

# Stability of genetic predictions for stayability using random regression models that include end points beyond 6 yr of age

Miguel A. Sánchez-Castro,<sup>✉</sup> Milton G. Thomas, R. Mark Enns, and Scott E. Speidel<sup>1,\*</sup>

<sup>\*</sup>Department of Animal Sciences, Colorado State University, Fort Collins, CO 80523-1171

---

© The Author(s) 2019. 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 Non-Commercial 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. 2019.3:1678–1682  
doi: 10.1093/tas/txz056

## INTRODUCTION

Stayability (STAY) in beef cattle, has been traditionally defined as the probability that a cow will remain in the herd until 6 yr of age, given she first calved as a 2-yr-old (Brigham et al., 2007). Cows that stay in production longer, benefit profitability of the herd by reducing the need of female replacements, decreasing the incidence of dystocia and increasing the average weaning weight of marketed calves (Garrick, 2006). Sire genetic evaluations for STAY are challenging due to the low heritability of the trait and the considerably long time (e.g., 6 yr) required to collect observations from the female progeny (Speidel et al., 2018). This delay in collection of phenotypes reduces the accuracy of sire's genetic predictions at early ages and therefore slows genetic progress. Traditional evaluations for STAY are typically performed with threshold models that only consider the success or failure of daughters reaching the age of 6 yr, explicitly ignoring the information from the female progeny that are not yet 6 yr of age or that are still producing beyond that age end point. Given these deficiencies, the application of a statistical approach designed to use information from both earlier and later ages is expected to improve the accuracy of sire evaluation. Within this context, random regression models (RRM) have been successfully applied to STAY data,

since binary observations can be assigned to any discrete point in time during a cow's lifetime and expected progeny differences (EPD) with higher accuracies can be generated for any particular age (Jamrozik et al., 2013). Prior research efforts using RRM have included earlier age end points in the estimation of an aggregated 6-yr STAY genetic prediction (Sánchez-Castro et al., 2017); however, predictions using regression equations have a tendency to be variable, particularly at the ends of the prediction range. For traits such as STAY, there is potential to extend the age end points beyond 6 yr of age in order to increase the stability of the 6-yr STAY EPD in comparison with the increases in accuracy associated with RRM that finalize their prediction range at the age of 6 yr. Therefore, the objectives of this study were to determine the efficacy associated with extending evaluated age end points beyond 6 yr of age in an effort to stabilize the prediction of 6-yr STAY and evaluate the accuracy gains using various RRM when compared to a traditional threshold evaluation.

## MATERIALS AND METHODS

The data used in this study were obtained from an existing database; therefore, the study was not subjected to animal care and use committee approval.

### *Data Collection and Editing*

Calving performance data consisting of 1,301 Angus females (progeny of 231 sires and 817

---

<sup>1</sup>Corresponding author: [scott.speidel@colostate.edu](mailto:scott.speidel@colostate.edu)

Received April 4, 2019.

Accepted May 14, 2019.

dams) collected from 1993 to 2016 at the Colorado State University Beef Improvement Center (CSU-BIC) were used for the study. Starting from the third calving, the value of 1 (successful) or 0 (unsuccessful) was attributed to cows that, respectively, calved or not. Consequently, a total of ten STAY end points were defined for the study, ranging from STAY03 through STAY12, with the total number of observations for the largest data set being 10,147.

### Statistical Analysis

Traditional EPD calculation for STAY06 was performed using a univariate BLUP threshold animal model along with a probit link function to convert binary observations to an underlying normal distribution. The model equation was:

$$y = Xb + Zu + e$$

where  $y$  corresponded to a vector of transformed observations of STAY06 on the underlying scale,  $b$  was a vector of unknown solutions for fixed effects, which included contemporary group (defined as a combination of dam and calf birth year),  $u$  corresponded to a vector of unknown solutions of animal random effects.  $X$  and  $Z$  were known incidence matrices relating observations in  $y$  to both fixed and random effects, and  $e$  was the vector of unknown residual errors. For this model, variances were assumed to be:

$$\text{Var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where  $A$  represents the Wright's numerator relationship matrix,  $I$  is an identity matrix and  $\sigma_a^2$  and  $\sigma_e^2$  are the additive and residual variances, respectively. As explained by Brigham et al. (2007), the additive variance ( $\sigma_a^2$ ) was unique for the evaluated age end point and the residual variance ( $\sigma_e^2$ ) was constrained to be equal to 1 in accordance to the specifications of the maximum "a posteriori" probit threshold model.

In addition, all STAY end points were evaluated using seven linear RRM with Legendre polynomials as their base function. In general, the only difference among the models was the age at the end of the prediction range (6, 7 and up to 12 yr of age); however, for all the models the equation was:

$$y = Xb + Qu + e$$

where  $y$  corresponded to a vector of STAY observations,  $b$  was a vector of unknown solutions for

fixed effects that included contemporary group (combination between dam and calf birth year) and a linear fixed regression,  $u$  corresponded to a vector of unknown solutions of random regressions for animal additive direct genetic effects.  $X$  and  $Q$  were known incidence matrices relating observations in  $y$  to both fixed and random effects, and  $e$  was the vector of unknown residual errors. Variances assumed for the models were:

$$\text{Var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} A \otimes G & 0 \\ 0 & R \end{bmatrix}$$

where  $A$  represents the Wright's numerator relationship matrix,  $G$  corresponds to a variance-covariance matrix of additive genetic random regression coefficients and,  $R = \text{diag}\{\sigma_{e_k}^2\}$  representing a diagonal matrix of temporary environmental variances that themselves vary depending on the  $k^{\text{th}}$  age end point. The model predicted the genetic merit of the presence of a weaned calf at each particular age end point; therefore, EPD were summed to obtain the individual's genetic merit for the presence of a calf at 3, 4, and up to 12 yr of age, depending on the RRM used. Accuracy estimations were performed according to the guidelines of the Beef Improvement Federation (2010). All the analyses were performed using ASREML 3.0 (Gilmour et al., 2009) and the Animal Breeder's Tool Kit (Golden et al., 1992).

## RESULTS AND DISCUSSION

The percentage of dams receiving a successful observation at each of the STAY end points included in the study is shown in Figure 1. The tendency of the average percentage of STAY to consecutive calvings is clearly negative, which was consistent with previous reports (Jamrozik et al., 2013) and the biology of cow production in general. Pearson and Spearman's correlations, as well as the regression coefficients of EPD for 6-yr STAY obtained with RRM that included age end points beyond 6, on the predictions obtained with the threshold animal model and the RRM that included only data up to 6 yr are shown in Tables 1 and 2, respectively. Overall, predictions among the threshold animal model and all the RRM were moderate and positively correlated, implying that EPD obtained with both methodologies were similar. Regarding to the Spearman's correlations, a moderate degree of re-ranking of individuals occurred between both methodologies. Similar results were reported by Lewis and Brotherstone (2002) when comparing the estimated breeding values for growth traits in

sheep obtained with random regression techniques with estimated breeding values obtained with a traditional method (univariate animal model) based on a Gompertz form.

With respect to the regressions of predictions obtained with the RRM on the traditional method, an underestimation of the genetic merit for STAY06 occurred with the traditional threshold model when compared to the RRM. Similar results were reported by Sánchez-Castro et al. (2017) when comparing EPD for STAY at consecutive ages. Among the seven RRM implemented in this study, all of them had essentially the same predictive power for the 6-yr EPD for STAY, since all correlations and regression coefficients obtained for these models were close to 1. This

result was expected as there was no age variability for STAY06, there were no missing records, and the analyses were not performed with independent data sets. Nobre et al. (2003) reported similar EPD estimated for birth weight (BW) using a multiple-trait model (MTM) and RRM; basically, authors explained that with an equal amount of information, and no age variability associated with BW, it was possible to estimate numerically accurate solutions and functionally identical parameters that lead to the MTM and RRM to provide identical results.

Regarding accuracy differences of STAY06 EPD predictions using various RRM when compared to a traditional threshold evaluation, results of this study suggest that mean accuracy for all RRM was considerably higher than the one obtained using an animal threshold model (Figure 2). The origin of this difference relies on the fact that different amounts of information are being incorporated with each method, in the case of the threshold animal model, only 6-yr-old females are being considered as they alone have an observation, whereas for the RRM, all available observations within the specific prediction range of each model are being incorporated into the analysis. In this context, Bohmanova et al. (2005) suggested that the accuracy of RRM increases when additional records were incorporated and Meyer (2004) reported that accuracies obtained

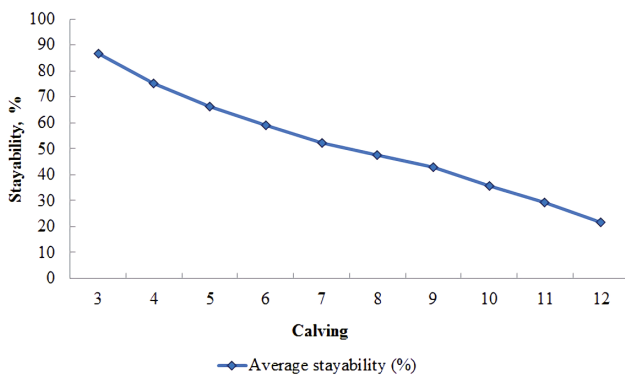


Figure 1. Average stayability (%) to consecutive calvings.

**Table 1.** Pearson correlations (above diagonal) and Spearman's rank correlations (below diagonal) of predictions for 6-yr stayability obtained with different statistical models

	TRAD	RRM06	RRM07	RRM08	RRM09	RRM10	RRM011	RRM12
TRAD		0.59	0.59	0.58	0.59	0.59	0.59	0.60
RRM06	0.65		0.99	0.99	0.99	0.99	0.99	0.98
RRM07	0.65	0.99		0.99	0.99	0.99	0.99	0.99
RRM08	0.64	0.99	0.99		0.99	0.99	0.99	0.99
RRM09	0.65	0.99	0.99	0.99		0.99	0.99	0.99
RRM10	0.65	0.99	0.99	0.99	0.99		0.99	0.99
RRM11	0.65	0.98	0.99	0.99	0.99	0.99		0.99
RRM12	0.65	0.98	0.99	0.99	0.99	0.99	0.99	

TRAD = traditional animal threshold model,  $RRM_k$  =  $k$ th random regression model.

**Table 2.** Regression coefficients of predictions for 6-yr stayability obtained with different statistical models

	TRAD	RRM06	RRM07	RRM08	RRM09	RRM10	RRM011
TRAD	1						
RRM06	8.17	1					
RRM07	7.90	0.96	1				
RRM08	7.77	0.95	0.99	1			
RRM09	7.79	0.95	0.99	0.99	1		
RRM10	7.88	0.95	0.99	1.00	1.01	1	
RRM11	8.02	0.96	0.99	1.01	1.01	1.01	1
RRM12	8.17	0.96	1.01	1.02	1.02	1.02	1.01

TRAD = traditional animal threshold model,  $RRM_k$  =  $k$ th random regression model.

with RRM were consistently higher than those estimated through MTM.

More interestingly, changes in the mean accuracy of EPD values for all STAY end points obtained with the seven RRM are shown in Figure 3. In general, for all the models mean accuracies of EPD were lower at the extremes of the prediction range than for the middle points. In this context, Schaeffer and Jamrozik (2008) reported that when using Legendre polynomials, the estimated covariance matrices used to calculate genetic variances over the range of data, tend to result in genetic variances that are much higher at the beginning and end of the data range than in the middle. Regardless of the previous problem, results of this study support the hypothesis that the use of older age end-point observations increased the stability of the 6-yr STAY EPD, and the accuracies obtained

with these models (RRM07 through RRM12) were higher than the accuracies obtained with the threshold animal model (Figure 2) and with the RRM that ends use of data at the age of 6 yr (Figure 3). Specifically, mean accuracies for 6-yr STAY EPD were 0.088, 0.386, 0.399, 0.407, 0.409, 0.409, 0.409 and 0.407, for the threshold animal model, RRM06, RRM07, RRM08, RRM09, RRM10, RRM11, and RRM12, respectively. Starting with RRM08 and up to RRM12, a stabilization of the mean accuracy for 6-yr STAY predictions becomes apparent because results are nearly identical between each end point (0.407 to 0.409). In conclusion, RRM have a better predictive power for 6-yr STAY EPD when compared to the traditional threshold evaluation because of their higher accuracies. When compared to each other, although all RRM implemented in this study

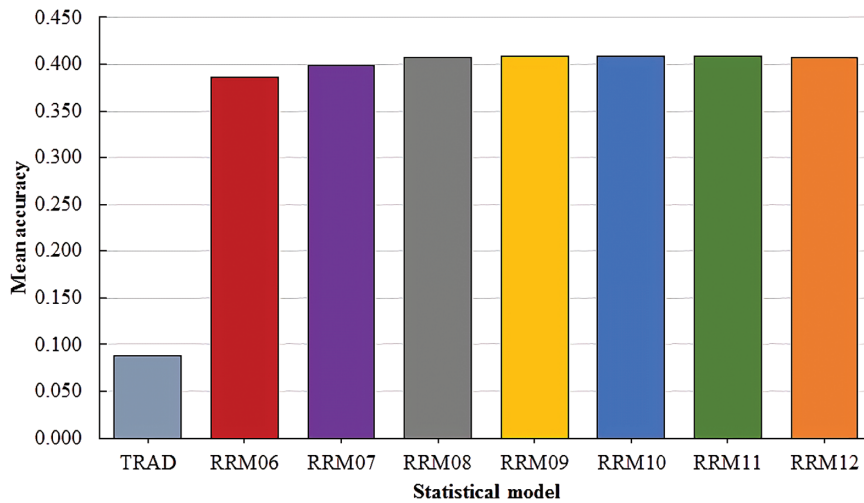


Figure 2. Mean accuracy for 6-yr stayability predictions obtained with all the statistical models (TRAD = traditional animal threshold model, RRMk = kth random regression model).

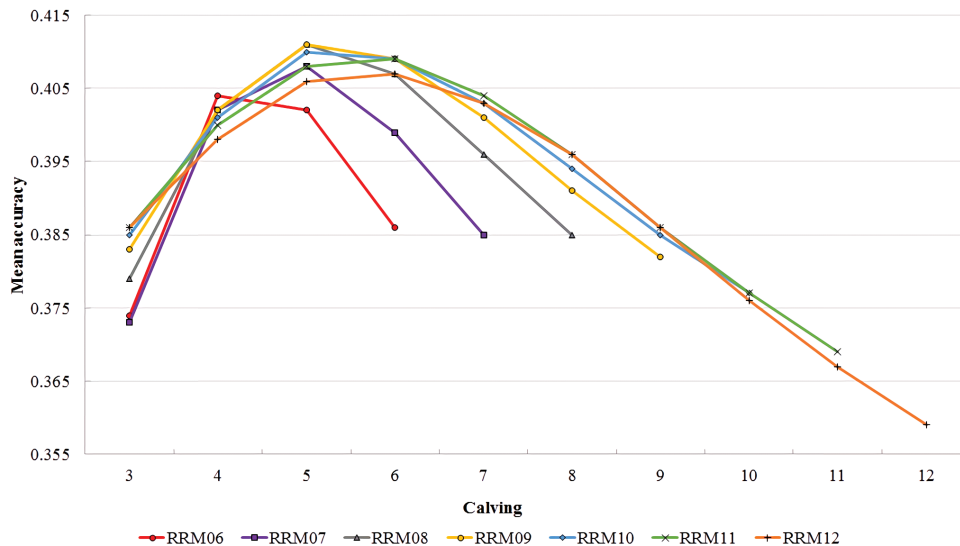


Figure 3. Changes in mean accuracy for all stayability end points depending on the random regression model used to analyze the data (RRMk = kth random regression model).

predicted the same genetic merit for individuals, an important stabilization of the genetic prediction for 6-yr STAY occurred when including observations of 8-yr-old females and beyond.

### ACKNOWLEDGMENTS

We would like to acknowledge to the Mexican National Council for Science and Technology (CONACYT) for providing the scholarship that allowed the first author to carry out his research at Colorado State University and, to the staff of the John E. Rouse Colorado State University Beef Improvement Center for their assistance in collecting data used in this study. This work was supported by U.S. Department of Agriculture, National Institute of Food and Agriculture, Hatch project COLO0607A, accession numbers 1006304 and COLO0681A, accession number 10110007. *Conflict of interest statement.* The authors declare that they have no conflict of interest.

### LITERATURE CITED

- Beef Improvement Federation (BIF). 2010. Guidelines for uniform beef improvement programs. 9th ed. North Carolina State University, Raleigh, NC.
- Bohmanova, J., I. Misztal, and J. K. Bertrand. 2005. Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. *J. Anim. Sci.* 83:62–67. doi:10.2527/2005.83162x
- Brigham, B. W., S. E. Speidel, R. M. Enns, and D. J. Garrick. 2007. Stayability to alternate ages. *Proc. West. Sec. Amer. Soc. Anim. Sci.* 58:27–30.
- Garrick, D. 2006. Evaluation, interpretation and relevance of stayability in beef cattle genetic improvement. *J. Anim. Sci.* 84(Suppl. 2):50.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, and D. Buttler. 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.
- Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's toolkit user's guide and reference manual. *Agric. Exp. Stn. Tech. Bull.* LTB92-2. Colorado State University, Fort Collins, CO.
- Jamrozik, J., S. McGrath, R. A. Kemp, and S. P. Miller. 2013. Estimates of genetic parameters for stayability to consecutive calvings of Canadian Simmentals by random regression models. *J. Anim. Sci.* 91:3634–3643. doi:10.2527/jas.2012-6126
- Lewis, R. M., and S. Brotherstone. 2002. A genetic evaluation of growth in sheep using random regression techniques. *Animal Sci.* 74:63–70. doi:10.1017/S1357729800052218
- Meyer, K. 2004. Scope for a random regression model in genetic evaluation of beef cattle for growth. *Livest. Prod. Sci.* 86:69–83. doi:10.1016/S0301-6226(03)00142-8
- Nobre, P. R., I. Misztal, S. Tsuruta, J. K. Bertrand, L. O. Silva, and P. S. Lopes. 2003. Genetic evaluation of growth in Nellore cattle by multiple-trait and random regression models. *J. Anim. Sci.* 81:927–932. doi:10.2527/2003.814927x
- Sánchez-Castro, M. A., R. J. Boldt, M. G. Thomas, R. M. Enns, and S. E. Speidel. 2017. Expected progeny differences for stayability in Angus cattle using a random regression model. *J. Anim. Sci.* 95(Suppl. 4):88. doi:10.2527/asasann.2017.179
- Schaeffer, L. R., and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. *J. Anim. Breed. Genet.* 125:145–146. doi:10.1111/j.1439-0388.2008.00748.x
- Speidel, S. E., B. A. Buckley, R. J. Boldt, R. M. Enns, J. Lee, M. L. Spangler, and M. G. Thomas. 2018. Genome-wide association study of stayability and heifer pregnancy in Red Angus cattle. *J. Anim. Sci.* 96:846–853. doi:10.1093/jas/sky041