

# Predicting ewe body condition score using adjusted liveweight for conceptus and fleece weight, height at withers, and previous body condition score record

Jimmy Semakula,<sup>†,‡,1,\*</sup> Rene A. Corner-Thomas,<sup>†,\*</sup> Steve T. Morris,<sup>†</sup> Hugh T. Blair,<sup>†</sup> and Paul R. Kenyon<sup>†</sup>

<sup>†</sup>Department of Animal Science, School of Agriculture and Environment, Massey University, Palmerston North, New Zealand <sup>‡</sup>National Agricultural Research Organization, Entebbe, Uganda

**ABSTRACT:** The relationship between ewe body condition score (BCS) and liveweight (LW) has been exploited previously to predict the former from LW, LW-change, and previous BCS records. It was hypothesized that if fleece weight and conceptus-free liveweight and LW-change, and in addition, height at withers were used, the accuracy of current approaches to predicting BCS would be enhanced. Ewes born in 2017 ( $n = 429$ ) were followed from 8 mo to approximately 42 mo of age in New Zealand. Individual ewe data were collected on LW and BCS at different stages of the annual production cycle (i.e., prebreeding, at pregnancy diagnosis, prelambling, and weaning). Additionally, individual lambing dates, ewe fleece weight, and height at withers data were collected. Linear regression models were fitted to predict current BCS at each ewe age and stage of the annual production cycle using two LW-based models, namely, unadjusted for conceptus weight and fleece weight (LW alone1) and adjusted (LW alone2) models. Furthermore, another two models based on a combination of LW, LW-change, previous BCS, and height at withers (combined models), namely, unadjusted (combined1) and

adjusted for conceptus and fleece weight (combined2), were fitted. Combined models gave more accurate (with lower root mean square error: RMSE) BCS predictions than models based on LW records alone. However, applying adjusted models did not improve BCS prediction accuracy (or reduce RMSE) or improve model goodness of fit ( $R^2$ ) ( $P > 0.05$ ). Furthermore, in all models, both LW-alone and combined models, a great proportion of variability in BCS, could not be accounted for ( $0.25 \geq R^2 \geq 0.83$ ) and there was substantial prediction error ( $0.33 \text{ BCS} \geq \text{RMSE} \geq 0.49 \text{ BCS}$ ) across age groups and stages of the annual production cycle and over time (years). Therefore, using additional ewe data which allowed for the correction of LW for fleece and conceptus weight and using height at withers as an additional predictor did not improve model accuracy. In fact, the findings suggest that adjusting LW data for conceptus and fleece weight offer no additional value to the BCS prediction models based on LW. Therefore, additional research to identify alternative methodologies to account for individual animal variability is still needed.

**Key words:** body condition score, height at withers, liveweight, prediction

© The Author(s) 2021. Published by Oxford University Press on behalf of the American Society of Animal Science.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Transl. Anim. Sci. 2021.5:1-12  
<https://doi.org/10.1093/tas/txab130>

<sup>1</sup>Corresponding author: [jynsemakula@yahoo.com](mailto:jynsemakula@yahoo.com)

Received May 20, 2021.

Accepted July 30, 2021.

## INTRODUCTION

Body condition score (BCS) in sheep is a commonly used subjective measure (Morris et al., 2002;

Vieira et al., 2015) to help make flock nutritional and management decisions. Devised by Jefferies (1961) and then revised by Russel et al. (1969), it subjectively quantifies the amount of soft tissue along the lumbar spine (Jefferies, 1961; Kenyon et al., 2014). Body condition score in sheep utilizes a 0.0–5.0 scale range with either half (0.5) units or quarter (0.25) units and is conducted through the palpation of the lumbar vertebrae immediately caudal to the last rib and above the kidneys (Kenyon et al., 2014).

Body condition score circumvents factors that can confound liveweight (LW) such as gut-fill, physiological status, fleece weight, and frame size (Coates and Penning, 2000; Kenyon et al., 2014). Despite the advantages of using BCS over LW to better manage flock nutrition, producers, especially under extensive flock management systems such as in the southern hemisphere, rarely utilize it (Jones et al., 2011; Corner-Thomas et al., 2016). Instead, farmers either depend on inaccurate visual inspection methods or utilize liveweight measures only (Besier and Hopkins, 1989). This low uptake among producers is driven by the procedure being subjective: relatively labor intensive and requiring training (Kenyon et al., 2014). Strategies to increase the adoption and use of BCS among producers, such as promotion of producer training and regular assessor recalibration workshops, have not yielded the desired change (Kenyon et al., 2014). This is likely because they do not address how to lessen the additional labor burden related to hands-on BCS, especially in large flocks under extensive management systems. Therefore, it could be reasoned that reliable and precise alternative methods to estimate BCS of sheep that involve reduced hands-on measurement would likely be useful and improve uptake and acceptance of the BCS technique. This indirect method would preferably be based on already existing and utilized on-farm management tools in order to reduce workload and be easily undertaken and not be subjective in nature.

The relationship between BCS and LW is well established in sheep (Sezenler et al., 2011; Kenyon et al., 2014; McHugh et al., 2019). In our previous study (Semakula et al., 2020a), it was demonstrated that BCS is positively and linearly related to LW. The relationship is known to differ by stage of the annual production cycle, age, and breed of ewes (Sezenler et al., 2011; McHugh et al., 2019). This relationship between BCS and LW was utilized to predict current BCS on a 5-point scale from lifetime liveweight (current and previous), liveweight change, and previous BCS based on linear regression models (Semakula et al., 2020b). It was

demonstrated that with a set of established equations it may be possible to calculate a predicted BCS instantly, at each live weighing, for each sheep. However, a great proportion of variability in BCS remained unaccounted for, leading to less robust models. Furthermore, in our previous study (Semakula et al., 2021b), machine learning classification algorithms were successfully (with up to 90% accuracy) used to predict BCS using LW predictors. However, these machine learning classification models were limited to a 3-point scale due to gross class imbalance in BCS data. Full scale BCS (5-point scale: 1.0–5.0) prediction based on linear regression does not require balanced data. In our study (Semakula et al., 2020b), it was hypothesized that greater accuracy could be achieved if key variables affecting the relationship between BCS and LW were also accounted for. Morphometric measurements such as height at withers are positively correlated with LW and BCS in sheep (Burke et al., 2004; Holman et al., 2012). Furthermore, pregnancy and fleece weight confound the relationship between BCS and LW (Kenyon et al., 2014; Brown et al., 2015). If these variables could be accounted for, BCS prediction accuracy may potentially be improved. Therefore, the aim of this study was to firstly determine if the ewe BCS prediction accuracies reported by Semakula et al. (2020b) can be reproduced in an independent dataset and secondly to investigate if the accuracy and scope of BCS prediction equations could be improved by adding information on the height at withers, fleece weight, and physiological state of a ewe.

## MATERIALS AND METHODS

### *Experimental Design*

The current study utilized data collected between 2017 and 2020 from one flock. Romney type ewes were initially raised at Riverside farm (2017–2018) and later (2019) transferred to Keeble farm as part of normal routine farm management. Riverside farm is located 11 km north to north-west of Masterton (40°50' S, 175°37' E) while Keeble farm was 5 km south of Palmerston North (40°24' S and 175°36' E), New Zealand. Ewes were maintained under commercial farming conditions from weaning to 42 mo of age (Pettigrew et al., 2018; Pettigrew et al., 2019). A total of 429 ewe lambs born in the same season (Aug–Sep 2017) were followed until maturity at 42 mo of age. Data were collected on whether study ewe lambs were born to mature or ewe lambs and in which breeding cycle.

Unfasted liveweights and BCS of ewes (born to ewe lambs or mature ewes) were recorded at 6 mo of age, prebreeding (PB), at pregnancy diagnosis (PD), and 8 d prior to the start of lambing (PL: prelambing) and at weaning (W: weaning; lambs on average of 3 mo of age) in each year. All weight measurement occasions were conducted when ewes were not wet. All ewes were followed for three productive full years. The ewes in this study were themselves presented for breeding at 8 mo of age. This study was approved by the Massey University animal ethics committee (protocol number: MUAEC 17/16).

All ewes were weighed (to the nearest 0.1 kg) using static digital weighing scales (Tru-Test group, model XR5000). Body condition score was undertaken by one experienced assessor using a 1.0–5.0 scale (1 = thin, 5 = obese) with sheep assessed to the nearest 0.5 of a BCS (Jefferies, 1961; Kenyon et al., 2014). Ewes were shorn each year during late pregnancy (47 to 49 d prior to the start of lambing), and fleece weights were recorded. Estimated fleece weights at the time of the weighing (equation 1) in each year were computed by multiplying the annual fleece weight at late pregnancy with the relative proportion of the fleece length (mm) at the corresponding time assuming a shorn fleece length of 150 mm and an amplitude of 19% of the mean (Cottle and Pacheco, 2017),

$$Y_t(\text{kg}) = \text{Fwt} * \text{Rl} \quad 1$$

where  $Y_t$  is the estimated fleece weight (kg) at a given time (month), Fwt was the actual fleece weight at the annual shearing (kg), and Rl is the proportion of wool length at a given time of the year relative to the wool length when shearing was last done (Length at shearing, mm). The minimum wool length left during shearing was 5.0 mm. All parameters were adapted from Cottle and Pacheco (2017).

The conceptus mass can confound accurate measurement of ewe conceptus-free liveweight, especially from mid-pregnancy onwards (Kenyon et al., 2008; Kenyon et al., 2011). Adjusted ewe liveweight can be obtained if the conceptus mass can be corrected for. Therefore, to allow for the computation of adjusted liveweights, lambing dates for each ewe were recorded. The dates were used to estimate days of pregnancy when the liveweight measurements were recorded at PD and PL. The gestation time (days of pregnancy at PD or PL) was computed as the difference between 147 d (gestation was assumed to be 147 d) and the time from the event (PD or PL liveweight measurement) to lambing. The predicted conceptus and gravid uterus

weight was determined using Gompertz equation (equation 2) below adapted by Freer et al. (2007),

$$Y = \text{SBW} \exp(A - B(\exp(-Ct))) \quad 2$$

where  $Y$  is the weight of the content of the gravid uterus, SBW is the scaled birth weight (the ratio of the actual birth weight to the standard birth weight of 5 kg assumed by Gompertz equation),  $t$  is the gestation length (days), and parameters  $A$ ,  $B$ , and  $C$  are constants 5.17, 8.38, and  $6.08 \times 10^{-3}$ , respectively. A 5 kg lamb at 147 d was used as the standard for scaling of birth weights. The final adjusted ewe liveweights excluded fleece weight and gravid uterus weight. To cater for both single- and twin-bearing ewes, a pooled lamb birth weight (overall weight of both lambs) was computed for twin-bearing ewes. Liveweights at pre-breeding and weaning were adjusted for fleece weight only, while at PD and PL, both conceptus and fleece weights were adjusted.

Height at withers (HW) was recorded every 6 mo using an automatic laser distance measurer (Stanley TLM130i distance meter, max range = 30 m  $\pm$  3 mm accuracy) attached to a sliding bar from above the weigh crate. The height of the ewe was computed using the formula:

$$\text{Unadjusted Height at withers (m)} = \text{XZ} \quad 3$$

where  $X$  is the distance from the laser meter ( $X$ ) to the floor of the weigh crate and  $Z$  is the distance from the laser meter to the ewe withers. Height at withers was later corrected based on predicted annual fleece growth to generate adjusted HW (Cottle and Pacheco, 2017).

### Statistical Analyses

Data were analyzed using R program version 3.3.4 (R Core Team, 2016) with package extensions in the caret package (Kuhn, 2008). Similar analytical procedures including variable formulation and selection, model building, cross-validation, and evaluation used by Semakula et al. (2020b) were followed. Consequently, both classification and multiple linear regression approaches were tested. Any missing values were imputed using the preProcess function and bagimput method from the caret package in R (Kuhn, 2008). Additionally, non-numerical data were made numerical and  $z$ -transformed (scaled and centered) during analysis using the same preProcess function above.  $Z$ -transformed values outside the 95% CI ( $z \pm 1.96$  range) were not used in the final analysis. Differences among correlation coefficients were tested for significance based on Fisher's  $r$ -to- $z$

transformation. In the present analysis, there was high-class BCS imbalance ([Supplementary Table A1](#)) making use of classification methods to predict individual BCS inappropriate ([Triguero et al., 2015](#)). In order to predict individual ewe BCS on a full scale (1.0–5.0), an alternative statistically robust method ([Norman, 2010](#)) to class imbalance was warranted. Consequently, the multivariate linear model which has been successfully utilized to predict BCS in cattle ([Martins et al., 2020](#)) and sheep ([Semakula et al., 2020b](#)) was applied.

### Variable Selection, Model Building, and Validation

The predictors for each BCS were selected through a variable selection technique executed in the R program ([R Core Team, 2016](#)) using the elastic net method in the glmnet extension ([Friedman et al., 2010](#)) in the caret package ([Kuhn, 2008](#)). Models were fitted and validated using a four-step procedure (data partitioning, resampling, model training, and validation) as described by [Semakula et al. \(2020a\)](#) and [Semakula et al. \(2020b\)](#). Using selected predictors, regression equations were fitted on a training dataset to predict BCS from lifetime ewe liveweight records (current and previous weights), liveweight change (difference in weight between two consecutive weights taken at different time points), height at withers, previous BCS scores (a record of all previous BCS scores), and their lamb birth and weaning weight data in one regression. Initially a total of eleven (11) regression

equations (each representing ewe age group and stage of the annual production cycle) were created for BCS prediction based on unadjusted lifetime LW measurements (Liveweight alone1 models). A previous measurement was that taken at a different time point (different stage of the annual production cycle) prior to the current one. Liveweight change refers to the change in liveweight between two time points. Furthermore, 11 more equations were generated incorporating liveweight change and previous BCS in addition to lifetime liveweight (combined1 models). The process of generating BCS prediction equations above was repeated based on adjusted LW (adjusting for conceptus weight and fleece weight) (Liveweight alone2 models) and based on adjusted LW, liveweight change, height (adjusted for fleece growth) at withers, and previous BCS (combined2 models). A description of variables is given in [Table 1](#).

### Model Evaluation

Models were evaluated as described by [Semakula et al. \(2020b\)](#). Model performance evaluation was conducted on training dataset using two metrics ([Theil, 1958](#); [Botchkarev, 2019](#)) adjusted coefficient of determination (adj.  $R^2$ ) and the root mean square error (RMSE). Each BCS prediction model validation was conducted on the testing dataset, with each replicated 1000-fold. The quality and success of the prediction models was assessed using the coefficient of

**Table 1.** Explanation of liveweight (LW), liveweight change, height at withers (H), and body condition score (BCS) variables by ewe age group and stage of the annual production cycle

Age (mo)	Stage of the annual production cycle	LW*	BCS <sup>†</sup>	Change in liveweight <sup>‡</sup>	HW <sup>‡</sup>
8–18	Pre-breeding	WP1	BP1		PH1
	Pregnancy diagnosis	WD1	BD1	WT11(WD1–WP1)	DH1
	Pre-lambing	WL1	BL1	WT12(WL1–WD1)	LH1
	Weaning	WW1	BW1	WT13(WW1–WL1)	WH1
19–30	Pre-breeding	WP2	BP2	DW-T1(WP2–WW1)	PH2
	Pregnancy diagnosis	WD2	BP2	WT21(WD2–WP2)	DH2
	Pre-lambing	WL2	BL2	WT22(WL2–WD2)	LH2
	Weaning	WW2	BW2	WT23(WW2–WL2)	WH2
31–42	Pre-breeding	WP3	BP3	DW-T2(WP3–WW2)	PH3
	Pregnancy diagnosis	WD3	BP3	WT31(WD3–WP3)	DH3
	Pre-lambing	WL3	BL3	WT32(WL3–WD3)	LH3
	Weaning	WW3	BW3	WT33(WW3–WL3)	WH3

\* Liveweight at pre-breeding (WP), pregnancy diagnosis (WD), pre-lambing (WL), and weaning (WW).

<sup>†</sup>Body condition score at pre-breeding (BP), pregnancy diagnosis (BD), pre-lambing (BL), and weaning (BW).

<sup>‡</sup>Change in liveweight between successive measurements within age groups, DW-T: change in liveweight between successive measurements between age groups.

<sup>‡</sup>Height at withers at pre-breeding (PH), pregnancy diagnosis (DH), pre-lambing (LH), and weaning (WH).

determination ( $r^2$ ), mean bias, root mean squared error (RMSE), residual prediction deviation (RPD), the ratio of performance to interquartile distance (RPIQ), and percent error (RPE) (McDowell et al., 2012); overall adjusted  $R^2$  value and error metrics between models were compared based on Wilcoxon signed-ranks test (Conover, 1973; Rahe, 1974) and a two-tailed paired t-test (Kim, 2015).

## RESULTS

### Descriptive Statistics

The frequency of ewe BCS score across age group and stage of the annual production cycle is presented in [Supplementary Table A1](#). The majority of the ewes had BCS ranging from 2.5 to 3.0, while the extreme BCS scale values (1.5 and 5.0) were the least common. Within age groups, the most frequent ewe BCS at 8–18 mo was 2.5 across stages of the annual production cycle. At 19–30 mo, the most frequent ewe BCS was 3.0 across all stages of the annual production cycle except at weaning and at 31–42 mo there was no clear pattern.

Summaries of ewe LW, BCS, and HW from 8 to 42 mo of age are presented in [Table 2](#). Body condition score did not significantly change ( $P > 0.05$ ) over time and across stages of the annual production cycle, while both LW ( $P < 0.05$ ) and HW ( $P < 0.01$ ) varied with annual production cycle and increased with ewe age. Unadjusted LW continued to increase with ewe age beyond 30 mo. However,

adjusted liveweight increased with age up to 30 mo before plateauing.

### Correlation Between Liveweights and Height at Withers

The relationship between ewe liveweight (LW) and height at withers (HW) was positive but weak to moderate across age groups and stages of the annual production cycle, regardless of whether unadjusted or adjusted LW was used ([Supplementary Table A2](#)). However, a negative association between LW and HW was observed at 19–30 mo at prebreeding. There was no pattern in the strength of BCS–HW association between same and different time points.

### Correlation Between BCS and Liveweights

There was a linear association between LW and BCS in all age groups and stages of the annual production cycle, but the association was weak to moderate, regardless of whether unadjusted or adjusted LW was used ([Supplementary Table A3](#)). Furthermore, this association was comparable ( $P > 0.05$ ) for both unadjusted and adjusted LW. Both the weakest and strongest relationships were observed at weaning. The relationships, however, were strongest when liveweight and BCS measurements were from the same time point (pair of LW–BCS measurements taken at the same time) except at PL 8–18 mo, compared to when lifetime (i.e., measurements taken at different time points) records were used.

**Table 2.** Mean liveweight unadjusted and adjusted for conceptus and fleece weight (LW), height at withers (HW), and body condition score (BCS) with respective standard deviations by ewe age group and stage of annual production cycle

Age (mo)	Stage of annual production cycle	<i>n</i>	LW			HW			BCS
			Unadjusted	Adjusted	<i>P</i> -value	Unadjusted	Adjusted	<i>P</i> -value	
8–18	Pre-breeding	428	43.7 (5.61)	41.5 (5.46)	<0.001				2.8 (0.42)
	PD	429	48.8 (5.83)	45.7 (5.42)	<0.001				2.7 (0.39)
	Pre-lambing	428	52.6 (7.49)	52.0 (7.47)	0.256	0.61 (0.032)	0.58 (0.032)	0.011	2.8 (0.41)
	Weaning	429	59.7 (7.10)	58.6 (7.05)	0.016				2.8 (0.53)
19–30	Pre-breeding	427	62.8 (6.67)	59.1 (6.73)	<0.001	0.61 (0.038)	0.59 (0.038)	0.006	3.0 (0.61)
	PD	426	63.0 (7.09)	60.2 (6.74)	<0.001	0.60 (0.036)	0.58 (0.036)	0.011	3.3 (0.63)
	Pre-lambing	424	70.8 (7.70)	62.0 (6.60)	<0.001				3.2 (0.63)
	Weaning	424	66.1 (8.67)	64.2 (8.67)	0.001	0.63 (0.033)	0.59 (0.033)	0.001	2.8 (0.67)
31–42	Pre-breeding	401	68.9 (7.71)	66.4 (7.74)	<0.001				
	PD	402	70.7 (7.76)	64.8 (7.57)	<0.001	0.62 (0.047)	0.59 (0.033)	0.001	3.1 (0.63)
	Pre-lambing	399	88.8 (9.32)	64.3 (8.27)	<0.001				3.4 (0.65)
	Weaning	402	69.0 (9.74)	66.8 (9.70)	0.002	0.64 (0.033)	0.61 (0.047)	0.001	2.8 (0.78)
	<i>P</i> -value		<0.001	<0.001		<0.001	<0.001		

Values in parentheses indicate the standard deviation. Adjusted indicates that variables were corrected for fleece conceptus weight (LW) and fleece growth (LW and HW). *P*-values based on t-tests.

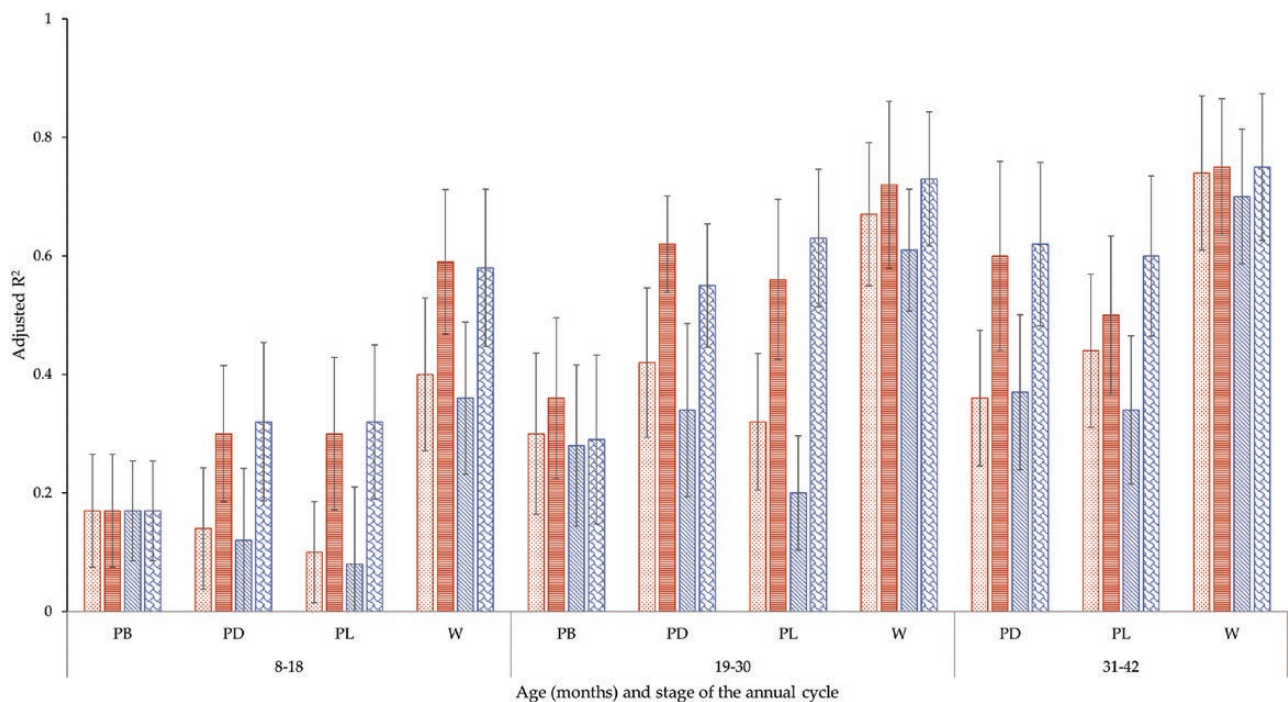
### Correlation Between BCS and Height at Withers

Generally, there was a poor linear association between ewe HW and BCS in all age groups and stages of the annual production cycle, regardless of whether unadjusted or adjusted HW (Supplementary Table A4). At any one time point, the relationship between BCS and HW did not vary ( $P > 0.05$ ) across age and stage of the annual production cycle except for 19–30 mo ewes at weaning ( $P < 0.01$ ) and 31–42 mo ewes at PL ( $P < 0.01$ ) and weaning ( $P < 0.05$ ). There was no clear pattern in the change of strength of BCS–HW association over time.

### Coefficient of Determination ( $R^2$ ) and Number of Predictors

To predict BCS at any given time, all current and previous individual liveweights (liveweight alone models) were included in linear regression equations. Separate models were formulated for unadjusted and adjusted LW (based on training dataset). The adjusted  $R^2$  values averaged across folds 0.38 (0.10 to 0.74), regardless of the time point. The adjusted  $R^2$  values were comparable ( $z = 0.37$ ,  $t_{10} = 0.56$ ,  $P > 0.05$ ) for both adjusted and unadjusted BCS prediction models across

age groups and stages of the annual production cycle (Figure 1). However, the average adjusted  $R^2$  value was greater for unadjusted than adjusted LW models ( $z = 2.40$ ,  $t_{10} = 2.23$ ,  $P < 0.05$ ). Within age groups, across stages of the annual production cycle, the adjusted  $R^2$  value varied with no clear pattern (Figure 1). There was a trend for adjusted  $R^2$  to improve at older ages, when a greater amount of previous liveweight information was known. In general, the adjusted  $R^2$  value was highest at weaning with no clear pattern in the lowest value. The average number of liveweight predictors (significant variables) for BCS prediction was comparable for models using unadjusted as well as adjusted LW (average: 6, 1 to 11) and with ewe age (Supplementary Figure A1). To improve the prediction of current BCS, the LW alone models were expanded by adding the unadjusted LW differences (change in liveweight measurements from adjacent time points) and all preceding BCS (combined unadjusted models) or by adding the adjusted LW differences (change in liveweight measurements from adjacent time points) and height at wither, and all preceding BCS (combined adjusted models). The overall proportion of variance explained (adjusted  $R^2$ ) improved ( $z = 3.62$ ,  $t_{21} = 5.71$ ,  $P < 0.001$ ) by approximately 1.3 times (from 0.38 to 0.50) in all



**Figure 1.** Adjusted coefficient of variation (adjusted  $R^2$ , with standard deviations) of models (dotted bar: unadjusted liveweight alone models; horizontal stripes: combined models based on unadjusted LW, liveweight change, and previous BCS; diagonal stripes: adjusted liveweight alone; and shingled: adjusted liveweight, liveweight change, height at withers, and previous BCS) for current BCS prediction across the stage of the annual production cycle and ewe age group. Colors (red indicates unadjusted liveweight while blue indicates adjusted liveweight was used). PB, PD, PL, and W indicate body condition score prior to pre-breeding, at pregnancy diagnosis, prior to lambing, and at weaning, respectively. In large samples where bootstrapping is applied, the standard deviations approximate the standard errors.

combined model categories compared to LW models (Figure 1). However, the adjusted  $R^2$  values were comparable ( $z = 1.07$ ,  $t_{10} = 0.99$ ,  $P > 0.05$ ) for both adjusted and unadjusted models across age groups and stages of the annual production cycle (Figure 1). Furthermore, the adjusted  $R^2$  values were marginally greater in combined models than liveweight alone models across age and stages of the annual production cycle. The highest adjusted  $R^2$  values were achieved at the weaning period with no clear pattern concerning the lowest value. The number of significant predictors for BCS was higher (average: 10, from 1 to 16 for unadjusted and 1 to 21 for unadjusted) in the combined models compared to liveweight alone models (Supplementary Figure A1). Overall, the number of predictors was increased 1.5 and 2.0 times for unadjusted and adjusted combined models, respectively, compared to LW alone models.

### Prediction Accuracy

To access the accuracy of predicting BCS, several prediction error metrics (MAE, RMSE, and

RPE) were computed. The error metrics appeared to vary across ( $P < 0.05$ ) age group but not ( $P > 0.05$ ) stage of the annual production cycle except for 19–30-mo-old ewes, when liveweight or combined models were used to predict BCS (Tables 3 and 4; Figure 2). Using adjusted LW did not affect BCS prediction accuracy ( $\pm 2SD$ ,  $P > 0.05$ ) except for the 19–30-mo-old ewes at PL. The average prediction error associated with BCS prediction in terms of MAE and RMSE were 0.38 and 0.45, and 0.32 and 0.40 body condition scores for liveweight alone and the combined models, respectively. In adjusted models, the average prediction error associated with BCS prediction in terms of MAE and RMSE was 0.37 and 0.45, and 0.33 and 0.41 body condition scores for liveweight alone and the combined models, respectively. However, combined models improved ( $z = 5.41$ ,  $t_{21} = 2.08$ ,  $P > 0.001$ ) the BCS prediction error by 10.7% (average RMSE: 0.45 vs 0.40) compared to LW alone models.

The magnitude of the BCS prediction error was moderate to high in both the liveweight and combined models, based on the smallest unit of measurement (0.5). The BCS predictions using the

**Table 3.** Coefficient of determination ( $r^2$ ), bias, root mean square error (RMSE), mean absolute error (MAE), relative prediction error (RPE), residual prediction deviation (RPD), and the ratio of performance to interquartile distance (RPIQ) based on testing data for the prediction of BCS in ewes between 8 and 42 mo by stage of the annual production cycle using unadjusted liveweight and adjusted liveweight (LW) alone models

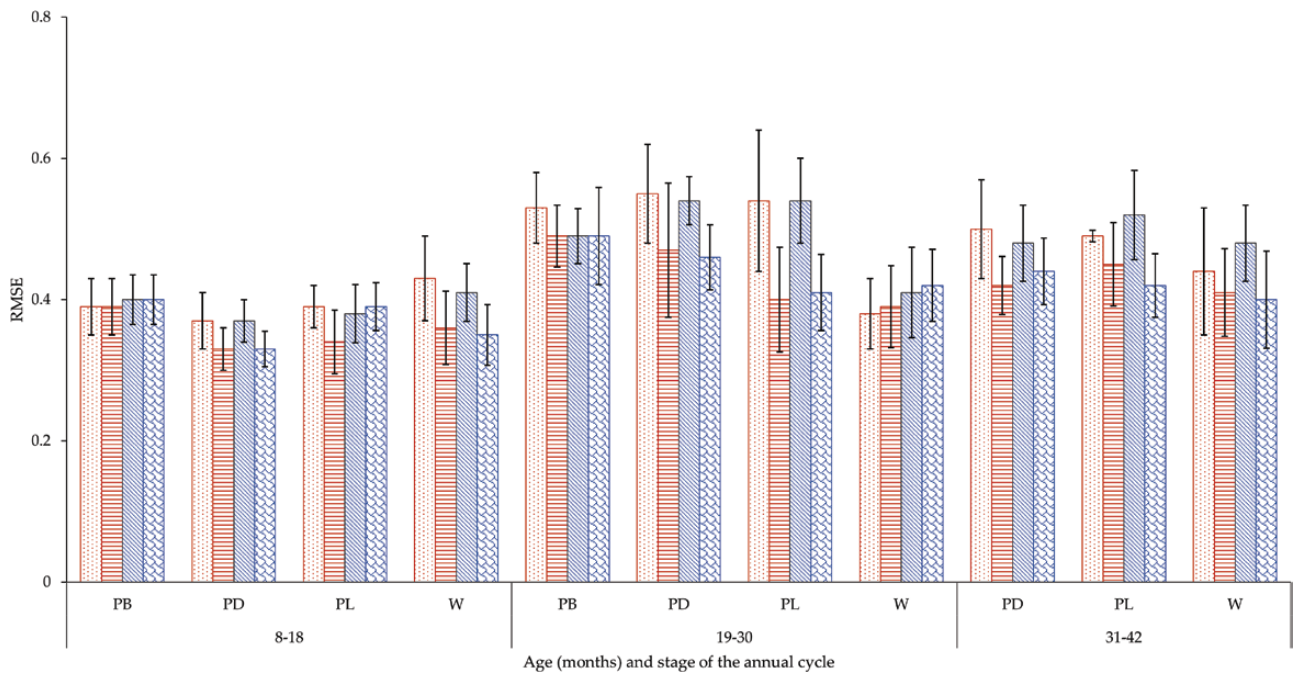
Metric	8–18				19–30				31–42		
	PB	PD	PL	W	PB	PD	PL	W	PD	PL	W
Liveweight alone <sup>1</sup> models (Unadjusted)											
$r^2$	12.90	13.89	10.30	36.70	25.50	26.61	17.50	64.02	33.20	20.33	71.10
BIAS	0.007	−0.043	0.004	−0.013	−0.05	−0.015	0.02	−0.02	0.204	−0.047	−0.152
RMSE	0.39	0.37	0.39	0.43	0.53	0.55	0.54	0.38	0.50	0.49	0.44
MAE	0.32	0.3	0.32	0.33	0.43	0.45	0.46	0.31	0.43	0.44	0.35
RPE	14.90	15.01	13.21	16.20	16.03	15.30	14.70	13.20	16.00	14.30	15.80
RPD	1.14	1.06	1.07	1.26	1.32	1.27	1.30	1.71	1.26	1.26	1.83
RPIQ	1.25	1.25	1.39	1.16	1.04	1.04	1.02	1.32	1.00	2.04	1.14
Liveweight alone <sup>2</sup> models (Adjusted)											
$r^2$	12.30	15.81	13.50	36.78	32.67	26.70	32.40	68.31	44.16	34.60	57.60
BIAS	0.006	0.019	0.003	−0.088	−0.002	−0.006	−0.008	−0.003	−0.037	0.038	0.063
RMSE	0.4	0.37	0.38	0.41	0.49	0.54	0.54	0.41	0.48	0.52	0.48
MAE	0.32	0.30	0.31	0.33	0.43	0.44	0.43	0.32	0.41	0.43	0.37
RPE	10.6	12.95	13.38	14.75	16.33	16.12	16.72	14.64	15.53	15.25	17.2
RPD	1.08	1.09	1.07	1.24	1.23	1.17	1.19	1.79	1.35	1.25	1.53
RPIQ	1.25	1.35	1.32	1.22	2.04	1.39	0.93	1.22	2.08	1.92	1.04

PB, PD, PL, and W indicate the four stages of the annual production cycle including pre-breeding, pregnancy diagnosis, pre-lambing, and weaning, respectively. Interpretation of measures: The best model has the highest  $r^2$ , RPD, and RPIQ, and the lowest RMSE and RPE. Ranges for values:  $r^2$  (0 indicates that the model accounts for none of the variability of the response data around its mean and 1.0 indicates that the model accounts for all the variability). RPD ( $< 1.4$ : Weak,  $1.4 < RPD < 2.0$ : Reasonable,  $> 2.0$ : Excellent). RPIQ ( $< 1.4$ : Very poor,  $1.4 < RPIQ < 1.7$ : Fair,  $1.7 < RPIQ < 2.0$ : Good,  $2.0 < RPIQ < 2.5$ : Very good,  $> 2.5$ : Excellent). Superscripts 1 and 2 indicate model based on unadjusted or adjusted liveweight, respectively. Bias (positive value indicates overestimation; negative sign indicates underestimation). Adjusted indicates that a model was based on liveweight corrected for conceptus and fleece weight.

**Table 4.** Coefficient of determination ( $r^2$ ), bias, root mean square error (RMSE), mean absolute error (MAE), relative prediction error (RPE), residual prediction deviation (RPD), and the ratio of performance to interquartile distance (RPIQ) based on testing data for the prediction of BCS in ewes between 8 and 42 mo by stage of the annual production cycle using unadjusted and adjusted combined models

Metric	8-18				19-30				31-42		
	PB	PD	PL	W	PB	PD	PL	W	PD	PL	W
Combined <sup>1</sup> models (Unadjusted)											
$r^2$	12.9	31.3	26.6	51.5	29.9	55.1	58.3	68.8	54.7	54	71
BIAS	0.007	-0.025	-0.005	0.009	0.051	-0.038	0.007	0.065	-0.014	-0.155	0.011
RMSE	0.39	0.33	0.34	0.36	0.49	0.47	0.40	0.39	0.42	0.45	0.41
MAE	0.32	0.27	0.28	0.29	0.41	0.31	0.31	0.31	0.34	0.35	0.31
RPE	14.90	12.13	11.97	12.86	16.69	12.18	12.42	14.08	13.64	13.20	14.96
RPD	1.14	1.20	1.18	1.42	1.20	1.48	1.56	1.73	1.50	1.49	1.84
RPIQ	1.25	1.52	1.47	1.39	2.00	1.83	2.50	1.28	1.19	1.11	1.22
Combined <sup>2</sup> models (Adjusted)											
$r^2$	13.0	31.9	18.7	53.7	36.4	55.5	57.1	67.2	51.7	57.8	71.3
BIAS	0.006	-0.001	-0.002	0.021	-0.01	0.005	0.054	-0.034	0.043	0.031	-0.054
RMSE	0.40	0.33	0.39	0.35	0.49	0.46	0.41	0.42	0.44	0.42	0.40
MAE	0.33	0.28	0.27	0.30	0.43	0.35	0.32	0.32	0.34	0.34	0.31
RPE	10.6	11.76	13.64	12.68	16.33	13.65	12.58	15.16	14.15	12.35	14.55
RPD	1.08	1.24	1.08	1.47	1.25	1.34	1.54	1.79	1.40	1.55	1.85
RPIQ	1.25	1.56	1.28	1.43	2.04	1.63	1.22	1.19	2.27	2.38	1.25

PB, PD, PL, and W indicate the four stages of the annual production cycle including pre-breeding, pregnancy diagnosis, pre-lambing, and weaning, respectively. Interpretation of measures: The best model has the highest  $r^2$ , RPD, and RPIQ, and the lowest RMSE and RPE. Ranges for values:  $r^2$  (0 indicates that the model accounts for none of the variability of the response data around its mean and 1.0 indicates that the model accounts for all the variability). RPD (< 1.4: Weak, 1.4 < RPD < 2.0: Reasonable, > 2.0: Excellent). RPIQ (< 1.4: Very poor, 1.4 < RPIQ < 1.7: Fair, 1.7 < RPIQ < 2.0: Good, 2.0 < RPIQ < 2.5: Very good, > 2.5: Excellent). Unadjusted indicates that models were based on all previous and current crude and previous liveweights, liveweight changes, and previous BCS). Adjusted indicates that models were based on all previous and current liveweights and liveweight changes corrected for conceptus and fleece weight, adjusted height at withers, and previous BCS. Superscripts 1 and 2 indicate without and with adjusted HW in the model, respectively. Bias (positive value indicates overestimation; negative sign indicates underestimation).



**Figure 2.** Root mean square error (RMSE with standard deviations) of models (dotted bar: unadjusted liveweight alone models; horizontal stripes: combined models based on unadjusted LW, liveweight change, and previous BCS; diagonal stripes: adjusted liveweight alone; and shingled: adjusted liveweight, liveweight change, height at withers, and previous BCS) for current BCS prediction across the stage of the annual production cycle and ewe age group. Colors (red indicates unadjusted liveweight while blue indicates adjusted liveweight was used). PB, PD, PL, and W indicate body condition score prior to prebreeding, at pregnancy diagnosis, prior to lambing, and at weaning, respectively.

unadjusted liveweight alone and combined models were, on average, 15.4% and 13.5%, respectively, from the actual value. In adjusted models, the predictions deviated by 15.9% and 13.4%, respectively, for LW alone and combined models. Therefore, combined models improved the BCS prediction error prevalence by 9.6% compared to LW alone models.

Models were categorized regardless of model type as weak (RPD: 1.06 to 1.35) or very poor to fair (RPIQ: 1.47 to 1.85). There was inconsistency in the BCS prediction model performance where a model with relatively good RPD ( $>1.4$ ) had a poor RPIQ ( $<1.4$ ) and vice versa. Using adjusted LW or unadjusted LW did not affect ( $P > 0.05$ ) both model RPD and RPIQ metrics. However, both RPD and RPIQ were improved ( $P < 0.05$ ) by 10% to 16% in the combined than LW alone models.

## DISCUSSION

The aim of this study was to explore the possibility of improving the prediction accuracy of BCS using a ewe's production characteristics as they aged from 8 to approximately 42 months. This was a follow-up study to Semakula et al. (2020b). Previously, using a different dataset, the relationship between liveweight and BCS at a given time point, and the possibility of using a linear combination of a ewe's unadjusted lifetime LW, liveweight change, and previous BCS data to predict BCS at a given time, was examined (Semakula et al., 2020a; Semakula et al., 2020b). Weak to moderate levels of BCS prediction accuracy were achieved. It was then postulated that if corrected liveweights (corrected for conceptus and fleece weight) and wither height (corrected for fleece length) data were used, BCS prediction accuracy would be improved.

In this study, the majority of the ewes had BCS between 2.5 and 3.0 which falls within the recommended BCS range (2.5–3.5) for optimal productivity (Kenyon et al., 2014). Additionally, there were few thin or obese ewes in the 8- to 18-mo-old group. These observations combined indicate that ewes were supplied with sufficient nutritional requirements through their first reproductive cycle. Furthermore, this study demonstrated unadjusted LW continued to increase beyond 30 mo of age but adjusted LW (adjusted for conceptus and fleece weight) plateaued. The observed trend in adjusted LW corroborates an earlier study which reported that mature Romney ewe weight was achieved by 33 mo (Pettigrew et al., 2019). It appears that the confounding effects of conceptus and fleece weight

increase with age, causing the apparent increase in weight unadjusted LW.

It was not clear why the relationship between LW and HW was negative for 18–30-mo-old ewes at prebreeding. Prior to breeding, farmers enhance their feeding strategies in a process known as flushing to ensure as many ewes reach the required breeding weight regardless of their frame sizes (Kenyon et al., 2011). Given that fact that this was the same cohort of ewes, it is possible that changes in nutritional effects could have randomly altered the relationship between LW and HW. With the moderate strength of association between LW and HW, height at withers was expected to significantly affect the relationship between LW and BCS. However, HW was poorly correlated with BCS. There was a weak to moderate correlation between LW and BCS as reported by Semakula et al. (2020a).

The observation that LW alone models were not as good as combined ones and, thus, likely to be unreliable in predicting future BCS based on linear regression, corroborate our previous findings (Semakula et al., 2020b). The variability in BCS explained for both liveweight and combined models increased with the number of predictors in the model. This was expected as it is known that as the number of predictors that significantly relate to the dependent variable increases, the proportion of the variance due to the regression increases (Li, 2017). However, in this study, a considerable amount of variability in BCS ( $0.26 \leq R^2 \leq 0.83$  and  $0.25 \leq R^2 \leq 0.72$ ) remained unaccounted for in both liveweight alone models and combined models, respectively. Some of the reasons for the apparent failure for both liveweight alone and combined models to account for more of the variability in BCS include 1) assessor consistency over time, 2) losses in liveweight due to gut-fill and urination when ewes were weighed at different times, and 3) confounding effects of fleece weight and conceptus weight (Semakula et al., 2020b). The consistency between BCS assessors varies between 5% to 27% and 40% to 60%, and within assessors from 16% to 44% and 60% to 90% for inexperienced and experienced assessors, respectively (Kenyon et al., 2014). In the current study, a single experienced assessor (with more than 30 yr of experience in BCS assessment) was used to determine all BCS to ensure consistency. It is, therefore, unlikely that the data used in this study were affected by assessor reliability. Liveweight losses resulting from fluctuations in gut-fill can account for between 5% and 23% of total liveweight in ruminants (Hughes,

1976; Moyo and Nsahlai, 2018). Thus, the duration between feeding and recording an individual's liveweight can influence the accuracy of the liveweight. Further, ewe fleece weight, pregnancy, and lambing data were collected and used to correct LW. Given that standard equations, with little known error rates and repeatability were used to adjust liveweight, it is possible that these strategies could have introduced some error cancelling the effect of adjusting for LW confounders. The study also did not measure individual time off feed prior to weighing, a function that many electronic weighing systems now have the potential to account for. Future studies should examine if the accuracy of BCS prediction can be improved by accounting for gut-fill fluctuations. In regression models, all residual errors are assumed to be contributed by the predictors, and thus, any inaccuracies in their measurement should be of concern (Dosne et al., 2016). Losses in liveweight due to gut-fill changes and urination in relation to when ewes were weighed and the effect of pregnancy on liveweight are therefore of concern, as they affect liveweight, a key variable for BCS prediction. When predictor variables are imprecise, estimations based on the standard model assumptions can lead to inaccurate parameter estimates even when large samples are used (Hausman, 2001; Pischke, 2007). Therefore, if errors in the measurement of liveweight could be minimized, the goodness-of-fit and accuracy of BCS prediction models should increase. In delayed weighing, accounting for liveweight losses with respect to time of delay (the duration from when the animal last fed to weight recording) using prediction equations offers a practical solution. These time-dependent, liveweight adjusting equations for ewes have been developed but are not regularly used (Burnham et al., 2009; Wishart et al., 2017).

The BCS prediction models using liveweight alone had larger error (MAE and RMSE) and lower RPD and RPIQ values, compared to combined models which led to high relative error prevalence (RPE). Combined models reduced the magnitude of all the prediction error metrics but were greater than those observed in our previous study (Semakula et al., 2020b). The model BCS prediction percentage error (RPE) was above the desired 10% (Hagerman et al., 2017; Lalic et al., 2018). The large BCS prediction error values (60% to 108% of the smallest unit on a 0.5 decimal scale) in the present study (where a 0.5-unit change in BCS changes the performance rank of a ewe) could lead to inaccurately predicted BCS values, thereby, greatly influencing management decisions. Ideally,

all prediction models should have had resolutions as low as 0.02 body condition scores. However, due to the intractable discrete nature of the BCS scale used, such resolutions cannot be achieved (Semakula et al., 2020b). It has been suggested that decisions concerning strategic feeding and management of ewes to maximize performance should be based on a critical range of BCS values (i.e., 2.5 to 3.5; Kenyon et al., 2014). The predictions found in this study may, therefore, overestimate or underestimate measures by 0.33 to 0.54 BCS, which could substantially change the ranking of a ewe, leading to less robust management decisions, which in turn could reduce flock productivity. The greater BCS prediction error than reported in our previous study (Semakula et al., 2020b) could be explained by the smaller sample size used in the current study leading to greater variability in the outcome and predictor variable measurements.

The findings suggest that using quantitative traits (physical and linear morphometric measurements) may not be sufficient to predict sheep BCS on a full range scale (1.0–5.0). Therefore, further studies using data such as image analysis (computed tomography: CT scans and dual-energy x-ray absorptiometry: DXA), and automated metabolic profiles, to account for individual animal variability may be warranted. Where a narrow range of BCS such as 1.0–3.0 is acceptable, further research should look at extending machine learning algorithms across all age groups and stages of the annual production cycle. Given the limitations of predicting BCS, itself a predictor of body composition. It would be worthwhile investigating how accurately liveweights and other predictors would predict total body fat and muscle weights, or proportions given they are more objective and continuous variables. The first step in these types of studies would require animals to be euthanized and/or tools such as CT scans.

## CONCLUSION

The combined models improved the proportion of variability in BCS that could be accounted for, as well as the accuracy metrics across all age groups and stages of the annual production cycle and over time (years), compared to the liveweight alone models. Using ewe data to correct LW (correct for fleece weight and conceptus weight) and height at withers as an additional predictor did not offer better model accuracy. The most common way of determining BCS is through a direct hands-on method; however, if it is not possible, the equations generated

in the current and previous study (Semakula et al., 2020b) could be used to predict BCS. These equations could potentially be incorporated in electronic weighing systems that utilize lifetime data, especially in large extensively run sheep flocks. However, the 30% to 90% variability in BCS that was unaccounted for, even in the combined models, coupled with the large prediction error associated with our equations dictates that they should be used with caution. Additional ways of accounting for individual variability in BCS could ameliorate the accuracy of BCS and warrant investigation.

### SUPPLEMENTARY DATA

Supplementary data are available at *Translational Animal Science* online.

### ACKNOWLEDGMENTS

The authors would like to thank Dr Emma Pettigrew, Ms Emmanuelle Haslin and Mr Dean Burnham for helping in data collection during the study.

*Conflict of interest statement.* The authors declare no conflict of interest with the research presented in this manuscript.

### LITERATURE CITED

- Besier, R. B., and D. Hopkins. 1989. Farmers' estimations of sheep weights to calculate drench dose. *J. Dep. Agric. West. Aust.* 1989(30):120–121. Available at: [https://researchlibrary.agric.wa.gov.au/journal\\_agriculture4/vol30/iss3/11](https://researchlibrary.agric.wa.gov.au/journal_agriculture4/vol30/iss3/11)
- Botchkarev, A. 2019. Performance metrics (error measures) in machine learning regression, forecasting, and prognostics: properties and typology. *Interdiscip. J. Inf. Knowl. Manag.* 14:45–79. doi: [10.28945/4184](https://doi.org/10.28945/4184)
- Brown, D. J., D. B. Savage, G. N. Hinch, and S. Hatcher. 2015. Monitoring liveweight in sheep is a valuable management strategy: a review of available technologies. *Anim. Prod. Sci.* 55(4):427–436. doi: [10.1071/AN13274](https://doi.org/10.1071/AN13274)
- Burke, J. A., P. L. Nuthall, and A. E. McKinnon. 2004. An analysis of the feasibility of using image processing to estimate the live weight of sheep. FHM research report. Canterbury, New Zealand: Lincoln University, . p. 1174–8796.
- Burnham, D. L., P. C. H. Morel, P. R. Kenyon, S. T. Morris, and K. J. Stafford. 2009. The effect of fasting on live weight in hogget wethers and pregnant mature ewes: Brief Communication. *Proc. N.Z. Soc. Anim. Prod.* 69:112–114.
- Coates, D. B., and P. Penning. 2000. Measuring animal performance. In: L. t. a. R. M. Jones, editor, *Field and laboratory methods for grassland and animal production research*. Wallingford, UK: CABI Publishing. p. 353–402.
- Conover, W. J. 1973. On Methods of Handling Ties in the Wilcoxon Signed-Rank Test. *J. Am. Stat. Assoc.* 68(344):985–988. doi: [10.2307/2284536](https://doi.org/10.2307/2284536)
- Corner-Thomas, R. A., P. R. Kenyon, S. T. Morris, A. L. Ridler, R. E. Hickson, A. W. Greer, C. M. Logan, and H. T. Blair. 2016. Brief communication: The use of farm-management tools by New Zealand sheep farmers: changes with time. *Proc. N.Z. Soc. Anim. Prod.* 76:78–80.
- Cottle, D., and D. Pacheco. 2017. Prediction of fleece insulation after shearing and its impact on maintenance energy requirements of Romney sheep. *Small. Ruminant. Res.* 157:14–22. doi: [10.1016/j.smallrumres.2017.10.004](https://doi.org/10.1016/j.smallrumres.2017.10.004)
- Dosne, A. G., M. Bergstrand, and M. O. Karlsson. 2016. A strategy for residual error modeling incorporating scedasticity of variance and distribution shape. *J. Pharmacokinet. Pharmacodyn.* 43:137–151. doi: [10.1007/s10928-015-9460-y](https://doi.org/10.1007/s10928-015-9460-y).
- Freer, M., H. Dive, and J. V. Nolan. 2007. *Nutrient requirements of domesticated ruminants*. Collingwood, Australia: CSIRO publishing.
- Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33:1–22.
- Hagerman, A. D., J. M. Thompson, C. Ham, and K. K. Johnson. 2017. Replacement beef cow valuation under data availability constraints. *Front. Vet. Sci.* 4:185. doi: [10.3389/fvets.2017.00185](https://doi.org/10.3389/fvets.2017.00185).
- Hausman, J. 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *J. Econ. Perspect.* 15(4):57–67. doi: [10.1257/jep.15.4.57](https://doi.org/10.1257/jep.15.4.57)
- Holman, B. W. B., A. Kashani, and A. E. O. Malau-Aduli. 2012. Growth and body conformation responses of genetically divergent Australian sheep to *Spirulina* (*Arthrospira platensis*) supplementation. *Am. J. Exp. Agric* 2:160–173. doi: [10.9734/AJEA/2012/992](https://doi.org/10.9734/AJEA/2012/992)
- Hughes, J. G. 1976. Short-term variation in animal live weight and reduction of its effect on weighing. *Anim. Breed. Abstr.* 44(3):11.
- Jefferies, B. 1961. Body condition scoring and its use in management. *Tasman. J. Agric.* 1961(32):19–21.
- Jones, A., A. J. van Burgel, R. Behrendt, M. Curnow, D. J. Gordon, C. M. Oldham, I. J. Rose, and A. N. Thompson. 2011. Evaluation of the impact of Lifetimewool on sheep producers. *Anim. Prod. Sci.* 51(9):857–865. doi: [10.1071/EA08303](https://doi.org/10.1071/EA08303)
- Kenyon, P. R., S. K. Maloney, and D. Blache. 2014. Review of sheep body condition score in relation to production characteristics. *N. Z. J. Agric. Res.* 57(1):38–64. doi: [10.1080/00288233.2013.857698](https://doi.org/10.1080/00288233.2013.857698)
- Kenyon, P. R., S. T. Morris, D. L. Burnham, and D. M. West. 2008. Effect of nutrition during pregnancy on hogget pregnancy outcome and birthweight and liveweight of lambs. *N. Z. J. Agric. Res.* 51(2):77–83. doi: [10.1080/00288230809510437](https://doi.org/10.1080/00288230809510437)
- Kenyon, P. R., S. J. Pain, P. G. Hutton, C. M. C. Jenkinson, S. T. Morris, S. W. Peterson, and H. T. Blair. 2011. Effects of twin-bearing ewe nutritional treatments on ewe and lamb performance to weaning. *Anim. Prod. Sci.* 51(5):406–415. doi: [10.1071/AN10184](https://doi.org/10.1071/AN10184)
- Kim, T. K. 2015. T test as a parametric statistic. *Korean J. Anesthesiol.* 68:540–546. doi: [10.4097/kjae.2015.68.6.540](https://doi.org/10.4097/kjae.2015.68.6.540).
- Kuhn, M. 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28(5):1–26. (Article) doi: [10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05)
- Lalic, B., J. Eitzinger, A. Dalla Marta, S. Orlandini, A. F. Sremac, and B. Pacher. 2018. *Agricultural Meteorology and Climatology*. Firenze, Italy: Firenze University Press.

- Li, J. 2017. Assessing the accuracy of predictive models for numerical data: Not  $r$  nor  $r^2$ , why not? Then what? *Plos One* 12:e0183250. doi:[10.1371/journal.pone.0183250](https://doi.org/10.1371/journal.pone.0183250).
- Martins, B., A. Mendes, L. Silva, T. Moreira, J. Costa, P. Rotta, M. Chizzotti, and M. Marcondes. 2020. Estimating body weight, body condition score, and type traits in dairy cows using three dimensional cameras and manual body measurements. *Livest Sci.* 236: 10454. doi: [10.1016/j.livsci.2020.104054](https://doi.org/10.1016/j.livsci.2020.104054)
- McDowell, M. L., G. L. Bruland, J. L. Deenik, and S. Grunwald. 2012. Effects of subsetting by carbon content, soil order, and spectral classification on prediction of soil total carbon with diffuse reflectance spectroscopy. *Appl. Environ. Soil. Sci.* 2012, Article ID 294121, 14 pages, 2012. doi:[10.1155/2012/294121](https://doi.org/10.1155/2012/294121)
- McHugh, N., F. McGovern, P. Creighton, T. Pabiou, K. McDermott, E. Wall, and D. P. Berry. 2019. Mean difference in live-weight per incremental difference in body condition score estimated in multiple sheep breeds and crossbreds. *Animal* 13:549–553. doi: [10.1017/S1751731118002148](https://doi.org/10.1017/S1751731118002148).
- Morris, S. T., P. R. Kenyon, and D. L. Burnham. 2002. A comparison of two scales of body condition scoring in Hereford x Friesian beef breeding cows. *Proc. N. Z. Grassl. Assoc.* 64:121–123.
- Moyo, M., and I. Nsahlai. 2018. Rate of Passage of Digesta in Ruminants; Are Goats Different? In: S. Kukovics, editor, *Goat Science No. Open access peer-reviewed Edited Volume*. IntechOpen. p. 39–74.
- Norman, G. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Adv. Health Sci. Educ. Theory Pract.* 15:625–632. doi:[10.1007/s10459-010-9222-y](https://doi.org/10.1007/s10459-010-9222-y).
- Pettigrew, E., R. Hickson, H. Blair, K. Griffiths, A. Ridlera, S. Morriss, and P. Kenyon. 2018. Differences in birth weight and neonatal survival rate of lambs born to ewe hoggets or mature ewes. *NZ. J. Anim. Sci. Prod.* 78:16–20. doi: [10.1080/00288233.2020.1713177](https://doi.org/10.1080/00288233.2020.1713177)
- Pettigrew, E. J., R. E. Hickson, S. T. Morris, N. Lopez-Villalobos, S. J. Pain, P. R. Kenyon, and H. T. Blair. 2019. The effects of birth rank (single or twin) and dam age on the lifetime productive performance of female dual purpose sheep (*Ovis aries*) offspring in New Zealand. *Plos One* 14:e0214021. doi:[10.1371/journal.pone.0214021](https://doi.org/10.1371/journal.pone.0214021).
- Pischke, S. 2007. Lecture notes on measurement error. London: London School of Economics. Available at: [https://economics.lse.ac.uk/staff/spischke/ec524/Merr\\_new.pdf](https://economics.lse.ac.uk/staff/spischke/ec524/Merr_new.pdf), Accessed on 16/11/2018.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, R version 3.4.4 (2018-03-15), Vienna, Austria. Available at: <https://cran.r-project.org>.
- Rahe, A. J. 1974. Tables of critical values for the Pratt matched pair signed rank statistic. *J. Am. Stat. Assoc.* 69(346):368–373. doi: [10.2307/2285660](https://doi.org/10.2307/2285660).
- Russel, A. J. F., J. M. Doney, and R. G. Gunn. 1969. Subjective assessment of body fat in live sheep. *J. Agric. Sci.* 72(3):451–454. doi: [10.1017/S0021859600024874](https://doi.org/10.1017/S0021859600024874).
- Semakula, J., R. A. Corner-Thomas, S. T. Morris, H. T. Blair, and P. R. Kenyon. 2020a. The Effect of Age, Stage of the Annual Production Cycle and Pregnancy-Rank on the Relationship between Liveweight and Body Condition Score in Extensively Managed Romney Ewes. *Animals* 10(5):784. doi: [10.3390/ani10050784](https://doi.org/10.3390/ani10050784).
- Semakula, J., R. A. Corner-Thomas, S. T. Morris, H. T. Blair, and P. R. Kenyon. 2020b. Predicting Ewe Body Condition Score Using Lifetime Liveweight and Liveweight Change, and Previous Body Condition Score Record. *Animals* 10(7):1182. doi: [10.3390/ani10071182](https://doi.org/10.3390/ani10071182).
- Semakula, J., R. A. Corner-Thomas, S. T. Morris, H. T. Blair, and P. R. Kenyon. 2021. Application of Machine Learning Algorithms to Predict Body Condition Score from Liveweight Records of Mature Romney Ewes. *Agriculture* 11(2):162. doi: [agriculture11020162](https://doi.org/10.3390/agriculture11020162).
- Sezenler, T., M. Özder, M. Yildirim, A. Ceyhan, and M. A. Yüksel. 2011. The relationship between body weight and body condition score in some indigenous sheep breeds in Turkey. *J. Anim. Plant. Sci.* 21(3):443–447.
- Theil, H. 1958. *Economic policy and forecasts*. Assisted by J.S. Cramer. H. Moerman and A. Russchen, Contributions to economic analysis. Amsterdam: North-Holland Pub.Co. p. 562.
- Triguero, I., S. del Río, V. López, J. Bacardit, J. M. Benítez, and F. Herrera. 2015. ROSEFW-RF: the winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem. *Knowl-Based. Syst.* 87:69–79. doi: [1854/LU-7066581](https://doi.org/10.1016/j.kbsys.2015.06.013).
- Vieira, A., S. Brandão, A. Monteiro, I. Ajuda, and G. Stilwell. 2015. Development and validation of a visual body condition scoring system for dairy goats with picture-based training. *J. Dairy Sci.* 98:6597–6608. doi:[10.3168/jds.2015-9428](https://doi.org/10.3168/jds.2015-9428).
- Wishart, H., C. Morgan-Davies, A. Stott, R. Wilson, and T. Waterhouse. 2017. Liveweight loss associated with handling and weighing of grazing sheep. *Small. Ruminant. Res.* 153:163–170. doi: [10.1016/j.smallrumres.2017.06.013](https://doi.org/10.1016/j.smallrumres.2017.06.013).