plot of effect of carcass weight on meat quality for different ossification categories: the top line corresponds to oss = 100, the bottom line to oss = 300+.

ossification score were erratic. However, even with these reduced datasets, the same general trends were evident. Instead of fitting ossification score as a fixed effect in the interaction with carcass weight, it was fitted as a covariate to obtain a smoothed function for the slope and intercept for different muscles. The variability of the coefficients for the muscle × ossification/carcass weight interaction was modelled as follows:

$$MQ = \cdots + k\,(\text{muscle}) \times g\,(\text{oss, cwt}) + \cdots$$

where g(oss,cwt) denotes the ossification/carcass weight score defined by Fig. 4, and the multiplier k (muscle) is allowed to vary from muscle to muscle. If the ossification/carcass weight is important for a muscle then k (muscle) is large, but if it is not important then k (muscle) is small: the average of the k-values was around 1. Again, the results appear to have a plausible meat science interpretation in that high k-values tend to occur for high activity and connective tissue muscles. For the ossification categories for which there were MFV animals (i.e. ossification score ≤140) there was an additional positive MFV effect of 4 MQ points on average.

Carcass suspension and aging

Carcass suspension and aging effects are discussed together due to their interaction, as highlighted by early analysis of the MSA database. This substantially increased the sample collection program as data for each carcass suspension method needed to be over a range of days aged for each of the muscles examined.

Carcass suspension

The model development database contained palatability results from several carcass suspension treatments. These included carcasses suspended by the Achilles tendon (AT), suspended from the obturator foramen (TX), suspended from the sacral ligament (TL) and carcasses which were prepared using the tendercut procedure (TC) described by Wang *et al.* (1994).

The majority (68%) of the dataset were AT hung, with a substantial minority (27%) TL hung, with relatively few TX and TC samples (3 and 2%, respectively), and not enough to allow incorporation of these methods in the model (Table 8). Work continues to extend the modelling range for the TX method.

Table 8. Number of samples available for aging estimates for all muscles by post-mortem aging period

Hang classifications are: suspended from the Achilles tendon (AT), the obturator foramen (TX), sacral ligament (TL) and carcasses which were prepared using the tendercut procedure (TC)

Hang	Days aged						Total
	5–6	7–10	11–14	15–21	22–28	29+	
AT	1727	2731	9512	6762	1534	379	22 645
TL	530	1206	3524	2759	832	68	8918
TX	0	752	138	11	3	2	906
TC	0	0	407	275	3	0	686
Total	2257	4689	13 581	9807	2372	449	33 155

Not surprisingly, it is found that the effect of carcass suspension method on palatability (MQ score) interacted with muscle. The effect of the TL carcass suspension method (at 5 days aging) is indicated in Table 9 by the TL intercept column, with the AT intercept column set to zero. The MQ score of *M. longissimus dorsi lumborum*, *M. biceps femoris* and *M. semimembranosus* were improved by the TL hanging method, while the *M. psoas major* and *M. spinalis dorsi* were negatively affected. In general, forequarter cuts tend not to be affected by the TL method compared with the AT method.

Hwang *et al.* (2002) compared the TX and TL carcass suspension methods against AT using consumer taste panels to evaluate stir-fry samples from 18 hindquarter muscle/muscle portions at 10 days aging. They showed that both tenderstretch

Table 9. Estimated aging effects (intercept and slope) for Achilles hung (AT) and tenderstretch (TL) carcass suspension methods

The regression lines are used to predict palatability from 5 days to 21 days post-mortem

Muscle type	AT		TL	
	Intercept	Slope	Intercept	Slope
M. spinalis dorsi	0.00	0.05	–1.52	0.08
M. psoas major	0.00	0.03	–3.56	0.11
M. longissimus dorsi lumborum (posterior)	0.00	0.30	7.62	0.20
M. longissimus dorsi lumborum (anterior)	0.00	0.32	8.06	0.22
M. triceps brachii caput longum	0.00	0.05	0.00	0.05
M. supraspinatus	0.00	0.05	0.00	0.05
M. biceps femoris	0.00	0.18	3.34	0.14
M. tensor fasciae latae	0.00	0.17	5.51	0.14
M. rectus femoris	0.00	0.06	5.90	0.00
M. vastus lateralis	0.00	0.06	5.90	0.00
M. biceps femoris (syn. *gluteobiceps*)	0.00	0.18	3.34	0.14
M. gastrocnemius	0.00	0.07	–3.71	0.24
M. semitendinosus	0.00	0.12	0.00	0.12
M. semimembranosus	0.00	0.15	8.92	0.00
M. semispinalis capitis	0.00	0.09	0.00	0.09
M. serratus ventralis cervicis	0.00	0.09	0.00	0.09
M. rectus abdominis	0.00	0.08	0.60	0.06
M. pectoralis profundus	0.00	0.00	0.00	0.00
M. pectoralis superficialis	0.00	0.00	0.00	0.00

(TL and TX) methods resulted in an improvement in hindquarter and loin muscles relative to AT hanging.

The limited TC data was obtained from three separate experiments utilising cattle from three abattoirs. The cattle included unweaned 10-month-old British breed steers, ~30-month-old British breed steers and ~30-month-old British breed, *Bos indicus* and Piedmontese cross steers. Tenderstretch and Achilles comparisons were made within each group with from 9 to 13 muscles tested under grill, roast and slow cooking methods.

Post-mortem muscle aging period

At an industry level, aging was generally believed to improve eating quality but knowledge of the degree and rate of improvement or of differences between muscles was limited at best. Differences in the aging potential of various muscles have been reported by Bouton and Harris (1972), Dransfield (1994) and others. Bouton and Harris (1972) also reported differences in aging rate between muscles and between the same muscles under different suspension treatments; and further, they reported a decline in aging rate over time. The potential for differential aging effects between muscles and suspension methods dictated a need for considerable data to develop plausible estimates for modelling.

The number of samples available in the model development dataset by aging period, muscle and hanging method is presented in Table 8. All aging was conducted at 1–4°C in a vacuum pack. Initially, the entire primal was aged, whereas later samples were fabricated from the primal and re-vacuued before freezing at the designated aging days.

While this is a large dataset, the numbers in individual (muscle × days aged × hanging method) cells were often small. All samples were aged for a minimum of 5 days, which was set as an MSA requirement. The majority of the aging data were in the range 7–21 days, with most in the 11–14 days category. There was very little information beyond 28 days aging, resulting in the model predictions being restricted to a 28-day maximum.

In general, aging tends to even things out: when the meat palatability improves with age, it improves at a faster rate for the hanging method that is behind at 5 days. However, within the range of the available data, it never catches up.

A common result for the effect of days aged on MQ is shown in Fig. 5, for the posterior striploin. Striploins from TL carcasses were substantially more palatable at 5 days than those from AT carcasses. However, the difference reduced with aging time at least up to ~25 days.

The three experiments conducted to compare hanging methods TL, AT and TX included 13 muscles derived from a range of cattle types sourced from different abattoirs. While this generated a useful dataset, all carcass muscles were not tested and aging comparisons were limited to a range of 13–20 days. Therefore, it was not possible to independently derive the coefficients for all the muscle × aging combinations. However, the available data show that the TC option is between the AT and the TL method in terms of the effects on palatability of the hindquarter muscles. Figure 5 presents the hang × age effect estimated for the AT and TL striploin data.
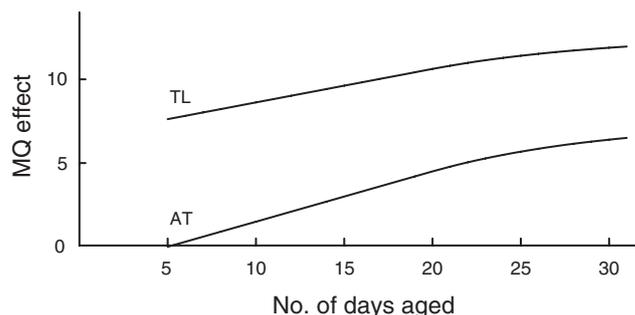

Fig. 5. Hang effects of Achilles hung and tenderstretch for posterior striploin.

Aging effects were derived for all muscle × hang combinations in the dataset, although the available data restricted the reliability for some cells. The fitted effect for each was obtained as a straight line. The value used in the model, for each hang method × muscle, was assumed to be linear up to days aged = 20. The literature would suggest a declining rate of increase in eating quality with aging, so an exponential curve was used for aging times greater than day 20, reflecting caution with prediction for higher aging times by ensuring that no matter what value for days aged was used the benefit could never exceed the fitted linear value at 30 days.

Marbling and fat: umb, amb, rbf, P8

The umb and amb (AUS-MEAT marbling score) are assessments of marbling. In addition rib fat depth (mm) at the quartering point and P8 fat depth (mm) are measurements of carcass fatness, primarily of the subcutaneous fat depot. The range of values, the slightly higher correlation with MQ, and ultimately the predictive value in concert with other model variables led to the choice of umb and rbf as grading model inputs.

The variable umb came to take on the role of a surrogate for at least part of the effects of several other variables and so, with this in mind, it is not that surprising that it appears consistently in the prediction model for all cuts, although with varying relative importance to other input variables.

Rib fat has less effect on the model; but, like pH_u, it has a censoring role. Animals with rbf less than 3 mm are rejected. The reasoning for including a minimum rib fat requirement relates to even chilling within the muscle (Polkinghorne *et al.* 2008*a*).

HGP implants

Initially, HGP use was not recorded in the database. After several years of data collection it was thought appropriate to add it to the list of variables and, where possible, it was added retrospectively.

The dataset relating to HGP was further boosted by data from several trials conducted to investigate HGP effects on eating quality. The use of HGPs has been studied extensively in the database (Thompson *et al.* 2008*a*; Watson 2008; Watson *et al.* 2008*b*), and in a range of associated experiments reported elsewhere. The result was an HGP penalty of the order of 3–6 MQ points on meat palatability, depending on the muscle which is further adjusted with aging.

The process of combining the results from several independent HGP experiments and deriving a final application for use in the grading model is further discussed by R. Polkinghorne, J. Thompson and R. Watson, unpubl. data.

Other variables considered for inclusion in the model

In addition to the primary variables selected for use within the prediction model, several others were evaluated at various points. These are briefly discussed below.

Feedlot variables: finishing system and days on feed

Finishing system indicated whether the animals had been finished on a high grain ration or on pasture. Within feedlot finish, the days on feed were also recorded. Both these variables were eventually omitted from the model as variance explained by finishing system was similar to that explained by the increased carcass weight relative to ossification and higher marbling scores associated with feedlot finishing systems. It seemed preferable to model using an outcome variable (carcass weight, ossification and marbling) rather than a treatment variable (feed) and so both finishing system and days on feed were excluded from the model.

pH_u, AUS-MEAT meat colour and USDA colour of lean scores

Several attempts were made to incorporate pH_u or a colour variable into the model. However, any such effects seemed always to be driven by the extreme values, which would be culled in practice anyway. It was decided to effectively leave these variables out of the model, except in their role as censoring variables and as a minor effect in the case of pH_u. If the pH or the colour was too extreme, then the meat was discarded. Any carcass with a pH_u beyond 5.7 or an AUS-MEAT meat colour above 3 is excluded from grading.

Establishment of a base cut × cook relationship

As the core retained variables were identified from muscle by muscle and overall analysis a base muscle × cook table of MQ scores was constructed (Table 10). Essentially, this comprised the data means.

In fitting the model based on the variables described in the previous section, it was decided to treat as baseline the 'standard' case given by:

Percentage *Bos indicus* = 0, carcass weight = 280,

ossification score = 150, sex = male, umb = 260,

rbf = 8, hang = AT, days aged = 5 and pH_u = 5.7

Table 10 specifies the model 'intercept', i.e. the predicted value with all the variables set at baseline level. Thus, these can be interpreted as the predicted values for the standard animal/treatment.

In applying the model, the predicted values for each muscle × cook combination are calculated as deviations from these standard scores, depending on the deviations of each of

Table 10. Estimated mean meat quality (MQ) scores for specified cut × cook combinations for the 'standard' animal (cooking methods include grill, roast, stir-fry, thin slice and slow cook)

Scores are on a 1–100 scale

Muscle type	Cooking method				
	Grill	Roast	Stir-fry	Thin slice	Slow cook
M. spinalis dorsi	77.4	67.1	76.8	72.8	–
M. psoas major	77.3	76.4	79.3	74.1	–
M. longissimus dorsi thoracis	62.2	62.0	61.8	64.2	–
M. longissimus dorsi et lumborum (posterior)	55.9	56.6	58.0	58.5	–
M. longissimus dorsi et lumborum (anterior)	53.8	54.5	57.1	57.6	–
M. triceps brachii caput longum	54.3	58.0	58.9	59.8	55.3
M. supraspinatus	–	48.1	50.4	52.6	53.5
M. biceps femoris	58.0	–	66.0	66.5	–
M. tensor fasciae latae	–	51.6	56.5	54.7	49.9
M. rectus femoris	47.0	60.1	55.0	58.6	42.8
M. vastus lateralis	36.8	47.9	44.6	51.5	47.9
M. biceps femoris (syn. *gluteobiceps*)	–	39.6	41.7	54.9	52.5
M. gastrocnemius	–	–	58.0	65.1	54.2
M. semitendinosus	43.6	47.4	45.2	47.7	44.5
M. semimembranosus	34.8	43.4	43.4	56.3	47.4
M. semispinalis capitis	60.8	53.8	58.8	64.6	64.3
M. serratus ventralis cervicis	53.7	55.0	55.8	59.2	62.0
M. rectus abdominis	–	–	62.7	59.0	56.7
M. pectoralis profundus	–	–	41.1	54.9	52.4
M. pectoralis superficialis	–	–	38.4	46.1	56.7

the predictor variables from their standard values, specified in Table 10.

Accuracy of the prediction model

An indication of the accuracy of the prediction model was given by applying it to the MSA dataset. For each meat sample, the observed MQ value was compared with the predicted MQ. This was applied to every sample for which a prediction was possible, even those with incomplete or doubtful data, or input values that would have excluded them from the grading process. The results are indicated in Table 11. In most cases, the average difference was less than 1 and in only 5/72 (7%) of the cells does a simple *t*-test indicate a non-zero mean at the 5% level.

Analysis indicates that the standard error for most of the predicted MQ scores will be less than 1. But these standard errors are based on the assumption that the model is correct, and so must be treated with some caution. It does suggest that, if the model is close to the truth, then the predicted mean scores will mostly be within 2 units of the population mean MQ score. However, predicting a mean MQ is different from predicting where the actual consumers'

Table 11. Numbers and average differences between predicted and observed meat quality (MQ) scores for the available data

Values in parentheses are the number of differences used to calculate the mean. *, significantly different from zero at the $P = 0.05$ level

Difference (predicted MQ – observed MQ)	Cooking method				
	Grill	Roast	Slow cook	Stir-fry	Thin slice
M. spinalis dorsi	–0.05 (137)	–	–	0.62 (289)	0.97 (36)
M. psoas major	0.00 (709)	0.54 (334)	–	–0.58 (120)	–
M. longissimus dorsi lumborum (anterior)	–0.34*(7631)	0.42 (272)	–	–0.65 (120)	–
M. longissimus dorsi lumborum (posterior)	0.29 (1366)	0.81 (329)	–	1.63 (114)	–
M. triceps brachii caput longum	–0.90*(657)	–0.77 (323)	–1.50 (193)	–0.47 (95)	–0.03 (37)
M. supraspinatus	–	2.03 (29)	–0.91 (22)	1.80 (65)	–3.40 (30)
M. biceps femoris (proximal)	–1.30 (127)	–	–	–0.67 (286)	–0.72 (39)
M. tensor fasciae latae	–	0.40 (42)	–	0.46 (190)	1.15 (13)
M. rectus femoris	–0.50 (358)	–0.96 (330)	0.35 (231)	–0.39 (377)	–1.44 (39)
M. vastus lateralis	–0.64 (82)	–0.64 (75)	–0.30 (53)	–0.97 (288)	–1.29 (45)
M. biceps femoris (syn. *gluteobiceps*)	–	–0.88 (735)	2.11*(367)	–0.59 (285)	–1.17 (70)
M. gastrocnemius	–	–	–0.46 (87)	1.44 (175)	0.41 (22)
M. semitendinosus	–0.35 (56)	–1.02*(583)	0.76 (255)	0.15 (259)	–0.85 (33)
M. semimembranosus	–1.46 (114)	–0.21 (982)	0.19 (416)	0.02 (522)	–2.06 (35)
M. semispinalis capitis	–2.27 (26)	–3.80 (62)	–	0.34 (272)	–0.28 (32)
M. serratus ventralis cervicis	–2.18 (61)	–1.25 (180)	1.09 (94)	0.67 (153)	0.41 (27)
M. rectus abdominis	–	–	–1.83 (18)	–0.35 (150)	–1.75 (12)
M. pectoralis profundus	–	–	2.57*(278)	1.00 (212)	–1.47 (72)
M. pectoralis superficialis	–	–	–1.56 (45)	1.14 (157)	2.34 (35)

MQ scores will be: i.e. actual MQ scores will be distributed around this mean with a standard deviation of the order of 8–10.

The results in Table 11 are consistent with the accuracy of the MSA model presented by Thompson (2002). In his analysis, Thompson (2002) calculated whether the consumer grade in terms of eating quality aligned with the grade assigned by the model. It was shown that the accuracy of the model was of the order of 50–70% and, if there were deviations between the assigned and allocated grades, the deviation was only of the order of one grade.

Based on an independent dataset of three muscles which had been taste-tested by Australian and Korean consumers, Thompson *et al*. (2008*b*) showed that accuracy of grade allocation was of the order of 59 and 53% for Korean and Australian consumers, respectively. When the data were examined, on the basis of residuals, the deviations were similar to those reported in this study, with the exception of the *M. semimembranosus* where deviations were of the order of 2–8 points on the MQ scale.

Model output for commercial application

While the primary use of the prediction model is as a grading tool, it is also an integral part of the extension and education efforts of explaining the basis of the MSA system to industry and in demonstrating the predicted eating quality impact of changes to the various input variables. Figure 6 presents a diagrammatic representation of the model predictions for a sample carcass. The input variables are displayed to the left hand top corner. The predicted grade for each muscle × cooking combination at the specified days aged is shown in the right hand portion. The grades displayed are

determined from an individual MQ points score for each of the cut × cook combinations.

While this form of display is used to demonstrate the model the relevant information is transferred as a data string in commercial grading applications. The data are transferred from the hand-held data capture unit (DCU) and can be uploaded to abattoir management information systems to produce mandatory feedback reporting to suppliers and can be integrated with traceability and management systems to allow eating quality scores to be modified with aging and tracked through inventory. Under current commercial practice, carcasses are most commonly assigned to stratification groups at the point of grading. Stratification groups, based on minimum MQ grade scores for any nominated cuts at specific days aged for specified cooking methods, are defined by the abattoir. The DCU is programmed to check each carcass against the stratification specification and allocate appropriately. This allows carcasses to be marshalled into runs for fabrication. Further discussion of commercial application of the model output is provided by Polkinghorne *et al*. (2008*a*).

Conclusions

This paper describes the statistical development of the MSA prediction model for beef palatability. This model is essentially a descriptive model, with the intention of describing the large amounts of data collected by MSA. The model is constrained to accord with accepted meat science principles and smoothness.

A broad range of input variables as discussed above are applied interactively on a muscle by muscle basis to produce a predicted consumer score for each muscle × cooking method combination. This score is used to allocate cuts into

Description	Format	Name	Input
Estimated % *Bos indicus*	% or X if doubt	EPBI	25
Animal sex type	M/F	Sex	M
Hormone Growth Promotant	Y or ? / N	HGP	N
Milk Fed Vealer	Y/N	MFV	N
Saleyard	Y/N	SlYrd	N
Rinse/Flush	Y/N	RnFl	N
Hot Std carcass weight	Weight in kg	HSCW	280
Hang method	AT/TS/TL/TC/TX	Hang	TS
Hump height	mm	Hump	5
Ossification USDA	USDA measure	uoss	120
Marbling USDA	USDA measure	umb	350
RibFat	mm	RbFt	5
Ultimate pH	Metered pH	UpH	5.5
Loin Temp at Grade	Metered Temp °C	Utmp	3
Days of aging from kill	Days aged	Age	21

cut	muscle	GRL	RST	SFR	TSL	SCT	CRN
spinalis	SPN081	5	4	5	4		
tenderloin	TDR034	5		4			
tenderloin	TDR062	5	4	5	4		
tenderloin	TDR063	4					
cube roll	CUB045	4	4	4	4		
striploin	STA045	4	4	4	4		
striploin	STP045	4	4	4	4		
oyster blade	OYS036	3	3	4	4		
blade	BLD095						
blade	BLD096	3	3	3	3	4	
chucktender	CTR085		3	3	3	4	
rump	RMP131	4	4	4	4	4	
rump	RMP231	4	4	4	4		
rump	RMP005	4		4	5		
rump	RMP032			4	5		
rump	RMP087		3	4	4	4	
knuckle	KNU066	3	4	4	4	3	
knuckle	KNU098		4	4	4		
knuckle	KNU099	3	3	3	3	3	
knuckle	KNU100		4	4	4		
outside flat	OUT005		3	3	4	4	3
outside flat	OUT029		3	4	3		
eye round	EYE075	3	3	3	3	3	3
topside	TOP001	3		3	4	3	
topside	TOP033	3		3	4	4	
topside	TOP073	3	3	3	4	4	
chuck	CHK068			3	3	4	
chuck	CHK074	4	3	3	4	4	
chuck	CHK078	3	3	3	3	4	
chuck	CHK081			3	4		
chuck	CHK082			3	3		
thin-flank	TFL051			3		3	
thin-flank	TFL052			4	4		
thin-flank	TFL064			4	3	4	
rib-blade	RIB041			3			
brisket	BRI056				3	3	
brisket	BRI057				3	4	
shin	FQshin					4	
shin	HQshin					4	
intercostal	INT037			3			

Fig. 6.   Example of Meat Standards Australia model output for a sample carcass.

three consumer defined acceptable quality grades or an unsatisfactory ungraded category. This prediction for individual muscles is a significant advance from applying a single quality grade to an entire carcass as a basis for grading.

Although we have derived a plausible and smooth model that described the available data well, this is not to say that the dataset could not be described at least as well by some other model. But what is true is that the model does provide a reasonable description of the data, which does appear to be useful in prediction. The words of George Box, Statistician, should be kept in mind here: 'All models are wrong . . . but some are useful' (Box and Draper 1987, p. 424).

Also worth repeating is the point that the model has been developed for active commercial use. While developed from a large number of controlled experiments and additional less controlled commercial collections the average performance of the model over a full range of commercial environments must take precedence to its accuracy on any specific set of experimental data. Additional data and ongoing analysis will enable further improvement and new versions are planned to maintain currency with consumer preference, which may change over time, and to make full use of additional data and analysis.

References

Bouton PE, Harris PV (1972) The effects of some post-slaughter treatments on the mechanical properties of bovine and ovine muscle. *Journal of Food Science* **37**, 539–543. doi: 10.1111/j.1365-2621.1972.tb02687.x

Box GEP, Draper NR (1987) 'Empirical model-building and response surfaces.' (Wiley: New York)

Canadian Beef Grading Agency (1997) 'Canadian beef carcass grading regulations.' (Canadian Beef Grading Agency: Calgary, Canada)

Carroll FD, Ellis KW, Lang MM, Noyes EV (1976) Influence of carcass maturity on the palatability of beef. *Journal of Animal Science* **43**(2), 413–417.

Dransfield E (1977) Intramuscular composition and texture of beef muscles. *Journal of the Science of Food and Agriculture* **28**, 833–842. doi: 10.1002/jsfa.2740280910

Dransfield E (1994) Optimisation of tenderisation, ageing and tenderness. *Meat Science* **36**, 105–121. doi: 10.1016/0309-1740(94)90037-X

Ferguson D, Thompson J, Polkinghorne R (1999) Meat Standards Australia: a 'PACCP'-based beef grading scheme for consumers. 3. PACCP requirements which apply to carcass processing. In 'Proceedings of the 45th International Congress of Meat Science and Technology, Yokohama, Japan.' p. 18.

Field R, McCormick IR, Balasubramanian V, Sanson D, Wise J, Hixon D, Riley M, Russell W (1997) Tenderness variation among loin steaks from A and C maturity carcasses of heifers similar in chronological age. *Journal of Animal Science* **75**, 693–699.

Gee A, Porter M, Polkinghorne R (2005) 'Protocols for the thawing, preparation and serving of beef for MSA trials for five different cooking methods.' (Meat and Livestock Australia: Sydney)

Hilton GG, Tatum JD, Williams SE, Belk KE, Williams FL, Wise JW, Smith GC (1998) An evaluation of current and alternative systems for grading quality carcasses of mature slaughter cows. *Journal of Animal Science* **76**, 2094–2103.

Hwang IH, Gee A, Polkinghorne R, Thompson JM (2002) The effect of different pelvic hanging techniques on meat quality in beef. In 'Proceedings of the 48th International Congress of Meat Science and Technology, Rome, 25–30 August 2002.' pp. 220–221.

JMGA (1988) 'New beef grading standards.' (Japanese Meat Grading Associations: Tokyo) [In Japanese]

Kim CJ, Lee ES (2003) Effects of quality grade on the chemical, physical and sensory characteristics of Hanwoo (Korean native cattle) beef. *Meat Science* **63**, 397–405. doi: 10.1016/S0309-1740(02)00099-2

Koch RM, Dikeman ME, Allen DM, May M, Crouse JD, Campion DR (1976) Characterization of biological types of cattle. III. Carcass composition, quality and palatability. *Journal of Animal Science* **43**, 48.

Koohmaraie M, Shackelford SD, Wheeler TL (1998) The biological basis of beef tenderness and potential approaches for its control and prediction. In 'Proceedings of the reciprocal meat conference, Fundepec symposium, Sao Palo, Brazil'.

Park BY, Hwang IH, Cho SH, Yoo YM, Kim JH, Lee JM, Polkinghorne R, Thompson JM (2008) Effect of carcass suspension and cooking method on the palatability of three beef muscles as assessed by Korean and Australian consumers. *Australian Journal of Experimental Agriculture* **48**, 1396–1404.

Polkinghorne R (2005) Does variation between muscles in sensory traits preclude carcass grading as a useful tool for consumers? In 'Proceedings of the 51st International Congress of Meat Science and Technology, 7–12 August 2005, Baltimore, USA'. p. 4.

Polkinghorne R, Watson R, Porter M, Gee A, Scott A, Thompson J (1999) Meat Standards Australia: a 'PACCP'-based beef grading scheme for consumers. 1. The use of consumer scores to set grade standards. In 'Proceedings of the 45th International Congress of Meat Science and Technology, Yokohama, Japan'. pp. 14–15.

Polkinghorne R, Philpott J, Gee A, Doljanin A, Innes J (2008a) Development of a commercial system to apply the Meat Standards Australia grading model to optimise the return on eating quality in a beef supply chain. *Australian Journal of Experimental Agriculture* **48**, 1451–1458.

Polkinghorne R, Thompson JM, Watson R, Gee A, Porter M (2008b) Evolution of the Meat Standards Australia (MSA) beef grading system. *Australian Journal of Experimental Agriculture* **48**, 1351–1359.

Rhee MS, Wheeler TL, Shackelford TL, Koohmaraie M (2004) Variation in palatability and biochemical traits within and among eleven beef muscles. *Journal of Animal Science* **82**, 534–550.

Romans JR, Tuma HJ, Tucker WL (1965) Influence of carcass maturity and marbling on the physical and chemical characteristics of beef. 1. Palatability, fiber diameter and proximate analysis. *Journal of Animal Science* **24**, 681.

Shackelford SD, Wheeler TL, Koohmaraie M (1995) Relationship between shear force and trained sensory panel tenderness ratings of 10 major muscles from *Bos indicus* and *Bos taurus* cattle. *Journal of Animal Science* **73**, 3333–3340.

Sherbeck JA, Tatum JD, Field TG, Morgan JB, Smith GC (1996) Effect of phenotypic expression of Brahman breeding on marbling and tenderness traits. *Journal of Animal Science* **74**, 304–309.

Shorthose WR (1996) A qualitative model of factors influencing beef tenderness. *Proceedings of the Australian Society of Animal Protection* **21**, 143.

Shorthose WR, Harris PV (1990) Effect of animal age on the tenderness of selected beef muscles. *Journal of Food Science* **55**, 1–8, 14. doi: 10.1111/j.1365-2621.1990.tb06004.x

Smith GC, Cross HR, Carpenter ZL, Murphey CE, Savell JW, Abraham HC, Davis GW (1982) Relationship of USDA maturity groups to palatability of cooked beef. *Journal of Food Science* **47**, 1100–1107, 1118. doi: 10.1111/j.1365-2621.1982.tb07627.x

Smith GC, Berry BW, Savell JW, Cross HR (1988) USDA maturity indices and palatability of beef rib steaks. *Journal of Food Quality* **11**, 1–13. doi: 10.1111/j.1745-4557.1988.tb00860.x

Thompson JM (2002) Managing meat tenderness. *Meat Science* **62**, 295–308. doi: 10.1016/S0309-1740(02)00126-2

Thompson JM, Polkinghorne R, Hearnshaw H, Ferguson DM (1999) Meat Standards Australia: a 'PACCP'-based beef grading scheme for consumers. 2. PACCP requirements which apply to the production sector. In 'Proceedings of the 45th International Congress of Meat Science and Technology, Yokohama, Japan'. p. 16.

Thompson JM, McIntyre BM, Tudor GD, Pethick DW, Polkinghorne R, Watson R (2008a) Effects of hormonal growth promotants (HGP) on growth, carcass characteristics, the palatability of different muscles in the beef carcass and their interaction with aging. *Australian Journal of Experimental Agriculture* **48**, 1405–1414.

Thompson JM, Polkinghorne R, Hwang IH, Gee AM, Cho SH, Park BY, Lee JM (2008b) Beef quality grades as determined by Korean and Australian consumers. *Australian Journal of Experimental Agriculture* **48**, 1380–1386.

Thompson JM, Polkinghorne R, Porter M, Burrow HM, Hunter RA, McCrabb GJ, Watson R (2008c) Effect of repeated implants of oestradiol-17β on beef palatability in Brahman and Brahman cross steers finished to different market end points. *Australian Journal of Experimental Agriculture* **48**, 1434–1441.

USDA (1989) 'Official United States standards for grades of carcass beef.' (USDA: Washington, DC)

Wang H, Claus JR, Marriott NG (1994) Selected skeletal alterations to improve tenderness of beef round muscles. *Journal of Muscle Foods* **5**, 137–147. doi: 10.1111/j.1745-4573.1994.tb00526.x

Watson R (2008) Meta-analysis of the published effects of HGP use on beef palatability in steers as measured by objective and sensory testing. *Australian Journal of Experimental Agriculture* **48**, 1425–1433.

Watson R, Gee A, Polkinghorne R, Porter M (2008a) Consumer assessment of eating quality – development of protocols for Meat Standards Australia (MSA) testing. *Australian Journal of Experimental Agriculture* **48**, 1360–1367.

Watson R, Polkinghorne R, Gee A, Porter M, Thompson JM, Ferguson D, Pethick D, McIntyre B (2008b) Effect of hormonal growth promotants on palatability and carcass traits of various muscles from steer and heifer carcasses from a *Bos indicus–Bos taurus* composite cross. *Australian Journal of Experimental Agriculture* **48**, 1415–1424.