This paper reviews recent research into predicting the eating qualities of beef. A range of instrumental and grading approaches have been discussed, highlighting implications for the European beef industry. Studies incorporating a number of instrumental and spectroscopic techniques illustrate the potential for online systems to non-destructively measure muscle pH, colour, fat and moisture content of beef with $R^2$ (coefficient of determination) values $>0.90$. Direct predictions of eating quality (tenderness, flavour, juiciness) and fatty acid content using these methods are also discussed though success is greatly variable. $R^2$ values for instrumental measures of tenderness have been quoted as high as 0.85 though $R^2$ values for sensory tenderness values can be as low as 0.01. Discriminant analysis models can improve prediction of variables such as pH and shear force, correctly classifying beef samples into categorical groups with $>90\%$ accuracy. Prediction of beef flavour continues to challenge researchers and the industry alike, with $R^2$ values rarely quoted above 0.50, regardless of instrumental or statistical analysis used. Beef grading systems such as EUROP and United States Department of Agriculture systems provide carcase classification and some indication of yield. Other systems attempt to classify the whole carcase according to expected eating quality. These are being supplemented by schemes such as Meat Standards Australia (MSA), based on consumer satisfaction for individual cuts. In Australia, MSA has grown steadily since its inception generating a 10% premium for the beef industry in 2015-16 of $187$ million. There is evidence that European consumers would respond to an eating quality guarantee provided it is simple and independently controlled. A European beef quality assurance system might encompass environmental and nutritional measures as well as eating quality and would need to be profitable, simple, effective and sufficiently flexible to allow companies to develop their own brands.

Keywords: bovine, palatability, quality assurance, grading, prediction

Implications

EU consumption of beef has been in decline for 20 years. This may be, at least in part, due to inconsistencies in eating quality, meaning that the customers cannot be sure of the quality they are purchasing. This review documents numerous examples of grading schemes and technologies that have been studied with the objective of managing and predicting beef-eating quality. The majority of investigations of instrumental methods have been carried out under experimental conditions and require further development and validation, but some grading systems have been proven to work. The research reported herein gives the European beef industry an overview of the development of methods to improve consumers’ experience of beef-eating quality.

Introduction

Beef production in Europe contributes to food security, sustainable land use, the socioeconomic well-being of rural communities, and the gastronomic pleasure of urban and rural consumers across the continent. Beef is also a high-value product that represents an expensive item in household shopping baskets.

Beef consumption in the EU is 10.9 kg/person per year, averaged over 28 countries in 2016 (Organisation for Economic Co-Operation and Development, 2017), with considerable variation between the member countries. This is lower than in South American countries (e.g. Argentina, 46.8 kg), North America (e.g. United States, 25.0 kg) and Australia (21.9 kg), but considerably higher than China (4.0 kg). However, consumption in developed countries has declined over the last 20 years, by 12% in the EU, 19% in the United States and 20% in Australia. Together with adverse publicity concerning environmental, health, authenticity and safety issues, inconsistent quality may have contributed to this decline.

The beef industry in Europe is exceedingly diverse. The breeds of cattle and the regimes for rearing them have been adapted over centuries to suit the climate and the availability of grazing and feed throughout the seasons. The proportion of beef derived from the dairy herd or bulls v. steers varies...
Quality assurance of European beef

The quality of beef first became a concern when beef ceased to be a local product and carcasses began to be transported long distances from where they were reared and slaughtered to where they were butchered and consumed (Polkinghorne and Thompson, 2010). This led to the development of ‘beef classification systems’, designed to describe the commercially important attributes of the carcasses and facilitate communication with distant primary producers. The main parameters were carcass weight, age/maturity, sex, fat cover and colour, conformation and freedom from bruising and blemishes. Such systems often came to include marbling, lean colour/appearance, fatness, estimates of yield such as conformation or eye muscle area and reporting of prices. Many of these parameters remain important in modern systems (Bonny et al., 2017).

‘Carcase grading’, in contrast, has been defined as ‘the placing of different values on carcases for pricing purposes, depending on the market and requirements of traders’ (AHDB Industry Consulting, 2008). Systems were developed in United Kingdom, Germany, Ireland and France in the 1960s and 1970s, and these evolved under the European Commission into EUROP grading, introduced in 1981. These grading systems are based upon the carcass and have become less useful as beef is traded increasingly as vacuum-packed cuts or boxed product rather than whole carcasses. The meat industry’s increasing desire to accurately predict the saleable meat yield (%) of a carcase further highlights the inadequacies of the EUROP system to meet current industry interests. As the prediction of saleable meat yield does not correlate with beef-eating quality, it is not further discussed in this review. However, recent literature regarding the improving saleable meat yield prediction has been extensively reviewed by Craigie et al. (2012). It has also become clear that the tacit assumption that there is a simple relationship between the eating qualities of the cuts within a carcase is not true, but depends on complex interactions with other factors such as hanging, breed, chill rate and ageing (Rhee et al., 2004; Polkinghorne, 2005; Farmer et al., unpublished data).

More recently, local ‘farm assurance schemes’ have been introduced, which involve product certification, to reassure the customer on issues such as animal welfare, production methods, traceability and good hygienic practice. Examples include Red Tractor Mark (UK), Farm Quality Assurance Scheme (NI) and similar national and regional schemes across Europe (Aragrande et al., 2005; Anonymous. European Commission, Agriculture and Rural Development, 2017). Despite these various schemes, there is evidence that the eating quality of beef in Europe is not consistent. Data from four countries across Europe show that 19% grilled sirloin, 25% grilled rump and 53% roast topside were deemed ‘unsatisfactory’ by consumers (Farmer et al., 2016). This indicates that beef is not always meeting consumer expectations.

Recent research on the management of beef eating quality

Much research on beef eating quality has examined the many production and processing factors which can affect eating quality and the scientific mechanisms for these effects. These studies are outwith the scope of this article.

Instead, this article will focus on those studies that address practical mechanisms by which beef eating quality might be managed and predicted in a commercial setting. It is notoriously difficult to judge eating quality from the appearance of the meat and the beef industry and researchers continue to seek methods to improve the consistency of their product. The development and increasing availability of new instrumental technologies have led to studies on their application to eating quality management.

A number of papers have already reviewed aspects of this work. It has been reported (Verbeke et al., 2010a) that European consumer groups considered muscle profiling to be beneficial, suggesting that there is potential for a beef eating quality guarantee system in Europe. Mullen and Troy (2005) highlighted the biochemical predictors for eating quality at early stages postmortem and reviewed the development of techniques for quality prediction. Barbut (2014) highlighted the impact of increasing automation on demand for sensors and control systems to manage quality during meat processing. Polkinghorne and Thompson (2010) have reviewed the development of grading systems.

Recent studies fall into two main categories: (a) Instrumental methods for predicting eating quality and (b) Grading methods for beef eating quality.

**Instrumental methods for predicting eating quality**

There has for some time been a demand from the industry for an instrument that could be placed on the beef production...
line, ideally very soon after slaughter, which is able to predict the eating quality of the final product at the point of consumption, some 7 to 21 days or more later. Given the many changes that occur as muscle matures into meat, this is an ambitious target. Nevertheless, work has been conducted in this area, with mixed success.

There are two approaches. The first is to measure online, and non-invasively, parameters that are known to predict eating quality, such as pH at various times post-slaughter and fat content or depth or marbling. The second aims to predict directly attributes such as tenderness, juiciness or flavour from detailed spectroscopic analyses. This presumes that the chemical parameters underpinning eating quality are sufficiently abundant to be measured by these techniques. Care is required to ensure that any predictions are not circumstantial, or apply only to a limited data set.

The following sections describe the many recent attempts to determine or predict meat quality using various types of instrumentation. These, together with a full list of references consulted, are listed in Supplementary Tables S1 and S2. These different authors use different methods and terminologies to describe their work. For the purpose of this report, we have defined ‘prediction’ as a forecast of a future result, for example, recording a spectrum at 24 h postmortem to predict a quality parameter at 14 days. In contrast, when the spectrum and the quality parameter are both recorded at the same time (or where the sample is frozen at that time), we will refer to this as ‘measurement’. In addition, a wide range of statistical methods has been used to evaluate the performance of the predictive models described. Some authors have used calibration and validation data sets while others use data independent of that used to create the model (full cross-validated predictions). Cross-validation is a process that assesses the prediction on a new set of samples. The ‘new set of samples’ can either be a sample set of the original data, that has not been used to develop the model (segmented cross-validation) or it can be a completely new set of data independent of that used to create the model (full cross-validation). Segmented cross-validation is commonly used when full cross-validation would be too time consuming or when all the samples are analysed together (Wu and Sun, 2013).

Robotic pH. Roehe et al. (2014) were first to report the automated measurement of muscle pH using a robotic system. Manual (commercial, hand held pH meters) and robotic techniques used to measure muscle pH at 45 min post-mortem (pH45min) were assessed against chemical pH measuring techniques. Correlation coefficients (R) between the automated methods tested and chemical assessment ranged between 0.38 and 0.47. The relationship between robotic pH and pH measured by chemical analysis explained only 22% of the variation between methods. Results also showed that the proposed system could operate in a commercial environment and could deal with variations in carcase size, classification and presentation. However, this work highlights several challenges facing robotic systems namely; accuracy, robustness and ease of cleaning of currently available pH probes.

Computer vision techniques. Jackman et al. (2008) have used a computer vision system to assess colour, marbling fat and also to predict the eating quality of beef aged between 2 and 21 days. In this case, sensory panels (14-day-aged) and Warner Bratzler Shear Force (WBSF days, 2, 7, 14 and 21) measurements were used to assess eating quality. Partial least squares regression techniques showed that sensory ‘overall acceptability’ and WBSF (day 7) were predicted with R2 values of 0.88 and 0.85, respectively. Other sensory attributes, ‘hard’ (0.48), ‘juice’ (0.60) and ‘flavour’ (0.49) were less well predicted. Jackman et al. (2009) also used multi-linear and partial least squares regression techniques to improve predictions for tenderness (0.72), hard (0.60) and flavour (0.78). Overall acceptability, juice and WBSF (day 21) were predicted with R2 values of 0.82, 0.46 and 0.83, respectively. A model was also created that correctly classified high- and low-quality carcasses with 90% accuracy.

Ultrasound. Ultrasound techniques have been used to predict a number of fat-related variables such as fat depth and intramuscular fat (IMF) with varied success. Aass et al. (2009) showed that ultrasound techniques could be used to predict IMF in lean cattle. A prediction model was developed using stepwise regression procedures to predict IMF with an R2 value (validation) of 0.80. Indurain et al. (2009) also illustrated that measurements of fat thickness, computer image measurement of fatness and ultrasound measurements in combination with carcase fatness or conformation score could improve the prediction of IMF in the longissimus muscle. However, Roehe et al. (2014) have been less successful in predicting fat depth from ultrasound measurements. Results show that the automated robotic measurement of fat depth correlates poorly (R2 = 0.43) with the reference method for measuring fat depth (steel ruler).

Computerised tomography imaging (computerised tomography scanning). Computerised tomography scanning techniques have yielded encouraging results in recent years for the prediction of fat, muscle and bone content of beef carcases from a scan of the live animal (Navajas et al., 2010; Prieto et al., 2010) but were less successful at predicting fatty acid profiles and sensory attributes of beef (Prieto et al., 2010).

Navajas et al. (2010) showed that CT carcase images (with entire carcase weight and primal weights) can be used to predict fat, muscle and bone content with R2 values between 0.89 to 0.99, depending on the tissue and whether the calibration or validation data set were analysed. Total carcase tissue was predicted with R2 values between 0.95
and 0.96. Prieto et al. (2010) have used CT scanning to predict the composition, fatty acid content and meat quality aspects of beef. Partial least squares regression techniques were used to analyse CT images, instrumental and sensory data from samples of crossbred Aberdeen Angus and Limousin cattle. $R^2$ values for calibration and validation sample sets were obtained for subcutaneous fat (0.94, 0.92), IMF (0.81, 0.86) total fat (0.89, 0.93) and muscle content (0.99, 0.97). Fatty acid profiles and IMF were moderately predicted with $R^2$ predictions for both sire breeds ranging between 0.61 to 0.75 and 0.71 to 0.76, respectively. Sensory traits were poorly predicted with $R^2$ values ranging between 0.01 and 0.26. The cost and size of this instrument limits its potential for online use, but it has a valuable role in research and for calibration of other methods.

**Magnetic resonance imaging.** Magnetic resonance imaging (MRI) has been used to determine the IMF content of beef (Lee et al., 2015). Statistical analysis identified a strong correlation ($R^2 = 0.98$) between MRI images and chemical measurements for percentage IMF. Again, the cost and size of this instrument will limit its potential for meat plant use.

**VIS–NIR spectroscopy.** The spectroscopic measurement of meat quality continues to draw interest (see references in this section and Supplementary Tables S1 and S2). Recent research regarding the eating quality of beef has focussed on prediction or measurement of colour, pH, chemical composition, instrumental, and sensory tenderness and other sensory attributes.

As expected, meat colour is measured well by Visible–Near Infrared (VIS–NIR) spectroscopy. In 2003, Liu et al. (2003) demonstrated that VIS–NIR (400 to 1800 nm) methods could measure Hunter $L^*$, $a^*$ and $b^*$ values for beef aged between 2 and 21 days with coefficients of determination ($R^2$) values ranging between 0.78 and 0.90. Hunter L values were measured to a lesser extent with $R^2$ ranging between 0.49 and 0.55. Similar measurements for beef colour (longissimus et lumborum) have been quoted ranging between 0.49 and 0.55. Related to beef colour, VIS–NIR instruments have also been used to classify dark cutting beef under non-oxygenated and bloomed conditions (Prieto et al., 2014). Partial least squares regression methods have created a model that correctly identified 80% to 95% of dark cutting beef samples presented depending on the instrument used and the degree of oxygenation.

The non-invasive pH measurement of beef muscle would be a valuable asset to meat producers and the application of VIS–NIR spectroscopy has been studied (Andres et al., 2008; Reis and Rosenvold, 2014). The pH of beef, *longissimus thoracis*, at 24 h postmortem (pH$_{24hours}$) has been successfully measured at day 1 with $R^2 = 0.97$ (Andres et al., 2008). Reis and Rosenvold (2014) were less successful in predicting ultimate pH (pHu) from beef samples scanned at 20 to 40 min postmortem. In the case of the later, partial least squares regression techniques could only predict pHu with $R^2$ between 0.20 and 0.36. However, further statistical analysis of the same data provided a prediction model that could segregate carcases as normal (pHu < 5.8) and high (pHu > 5.8) with 90% of the high pHu carcases correctly classified.

The chemical composition of beef plays an important role in its eating quality (Su et al., 2014). VIS–NIR spectroscopy has been used to predict beef composition with varying degrees of success. Su et al. (2014) have developed a number of partial least squares regression models from NIR spectroscopic data (1000 to 1800 nm) obtained from minced beef samples. These models predict fat, protein and moisture content with $R^2$ values for calibration and validation data sets in excess of 0.98. This finding is supported by Prieto et al. (2011) who have used VIS–NIR spectroscopy (350 to 1800 nm) to predict the IMF content in *m. longissimus thoracis* and *longissimus lumborum* samples from Limousin and Aberdeen Angus cattle. In this case, IMF was predicted with a cross-validation $R^2$ value of 0.75.

VIS–NIR spectroscopy has been extensively used to both measures and predict the fatty acid content of beef. Sierra et al. (2008) applied NIR spectroscopy (850 to 1050 nm) to measure individual and groups of fatty acids present in ground beef samples taken from the *longissimus thoracis* of weaning bulls. These authors report that prominent fatty acids such as C14:0, C16:0, C16:1 cis9, C17:0, C18:1 cis9 and C18: cis11 were measured with cross-validation $R^2$ values greater than 0.76. Saturated, branched and monounsaturated fatty acids were measured with cross-validation $R^2$ values of 0.84, 0.70 and 0.85, respectively. Similar results for the fatty acid content of beef (*m. longissimus thoracis* and *longissimus lumborum*) have been documented by Prieto et al. (2011) using VIS–NIR spectroscopic measurements (350 to 1800 nm). Saturated, monounsaturated and polyunsaturated fatty acids were measured with cross-validation $R^2$ values of 0.68, 0.75 and 0.64, respectively. Individual fatty acids (C16:0, C16:1, C18:0, trans 11 C18:1, C18:2 n-6, C20:1, and cis9 trans 11 C18:2) were measured with cross-validation $R^2$ values between 0.69 and 0.90. However, poorer cross-validation $R^2$ values ranging between 0.12 and 0.62 were observed for fatty acids, C14:0, C18:3 n-3, C20:4 n-6, C20:5 n-3, C22:6 n-3, n-6 and n-3. The same authors have also investigated the potential for VIS–NIR spectroscopy (400 to 2498 nm) to measure concentrations of polyunsaturated fatty acids and their bio-hydrogenation products in the subcutaneous fat of beef cows that were fed flaxseed (Prieto et al., 2012). They suggested that the n-3 fatty acids were measured with $R^2$ (coefficient of determination) values ranging between 0.81 and 0.86 and that conjugated linolenic, linoleic and trans-monounsaturated fatty acids were measured with respective $R^2$ values of 0.85, 0.90 and 0.84 to 0.90. Both papers reported difficulties in determining (prediction and measurement) individual polyunsaturated fatty acids. Sierra et al. (2008) suggests that the failure to accurately determine individual fatty acids could be related to similarities in the NIR absorption spectra of all fatty acids. Prieto et al. (2012) proposed that high levels of unsaturation in some fatty acids reduce the number of C–H bonds.
bonds and therefore make a determination by NIR more difficult. However, the same reasoning could also explain why some of the other fatty acid groups are determined with high $R^2$ values. It is unclear from these reports whether individual fatty acids are actually being quantified or whether chemically related groups of compounds (e.g. saturated, n-3, n-6 fatty acids) are measured with strong correlation within these groups.

The accurate online prediction of beef tenderness continues to elude researchers and processors. Recent research indicates that VIS–NIR spectroscopy techniques only moderately predict or measure sensory tenderness, with slightly better results for instrumental tenderness. Venel et al. (2001) have studied the potential for NIR spectra (750 to 1100 nm) to ‘predict’ WBSF of 14-day-aged beef (seminembranosus). Here, $R^2$ prediction values ranged between 0.54 and 0.74 depending on sample segregation. However, a ‘prediction’ as defined in this review is a forecast of a future result. In this particular case, NIR spectra of steak samples were collected on the same day as the WBSF assessment thus indicating measurement rather than prediction. Park et al. (2001) recorded improved $R^2$ prediction values between 0.61 and 0.69 depending on predictive model and the spectral range studied. The sensory and instrumental tenderness of aged beef (2 to 21 days) has been measured with $R^2$ values varying between 0.22 and 0.72 (Liu et al., 2003). The same authors quote a moderately poor $R^2$ value of 0.58 for sensory chewiness, but they do discuss prediction models capable of correctly classifying 83% to 96% of beef samples into ‘tender’ and ‘tough’ categories, on the basis of WBSF. Although the authors claim to predict WBSF at various days of ageing, by the definitions used in this paper, this is again a ‘measurement’ and not a ‘prediction’, as NIR spectra were recorded on the aged beef samples and not on samples before ageing. Similar $R^2$ values (0.65) for measurement of WBSF have been reported by Andres et al. (2008), although $R^2$ values for measurements of sarcomere length (0.16) and cooking loss (0.20), indicators of meat tenderness, were much lower than that for WBSF measurement. Prieto et al. (2009) have studied the shear force prediction of beef, aged 3 to 14 days, using both Volodkevitch and slice shear force methods. However, $R^2$ values of shear force prediction for these methods were much lower than previously seen for WBSF. Cooking loss and sensory tenderness were only predicted with $R^2$ values of 0.35 and 0.28, respectively.

VIS–NIR spectroscopy has also been used to determine other sensory attributes, but predictions are poor. Authors explain these low measurements and predictions are due to a lack of variation in sensory scores for specific attributes. The juiciness of cooked beef has been determined with $R^2$ values of 0.50 (Liu et al., 2003) and 0.21 (Prieto et al., 2009) while flavour, abnormal flavour and overall liking have been predicted with $R^2$ values of 0.59, 0.22, 0.25 (Prieto et al., 2009).

**Hyperspectral imaging.** Hyperspectral imaging is a relatively new spectroscopic technique that provides spectral information on a pixel scale in both the visible and short wave IR regions. A recent review (Xiong et al., 2017) and the references in this section (and Supplementary Tables S1 and S2) highlight the increasing popularity of this technique to determine aspects of beef eating quality such as colour, pH, composition and tenderness.

In general, hyperspectral imaging predicts beef colour very well. Wu et al. (2010) have shown how stepwise regression analysis techniques can be used to identify key wavelengths that predict Hunter $L$, $a^*$ and $b^*$ values. Multi-linear regression methods were then used to predict $L$, $a^*$ and $b^*$ values with $R^2$ cross-validation values of 0.92, 0.90 and 0.88, respectively. These findings have been subsequently improved by the same research group who observed $R^2$ cross-validation values of 0.96, 0.96 and 0.97 for prediction of $L$, $a^*$ and $b^*$ values. ElMasry et al. (2012) have also investigated the potential of hyperspectral imaging (900 to 1700 nm) to measure beef colour. However, $R^2$ cross-validation values for $L$ (0.88) and $b^*$ (0.81) were lower than values reported by Wu et al. (2010). Furthermore, the measurement of $a^*$ values was not reported. These lower levels of colour prediction may be related to the scanning range (short wave IR region) used by ElMasry et al.

Prediction and measurement of pH have not been as successful as those for colour with $R^2$ cross-validation values of 0.86 and 0.73 (Wu et al., 2010; ElMasry et al., 2012). In both cases, partial least squares regression techniques identified specific wavelengths from which multi-linear regression tools were used to create pH prediction/measurement models. It is interesting to note that discriminant analysis methods have not yet been applied to hyperspectral data in order to predict high and low pH categories, as was applied successfully to pH predictions using NIR data (Reis and Rosenvold, 2014).

Partial least square regression analysis of hyperspectral imaging data (900 to 1700 nm) has been used to measure water, fat and protein content of beef with $R^2$ values of 0.89, 0.84 and 0.86, respectively (ElMasry et al., 2013). Total fat and fatty acid contents of Japanese Wagyu beef have been determined using hyperspectral imaging (1000 to 2300 nm) and partial least squares regression techniques with varying degrees of success (Kobayashi et al., 2010). Total fat, saturated fatty acids and total unsaturated fatty acids were predicted with $R^2$ values of 0.90, 0.87 and 0.89. Predictions for individual fatty acids ranged from $R^2 = 0.68$ to 0.89 depending on the fatty acid. However, as previously noted with VIS–NIR studies, this work does not report whether individual fatty acids are predicted due to cross-correlation with the main fatty acid groups, saturated, monounsaturated and polyunsaturated fatty acids.

In recent years several authors have documented the prediction and measurement of beef tenderness from the statistical analysis of hyperspectral data. WBSF values of 7-day-aged beef have been predicted from hyperspectral images acquired at 2 days postmortem with an $R^2$ cross-validation value of 0.86 (Wu et al., 2010). Drip loss, a contributing factor to beef tenderness, has been measured
from hyperspectral images of beef (taken on same day as reference measurement) with $R^2$ values ranging between 0.87 and 0.89 (ElMasry et al., 2011). The prediction of beef tenderness improves when discriminant analysis techniques are used to categorise beef samples into groups. Naganathan et al. (2008) have reported a hyperspectral imaging (400 to 1000 nm) model capable of discriminating, but not predicting, 14-day-aged rib eye beef steaks into one of three tenderness categories with 96% accuracy. These results are supported by Wu et al. (2010) who predicts correct classification of 91% of tender beef samples applying similar techniques to hyperspectral and WBSF data. Cluff et al. (2013) also showed that linear discriminant models could correctly classify 83% and 75% of tough and tender (on slice shear force) samples of longissimus dorsi based on hyperspectral data recorded between 922 and 1739 nm before sample cooking. Recently, Naganathan et al. (2015) have tested a prototype online hyperspectral instrument (450 to 950 nm) for tender prediction of beef. Here, beef samples were scanned at 2 days postmortem after which they were aged to 14 days and evaluated for slice shear force. Results show that 93% of tender (low shear force) beef samples were correctly classified and that models were also cross-validated with 88% of validation samples correctly classified. Further field testing of this system using Fisher’s linear discriminant, support vector machine and decision tree analyses predict slice shear force of 14-day-aged rib eye beef with test certification accuracy of 87%.

This review highlights the volume of research that has been carried out to assess the potential of hyperspectral imaging to predict instrumental measures of beef-eating quality. However, during this review, we did not find any examples where hyperspectral imaging has been used to accurately predict the sensory qualities of beef.

**Raman spectroscopy.** The potential to predict sensory eating qualities of beef silverside using Raman spectroscopy has been investigated by Beattie et al. (2004). Partial least

---

Table 1  Summary of grading systems currently in use to predict beef-quality/eating quality (Farmer et al., 2010a; Polkinghorne and Thompson, 2010; Bonny et al., 2017)

<table>
<thead>
<tr>
<th>Grading scheme</th>
<th>Country</th>
<th>Grading unit</th>
<th>Number of grades*</th>
<th>Basis of grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>USDA</td>
<td>USA</td>
<td>Carcase</td>
<td>8 Quality grades</td>
<td>Sex; carcase weight; marbling; ossification; meat colour, texture; eye muscle area; ribfat; kidney and perirenal fat</td>
</tr>
<tr>
<td>Canada</td>
<td>Canada</td>
<td>Carcase</td>
<td>5 Quality grades</td>
<td>Sex; marbling; meat colour; texture; fat colour, thickness</td>
</tr>
<tr>
<td>EUROP</td>
<td>Europe</td>
<td>Carcase</td>
<td>5 Classification</td>
<td>Sex; marbling; meat colour; texture; fat colour, thickness</td>
</tr>
<tr>
<td>JMGA</td>
<td>Japan</td>
<td>Carcase</td>
<td>5 Quality grades</td>
<td>Sex; marbling; meat colour; brightness, texture; fat colour, lustre, texture, firmness, thickness; eye muscle area, rib thickness</td>
</tr>
<tr>
<td>Korea</td>
<td>South Korea</td>
<td>Carcase</td>
<td>5 Quality grades</td>
<td>Sex; marbling; meat colour; texture; fat colour, thickness; lean maturity; eye muscle area, rib thickness</td>
</tr>
<tr>
<td>South Africa</td>
<td>South Africa</td>
<td>Carcase</td>
<td>3 Classification</td>
<td>Sex; marbling; meat colour; texture; fat colour, thickness; lean maturity; eye muscle area, rib thickness</td>
</tr>
<tr>
<td>Quality Mark</td>
<td>South Africa</td>
<td>Carcase</td>
<td>5 Quality grades</td>
<td>Country of origin; age; handling; absence of growth promoters; licensed plant; ultimate pH</td>
</tr>
<tr>
<td>MLC Blueprints (+ updates)</td>
<td>United Kingdom</td>
<td>Carcase</td>
<td>Pass/fail quality grades</td>
<td>Age/sex; growth rate; diet; EUROP grade/fat class; transport and lairage handling; slaughter techniques, defects; hanging; electrical stimulation, chilling and pH/T decline; maturation</td>
</tr>
<tr>
<td>Red Tractor and Quality Standard Marks</td>
<td>United Kingdom</td>
<td>Carcase</td>
<td>Pass/fail quality grades</td>
<td>Age/sex; EUROP grade/fat class; maturation</td>
</tr>
<tr>
<td>AUS-MEAT</td>
<td>Australia</td>
<td>Carcase</td>
<td>Classification grades</td>
<td>Diet; marbling; meat colour; fat colour; shape; texture; ribfat; damage</td>
</tr>
<tr>
<td>MSA</td>
<td>Australia</td>
<td>Cut</td>
<td>3 Quality grades</td>
<td>Bos indicus %; hormonal growth promoter implants; carcase weight; sex; hump height; electrical stimulation; hang; marbling; ossification; meat colour; pHu; ageing time; cooking method</td>
</tr>
</tbody>
</table>

USDA = United States Department of Agriculture; JMGA = Japanese Meat Grading Association; MLC = Meat and Livestock Commission (now Agriculture and Horticulture Development Board); MSA = Meat Standards Australia; pHu = ultimate pH.

*Classification grades are descriptive terms for the carcass to aid trading while quality grades aim to place a value on the carcass on the basis of its perceived quality. Grades may also indicate yield (Polkinghorne and Thompson, 2010) but this aspect is not discussed in this paper.
Grading for beef-eating quality

A number of grading schemes have been applied to eating quality prediction, including those developed by national industry representative bodies in the United Kingdom, Europe, United States, Canada, New Zealand, Japan and South Korea (AHDB Industry Consulting, 2008; Polkinghorne and Thompson, 2010). These are all based upon the grading of carcases. In contrast, the national industry organisation in Australia, Meat and Livestock Australia, has developed a cuts-based grading scheme known as Meat Standards Australia (MSA). Table 1 lists the main grading systems developed by beef organisations across the world. Systems developed by supermarkets and other commercial organisations are generally confidential and are not included.

The United States Department of Agriculture (USDA) grading system dates back to the 1920s and was not originally designed to indicate eating quality (Polkinghorne and Thompson, 2010), but is generally assumed to do so. This relationship appears to be marked only for extremes. Consumers did not differentiate between beef of intermediate quality USDA grades, but sirloins (m. longissimus lumborum) from higher quality grades were better than lower quality grades (Tedford et al., 2014). These authors observed the same trend for beef graded with the Canadian system. O’Quinn et al. (2015) reported that American consumers were unable to detect differences in eating quality among tenderloin steaks from USDA and select grades. Thus large differences in grade give eating quality differences, but smaller differences do not. It has been shown that USDA maturity bands do not influence eating quality in grain-finished cattle up to 30 months and that only marbling is important (Acheson et al., 2014). Mateescu et al. (2016) reported that WBSF, IMF, hot carcase weight and marbling score predicted eating quality of sirloin better than did the USDA grade. Likewise, the Canadian quality grades, A, AA, AAA did not differentiate beef on cooked tenderness, as measured by shear force (Puente et al., 2016).

The application of the Australian MSA system to European beef has been reviewed by Bonny et al. (2017). The model correctly classifies 50% to 70% of the samples with 95% to 97% of the predicted scores being within one grade of their consumer scores (Thompson, 2002). The MSA system has been widely tested and validated across Europe since 2003 and has been shown to be applicable to European beef and consumers (Farmer et al., 2009a and 2009b; Legrand et al., 2013; Guzek et al., 2015; Bonny et al., 2017). Studies have also been conducted in South Korea, South Africa, Japan and United States as well as Australia (Polkinghorne, 2007; Polkinghorne et al., 2011). Meat Standards Australia has been shown to be effective in all these countries with some differences due to different expectations and cooking methods. Studies in Northern Ireland showed that MSA could be adapted to improve further the prediction (Farmer et al., 2010b). As all the consumer panels were conducted using the MSA protocol, it has been possible to combine the consumer data from Northern Ireland, Ireland, France and Poland to allow further analysis. This work has been reviewed by Bonny et al. (2018).

Few authors have compared quality assurance systems. Smith et al. (2008) concluded that both the USDA and MSA system deliver palatability prediction for consumers in the United States and Australia, respectively. Tedford et al. (2014) reported that the USDA and Canadian grading systems only partially predicted eating quality. Farmer et al. (2010a) reported that MSA and the UK Blueprint system appeared better than USDA and the New Zealand system at assuring eating quality for Northern Ireland’s beef, but that, due to the cuts-based system, MSA gave a lower proportion of non-qualifying beef.

Most eating quality systems generally focus primarily on tenderness. However, the flavour is an important factor in acceptability and has been shown to contribute highly to acceptability (Oliver et al., 2006; Polkinghorne, 2007). It is known that consumers can differ in their liking for specific flavours; for instance, American consumers prefer beef from grain-fed animals and rate Australian grass-fed beef lower than Australian consumers (Polkinghorne, 2007). There is also some evidence for flavour differences between muscles (Kukowski et al., 2004; Meisinger et al., 2006). Some muscles from carcases receiving higher USDA grades achieved higher consumer scores for flavour (Legako et al., 2015; Hunt et al., 2016) and that this was associated with changes in volatile flavour compounds (Legako et al., 2015). Recent research has demonstrated clear flavour differences between muscles and packaging treatments which may be explained by the composition of volatile aroma compounds from the cooked beef (Farmer et al., unpublished data). Many of the compounds responsible for flavour are difficult to measure routinely, due to their low concentrations, but marker compounds have been identified that are associated with flavour liking (Farmer et al., 2013). It may be possible to identify specific flavour notes which can enhance liking for some groups of consumers.

Way forward for the European beef industry?

The European beef industry is conscious of pressures to deliver a safe, environmentally sustainable product with high nutritional value, which also meets the gastronomic
expectations of the consumer. Individual companies are investing in numerous initiatives to deliver the expected quality and consistency to their retail customers and the final consumer, often working in partnership with beef scientists. Two international workshops between scientists and industry have been convened to discuss these issues and have been reported (Farmer et al., 2016; Farmer et al., 2017). Workshops seeking industry views identified the need for reliable and robust techniques to monitor eating quality and reduce inconsistency as key for the future management of beef quality (Farmer et al., 2016).

The MSA system has been shown to be widely effective in improving eating quality assurance in many countries. Nevertheless, the European beef industry has been slow to implement this system. Some of the reasons are highlighted in a report on the perceptions of the French beef industry (Hocquette et al., 2011), where the system is seen as difficult to implement given the complexity of the beef industry and market. Similar responses have been received anecdotally from the industry in Northern Ireland and Poland. However, uptake of MSA or related models is continuing, with the development of MSA-type systems in New Zealand and Poland.

It has been proposed that a global version of MSA could be evolved which would combine the advances made in Australia with developments in genetic and other markers for quality (Hocquette et al., 2014). Such a global eating quality model could also incorporate indicators for flavour characteristics, nutritional quality, environmental considerations and economic efficiency for the benefit of not only the consumer but the entire supply chain (Hocquette et al., 2014). Rapid developments in instrumental technologies raise the possibility that this global model could incorporate these instrumental methods when they become sufficiently effective and robust that they can be used on the slaughter or processing line.

Over the next few years, the European beef industry will decide how it plans to meet the challenge of delivering high-eating quality to an increasingly perceptive consumer base. It is not yet clear whether the current systems of retailer and company ‘specs’ will continue or whether new systems will develop which could comprise instrumental monitoring and prediction, a version of MSA, or a new global-eating quality assurance method that incorporates new elements. Whatever system or systems are used they will need to be:

- Profitable – to be commercially viable.
- Simple – at the point of operation.
- Effective – proven to deliver better-eating quality to consumers.
- Flexible – to allow evolution, development and support of existing and new brands.

At the moment, the best chance of predicting eating quality seems likely to be realised by combining the benefits of an eating quality grading system like MSA with the additional measurements provided by an advanced spectroscopic technique such as hyperspectral imaging. What is certain is that the rapid pace of technological development both in terms of monitoring of composition and markers for quality, and the identification of genetic markers, means that there is considerable potential for the European beef supply chain to find new ways to maintain and enhance the quality of its end product.

Acknowledgements
The authors acknowledge funding for research projects and knowledge exchange workshops from Department of Agriculture, Environment and Rural Affairs (Northern Ireland), InnovateUK, InvestNI, Meat and Livestock Australia, UK Science and Innovation Network and individual meat companies.

Declaration of Interest
The authors declare that they have no conflict of interest.

Ethics statement
This paper is a review of the existing published literature and, therefore, no approval by an ethics committee or compliance with national legislation is required.

Software and data repository resources
This paper is a review of the existing published literature and the data reported is already available from the journals cited.

Supplementary material
To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731118001672

References


