finalreport

SUPPLY CHAIN MANAGEMENT

Project code:	**SCSB.086**
Prepared by:	**Dr Chris Oldham**
	The Department of Agriculture of Western Australia
Date published:	**September 2005**
ISBN:	**1 7403 69157**

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

The comparison of predictive methods for the determination of fatness and condition in sheep

In submitting this report, you agree that Meat & Livestock Australia Limited may publish the report in whole or in part as it considers appropriate.

This publication is published by Meat & Livestock Australia Limited ABN 39 081 678 364 (MLA). Care is taken to ensure the accuracy of information in the publication. Reproduction in whole or in part of this publication is prohibited without the prior written consent of MLA.

Abstract

In 2004-5 during the MLA Prime Time sheep forums, it became apparent that both condition score (estimated by palpating the short ribs) and fat score (estimated by palpating the long ribs) were being recommended and that producers and researchers were confused about the terminology and methodologies of both techniques discouraging the application of either technique. Hence, in an experiment commissioned by MLA the condition score, fat score and C-site ultrasound fat and C-site ultrasound eye muscle depth were recorded for a group of 92 live Merino sheep in 6 months wool (45 young and 47 mature). All scores and measurements were done by two to four experienced assessors and repeated three times (runs) in a random order. Approximately 24 hr after the above measurements, the sheep were slaughtered at an abattoir where the tissue depth at the GR site over the 12th rib (fat + muscle) was recorded on the hot carcass for 89 of these sheep. This measurement was also done by two experienced assessors using standard GR knives and repeated three times in random order.

The results showed that:

- All condition and fat score assessors showed very high repeatability.
- The condition score assessors scored similarly with only very small between assessor biases.
- By contrast, there were considerable between assessor biases for the fat scorers.
- Rather than score estimated GR tissue depth in whole numbers 1 to 5 our experts chose to estimate GR tissue depth in mm.
- There is a strong significant relationship between both condition and fat scores and the more objective measurements at the C and GR sites. Both methods could be used to predict actual GR tissue depth to within 2 mm.
- However, while the mean GR tissue depth at Fat Scores 1, 2 and 3 were similar to the estimates given in the original scale the variance about the means was greater than suggested in the original scale causing ‘overlap’ of the original GR tissue depth ranges.
- There is a strong significant relationship between condition score and fat score. However, in the range condition score of 2 to 3 that is considered to be critical for the management of ewes in the Lifetime Wool project there was very little variation in estimated or actual GR tissue depth.
- There was a small but significant effect of Age on some of the above relationships, including the relationship between fat and condition score.

It was concluded that while a skilled operator could accurately estimate the total tissue depth (mm of fat + muscle) at the GR site over the 12th rib, the variance about the fat scores and their related 5 mm ranges in GR tissue depth, caused significant overlap in the liveweight x fat score grids currently used by the industry for setting price. This issue could be addressed by replacing the current scoring system with training in assessing GR tissue depth in mm. Timely and accurate abattoir feedback on actual GR tissue depth would be an essential part of ongoing calibration of all assessors.

In this experiment the repeatability of all skilled operators was commercially acceptable. However, while the relationship between condition score and estimated GR tissue depth explained 88% of the variance, the relationship was not linear. In fact, in the condition score range of 2 to 3, considered to be critical for the management of ewes in the Lifetime Wool project, there was very little variation in estimated or actual GR tissue depth. Therefore in future recommendations for management of ewe flocks it is our conclusion that targets framed in terms of condition score would offer far greater control over predicted effects on the performance of the ewes and future performance of their progeny than targets framed in terms of estimated GR tissue depth.

Executive Summary

Condition scoring by manual palpation of fat and muscle over the loin (short ribs) has been used to estimate the 'energy status' or 'nutritional well-being' of adult ewes. Alternatively, fat scoring by manual palpation at the midside over the long ribs has been used to estimate the yield of saleable meat (%) of young sheep being marketed for meat; whole scores 1 to 5 are related to 5 mm ranges in total tissue (fat + muscle) at the GR site over the 12th rib. MLA, which continues to advocate the use of fat score for prime lamb marketing and has also endorsed fat score as a method for ewe management, was concerned that the increased use of both terminologies is confusing producers and may actually be discouraging the application of either technique. Hence, MLA commissioned an experiment to test:

1. That skilled industry personal can repeatability assess condition score or fat scores
2. That the relationship between condition score and fat score is such that industry can confidently convert flock estimates between the two methods.
3. That the relationship will be similar in hogget and adult sheep
4. That methods of subjective assessment bear a statistically significant relationship with measured ultrasonic C site fat and eye muscle depth and carcase GR site tissue depth (as measured on hot carcase by a GR knife).

In the experiment the condition score, fat score and C-site ultrasound fat and C-site ultrasound eye muscle depth were recorded for a group of 92 live Merino sheep in 6 months wool (45 young and 47 mature). All scores and measurements were done by two to four experienced assessors and repeated three times (runs) in a random order. Approximately 24 hr after the above measurements, the sheep were slaughtered at an abattoir (WAMMCO, Katanning) where the GR site tissue depth (fat + muscle) was recorded on the hot carcass for 89 of these sheep. This measurement was also done by two experienced assessors using standard GR knives and repeated three times in random order.

The results showed that;

- All condition and fat score assessors showed very high repeatability (av. correlation between runs = 0.90 to 0.95).
- The condition score assessors scored similarly with only very small assessor biases.
- By contrast, there was considerable assessor bias for the fat scorers.
    - Rather than score estimated GR tissue depth as 'fat scores' 1 to 5 our experts chose to estimate GR tissue depth in mm.
    - Fat score assessor number 2 was the only assessor to accurately estimate GR tissue depth.
    - The other three fat score assessors over estimated the tissue depth by an average 3 to 6 mm.
- There was a strong significant relationship between both condition and fat scores and the more objective measurements at the C and GR sites.
    - The error was smaller for the low condition score sheep (≤ CS 2.5) maintained in their original paddock in the 3 weeks leading to the experiment (thinner) than for the higher

condition score sheep (≥ CS 2.5) fed a supplement in a paddock with more green feed in an attempt to gain condition in the 3 weeks leading to the experiment (fatter).

  - Prediction of the average GR tissue depth from a single condition score or fat score could be done with average error of about 0.6mm for the thinner sheep and 2mm for the fatter sheep.
  - The original method of fat scoring relates whole scores 1 to 5 to measured GR tissue depth in 5 mm increments (eg, FS 1 = 0-5 mm, FS 2 = 6-10mm etc). However, using the most accurate FS assessor from the current experiment or the WAMMCO commercial FS assessor the measured GR tissue depth ranges for carcasses with FS = 2 or 3 were much wider and overlapping (FS 2 = 3 to 12 mm and FS 3 = 5 to 16 mm).
  - Nonetheless, the WAMMCO assessor correctly assigned fat scores to 68/80 carcasses and a further 7/12 carcasses were within 1 mm of a category boundary and the nearby boundary was selected (ie. A 9.5 mm sheep was assigned FS = 3, rather than FS = 2).

- There was a strong significant relationship between condition score and fat score.
  - Prediction of the average condition score from a single fat score could be done with average error of about 0.25 CS units and 95% confidence to within three quarters of a condition score unit.
  - However, because the relationship was not linear, in the range CS of 2 to 3 that is considered to be critical for the management of ewes in the Lifetime Wool project there was very little variation in estimated or actual GR tissue depth (1 to 4 mm and 1 to 3 mm, respectively).
- There was a significant effect of age on some of the above relationships, including the relationship between fat and condition score.

It was concluded that while a skilled operator could accurately estimate the total tissue depth (mm of fat + muscle) at the GR site over the 12th rib, the variance about the fat scores and their related 5 mm ranges in GR tissue depth, caused significant overlap in the liveweight x fat score grids currently used by the industry for setting price. This issue could be addressed by replacing the current scoring system with training in assessing GR tissue depth in mm. Timely and accurate abattoir feedback on actual GR tissue depth would be an essential part of the training and ongoing calibration of all assessors.

The high repeatability of all GR tissue depth assessors but consistent and significant bias shown by 3 of the 4 assessors also suggests that the accuracy between assessors could be readily improved if a series of physical ‘benchmark’ models were developed to aid/guide the estimation of GR tissue depth in live sheep in much the same way as the Lifetime Wool project has developed physical ‘benchmark’ models for condition scoring.

In this experiment the repeatability of all skilled operators (both subjective and objective) was very high and commercially acceptable. However, while the relationship between condition score and estimated GR tissue depth explained 88% of the variance, the relationship was not linear. In fact, in the condition score range of 2 to 3, considered to be critical for the management of ewes in the

Lifetime Wool project, there was very little variation in estimated or actual GR tissue depth (about 1 to 4 mm in both cases). Therefore in future recommendations for management of ewe flocks it is our conclusion that targets framed in terms of condition score would offer far greater control over predicted effects on the performance of the ewes and future performance of their progeny than targets framed in terms of estimated GR tissue depth.

The curvilinear relationship between condition score and estimated GR tissue depth maybe explained by the fact that estimation of condition score includes changes in eye muscle as well as tissue cover (fat & muscle) over the spine and short ribs. In this experiment eye muscle depth increases linearly with increasing condition whereas it there is a clear lag of at least 5 mm in eyemuscle depth before GR tissue depth increases.

There were a number of small differences in the various relationships between mature and young ewes. For example the estimated condition score of adult ewes was about 0.25 of a score greater than young ewes at the same estimated GR tissue depth (mm) over the critical range of 3 to 12 mm. However, none of these small differences would have any great impact on how the industry is currently or contemplates using condition score or estimated GR tissue depth in the future.

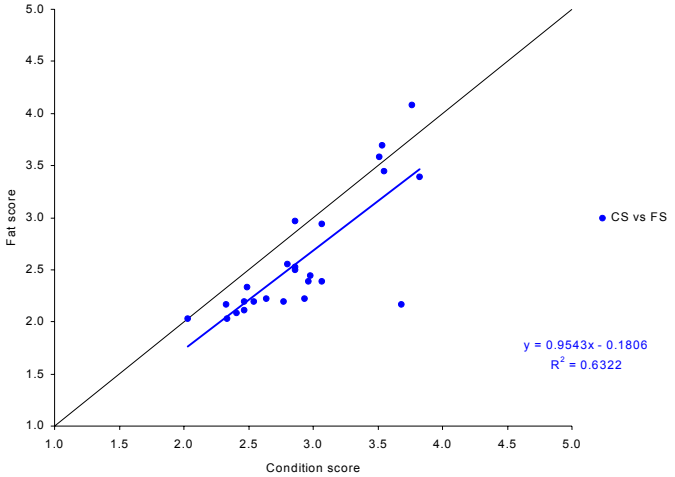
Contents

1 Background

Condition score (CS) relates to the tissue cover (fat + muscle) as manually palpated over the loin (short rib) area of sheep. Whereas fat score (FS) relates to the tissue cover (fat + muscle) as manually palpated over the 12th rib (1st long rib from the short loin) at the GR site, approximately 110 mm from the vertical processes of the spine.

Traditionally, CS has been used to estimate the 'energy status' or 'nutritional well-being' of adult ewes (Russell et al. 1969, J. agric Sci. Camb.72, 451-454; Feeding standards for Australian livestock, Ruminants p 58-68). While FS has been used to estimate the yield of saleable meat (%) of young sheep being marketed for meat; whole scores are related to 5 mm ranges in total tissue (fat + muscle) at the GR site over the 12th rib (see Fat Score on NSW DPI website).

The Lifetime Wool project (LTW; AWI-EC298) is a National project writing new nutritional guidelines for ewe flocks that will need to be framed in terms of 'fatness' and/or liveweight (LW) targets. LTW staff have variously been using CS (Vic, WA & SA) or FS (NSW) to manage flocks. Within LTW it has become essential to resolve the rigorous internal debate between the condition scorers and the fat scorers. Similarly during the MLA Prime Time Roadshow last year it became very clear that both academics and producers were confused with respect to the pros and cons of the 2 approaches to estimating fatness. Recently, LTW conducted a comparison that strongly suggested that CS & FS are equally repeatable and measuring the same thing (relative fatness) in the hands of trained assessors (see figure below; mean of 3 replicates on 25 sheep by 3 trained assessors of either CS or FS).



Hence, Ian Ross (MLA) convened a meeting in Sydney in November 2004 to discuss the issues. As a result of the discussions Chris Oldham (DAWA and Lifetime Wool) was asked to design an experiment that would resolve the issues and for the first time calibrate both CS & FS against a continuous scale of tissue depth measured at the GR site in both mature and young Merino ewes.

2 Project Objectives

1. To establish the relationship between CS and FS in both adult and young Merino ewes as assesses by skilled industry recognised assessors.
2. To establish the relationship between subjectively assessed CS, FS and objectively measured C site fat and eye muscle depth in the live animal and measured GR site tissue depth in the hot carcass.

3 Methodology

Selection and feeding of the experimental sheep - On 3rd August 2005, three weeks prior to the experiment approximately 100 sheep (50 mature aged ewes and 50 young ewes approximately 12 months old) were selected to equally represent the range of condition scores (CS) present in their flock (CS assessor 4 in Table 1; range 1.0 to 4.0; thin to fat). Thereafter, the thinner half of the flock (both ages combined; average CS = 1.8; range 1 to 2.5) continued to be run on a green pasture with approximately 700 kg DM on offer. In order to at least maintain their superior condition and if possible increase the condition of the fatter ewes to create a group at condition score 4.5 or greater by the experiment on 23rd August, the fatter half of the flock (average CS = 3.2; range 2.5 to 4) were given access to Macco Feeds 707 lamb finisher pellet (John Milton pers. com.) in a self feeder and pasture of approximately 1000 kg DM on offer. However, the average change in condition score was similar (thin flock CS 1.8 to 2.0 and fat flock CS 3.2 to 3.5) in both flocks and no sheep were classified as CS 4.5 by 23rd August. The distribution of CS in the two age groups of ewes was very similar; CS ≥1≤2 = 9 ewes; CS ≥1≤2 = 14/17 ewes for young v mature; CS ≥2≤3 = 14 ewes; CS ≥4 = 8 ewes.

Measurements– In this experiment the condition score ('CS'), fat score ('FS') and C-site ultrasound fat ('C-site fat') and C-site ultrasound eye muscle depth ('C-site EM') were recorded for a group of 92 live Merino sheep in 6 months wool (45 young and 47 mature). All scores and measurements were done by two to four experienced assessors and repeated three times (runs) in a random order.

Table 1. Details of experienced assessors used in the experiment

Measure	Assessor	Details	Address	Expertise
Condition Score (subjective 1=thin & 5=fat)	CS-1	Katrina Copping	SARDI, Straun S Aust	TO Lifetime wool project
	CS-2	Barbra Sage	DAWA, Northam, WA	TO Lifetime wool project
	CS-3	Andrew Kennedy	Vic DPI, Hamilton, Vic	RO Vic DPI
	CS-4	Tom Plaisted	DAWA, Albany, WA	TO Lifetime wool project
Fat Score (subjective mm of tissue depth at the GR site; 1=thin to 20=fat)	FS-1	John Sullivan	Cowra, NSW	Stock agent
	FS-2	Geoff Duddy	NSW DPI, Yanco, NSW	DLO NSW DPI
	FS-3	Rob Davidson	WAMMCO, Perth, WA	Supply Development Manager WAMMCO
	FS-4	Bob Marchant	NSW DPI, Armidale, NSW	DLO NSW DPI
C-site Fat depth (Objective Real-time ultrasound in mm)	C-site F-1	Stephan Spiker	Advanced Livestock Services, Hamilton, Vic	MLA accredited
	C-site F-2	Peter Moore	Scan West, Williams, WA	MLA accredited
C-site Eye muscle depth (Objective Real-time ultrasound in mm)	C-site EM-1	Stephan Spiker	Advanced Livestock Services, Hamilton, Vic	MLA accredited
	C-site EM-2	Peter Moore	Scan West, Williams, WA	MLA accredited
GR tissue depth of carcass (Objective GR-knife in mm)	GR-1	Rob Davidson	WAMMCO, Perth, WA	Manager WAMMCO
	GR-2	Geoff Duddy	NSW DPI, Cowra, NSW	Sen DO NSW DPI

Approximately 24 hr after the above measurements, the sheep were slaughtered at an abattoir (WAMMCO, Katanning) where the GR site tissue depth (fat + muscle) was recorded on the hot carcass for 89 of these sheep. This measurement was also done by two experienced assessors using standard GR knives and repeated three times in random order. Each assessor used either the left or right side of each carcass but the same cut for each of the three runs.

Analysis - Regression analysis and other statistical techniques were used to first examine repeatability of the assessors, then the relationships between the various scores and measurements. Bias between assessors was examined, and the prediction error and 95% confidence intervals were calculated for estimation based on a single condition or fat score.

4 Results and Discussion

While the differential feeding of the 'thin' and 'fat' flocks did not have the desired effect of stretching the range of CS it was associated with a subtle shift of a number of the sheep from CS 3 to CS 3.5 or CS 4. This meant that there were only 5 young sheep and 3 mature sheep in medium condition (around CS=3). This slight distortion in the population was even more pronounced with the GR tissue depth measurements with no sheep having an average carcase GR tissue depth between 3 and 5 mm.
In the analysis described below, the presence of these two groups has been assumed to not have an effect on the statistical relationships, however some of the results are reported separately to show differences between the groups, labelled 'fat' and 'thin'.

Repeatability - Before examining the relationship between the various people or different measurements, this section looks at the similarity of each persons repeated measures on the same animals.
Repeatability *- refers to the agreement or variation between repeated independent measurements by the same person on the same animal.*
In this exercise the repeatability for a given assessor and measurement is assessed by examining the results from the 3 runs, with a highly repeatable assessor often giving the same or very similar result in each run. The below table summarises repeatability using the following measures:
Correlation – the average correlation between two runs; and
Mean Absolute Error (MAE) – the average difference between measurements on two runs.
Table 2 contains the average results from comparing runs 1 and 2, 1 and 3, and 2 and 3.

Table 2. Assessor repeatability

Measure	Assessor	Correlation	Mean Abs. Error		Range[1]	MAE /
			thin	*fat*	*(av. thin – av. fat)*	Range
CS	CS-1	0.95	0.18	0.19	1.5 (2.1 to 3.6)	12%
	CS-2	0.91	0.18	0.15	1.2 (2.3 to 3.6)	14%
	CS-3	0.93	0.36	0.24	1.9 (1.9 to 3.8)	16%
	CS-4	0.93	0.21	0.29	1.7 (2.1 to 3.8)	14%
FS	FS-1	0.95	0.65	1.71	9.4 (3.2 to 12.6)	12%
	FS-2	0.91	0.43	2.43	8.4 (1.5 to 9.9)	16%
	FS-3	0.95	0.95	1.47	9.6 (5.1 to 14.7)	12%
	FS-4	0.90	1.63	1.74	7.8 (6.5 to 14.4)	21%
C-site fat	C-site F-1	0.94	0.22	0.53	2.6 (1.1 to 3.7)	14%
	C-site F-2	0.87	0.28	0.54	1.9 (1.2 to 3.1)	22%
C-site EM	C-site EM-1	0.92	1.32	1.41	7.9 (17.5 to 25.5)	17%
	C-site EM-2	0.95	0.90	1.38	7.3 (15.2 to 22.5)	15%
GR tissue	GR-1	0.99	0.20	0.47	7.0 (1.6 to 8.6)	5%
	GR-2	0.99	0.12	0.44	7.4 (1.2 to 8.7)	4%

[1] *a measure of the Range (spread; average of the thin – average of the fat group of sheep) of the data is given to put the Correlation and Mean Absolute Error in context. An assessor that has a wider range is able to more easily obtain a higher Correlation, but is also more likely to have a higher Mean Absolute Error.*

A few comments on Table 2:

- All assessors were highly repeatable with all but one correlation above 0.9 and with the average difference between repeated measurements very small compared to the range. The repeatability of all the measures was similar, except for GR tissue (see comments below).
- There was a lot lower error in the estimation of C-site fat, FS and GR tissue for the thin animals compared to the fat animals. This however is simply explained by the narrower range of values assigned to the thinner animals for these measurements.
    - The ratio of Mean Absolute Error for thin animals to the range (max-min) of thin animals was similar to the same ratio for fat animals (and also similar to the last column in the above table, which was calculated based on all the animals).
- Assessor C-site F-1 was a little more repeatable with **C-site** fat than assessor C-site F-2, while assessor C-site EM-2 was more repeatable with C-site EM on thin animals than assessor C-site EM-1.
- The average error between repeated **condition scores** was remarkably low at 0.22.
    - For assessor CS-2 and assessor CS-4 who measured in intervals of 0.5, the proportion of repeated scores that were the same was 61%, the proportion that were different by 0.5 was 38% and the remaining 2% differed by 1.
    - Comparing the four assessors by Correlation and Mean Absolute Error gives different conclusions (see footnote below the table for more details). For example assessor CS-2 has the best (lowest) average error but the worst (lowest) correlation. A better comparison is Mean Absolute Error as a proportion of the range giving the following (all commendable) ratios: Assessor CS-1(12.5%), assessor CS-2(13.6%), assessor CS-4(14.4%), assessor CS-3(15.9%).
- The repeatability of **fat scores** by all four assessors was also commendable, with the following ratios of average error as a proportion of the range: Assessor FS-1(12.1%), assessor FS-3(12.3%), assessor FS-2(16.1%) and assessor FS-4(21.4%).
    - Especially in this case it is worth noting that repeatability is not the ultimate measure of what makes a good assessor. As explained later in this report, assessor FS-2 scored very differently to the other three assessors, estimating GR tissue with much lower bias.
- Both assessors of **GR tissue** were exceptionally repeatable with the highest correlation of 0.99 and the smallest average error of only 5% compared to the range. Assessor GR-2 was a little more repeatable than assessor GR-1.
    - However, it should be noted the assessors used the same cut for each of the three runs (each assessor using their chosen side) and thus the runs were not completely independent and the repeatability is overstated.

- The more subjective measures of condition and fat scores were just as repeatable as the more objective measures of C-site fat and eye muscle depth.

Condition Score - The below graphs plot the average condition score assigned to each sheep (average of the three runs) for each pair of assessors. The dotted lines represent perfect agreement.

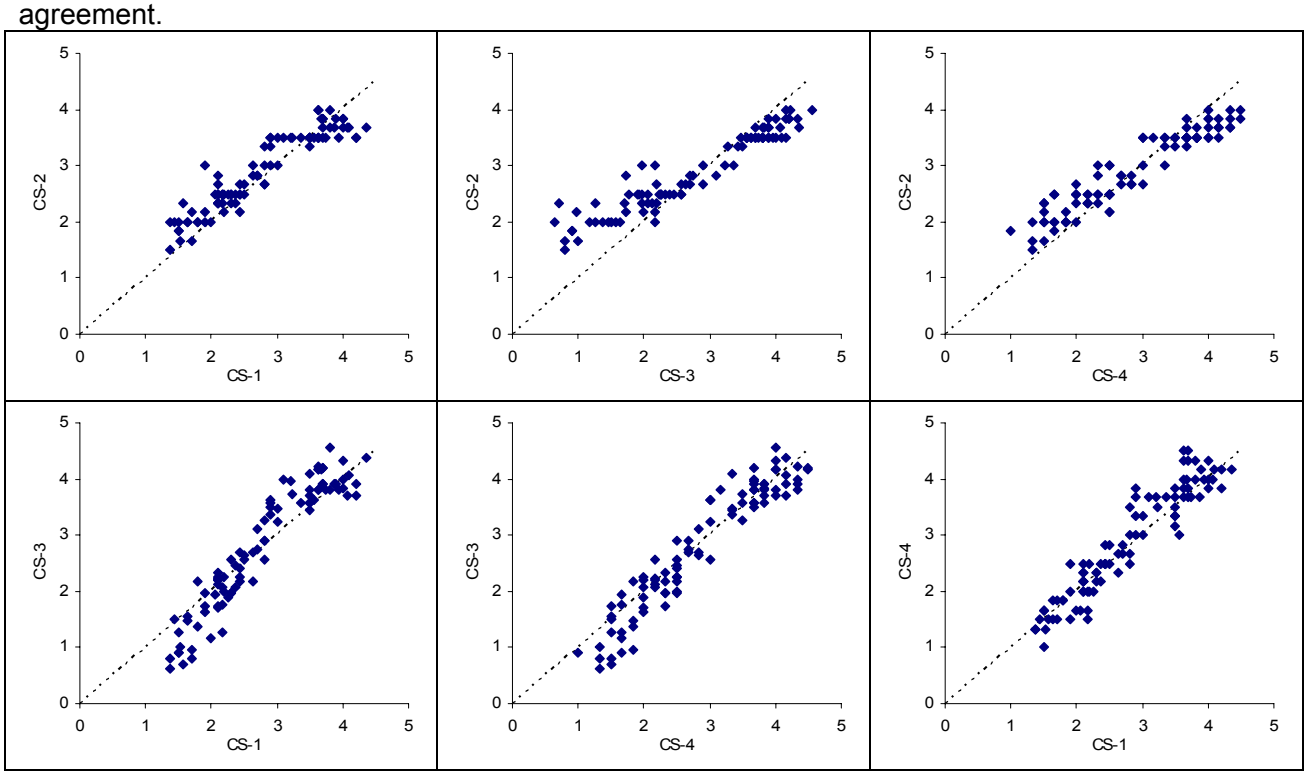


Figure 1. Scatter plots comparing average condition scores given by different assessors

From these graphs we can see that condition score assessors are generally consistent, with most points near the dotted line. The largest difference occurred with the low condition sheep, with Assessor CS-2 tending to give a whole score higher than assessor CS-3, with the other two assessors about mid-way between.
The correlations between each pair of assessors are high as summarised below:

Table 3. Correlations between condition score assessors

	CS-1	*CS-2*	*CS-3*	*CS-4*
CS-1	1			
CS-2	0.94	1		
CS-3	0.95	0.95	1	
CS-4	0.95	0.94	0.95	1

The relationships are combined in the next graph, which plots the average score assigned to the sheep (X axis) against how much each assessor's average differs from this (Y-axis). A best fitting cubic curve is included for each person to show how the bias changed across a range of average condition scores.

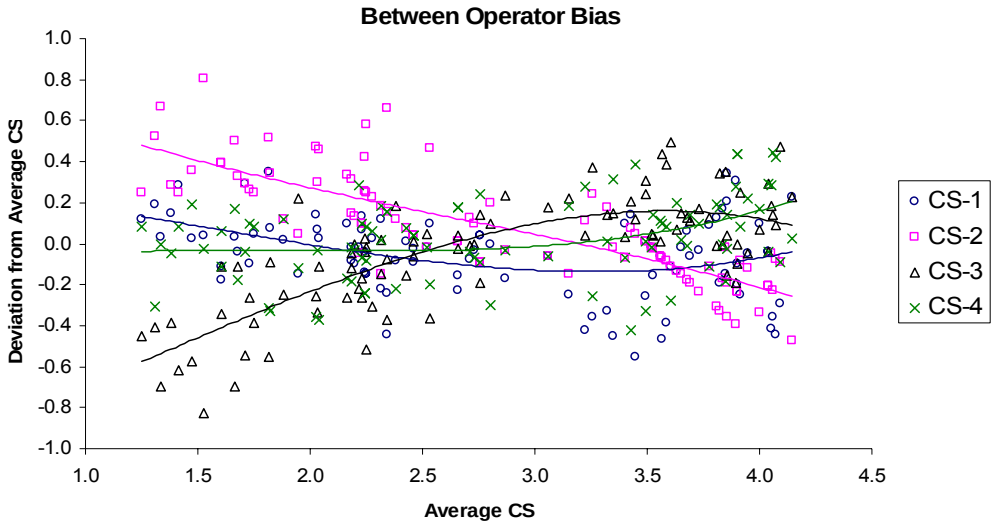
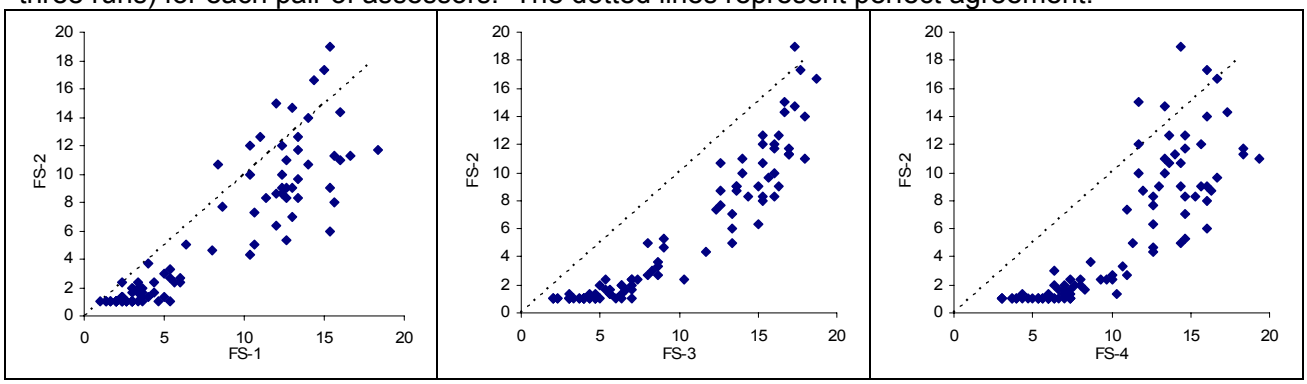


Figure 2. Bias between condition score assessors

The above graph is useful in comparing the different assessors, with the vertical difference between the fitted lines of two assessors providing an estimate of the bias in estimation at a given condition score. It is encouraging to observe that the bias remains below half a condition score, except for comparisons between assessor CS-2 and assessor CS-3 at condition score below about 2.
The above graph should not be used to make comment about assessor bias with respect to the true condition score, as this is unknown.
Later in this report, the relationships between condition score and the more objective measures of 'C-site fat', 'C-site EM' and 'GR tissue' are examined

Fat Score - The below graphs plot the average fat score assigned to each sheep (average of the three runs) for each pair of assessors. The dotted lines represent perfect agreement.



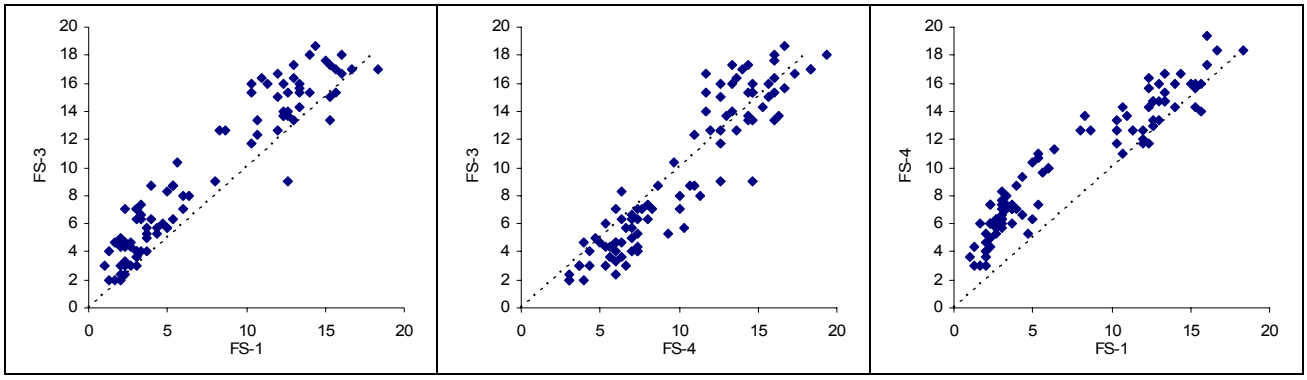


Figure 3. Scatter plots comparing average fat scores given by different assessors.

From these graphs we can see that fat score assessors are quite different, with only Assessor FS-3 and Assessor FS-4 showing some similarity. The other plots show the points away from the dotted line, indicating considerable bias between assessors. Assessor FS-2 in particular was very different to the other assessors.

The correlations between each pair of assessors are summarised below. Despite the large bias, the correlations are quite high. This highlights the problem with simply looking at a correlation coefficient when examining the relationship between variables.

Table 4. Correlations between fat score assessors.

	FS-1	*FS-2*	*FS-3*	*FS-4*
FS-1	1			
FS-2	0.89	1		
FS-3	0.95	0.93	1	
FS-4	0.95	0.84	0.93	1

The relationships are combined in the next graph, which plots the average score assigned to the sheep (X axis) against how much each assessor's average differs from this (Y-axis). A best fitting cubic curve is included for each person to show how the bias changed across a range of average fat scores.

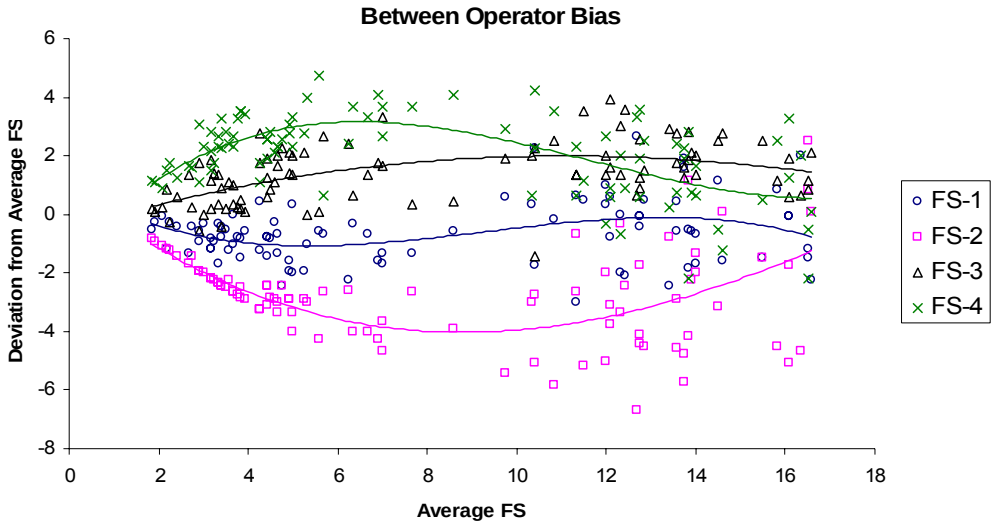
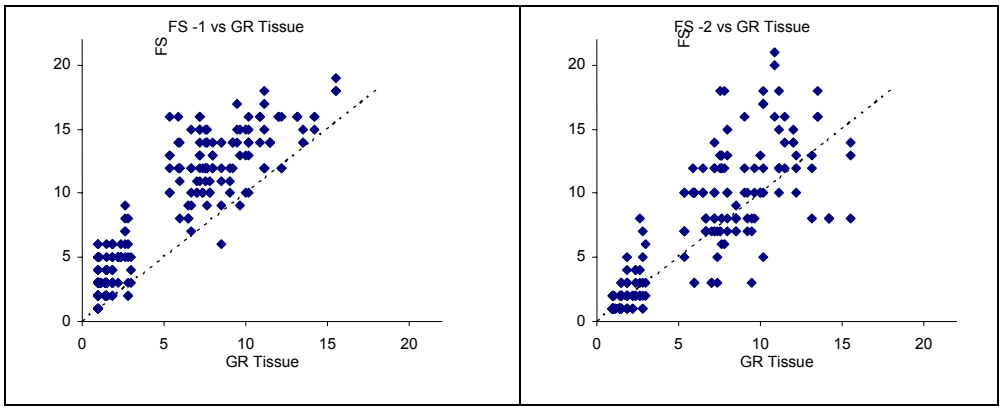


Figure 4. Bias between fat score assessors

The above graph is useful in comparing the different assessors, with the vertical difference between the fitted lines of two assessors providing an estimate of the bias in estimation at a given fat score. As discussed above, the bias between assessors is mostly very large, with the bias between assessor FS-4 and assessor FS-2 approaching 7mm (for example a sheep assessor FS-2 scored about 4mm, assessor FS-4 would score about 11mm).

The above graph should not be used to make comment about assessor bias with respect to the true fat score, as this is better done by comparing to the GR tissue results. The below graphs show this comparison by plotting each assessors individual fat scores against the average GR tissue result. The dotted lines represent perfect agreement.



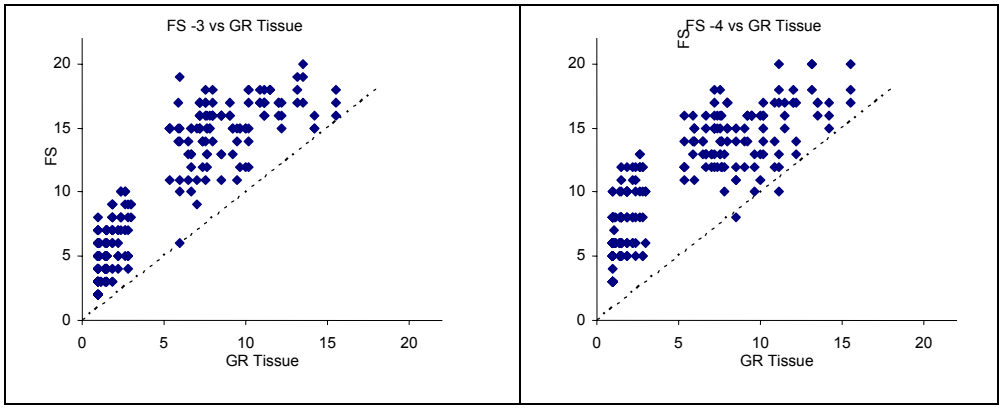


Figure 5. Predicted verses actual (GR tissue) depth for fat score assessors

From these graphs it is clear that assessor FS-1, assessor FS-3 and assessor FS-4 have large biases in their fat score as an estimation of GR tissue. They consistently (more than 95% of the time) overestimated the GR tissue depth. Assessor FS-2 is the only assessor whose bias is small (points centred about the dotted line). The below table summarises the bias and root mean squared error for the four assessors and shows again that assessor FS-2 fat scores provide the closest relationship to GR tissue.

Table 5. Bias and root mean squared error for fat score assessors.

	Average Bias		**Root Mean Squared Error**	
Assessor	'fat'	'thin'	'fat'	'thin'
FS – 1	4.1	1.8	4.7	2.1
FS – 2	1.5	0.1	3.9	0.7
FS – 3	6.2	3.6	6.6	4.0
FS – 4	5.8	5.1	6.4	5.5

Later in this report, the relationships between FS and GR tissue is examined in more detail, together with the relationships between FS and C-site fat and C-site EM.

C-site fat vs Condition/Fat Score - The below graphs plot the relationships between the average score from assessors CS-4 and FS-2 (each an average of 3 condition/fat scores) and the average C-site fat (an average of 6 measurements). Assessor FS-2 was selected as it was the only fat score operator with small bias. An example curve fit is shown, but note that this is not appropriate for use on uncorrected individual scores and it does not take into account the presence of x-axis error.

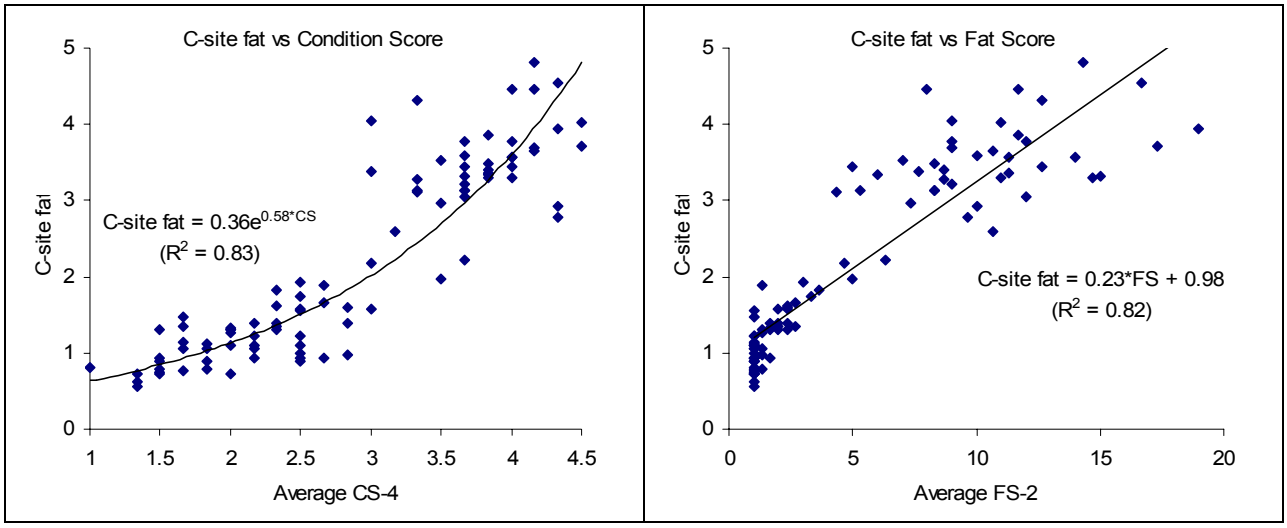


Figure 6. Relationship between C-site fat and condition score (CS-4) or fat score (FS-2).

These graphs show that there is a good relationship ($p<0.001$) between C-site fat and both condition score and fat score. The relationship is better (lower variation) for the lower condition sheep.
To compare the relationship for each of the four condition score and fat score assessors, a smoothing spline was fitted separately for each assessor:

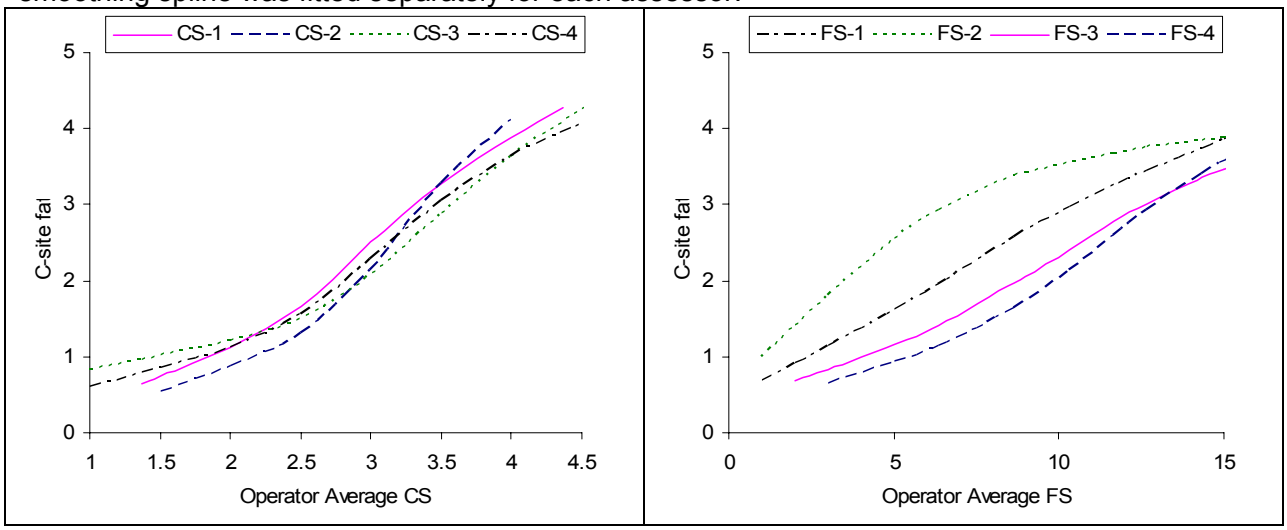


Figure 7. Fitted spline curves to C-site fat verses assessor average condition/fat score

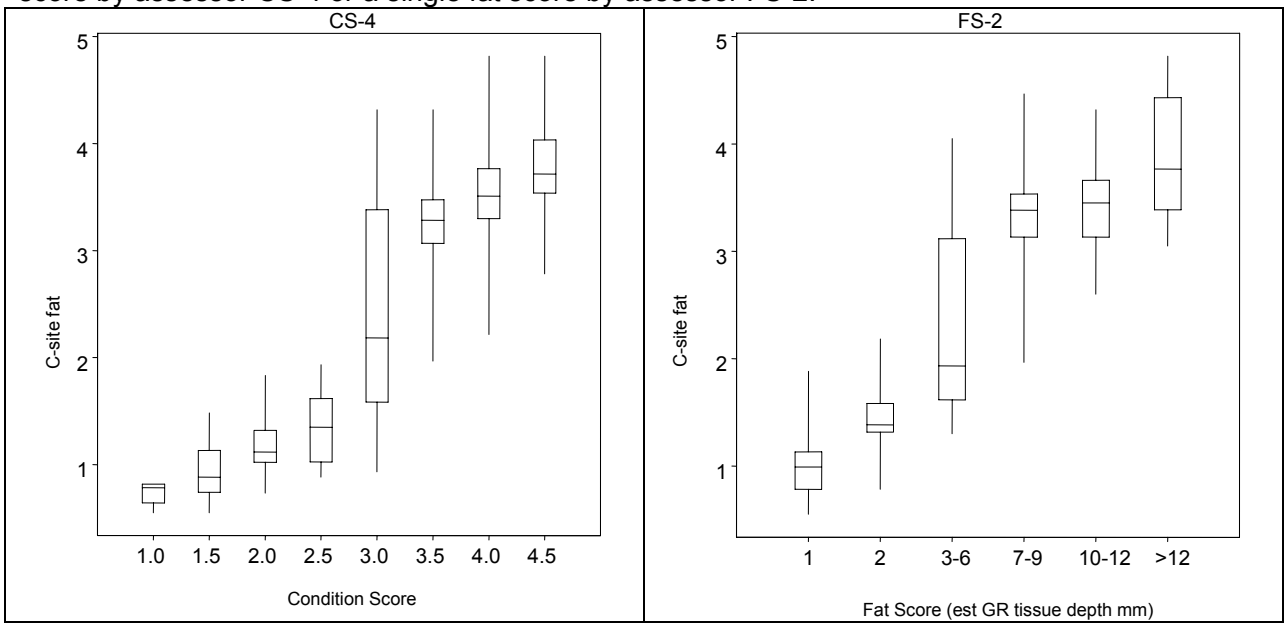
The relationships are more similar for the condition scorers than for the fat scorers. This is a result of the greater consistency between condition score assessors, as shown earlier.
In practice sheep are often condition or fat scored once by one assessor rather than three times by four assessors. The following analysis (tables and graphs) focus on the relationship between an individual (single run) condition score or fat score and the average C-site fat.
The below table summarises the prediction error and confidence interval for estimation from a single score by each assessor. For each individual score, the error was calculated as the difference between the predicted C-site fat (from the above spline fit) compared to the average C-site fat value

assigned to the animal. Thus the average error and confidence interval are based on precision, ignoring bias (corrected via the fitted curve) and are therefore optimistic. Because there was more error associated with the fatter animals, the table provides estimates for both the 'fat' and 'thin' groups.

Table 6. Error and confidence interval for prediction of C-site fat (mm) based on a single condition/fat score

	Average Prediction Error		**Approx. 95% C.I. (+/- 2*SD)**	
Assessor	'fat'	'thin'	'fat'	'thin'
CS – 1	0.50	0.28	+/- 1.18	+/- 0.74
CS – 2	0.47	0.27	+/- 1.22	+/- 0.71
CS – 3	0.49	0.28	+/- 1.16	+/- 0.72
CS – 4	0.49	0.30	+/- 1.31	+/- 0.76
FS – 1	0.38	0.23	+/- 0.98	+/- 0.58
FS – 2	0.47	0.23	+/- 1.20	+/- 0.60
FS – 3	0.45	0.21	+/- 1.14	+/- 0.54
FS – 4	0.51	0.31	+/- 1.19	+/- 0.81

From the above we can observe that the fat scorers (except assessor FS-4) showed a little better relationship (evidenced by generally lower error and narrower confidence interval) with C-site fat compared to the condition scorers. Otherwise the results are quite similar between the assessors.
The box-plots below examine how the distribution of C-site fat varies according to a single condition score by assessor CS-4 or a single fat score by assessor FS-2.



[Note: the boxes represent the middle half of the values; the lines in the boxes represent the median; and the vertical lines reach out to the maximum and minimum values.]

Figure 8. Distribution of C-site fat vs. condition score (Assessor CS-4) or fat score (Assessor FS-2)

From the box plots it can be observed that there is a good relationship between an individual condition score or fat score and C-site fat. (Often the boxes one condition score apart or a few fat scores apart, overlap very little.)
Also you may notice that condition score 3 has by far the widest distribution of C-site fat. While this might indicate that a score of 3 is typically given to a wider range of sheep (some quite fat and some quite thin), it could also be largely a consequence of the experimental design, where condition score 3 is about half way between the two main groups of sheep used (the typical condition score of the thin and fat groups is near 2 and 4 respectively.)

C-site EM vs Condition/Fat Score - The below graphs plot the relationships between the average score from assessors CS-4 and FS-2 (each an average of 3 condition/fat scores) and the average C-site EM (an average of 6 measurements). Assessor FS-2 was selected as it was the only fat score operator with small bias. An example curve fit is shown, but note that this is not appropriate for use on uncorrected individual scores and it does not take into account the presence of x-axis error.

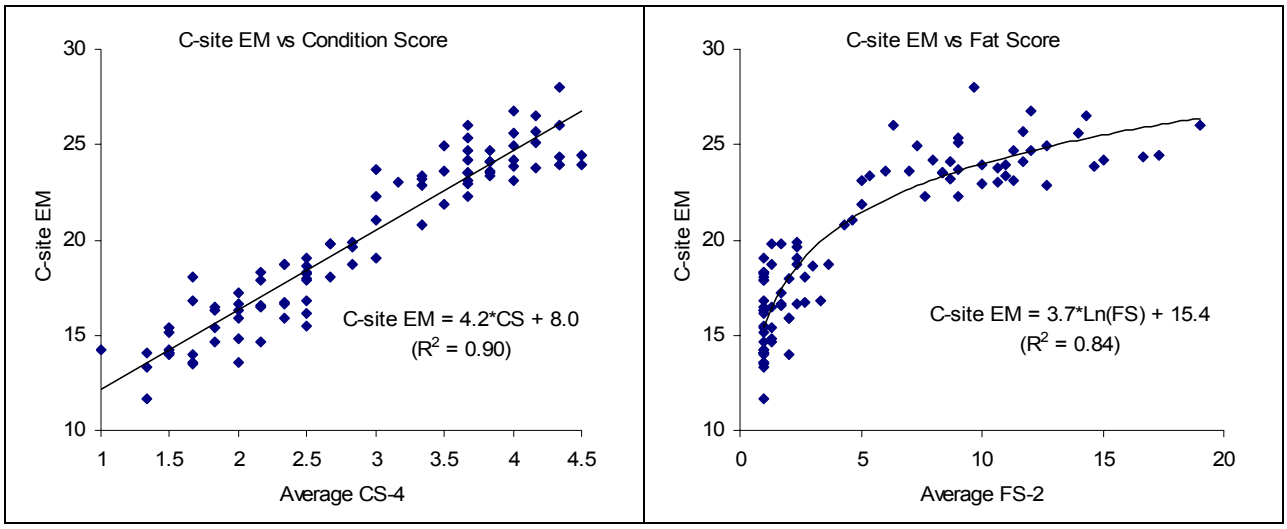


Figure 9. Relationship between C-site eye muscle and condition score (CS-4) or fat score (FS-2).

These graphs show that there is a good relationship ($p<0.001$) between C-site EM and both condition score and fat score. The relationship with condition score is more linear and explains a higher percentage of the variation.
To compare the relationship for each of the four condition score and fat score assessors, a smoothing spline was fitted separately for each assessor:

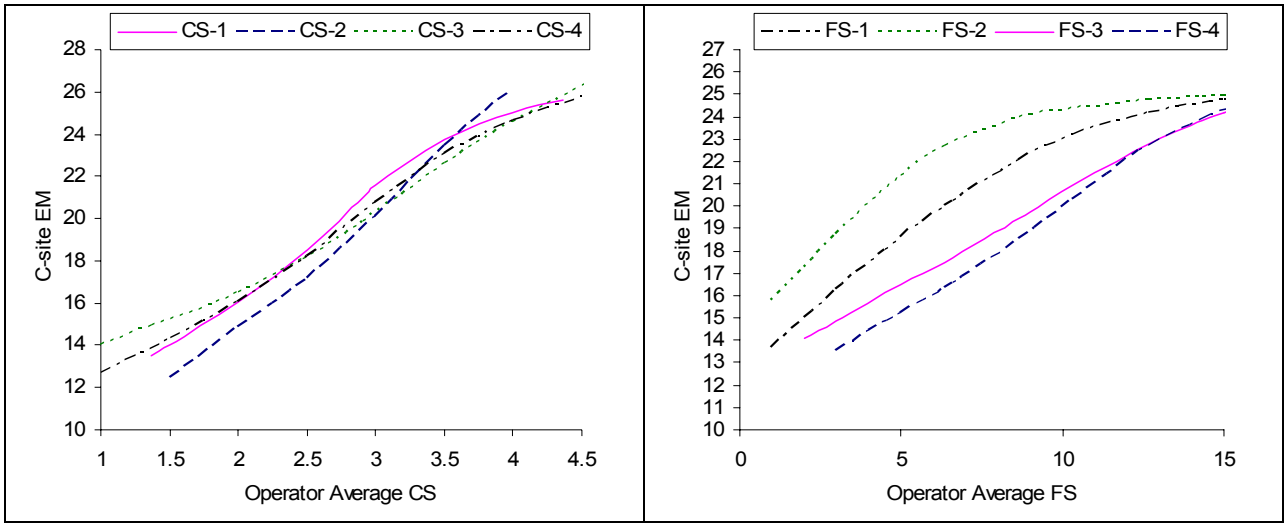


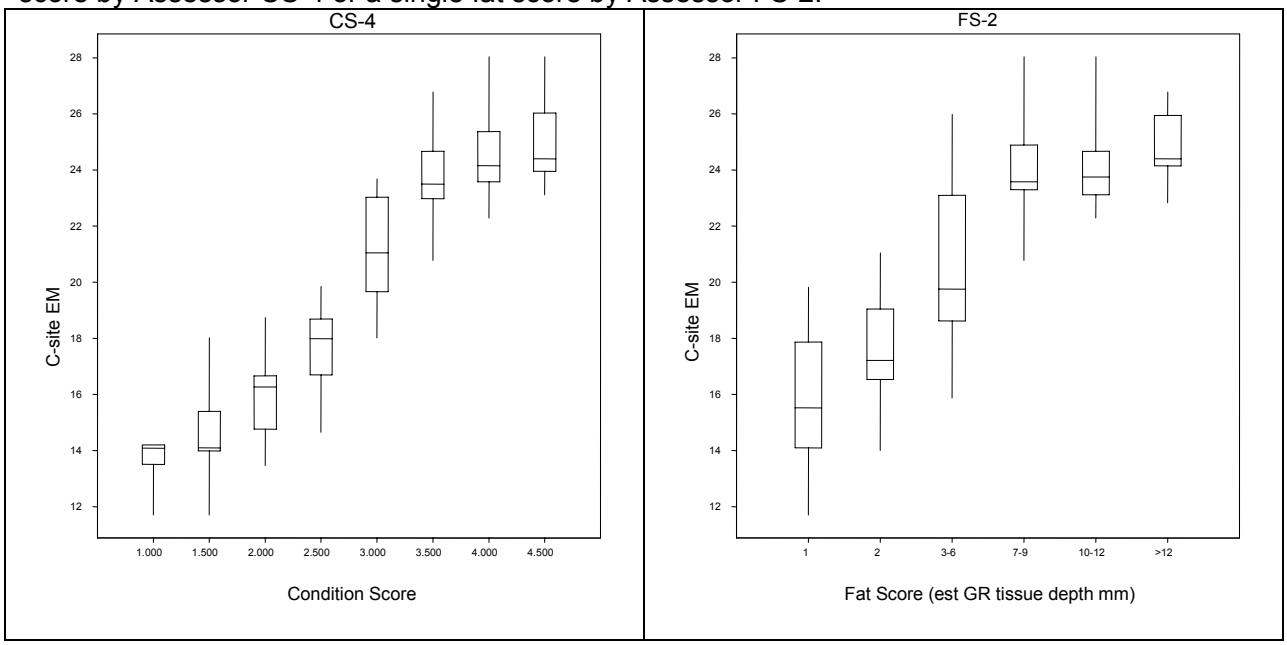
Figure 10. Fitted spline curves to C-site eye muscle depth vs. assessor average condition/fat score

Again the relationships are more similar for the condition scorers than for the fat scorers. This is a result of the greater consistency between condition score assessors, as shown earlier.
In practice sheep are often condition or fat scored once by one assessor rather than three times by four assessors. The following analysis (tables and graphs) focus on the relationship between an individual (single run) condition score or fat score and the average C-site EM.
The below table summarises the prediction error and confidence interval for estimation from a single score by each assessor. For each individual score, the error was calculated as the difference between the predicted C-site EM (from the above spline fit) compared to the average C-site EM value assigned to the animal. Thus the average error and confidence interval are based on precision, ignoring bias (corrected via the fitted curve) and are therefore optimistic. To be consistent with the previous analysis, the table provides estimates for both the 'fat' and 'thin' groups.

Table 7. Error and confidence interval for prediction of C-site EM (mm) based on a single condition/fat score

	Average Prediction Error		**Approx. 95% C.I. (+/- 2*SD)**	
Assessor	'fat'	'thin'	'fat'	'thin'
CS – 1	1.17	1.35	+/- 3.04	+/- 3.38
CS – 2	1.13	1.49	+/- 2.99	+/- 3.55
CS – 3	1.01	1.23	+/- 2.77	+/- 3.08
CS – 4	1.18	1.14	+/- 2.88	+/- 2.88
FS – 1	1.08	1.25	+/- 2.82	+/- 3.21
FS – 2	1.30	1.63	+/- 3.34	+/- 3.89
FS – 3	1.15	1.46	+/- 3.03	+/- 3.57
FS – 4	1.34	1.69	+/- 3.43	+/- 4.10

From the above we can again observe that the condition scorers showed on average a slightly better relationship (evidenced by generally lower error and narrower confidence interval) with C-site EM compared to the fat scorers. Overall the results are quite similar between the assessors.
The below box-plots examine how the distribution of C-site EM varies according to a single condition score by Assessor CS-4 or a single fat score by Assessor FS-2.



[Note: the boxes represent the middle half of the values; the lines in the boxes represent the median; and the vertical lines reach out to the maximum and minimum values.]

Figure 11. Distribution of C-site eye muscle depth verses condition score (Assessor CS-4) or fat score (Assessor FS-2)

From the above box plots it can be observed that there is a good relationship between an individual condition score or fat score and C-site EM. (The boxes one condition score apart or a few fat scores apart, overlap very little, except at the very upper end.)

GR tissue vs Condition/Fat Score - The below graphs plot the relationships between the average score from assessors CS-4 and FS-2 (each an average of 3 condition/fat scores) and the average GR tissue (an average of 6 measurements). Assessor FS-2 was selected as it was the only fat score operator with small bias. An example curve fit is shown, but note that this is not appropriate

for use on uncorrected individual scores and it does not take into account the presence of x-axis error.

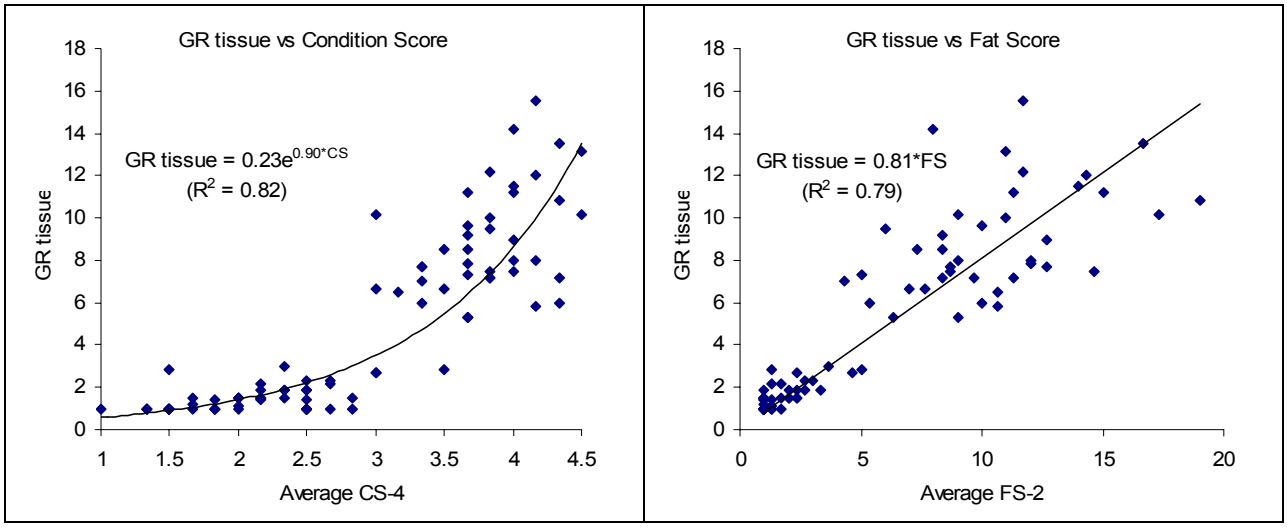


Figure 12. Relationship between GR-site tissue and condition score (CS-4) or fat score (FS-2).

These graphs show that there is a good relationship ($p<0.001$) between GR tissue and both condition score and fat score. There is a lot more variability in the relationship for the fatter sheep. As might have been expected, the relationship with fat score is more linear than the relationship with condition score.

To compare the relationship for each of the four condition score and fat score assessors, a smoothing spline was fitted separately for each assessor:

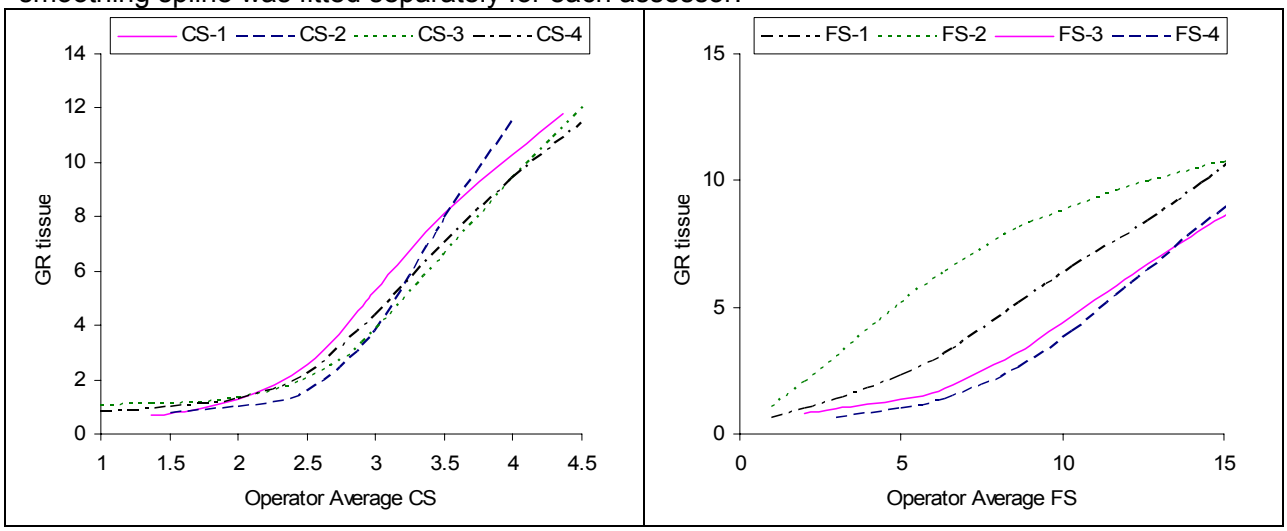


Figure 13. Fitted spline curves to GR-site tissue depth verses assessor average condition/fat score

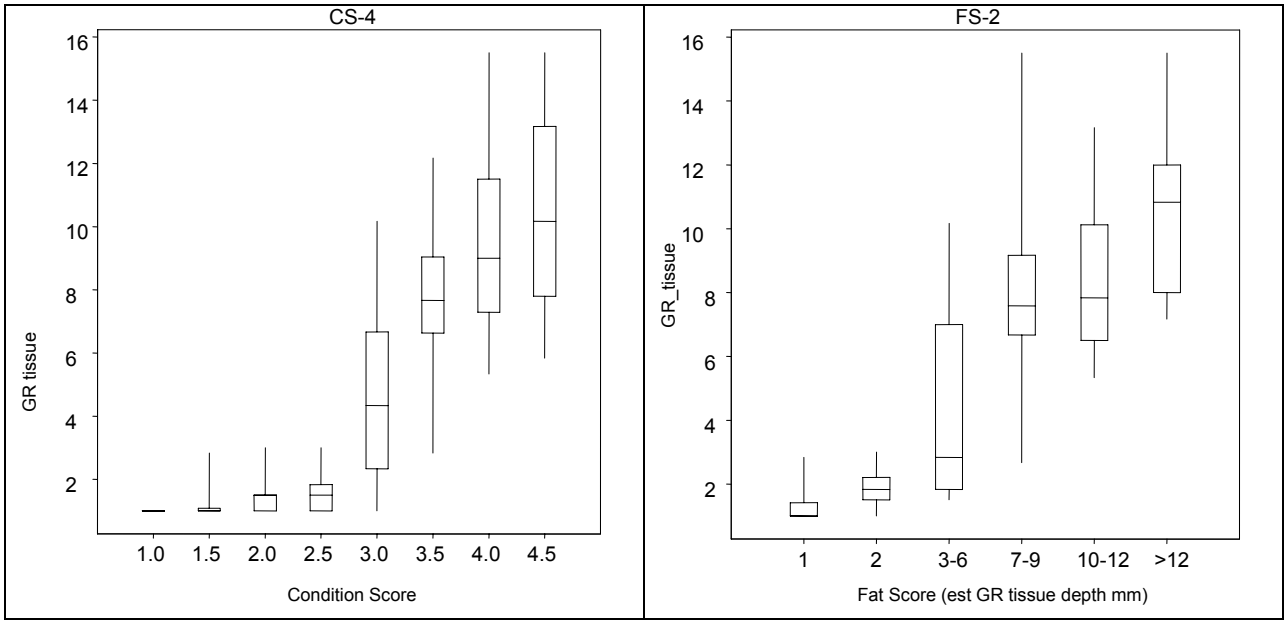
The relationships are more similar for the condition scorers than for the fat scorers. This is a result of the greater consistency between condition score assessors, as shown earlier.

In practice sheep are often condition or fat scored once by one assessor rather than three times by four assessors. The following analysis (tables and graphs) focus on the relationship between an individual (single run) condition score or fat score and the average GR tissue depth.
The below table summarises the prediction error and confidence interval for estimation from a single score by each assessor. For each individual score, the error was calculated as the difference between the predicted GR tissue depth (from the above spline fit) compared to the average GR tissue value assigned to the animal. Thus the average error and confidence interval are based on precision, ignoring bias (corrected via the fitted curve) and are therefore optimistic (for error including bias see Table 5). Because there was more error associated with the fatter animals, the table provides estimates for both the 'fat' and 'thin' groups.

Table 8. Error and confidence interval for prediction of GR tissue depth based on a single condition/fat score

	Average Prediction Error		**Approx. 95% C.I. (+/- 2*SD)**	
Assessor	'fat'	'thin'	'fat'	'thin'
CS – 1	2.19	0.69	+/- 5.17	+/- 1.87
CS – 2	1.98	0.60	+/- 5.26	+/- 1.77
CS – 3	2.05	0.57	+/- 5.07	+/- 1.68
CS – 4	2.25	0.61	+/- 5.43	+/- 1.80
FS – 1	1.79	0.40	+/- 4.50	+/- 1.22
FS – 2	2.04	0.43	+/- 5.19	+/- 1.44
FS – 3	1.87	0.47	+/- 4.67	+/- 1.25
FS – 4	2.10	0.77	+/- 5.31	+/- 2.22

From the above we can observe that the fat scorers (except assessor FS-4) showed a better relationship (evidenced by generally lower error and narrower confidence interval) with GR tissue compared to the condition scorers. Overall the results are quite similar between the assessors.
The below box-plots examine how the distribution of GR tissue varies according to a single condition score by Assessor CS-4 or a single fat score by Assessor FS-2.



[Note: the boxes represent the middle half of the values; the lines in the boxes represent the median; and the vertical lines reach out to the maximum and minimum values.]

Figure 14. Distribution of GR-site tissue depth verses condition score (Assessor CS-4) or fat score (Assessor FS-2)

From the above box plots it can be observed that there is a good relationship between an individual condition score or fat score and GR tissue. (The boxes one condition score apart or a few fat scores apart, do no overlap much except for at the very high end and at the low condition scores)

The original method of fat score relate whole scores 1 to 5 to measured GR tissue depth ranges as shown in Table 9 (reference Bill O'Halloran NSW DPI and MLA-ProGraze manual). However, it is interesting to note that when the estimated GR tissue depth in mm from FS-2 was converted to **the original scale FS 1, 2, 3 etc** (FS-2; Table 2) or the WAMMCO commercial FS assessor the measured GR tissue depth ranges for carcasses for FS = 2 or 3 were much wider and overlapping. Nonetheless, the mean GR tissue depth for sheep in FS=2 (7.6mm WAMMCO, 8.0mm FS-2) is significantly different to both FS=1 (1.5mm WAMMCO, 1.8mm FS-2) and FS=3 (10.9mm WAMMCO, 9.9mm FS-2) with $p<0.001$. This finding supports the shift of modern day FS assessors to assigning an estimated GR tissue depth to live sheep rather than a fat score. The high repeatability of all FS assessors but consistent and significant bias shown by 3 of the 4 assessors also suggests that the accuracy between assessors could be readily improved if a series of physical 'benchmark' models were developed to aid/guide the estimation of GR tissue depth in live sheep.

Table 9. Relationship between original Fat Score system and the range in GR tissue depth (mm) for the most accurate assessor (FS-2) and the WAMMCO commercial assessor

Fat Score	Estimated GR Tissue depth (mm) Original Scale (Bill O'Halloran; ProGraze)	[a]FS-2 Measured GR tissue depth mm ~95% confidence	WAMMCO Measured GR tissue depth mm ~95% confidence
1	0 - 5	1 – 4.7	1 – 2.7
2	6 - 10	2.8 – 13.1	3.6 – 11.7
3	11 - 15	4.7 – 15.0	4.5 – 17.2
4	16 - 20		
5	>20		

[a] The estimated GR tissue depth in mm from FS-2 was converted to **the original scale FS 1, 2, 3** and then compared with the measured GR tissue depth associated to estimate the GR tissue depth range for each **original FS** with 95% confidence

Condition Score vs Fat Score – Figure 15 shows the relationship between the average condition scores (an average of 12 measurements) and the average fat scores from assessor FS-2 (an average of 3 measurements). Assessor FS-2 was selected as it was the only fat score operator with small bias. An example curve fit is shown, but note that this is not appropriate for use on uncorrected individual scores and it does not take into account the presence of x-axis error.

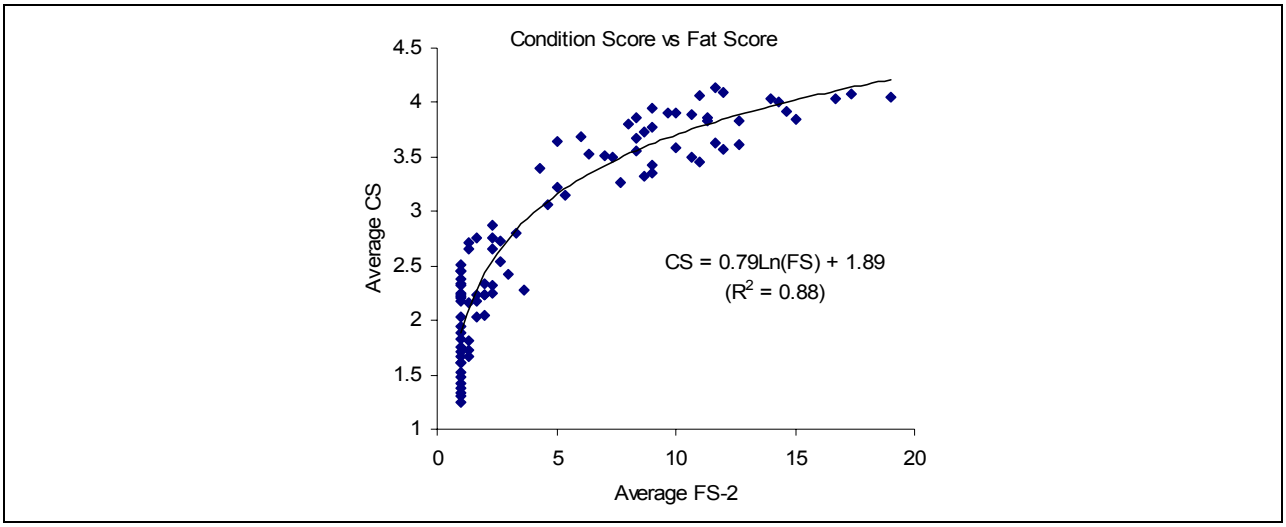


Figure 15. Relationship between average fat score (Assessor FS-2) and average condition score

These graphs show that there is a good relationship ($p<0.001$) between condition score and fat score. To see how the relationship differed between fat score assessors, a smoothing spline was fitted separately for each fat score assessor:

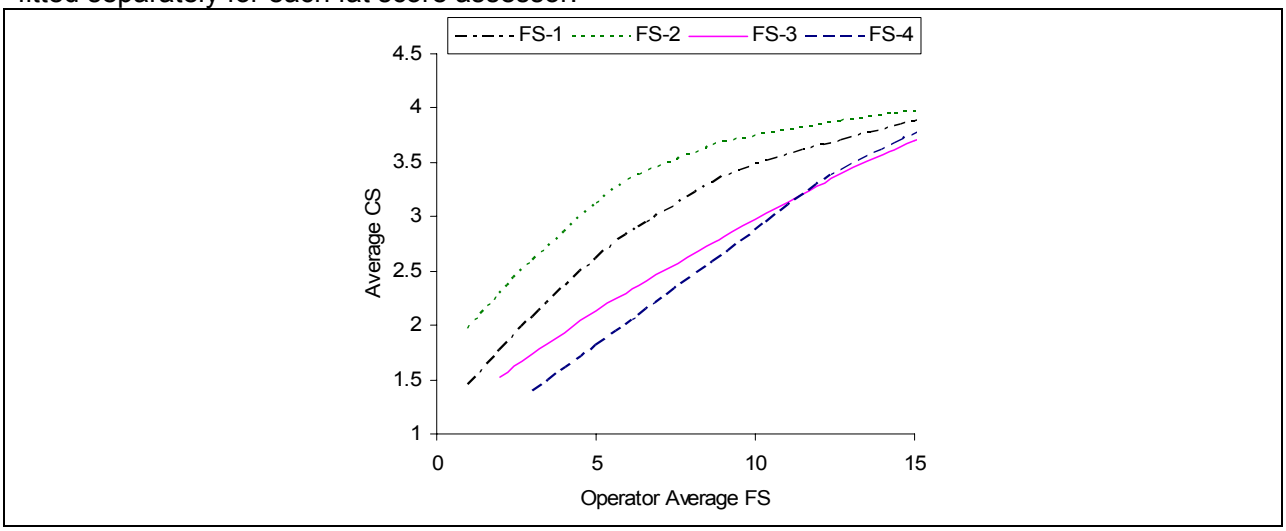


Figure 16. Fitted spline curve to assessor average fat score verses average condition score

The relationships are quite different for the fat scorers, resulting from the large differences in bias, as shown earlier.

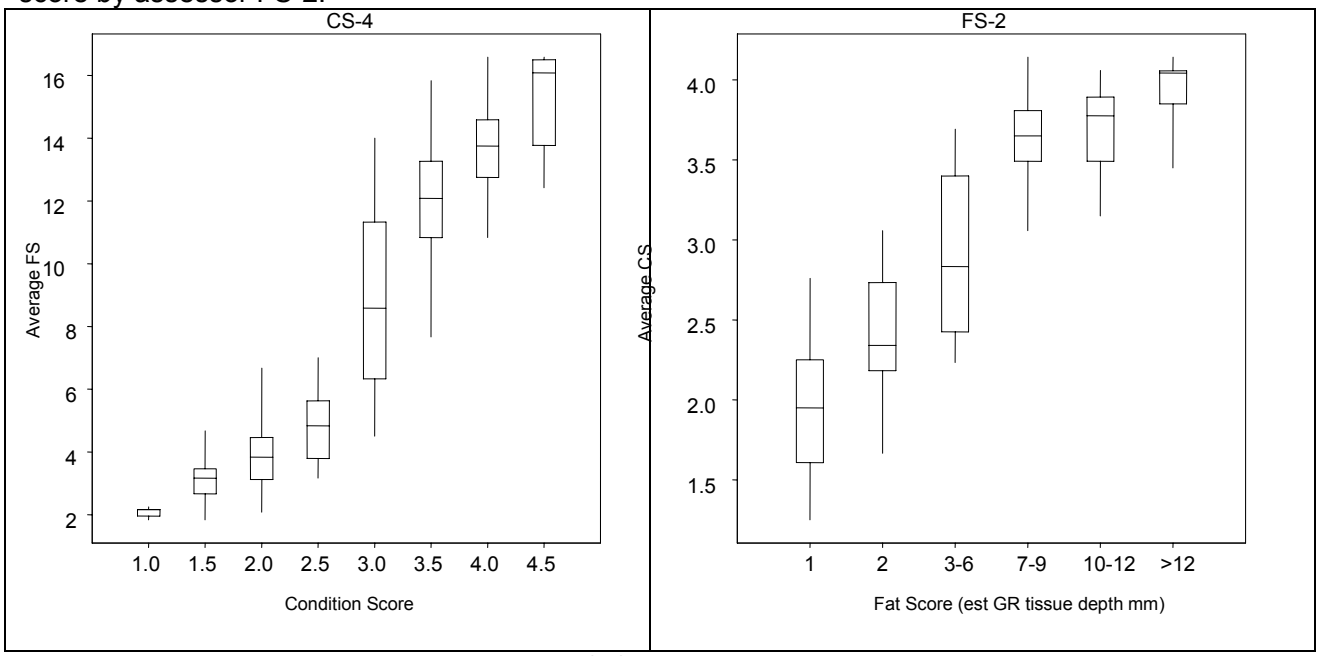
In practice sheep are often condition or fat scored once by one assessor rather than three times by four assessors. The following analysis (tables and graphs) focus on the relationship between an individual (single run) fat score and the average condition score.

The below tables summarise the prediction error and confidence interval for estimation from a single score by each assessor. For each individual fat score, the error was calculated as the difference between the predicted condition score (from the above spline fit) compared to the average condition score assigned to the animal.

Table 10. Error & confidence for a prediction of condition score based on a single fat score

	Average Prediction Error		**Approx. 95% C.I. (+/- 2*SD)**	
Assessor	'fat'	'thin'	'fat'	'thin'
FS – 1	+/- 0.20	+/- 0.23	+/- 0.50	+/- 0.57
FS – 2	+/- 0.23	+/- 0.34	+/- 0.61	+/- 0.78
FS – 3	+/- 0.18	+/- 0.28	+/- 0.47	+/- 0.67
FS – 4	+/- 0.23	+/- 0.31	+/- 0.61	+/- 0.75

In a similar way, we can see from the above table that a single fat score predicts condition score with an average error of about one quarter of a condition score. This is encouraging when considering that some condition score assessors score in intervals of half a condition score.
The below box-plots examine how the distribution of fat score varies according to a single condition score by assessor CS-4; and how the distribution of condition score varies according to a single fat score by assessor FS-2.



[Note: the boxes represent the middle half of the values; the lines in the boxes represent the median; and the vertical lines reach out to the maximum and minimum values.]

Figure 17. Distribution of average fat/condition score verses condition score(Assessor CS-4)/fat score(Assessor FS-2)

From the above box plots it can be observed that there is a good relationship between an individual condition score and the average fat score; and between an individual fat score and the average condition score. (The boxes one condition score apart or a few fat scores apart, overlap very little).

Young vs Mature sheep - The following relationships were examined to see if there was any significant difference for young compared to mature sheep:

- Average 'C-site fat' vs. Average CS and Average FS
- Average 'C-site EM' vs. Average CS and Average FS
- Average 'GR tissue' vs. Average CS and Average FS
- Average CS vs. Average FS

Two of the above 7 relationships showed a highly significant difference ($p<0.001$) between young and mature sheep. These two relationships are graphed below.

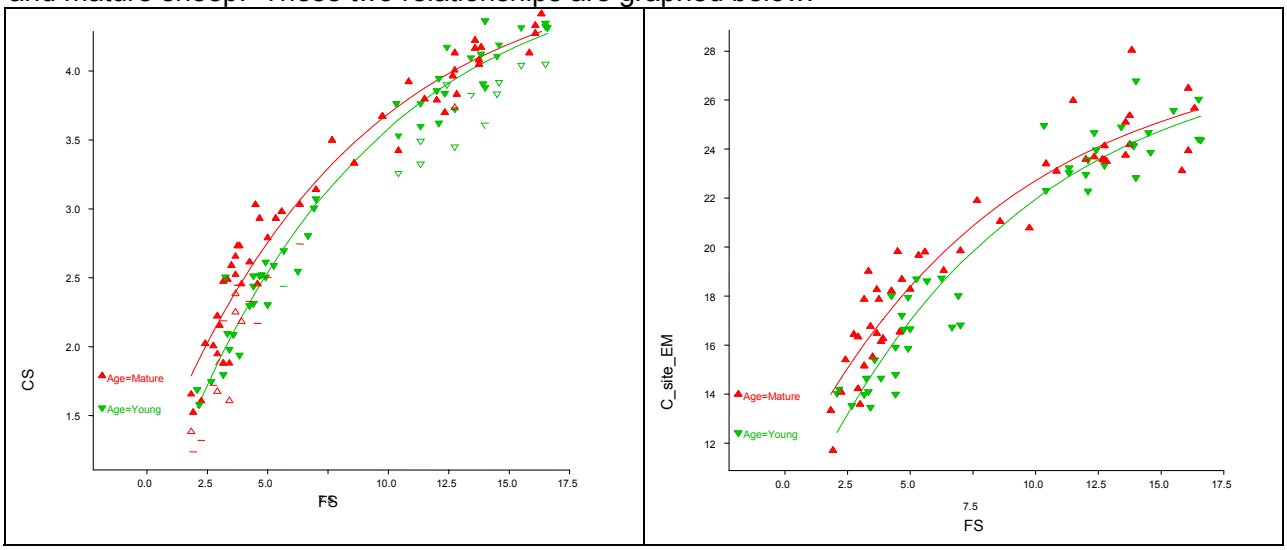


Figure 18. Age effect on the relationship between condition score/C-site EM verses fat score

Two other relationships also showed a significant difference ($p=0.03$) between young and mature sheep. In both cases the difference was mainly evident for the lower condition sheep. The significant age effect on the CS verses GR tissue relationship was determined on a log scale (not graphed).

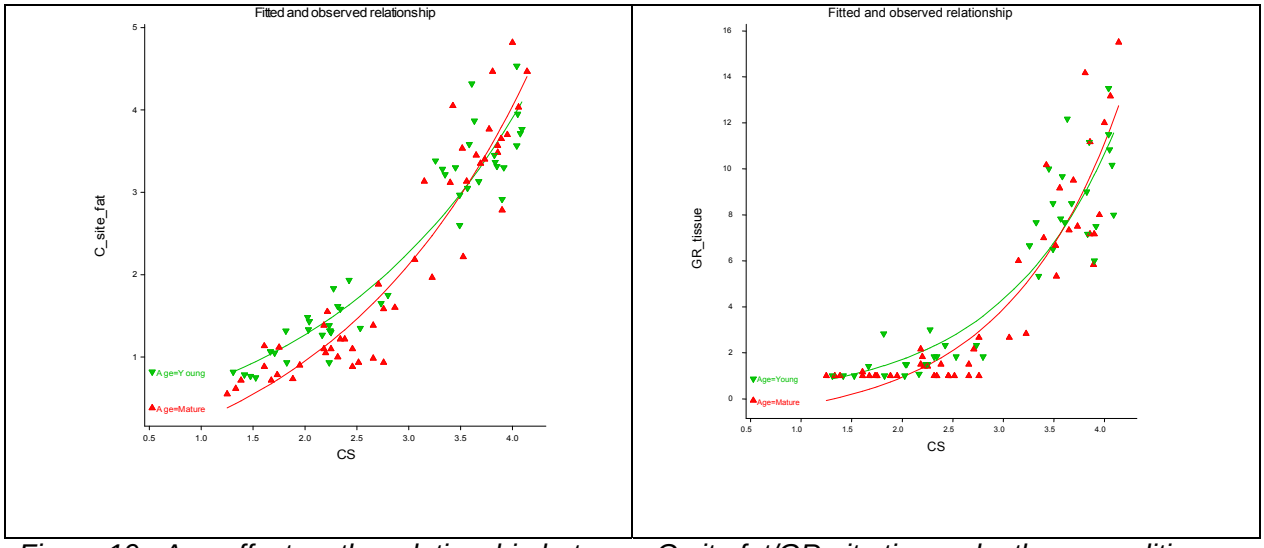


Figure 19. Age effect on the relationship between C-site fat/GR-site tissue depth vs. condition score
The other 3 relationships (not graphed) did not show any significant difference.

5 Success in Achieving Objectives

All the objectives of this experiment were successfully achieved.

6 Impact on Meat and Livestock Industry – now & in five years time

"Measure to manage" is the message that can be heard loud and clear from the MLA's 'Wean-more-lambs' and AWI's 'Lifetime Wool' projects. Every ewe management decision that is made on-farm can be significantly enhanced through an understanding of the current nutritional status of the ewes. This can only be achieved through regular monitoring of liveweight, condition score or fat score.

The link between ewe nutrition and production and reproduction is well established. Ewes that receive less nutrition will produce less wool and wean fewer, smaller lambs that will be less productive throughout their lifetime. However improving pasture utilisation through maintaining high stocking rates is a key profit driver of a Merino production business. The challenge is therefore to achieve the appropriate balance between stocking rate and individual ewe performance. The 150 wool producers involved nationally in Lifetime Wool are discovering the importance of monitoring breeding ewes to enable timely management. They have seen that although pasture assessments and knowledge of the nutritional value of any supplements being used are a guide to ewe nutrition, the only accurate description of ewe nutrition is how the ewes are performing which is assessed with regular monitoring of liveweight, condition score or fat score – let the sheep do the talking! However, until this experiment confusion in the minds of farmers, consultants and scientists with respect to the relationship between the various methods of assessing the 'nutrition wellbeing' of ewe flocks has hindered the successful adoption of these messages. Experience with farmers over many years is that they are reluctant to invest the money and time to accurately weigh sheep but now that we have established strong usable relationships between condition score, and estimated GR tissue depth

(the old fat score) farmers may have more confidence to use these methods to manage their ewes to more profitable condition or GR tissue depth targets.

Increasingly, all classes of sheep are being sold on a weight and grade basis where minimising the proportion of the sale flock that is heavily discounted for being outside the carcass weight x fat score grid has a large impact on profit. This experiment strongly suggests that while a skilled operator can accurately estimate the total tissue depth (mm of fat + muscle) at the GR site over the 12th rib (original fat score), the variance about the estimate was greater than indicated by the rigid 5 mm ranges in GR tissue depth currently used by the industry for setting price grids. This issue could be addressed by replacing the current scoring system with training in assessing GR tissue depth in mm. Timely and accurate abattoir feedback on actual GR tissue depth would be an essential part of the training and ongoing calibration of all assessors.

7 Conclusions and Recommendations

It was concluded that while a skilled operator could accurately estimate the total tissue depth (mm of fat + muscle) at the GR site over the 12th rib, this measurement was poorly related to the whole fat scores and their related 5 mm ranges in GR tissue depth currently used by the industry for setting price grids. This issue could be addressed by replacing the current scoring system with training in assessing GR tissue depth in mm. Timely and accurate abattoir feedback on actual GR tissue depth would be an essential part of the training and ongoing calibration of all assessors.

The high repeatability of all GR tissue depth assessors but consistent and significant bias shown by 3 of the 4 assessors also suggests that the accuracy between assessors could be readily improved if a series of physical ‘benchmark’ models were developed to aid/guide the estimation of GR tissue depth in live sheep in much the same way as the Lifetime Wool project has developed physical ‘benchmark’ models for condition scoring.

The repeatability of all skilled operators (both subjective and objective) was very high and commercially acceptable. However, while the relationship between condition score and estimated GR tissue depth (old Fat Score) was very strong and predictable, the relationship was not linear. In fact, in the condition score range of 2 to 3, considered to be critical for the management of ewes in the Lifetime Wool project, there was very little variation in estimated or actual GR tissue depth (1 to 4 mm and 1 to 3 mm, respectively). Therefore in future recommendations for management of ewe flocks it is our conclusion that targets framed in terms of CS targets would offer far greater control over predicted effects on the performance of the ewes and future performance of her progeny than targets framed in terms of estimated GR tissue depth.

There were a number of small differences in the various relationships between mature and young ewes. For example the estimated condition score of adult ewes was about 0.25 of a score greater than young ewes at the same estimated GR tissue depth (mm) over the critical range of 3 to 12 mm. However, none of these small differences would have any great impact on how the industry is currently or contemplates using condition score or estimated GR tissue depth in the future.

8 Bibliography

Feeding standards for Australian livestock, Ruminants p 58-68

Russell et al. 1969, J. agric Sci. Camb.72, 451-454