Beef quality grades as determined by Korean and Australian consumers

J. M. Thompson[A,F], *R. Polkinghorne*[B], *I. H. Hwang*[C], *A. M. Gee*[D], *S. H. Cho*[E],
B. Y. Park[E] *and J. M. Lee*[E]

[A]Cooperative Research Centre for Beef Genetic Technologies, School of Environmental and Rural Sciences,
  University of New England, NSW 2351, Australia.
[B]Marrinya Agricultural Enterprises, 70 Vigilantis Road, Wuk Wuk, Vic. 3875, Australia.
[C]Department of Animal Resources and Biotechnology, Chonbuk National University, Jonju 561-756, Korea.
[D]Cosign, 20 Eleventh Avenue, Sawtell, NSW 2452, Australia.
[E]National Livestock Research Institute, Rural Development Administration, Suwon 441-350, Korea.
[F]Corresponding author. Email: jthompso@une.edu.au

Abstract.   Consumer responses were examined in an incomplete factorial design where Australian consumers evaluated 216 beef samples derived from 18 cattle killed in Australia and Korean consumers evaluated 216 samples from the same 18 cattle, plus 216 similar samples from 18 Korean cattle. Samples of the *Mm. triceps brachii, longissimus dorsi* and *semimembranosus* were cooked using grill and Korean barbeque methods. Each sample was sensory tested by 10 consumers, who scored it for tenderness, juiciness, like flavour, and overall liking. Consumers then graded each sample as either unsatisfactory (2 star), good every day (3 star), better than every day (4 star), or premium (5 star) quality.

For those samples assessed by both Australian and Korean consumers, the Korean consumers graded a higher proportion of samples 'unsatisfactory' and a lower proportion of samples 'premium' grade product than Australian consumers. Using a composite meat quality score (MQ4) to predict grade, a discriminant analysis showed that the Korean consumers had boundary cut-offs for the lower grades, which were ~4–10 palatability units higher than the Australian consumers.

Analysis of the residuals between actual and predicted palatability scores showed that the Meat Standards Australia (MSA) grading model produced relatively unbiased estimates within ±2 MQ4 units for the different consumer groups, muscle and carcass suspension treatments, with the exception of the *M. semimembranosus* samples. Implications of the results for both Korean and Australian beef markets through the use of an empirical grading model to predict palatability are discussed.

Introduction

The Meat Standards Australia (MSA) beef grading scheme uses a total quality management approach to predict beef palatability (Polkinghorne *et al*. 1999, 2008; Thompson 2002). A unique feature of the MSA scheme is the use of consumer taste panels to identify and quantify the critical control points (CCPs) to include in a beef grading model to predict palatability. Consumer sensory scores for tenderness, juiciness, like flavour and overall acceptability were combined into a single meat quality score. Boundaries between grades were estimated from the consumer sensory data. Data on consumer responses from over 60 000 muscle samples, which were sourced from a variety of production, processing, value adding and cooking treatments, have been used to progressively develop the MSA grading prediction model (Watson *et al*. 2008a). This model quantifies both the direct effects and interactions of the CCPs on the palatability of individual muscles prepared using a variety of cooking methods. Using data from the MSA database, Thompson (2002) showed that the MSA grading model allocated between 50 and 70% of the samples to their correct consumer grades.

In addition to scoring samples for tenderness, juiciness, like flavour and overall liking, consumers in the MSA taste panels were also asked to grade beef samples on palatability. A discriminant analysis was used to form a composite meat quality score (MQ4), which maximised the allocation of samples to the correct palatability grade by optimising the MQ4 boundaries between the grades (Polkinghorne *et al*. 1999; Watson *et al*. 2008b). If the MSA grading model and consumer testing protocol are to be extended internationally to underpin grading systems in other markets, there is a need to validate the accuracy of the model using diverse consumer groups. In particular, there was interest in using consumers from diverse cultural backgrounds to determine any differences in the way they graded individual beef samples and the boundaries between the grades.

There have been several MSA investigations of the effect of using different consumer groups to score palatability of beef (R. Polkinghorne, unpubl. data). The results showed little difference in the sensory scores given by consumers who originated from an urban or rural background, or in Japanese consumers who had recently arrived in Australia *v*. long-term residents of European descent. If this was confirmed for consumers from overseas markets, the consumer testing protocol and the MSA MQ4 score may have value as an international descriptor of beef quality.

A large cooperative project was undertaken by Meat and Livestock Australia in Australia and the National Livestock Research Institute (NLRI) in Korea to compare how Korean and Australian consumers scored the palatability of beef. Specifically, the project aimed to determine the distribution of grades given by Korean and Australian consumers and the accuracy of an MQ4 score as a basis to grade beef samples. For Korean consumers it was appropriate that they taste Australian and Korean beef; however, due to import restrictions Australian consumers only tasted Australian beef. The accuracy of the MSA grading model to predict consumer scores for individual samples from the Korean and Australian consumer groups was also tested. The project also aimed to determine the distribution of grades given by different consumer groups and the accuracy of an MQ4 score as a basis to grade beef samples. The accuracy of the MSA grading model to predict consumer scores for individual samples from the Korean and Australian consumer groups was also tested. Polkinghorne *et al.* (2003) have presented a preliminary report of the grading results.

Materials and methods

Experimental design

The experimental design was an incomplete factorial, whereby samples from 18 Australian carcasses cooked using both grill and Korean barbeque (BBQ) methods were tasted by Australian and Korean consumers, and samples from 18 Korean carcasses cooked using both grill and Korean BBQ methods were tasted by Korean consumers. An incomplete factorial was necessary as import restrictions would not allow Korean beef samples to be imported into Australia for testing with Australian consumers. To generate variation in palatability, alternate carcass sides were hung by the Achilles tendon or hip suspended. At boning, three primals comprising the blade, striploin and topside were removed from each carcass side at both the Korean and Australian slaughters. These were stored at 1°C for 7 days before dissecting out the *Mm. triceps brachii*, *longissimus dorsi* and *semimembranosus* for the preparation of sensory samples.

The 18 Korean and 18 Australian carcasses × two carcass suspension treatments × three muscles × two cooking techniques provided a total of 216 samples for sensory testing by Australian consumers (the Australian consumers tested only Australian samples) and 432 samples for sensory testing by Korean consumers (the Korean consumers tested both Australian and Korean samples). It should be noted that this experiment was not aimed at comparing the eating quality of Korean and Australian carcasses, due to both the low number of carcasses sampled from each country, and the origin of carcasses being confounded with production and processing effects. Rather this experiment focussed on comparing consumer responses between Australian and Korean consumers.

Animals and the slaughter and boning procedures

In Australia, 18 grain-fed Angus steers, which were ~2 years old, and had been fed for 150 days on a high concentrate ration, were transported 1.5 h from the feedlot to a commercial abattoir where they were kept off feed with access to water, before being slaughtered the following day. At this abattoir an immobiliser

was installed at sticking and a rigidity probe at the downward hide puller. Both these electrical inputs used low voltage current at low frequency and in effect accelerated glycolysis. Data from previous slaughters of similar cattle showed that the glycolytic rate was very fast and so the immobiliser and rigidity probe were switched off during the slaughter of the experimental cattle. At the end of the slaughter chain, alternate sides were tenderstretched, or hung by the Achilles tendon before being placed in the chiller.

In Korea, 18 grain-fed Hanwoo steers from a grain feeding program at the NLRI were slaughtered in three groups, each comprising six animals. Before slaughter, animals were transported a short distance to the research abattoir where they were kept off feed, but were given access to water, before being slaughtered the following day. Low voltage stimulation was applied (45 V, 100 ms on and 12 ms off, 36 pulses per second, applied during bleeding for 10–20 s). As for the Australian carcasses, alternate sides were either tenderstretched or hung by the Achilles tendon before being placed in the chiller. At both the Australian and Korean slaughters, the tenderstretch treatment was rotated between left and right sides, each suspended by a cord or hook inserted through the obturator foramen of the pelvic bone. Both Australian and Korean carcasses were boned the day after slaughter and the blade, striploin and topside primals removed, vacuum packed and held at 1°C for 7 days before sample preparation.

Sample preparation

After aging for 7 days, the *Mm. triceps brachii*, *longissimus dorsi et lumborum* and *semimembranosus* were dissected out from the primals and trimmed of all fat and epimysium to prepare steaks for grill testing and blocks for later conversion to BBQ strips for cooking. Muscles from Australian carcasses had two sets of steaks and two BBQ blocks prepared, whereas muscles from Korean carcasses only had one set of steaks and one BBQ block prepared. Five steaks were prepared for each grill test measuring ~100 by 50 mm by 25 mm thick (cut across the grain), whereas the BBQ blocks measured ~75 by 20 by 50 mm [a summary of the sensory protocols is provided in Accessory Publication to the online version of the paper by Watson *et al.* (2008*b*)].

Sensory design and testing

For the grill panels each consumer received seven half steaks, comprising a starter steak followed by six experimental subsamples. Similarly, for the BBQ panels, each consumer received seven BBQ strips, comprising a starter strip followed by six experimental subsamples.

The 216 grill steak portions tasted by Korean consumers were tested in two grill panels, where each of the 360 consumers received subsamples from three Korean and three Australian carcasses. The 108 grill steak portions tasted by Australian consumers were tested in two grill taste panels, where each of the 360 consumers received three subsamples from the present experiment and three from other experiments being undertaken as part of the MSA consumer testing program.

The 216 BBQ strips tasted by Korean consumers were tasted in six BBQ panels, where each of the 360 consumers received BBQ

strips from three Korean and three Australian carcasses. The 108 BBQ strips tasted by Australian consumers were tested in three BBQ panels where each of the 180 consumers received six samples from Australian carcasses.

The recruitment of consumers was described in detail by Hwang *et al*. (2008). The sensory questionnaire and sensory score sheets for Korean consumers were translated from English into Korean. After translation these sheets were given to independent Korean groups, back translated and where necessary, adjusted several times to ensure that the correct meaning was being conveyed.

In Australia, sensory tests were undertaken by Sensory Solutions Pty Ltd whereas in Korea, the sensory tests were conducted by staff at the NLRI. A supervisor from Sensory Solutions Pty Ltd travelled to Korea to train the NLRI staff in the MSA consumer testing protocols and was present during testing to ensure uniform testing in both countries.

Thawed steaks were cooked in lots of 10 for a set time on a grill (Silex, Stanmore, NSW; 220–230°C) to achieve a medium degree of doneness that was assessed subjectively against the MSA photographic doneness standards (the photographic standards are available from MSA, via MLA, on request). The BBQ strips were thawed for 0.5 h at room temperature and then cooked on a gas heated tin plate (~225°C) with a water jacket. The strips were individually cooked by placing them on the tin plate and turning at the first pooling of liquid on the surface of the sample.

The testing protocol has been described in detail by Watson *et al*. (2008b). Briefly, consumers were asked to score each sample for tenderness, juiciness, like flavour and overall liking by placing a mark on a 100-mm line scale. These scales were anchored with the following definitions: tenderness – very tough to very tender; juiciness – very dry to very juicy; like flavour – dislike extremely to like extremely; overall liking – dislike extremely to like extremely. Consumers were also asked to grade each sample by ticking the box that best described the sample, based on the following word associations; unsatisfactory (2 star), good every day (3 star), better than every day (4 star), or premium quality (5 star).

Statistical analyses

Standard error for proportions of samples allocated to the four eating quality grades were calculated as $[p(1-p)/N]^{0.5}$, where p is the proportion and N is the number of observations that make up the proportion (Steel and Torrie 1980).

A linear discriminant function (PROC DISCRIM, SAS 1997) was used to assess the accuracy of using the four sensory scores to predict the consumer grades (unsatisfactory, good every day, better than every day and premium). The four sensory scores were then combined into a single MQ4 score and the boundaries between the grades were calculated. The analyses were performed using six data subsets from Korean and Australian consumers, cooked using grill and Korean BBQ methods and sampled from both Korean and Australian carcasses.

The accuracy of the four variable linear discriminate functions for grade (termed the 'optimal' weightings) was compared with a linear discriminate function for grade estimated using the MQ4 score, [termed the 'fixed' weightings calculated using the

weightings for tenderness, juiciness, like flavour and overall liking scores of 0.4, 0.1, 0.2 and 0.3, respectively; Watson *et al*. (2008b)]. Boundaries between the grades for the fixed linear discriminate function were estimated by assuming that the boundaries occurred where the adjacent discriminant functions were equal.

Before calculating the mean sensory score for each sample, the individual consumers' scores for the 10 subsamples were ranked and the two highest and two lowest scores were clipped [to reduce the bias and the variance of the estimate; see Watson *et al*. (2008b)]. Biases in the prediction of the MQ4 score using the MSA grading model (Watson *et al*. 2008a) were examined by calculating the residual MQ4 scores (i.e. the actual MQ4 score minus the predicted MQ4 score from the MSA grading model) and subjecting these to an analysis using the final model used by Park *et al*. (2008), which contained terms for origin of the cattle, carcass suspension, muscle, position(muscle), cook, carcass suspension × muscle, carcass suspension × cook, carcass suspension × cook × muscle, consumer and consumer × muscle.

Results

Sensory scores and allocation to grades

Raw means in Table 1 showed that Australian consumers scoring Australian samples had the highest mean scores for all sensory traits, cooked by either grill or BBQ methods. Korean consumers gave Korean samples lower tenderness scores compared with Australian samples for both grill and BBQ cooking techniques, although there was little difference between juiciness, like flavour and overall liking scores. The standard deviation of the tenderness scores tended to be slightly higher than the other sensory scores given by both Australian and Korean consumers, although like flavour scores given by Korean consumers had a lower standard deviation for both grill and BBQ samples, compared with those given by Australian consumers.

Table 2 showed that Korean consumers had very similar correlations between sensory traits when assessing Korean and

Table 1.   Raw means and variance for sensory scores (on a 1–100 scale) of three muscles from hip-suspended and Achilles-hung sides given by Australian and Korean consumers evaluating grill and barbeque (BBQ) samples from Australian and Korean carcasses (each mean is the average of 1080 consumer scores)

Sensory scores	Australian consumers Australian samples	Korean consumers Australian samples	Korean samples
Grill cooking technique			
Tenderness	60.3 ± 25.3	57.3 ± 23.4	51.6 ± 24.8
Juiciness	64.7 ± 23.5	54.5 ± 21.1	58.3 ± 21.7
Like flavour	63.8 ± 22.4	57.4 ± 17.6	57.6 ± 18.3
Overall liking	63.5 ± 23.4	56.7 ± 20.2	55.1 ± 21.9
Korean BBQ cooking technique			
Tenderness	65.2 ± 22.1	64.5 ± 23.2	61.0 ± 23.9
Juiciness	67.4 ± 20.2	62.7 ± 20.3	62.0 ± 20.0
Like flavour	65.6 ± 21.5	63.5 ± 17.7	61.0 ± 17.6
Overall liking	66.2 ± 21.4	62.9 ± 20.4	60.7 ± 20.1

Table 2. Simple correlation coefficients between sensory scores for samples that had been cooked using grill or Korean barbeque methods and served to Korean and Australian consumers

The Korean consumers were given samples from both Korean and Australian carcasses. Within consumer groups, grilled samples were above and barbeque samples were below the diagonal

Sensory score	Tenderness	Juiciness	Like flavour	Overall liking
Korean consumers assessing Korean carcasses (n = 1080)				
Tenderness score	1.00	0.58	0.42	0.76
Juiciness	0.51	1.00	0.43	0.61
Like flavour	0.33	0.33	1.00	0.66
Overall liking	0.74	0.58	0.51	1.00
Korean consumers assessing Australian carcasses (n = 1080)				
Tenderness score	1.00	0.59	0.38	0.75
Juiciness	0.55	1.00	0.37	0.64
Like flavour	0.37	0.31	1.00	0.58
Overall liking	0.74	0.59	0.54	1.00
Australian consumers assessing Australian carcasses (n = 1080)				
Tenderness score	1.00	0.70	0.72	0.81
Juiciness	0.77	1.00	0.71	0.77
Like flavour	0.73	0.73	1.00	0.90
Overall liking	0.81	0.79	0.93	1.00

Fig. 1. Frequency distributions of eating quality grades allocated by Korean and Australian consumers for both the (*a*) grill and (*b*) barbeque samples. Vertical bars indicate standard errors.

Australian samples. Korean consumers also had lower correlations between sensory scores than Australian consumers for both the grill and BBQ cooking methods; i.e. the different sensory scores given by Korean consumers were less dependent than the scores given by Australian consumers. Within Korean and Australian consumer groups there was little difference between cooking methods in the magnitude of the correlations.

Figure 1*a*, *b* showed the frequency distributions of eating quality grades allocated by Korean and Australian consumers for both the grill and BBQ samples. For both cooking methods, Korean consumers (testing both Korean and Australian samples) allocated a higher proportion of samples as 2-star grade (or unsatisfactory), compared with Australian consumers. For samples prepared as grills there was little difference in the proportions of samples allocated to 3-star grade (or good every day), whereas Australian consumers allocated a higher proportion of samples to 4-star (better than every day) and 5-star grades (premium), than did the Korean consumers. A similar pattern was observed for BBQ samples, except there was little difference between consumer groups in the proportions allocated to grades 3 and 4.

Accuracy of using composite scores to allocate samples to grades and the boundaries between grades

Table 3 showed the accuracy of two discriminant functions at allocating samples to the correct eating quality grades for both
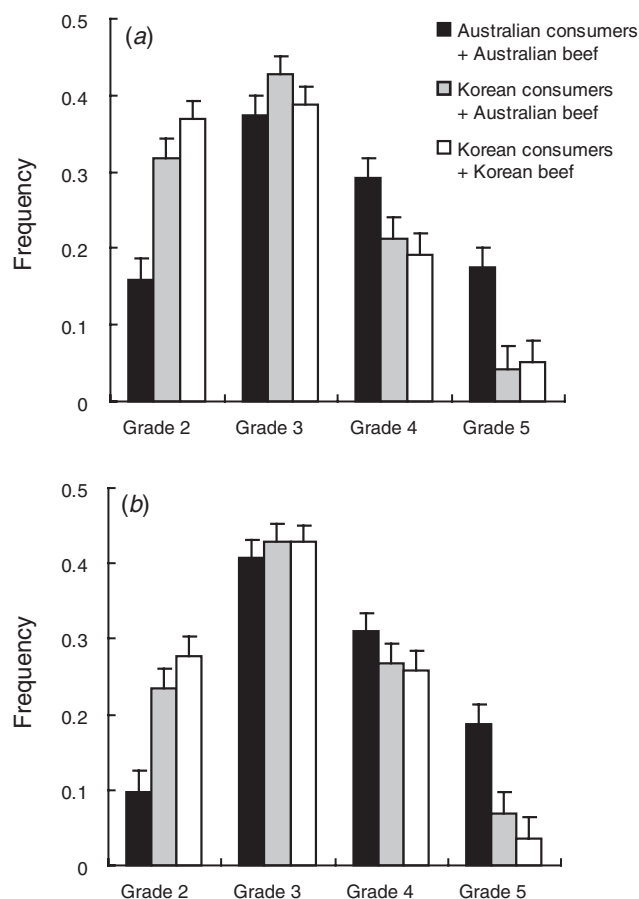
Korean and Australian consumers. For Korean consumers presented with grill or BBQ samples, the first discriminant function estimated from tenderness, juiciness, like flavour and overall liking scores was more accurate at allocating samples to the correct grade for the higher eating quality grades compared with the lower eating quality grades, regardless of the origin of the sample. For Australian consumers testing grilled samples, the accuracy of the linear discriminant function at allocating samples to the correct grades was similar across all grades. For Australian consumers testing the BBQ samples, the accuracy was high for the 3-star product. For both consumer groups there was a trend for a higher accuracy in correctly allocating samples to eating quality grades for samples that had been grilled rather than cooked as BBQ. Overall accuracy at allocating samples to the correct grades using the four variable discriminant functions was ~68% for both the Australian and Korean consumers.

The second discriminant function in Table 3 examined the accuracy of correctly allocating samples to the correct grade using a linear discriminant function based on an MQ4 score [calculated using fixed weightings for the four sensory scores, see Watson *et al*. (2008*b*)], to predict eating quality grade. Table 3 showed that for the different consumer group, cooking method and sample origin subclasses the losses in accuracy

Table 3. Percentage accuracy of two discriminant functions to allocate samples to the quality grade assigned by the consumer (expressed as a percentage of the number allocated to the correct grade relative to the number predicted for a particular grade)

Discriminant function 1 used optimal weightings for the four sensory scores compared with function 2, which used fixed weightings to calculate a composite meat quality (MQ4) score to allocate samples to grades. T, tenderness; J, juiciness; LF, like flavour; OL, overall liking

Consumer group	Cooking method	Carcass origin	Unsatisfactory (2 star)	Good every day (3 star)	Better than every day (4 star)	Premium (5 star)	Overall
1. Boundaries estimated using the linear discriminant function of the four sensory scores (T, J, LF, OL)							
Korean	Grill	Korean carcasses	88	70	45	38	68
Korean	Grill	Australian carcasses	79	73	51	32	65
Korean	Korean BBQ	Korean carcasses	77	74	56	25	65
Korean	Korean BBQ	Australian carcasses	72	73	57	43	65
Australian	Grill	Australian carcasses	71	76	64	67	70
Australian	Korean BBQ	Australian carcasses	54	84	58	59	65
2. Boundaries estimated using the MQ4 score (with fixed weightings for the four sensory scores)							
Korean	Grill	Korean carcasses	80	75	48	32	67
Korean	Grill	Australian carcasses	72	65	45	29	59
Korean	Korean BBQ	Korean carcasses	73	70	53	21	60
Korean	Korean BBQ	Australian carcasses	67	68	51	31	58
Australian	Grill	Australian carcasses	67	71	60	67	66
Australian	Korean BBQ	Australian carcasses	50	80	56	58	63

from using a combined MQ4 score, compared with the four variable discriminant functions for each consumer and cooking method subgroup were small, ranging between 7 and 1 percentage units.

Boundaries between the eating quality grades for the MQ4 score for the various consumer group, cooking method and sample origin subclasses are shown in Table 4. There was little difference in the grade boundaries estimated for Korean consumers who tasted samples from Korean and Australian carcasses. Across cooking method and consumer groups, boundaries tended to be similar for grades 3 and 4 and grades 4 and 5 boundaries. However, at the lower grade Korean consumers tended to have a boundary which was 4–10 points greater than for Australian consumers.

Table 4. Grade boundaries calculated from the discriminant function of the composite meat quality score for the different consumer groups, cooking methods and carcass origin

Consumer group	Cooking method	Carcass origin	Grade boundaries		
			2–3	3–4	4–5
Korean	Grill	Korean carcasses	48	66	78
Korean	Grill	Australian carcasses	49	66	79
Korean	Korean BBQ	Korean carcasses	52	69	81
Korean	Korean BBQ	Australian carcasses	53	68	81
Australian	Grill	Australian carcasses	43	64	81
Australian	Korean BBQ	Australian carcasses	45	66	81

Accuracy of the MSA model to predict an MQ4 score for both Australian and Korean consumers

The statistical model used by Park *et al*. (2008) was fitted to describe treatment effects on the residual sensory scores (Table 5). The interaction cook × carcass suspension × muscle was significant ($P < 0.01$) due to the MSA model underestimating the MQ4 score for the grilled *M. semimembranosus* for both carcass suspension methods and underestimating eating quality for the BBQ samples from the Achilles hung side (see residuals in Fig. 2). The effect of carcass origin was significant ($P < 0.001$) with the MSA model overestimating the MQ4 scores for Australian cattle by 3.1 units, and underestimating the MQ4 scores for Korean cattle by 2.8 units. However, as discussed earlier this difference should be treated with caution as the meat

Table 5. *F*-ratios for the effect of cattle, carcass suspension, muscle, cooking method and consumer group on the residuals for the meat palatability score (actual consumer composite meat quality score – score predicted from the Meat Standards Australia model)

NDF, numerator degrees of freedom; DDF, denominator degrees of freedom; *, $P < 0.05$; **, $P < 0.01$ and ***, $P < 0.001$, respectively. n.s., not significant

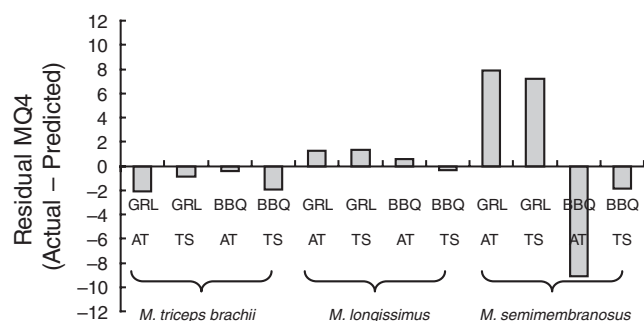| Independent variables	NDF	DDF	*F*-ratio
Cattle	1	629	55.37***
Carcass suspension	1	629	1.87n.s.
Muscle	2	629	4.68n.s.
Position(muscle)	1	629	1.11n.s.
Cook	1	629	52.02***
Carcass suspension×muscle	2	629	3.42*
Muscle×cook	2	629	42.69***
Carcass suspension×cook	1	629	1.21n.s.
Cook×carcass suspension×muscle	2	629	6.63**
Consumer	1	629	32.75***
Consumer×muscle	2	629	2.34n.s.

Fig. 2. Residual composite meat quality (MQ4) scores (i.e. actual consumer MQ4 score – MQ4 score predicted from the Meat Standards Australia model) for the interaction between cook (grill, GRL and barbeque, BBQ methods), carcass suspension (TS, tenderstretch; AT, achilles hung) and muscle (*Mm. triceps brachii*, *longissimus dorsi* and *semimembranosus*). Predicted MQ4 scores were from the SP2004 Meat Standards Australia model (Watson *et al*. 2008*a*).

quality of the carcasses was confounded with processing and was based on very low numbers. The effect of consumer on the residuals was also significant ($P < 0.001$) with the MSA model underestimating MQ4 scores for Australian consumers by 2.4 palatability units and overestimating MQ4 scores for Korean consumers by 2.1 palatability units.

Discussion

Sensory scores and allocation to grades

This study showed that Korean consumers had a more discriminating attitude to both overall sensory scores given to beef samples and the allocation of these samples to eating quality grades than Australian consumers. Korean consumers tended to give samples lower sensory scores for beef than Australian consumers. In addition, Korean consumers downgraded a higher proportion of samples to 2-star grade (or unsatisfactory) with a correspondingly lower proportion to a 5-star grade (premium), than did Australian consumers. This pattern was evident for samples prepared by both grill and BBQ methods and so was unlikely to be associated with unfamiliarity of the Korean consumers with the grill method, or Australian consumers with the BBQ cooking method. Finally, the more discriminating attitude of Korean consumers was also reflected in the higher grade boundaries between grades 2 and 3 (good every day), where within cooking techniques Korean consumers had a boundary which was 4–10 MQ4 units higher than for Australian consumers.

This difference in attitude to beef quality between Korean and Australian consumers has implications for both the domestic Korean market and for countries that are importing beef into Korea. For Korean beef producers these results indicate the need to avoid the production and supply of low quality beef, if the Korean consumer is to consider beef of satisfactory eating quality. Countries that export beef to Korea should be aware that product at the lower end of the quality scale has a higher risk of being classed as 'unsatisfactory' by Korean consumers.

For Korean consumers the discriminant function calculated using either the four sensory scores, or a composite MQ4 score, which used fixed weightings for the sensory traits, was more accurate at the allocation of samples to the correct grade at the lower than to the higher palatability grades. This suggested that for Korean consumers other factors not described by the four sensory dimensions may play an increasingly important role in describing eating quality at the higher grades. It was interesting that this same trend was not evident for Australian consumers, so that a linear combination of the four sensory scores was capable of describing the eating quality grades across a wide range in quality. This decline in accuracy at the higher eating quality grades for Korean consumers was not simply due to the addition of linear or curvilinear interactions between the sensory scores being added to the discriminant function, but possibly a result of other factors not captured by the four sensory dimensions (J. M. Thompson, unpubl. data). It was also possible that the translation of consumer questions from English to Korean provided descriptions that were not directly analogous to the original English meaning. In this experiment, this risk was minimised before the start of the experiment by the repeated back translation and adjustment of the Korean language questionnaire sheets before the final wording was agreed.

There are few studies with which to compare the accuracy of consumer sensory scores to predict quality grades. In a recent study, Platter *et al*. (2003) calculated the accuracy of sensory scores for tenderness, juiciness and flavour to predict 'unsatisfied' or 'satisfied' categories for beef striploin samples. In the present study, the accuracy for tenderness, juiciness, like flavour and overall acceptability to predict 'unsatisfactory' or better than 'unsatisfactory' grades varied between 77 and 83% correctly classified for Australian consumer, whereas for Korean consumers, the corresponding accuracies ranged from 67 to 80% correctly classified (J. M. Thompson, unpubl. data). These results were similar to those of Platter *et al*. (2003), where a logit analysis estimated that sensory scores correctly classified 78 to 79% of beef samples into 'unsatisfied' or 'satisfied' categories.

The loss in overall accuracy at allocating samples to grades using an MQ4 score with fixed weightings of 0.4, 0.1, 0.2 and 0.3 for tenderness, juiciness, like flavour and overall liking (Watson *et al*. 2008*b*) compared with using the optimal weightings from the disciminant analysis was of the order of 5% for Korean and 2% for Australian consumers. The slightly lower loss in accuracy for Australian consumers was not surprising given the higher correlations between sensory scores for Australian consumers. The higher correlations between sensory scores for Australian consumers suggested that they were more prone to giving similar scores for all sensory dimensions, in other words they either liked or disliked the samples and rated all four sensory dimensions accordingly, whereas the Korean consumers tended to be more independent in the manner in which they scored each of the sensory dimensions. The higher the correlations between the sensory scores, the less important it was to optimise the weightings of the discriminant function to determine the weightings to calculate the MQ4 score.

As discussed the sensory scores given by Korean consumers tended to be slightly lower than those given by Australian consumers. In addition, like flavour scores given by Korean

consumers had a lower standard deviation. A possible reason may be Koreans consumers normally eat beef with sauces and marinades, and, therefore, in this study where samples were eaten without any additives, Korean consumers felt that the beef samples did not cover the range of flavours normally experienced.

Accuracy of the MSA prediction model

Residuals, calculated from the actual minus predicted MQ4 scores, showed that accuracy of the MSA model was ±2 MQ4 scores for the *Mm. longissimus dorsi* and *triceps brachii* for both Australian and Korean consumers. The exception was the *M. semimembranosus*, which showed relatively large biases of ±8 MQ4 units depending upon carcass suspension and cooking treatments. The number of *M. semimembranosus* samples used in the construction of the MSA model reported by Watson *et al.* (2008*a*) was low compared with the other muscle and cooking combinations and so the present data were a useful contribution. It needs to be stressed that the development of the MSA model is a dynamic process, whereby the relationship between input variables and palatability scores has been built up using the data from many different experiments. It is intended that the biases identified in this experiment will be considered along with other additional experimental data and incorporated into the model to reestimate the coefficients and refine the MSA model in future versions.

Conclusions

Korean consumers were more discriminating when grading beef samples than Australian consumers. Differences between the consumers were evident in several ways. First, Korean consumers graded a larger proportion of the samples as 2 star (unsatisfactory) and a lower proportion as 5 star (premium). Second, the Korean consumer had a higher boundary between eating quality grades 2 and 3 than Australian consumers. This has implications for both Korean producers supplying the Korean domestic market and the importers who supply beef to Korea. Third, the lower correlations between sensory scores showed that Korean consumers were more independent in their scores for the different sensory dimensions compared with Australian consumers.

The current MSA model accurately predicted palatability for both consumer groups and cooking methods for the *Mm. triceps brachii* and *longissimus dorsi*. However, some adjustment to the MSA model is required for *M. semimembranosus* as the current model underestimated palatability scores for the grilled *M. semimembranosus* samples from Achilles hung and hip suspension sides, whereas it overestimated palatability in BBQ samples from Achilles hung sides.

These results suggested that an empirical model that uses commercial inputs can predict consumer palatability for both Australian and Korean consumers. However, some adjustment may be required to minimise current bias in the MSA model. This could include adjusting the boundaries of the lower quality grades for Korean consumers and improving the prediction of the

M. semimembranosus. The fixed weightings used by MSA used to calculate the palatability score from the four taste panel scores appeared robust for the different consumer groups and cooking methods, due largely to the high correlation between sensory scores.

Acknowledgements

References

Hwang IH, Polkinghorne R, Lee JM, Thompson JM (2008) Demographic and design effects on beef sensory scores given by Korean and Australian consumers. *Australian Journal of Experimental Agriculture* **48**, 1387–1395.

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.

Platter WJ, Tatum JD, Belk KE, Chapman PL, Scanga JA, Smith GC (2003) Relationships of consumer sensory ratings, marbling score, and shear force value to consumer acceptance of beef strip loin steaks. *Journal of Animal Science* **81**, 2741–2750.

Polkinghorne R, Watson R, Porter M, Gee A, Scott J, 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, 1–6 August 1999, Yokohama, Japan.' pp. 14–15.

Polkinghorne R, Gee AM, Hwang IH, Park BY, Cho SH, Lee JM , Kim KN, Thompson JM (2003) A comparison of beef palatability in Korean and Australian consumers. In 'Proceedings of the 49th international congress of meat science and technology, 31 August–5 September 2003, Campinas, Brazil.' pp. 227–228.

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

SAS (1997) 'Applied statistics and the SAS programming language.' (SAS Institute: Cary, NC)

Steel RGD, Torrie JH (1980) 'Principles and procedures of statistics. A biometrical approach.' (McGraw-Hill International Book Company: Sydney)

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

Watson R, Polkinghorne R, Thompson JM (2008*a*) Development of the Meat Standards Australia (MSA) prediction model for beef palatability. *Australian Journal of Experimental Agriculture* **48**, 1368–1379.

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