THE VALUE OF RESEARCH AND INDUSTRY FLOCKS FOR PREDICTING BREECH STRIKE RESISTANCE IN AUSTRALIAN MERINO SHEEP

E. Dehnavi[1], A.A. Swan[1], J.L. Smith[2], T.L. Bird-Gardiner[3], G. Burbidge[4] and D.J. Brown[1]

[1] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia
[2] CSIRO Agriculture and Food, Armidale, NSW, 2350 Australia
[3] NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW, 2823 Australia
[4] Burbidge Farms, Tarcutta, NSW, 2652 Australia

SUMMARY

This study used data from the Sheep CRC Information Nucleus Flock (INF), a Merino Lifetime Productivity (MLP) flock, and a ram breeding flock (Connemara) to evaluate the prediction of breeding values for breech strike between and within datasets. Cross-validation was used to evaluate the accuracy, predictability and dispersion of estimated breeding values. Validation between datasets had low predictability due to low linkage (pedigree-based) across flocks, but validation within datasets was encouraging. Considering the poor linkage between the three datasets and the low incidence of breech strike across flocks, the industry needs to continue investing in building and maintaining suitable sheep reference populations with a wide range of traits, including flystrike observations, to develop accurate predictions required to underpin direct and indirect selection. In addition, quantifying the value of genomic information to improve the accuracy of predictions will be the subject of ongoing research.

INTRODUCTION

Flystrike is estimated to be the fifth highest cost to the Australian sheep industry ($170 million per year, Lane *et al.* 2015), with breech strike identified as the most common type. Resistance to flystrike is a priority research area for Australian Wool Innovation (AWI). To make genetic progress in flystrike resistance, accurate and standardised data collection of phenotypes for flystrike, probably combined with genotyping is the first step. Establishing a well-designed sheep reference population is a crucial step (van der Werf *et al.* 2010) for developing Australian sheep breeding values (ASBVs; Brown *et al.* 2010), especially considering the different incidence rates of flystrike in various environments (Bird-Gardiner *et al.* 2013; Greeff *et al*. 2014; Smith *et al.* 2009). Therefore, this study used data from the Sheep CRC Information Nucleus Flock (INF), one of the Merino Lifetime Productivity (MLP) flocks, and a ram breeding flock (Connemara) to estimate the accuracy, predictability and dispersion of pedigree-based breeding values within and across datasets.

MATERIALS AND METHODS

Data. A phenotype for breech strike was defined as a binary trait with 0/1 indicating "struck"/"not struck" within a defined shearing period (described in detail by Dehnavi *et al.* 2023). Three datasets including animals phenotyped for breech strike (struck or non-struck) were used for this study. The first dataset was from the Sheep CRC Information Nucleus Flock (INF), including 1,335 Merino lambs born between 2008 and 2011, recorded across six research stations (Trangie, NSW; Cowra, NSW; Rutherglen, VIC; Hamilton, VIC; Struan, SA and Turretfield, SA). A second dataset with 2,115 animals from 28 sires from the New England sire evaluation site hosting a Merino Lifetime Productivity (MLP) flock at the CSIRO "Chiswick" research station at Uralla,

---

* a joint venture of NSW Department of Primary Industries and University of New England

NSW. The MLP animals were born in 2017 and 2018. The last dataset was 1,941 lambs born between 2017 and 2021 in the "Connemara" Merino ram breeder flock, which were the progeny of 31 sires. Routine screening for flystrike is done primarily from birth to crutching time (6-7 months) for the Connemara flock but continued for a longer period in the other flocks. For this study, all animals were assessed and considered for flystrike up to yearling age. Pedigree data for animals with phenotypes were extracted from the Sheep Genetics MERINOSELECT database.

Cross-validation analyses. Breeding values (EBVs) were estimated using a binomial model with a probit link function in ASReml (Gilmour *et al.* 2015). Models included contemporary groups (CGs) and the interaction of birth type and rearing type as significant fixed effects, and the direct animal genetic effect was considered a random effect (Dehnavi *et al.* 2023). To estimate the differences in the accuracy of predictions, an internal cross-validation procedure within each dataset and external cross-validation between datasets were tested as described by Legarra and Reverter (2018). The MLP and Connemara (CON) datasets were separated into four cross-validation groups, and INF data were grouped into three groups of approximately the same size. All animals were randomly assigned to subgroups based on their CGs (Table 1).

Table 1. The number of animals (N), sires (Sire), sires per contemporary group (Sire/CG), average incidence (Mean) and standard deviation (SD) for breech strike (0/1) for subgroups used in the cross-validation analysis

Group	N	Sire	Sire/CG	Mean	SD
INF dataset					
INF1	473	66	15.67	0.19	0.40
INF2	383	80	13.89	0.07	0.26
INF3	479	109	16.10	0.13	0.34
MLP dataset					
MLP1	579	28	11.91	0.05	0.21
MLP2	459	28	12.20	0.05	0.23
MLP3	692	28	11.09	0.05	0.22
MLP4	385	28	10.55	0.09	0.28
Connemara dataset					
CON1	368	23	6.25	0.24	0.43
CON2	562	20	6.75	0.21	0.41
CON3	511	28	7.50	0.25	0.43
CON4	500	21	7.67	0.16	0.37

Prior to generating EBVs, variance components were computed separately within each dataset, and in the combined dataset. These components were then used for the best linear unbiased prediction (BLUP) analysis and calculation of accuracy. Breeding values were estimated in the full dataset using pedigree and phenotype information for all animals. Following the analysis of the full dataset, six validation scenarios were investigated (Table 2). First, EBVs were calculated for each internal validation group (Table 1) after their phenotypes were removed, using data from the other groups of that dataset as a training population (INF – INF analysis, replicated three times; MLP – MLP and CON – CON analyses, replicated four times). Second, the prediction of each dataset was carried out using two other grouped datasets as a training population (INF+MLP – CON analysis and INF+CON – MLP analysis, each replicated four times; MLP+CON – INF analysis, replicated three times).

For each scenario, validation metrics were calculated and averaged across replicates. Accuracy and dispersion metrics were computed using the LR method (Legarra and Reverter 2018). The accuracy (LR_{acc}) was calculated as the covariance between EBVs from the full and part analysis corrected for kinship matrix and genetic variance. The dispersion (LR_{slop}) was calculated as regression slopes between the EBVs from each of the analyses (partial) with EBVs from the full analysis (whole) in the validation set. The LR_{slop} parameter is expected to have a value close to one if there is no over or under-dispersion. Pearson correlation between partial and whole EBVs was also considered as an indicator of the predictability of EBVs.

RESULTS AND DISCUSSION

Validation results followed similar patterns for two of the three different internal validation scenarios (Table 2). For the CON – CON and MLP – MLP scenarios, correlations between EBVs of each analysis and EBVs of full analysis in the validation set were 0.89 and 0.83, respectively and for the INF – INF scenario the correlation was 0.53. The LR_{acc} was low ranging from 0.14 for INF – INF to 0.37 for MLP – MLP internal scenarios. The LR_{slop} for all internal validation scenarios was more than one. However, validation within the INF and MLP datasets was closer to one (1.06 and 1.05, respectively). CON and MLP could not predict breech strike in the INF dataset accurately. This scenario had a low correlation (0.35), very low accuracy (0.08) and a high LR_{slop} (1.21). INF alone could not predict animals externally (results not shown in Table 2). INF with CON predicted MLP (INF+CON – MLP scenario) with a correlation of 0.51, accuracy of 0.13 and dispersion of 0.83 compared to the prediction of CON using INF and MLP (INF+MLP – CON scenario) with a correlation, accuracy and dispersion of 0.16, 0.06 and 0.59, respectively (Table 2).

Table 2. The number of records (N_{Train} and N_{Valid}), the percentage of progeny in the training group having common sires with the validation group (F_{Prog}), genetic variance (σ^2_g), Pearson correlation, linear regression coefficient (LR_{slop}) and accuracy (LR_{acc}) for each validation scenario (training – validation) averaged across replicates

Scenario	N_{Train}	N_{Valid}	F_{Prog}	σ^2_g	Correlation[1]	LR_{acc}[1]	LR_{slop}[1]
INF – INF	890	445	65.76	0.12	0.53 (0.09)	0.14 (0.01)	1.06 (0.14)
CON – CON	1456	485	71.40	0.09	0.89 (0.04)	0.37 (0.03)	1.12 (0.06)
MLP – MLP	1586	529	100	0.09	0.83 (0.02)	0.24 (0.01)	1.05 (0.02)
MLP+CON – INF	4056	445	1.25	0.10	0.35 (0.07)	0.08 (0.03)	1.21 (0.35)
INF+MLP – CON	3450	485	4.45	0.10	0.16 (0.12)	0.06 (0.03)	0.59 (0.05)
INF+CON – MLP	3276	529	4.49	0.10	0.51 (0.10)	0.13 (0.02)	0.83 (0.11)

[1] Standard deviation for evaluation metrics is presented within parenthesis.

The internal-validation scenarios for MLP and CON resulted in higher prediction accuracy compared to the INF dataset. This may be because the INF dataset consists of different flocks subjected to different fly control regimes across a range of environments with a large degree of between-strain genetic variances (Swan *et al.* 2016), whereas the other two scenarios were performed within one flock (Connemara and New England sites), and in the case of CON, without pre-emptive fly control. However, the genetic variance of breech strike was low for all datasets with slightly more variation for INF (Table 2). Additionally, INF had a lower percentage of link progeny from common sires between training and validation data (66% for INF compared to 71% and 100% for CON and MLP, respectively; Table 2).

The accuracy of genomic predictions (Habier *et al.* 2010) and parameter estimation (van der Werf *et al.* 2010) can benefit from larger reference populations. Accurate and consistent data recording in seed stock flocks can contribute to establishing a reference population for the industry (Alexandri *et al.* 2022). In this study, there were low levels of linkage which contributed to low correlations and accuracy when predicting breeding values between datasets.

Overall, the predictability of breeding values for validation animals was lower between datasets than within datasets. This shows the necessity of strategic data collection, especially from flocks that are well-linked externally to be able to predict animals across flocks with different incidence rates accurately. It is important to note that the effectiveness of data also depends on the quality of the trait measured, its incidence rate and diversity within and between flocks as well as the influence of environmental effects recorded on the flock. Genomic information can fill the gaps in the pedigree-based relationship matrix and this is likely to lead to better genetic connections between data sets. Therefore, investigating the impact of genomic versus pedigree information on predictions between datasets will be a focus of ongoing research.

CONCLUSIONS

This study demonstrated that flystrike was predictable within each of the three datasets used for this study, but predictions between datasets were not feasible due to the low genetic linkage established through pedigree alone. In order to build a reference population for predicting flystrike it is critical to establish well-linked flocks across environments. The ideal flock has accurate and standardised data collection including phenotypes for different flystrike types (breech and body strike), along with phenotypes for production and indicator traits and genotypes for a large number of animals.

ACKNOWLEDGEMENTS

This work was funded by Australian Wool Innovation Ltd (project ON-00775) and used data provided by the Sheep CRC, Merino Lifetime Productivity project and Burbidge Farms. The Australian Government supports the research, development and marketing of Australian wool through Australian Wool Innovation Ltd.

REFERENCES

Alexandri P., Walkom S. F., Swan A. A., van der Werf J.H.J. and Brown D. J. (2022) *Proc.* 12th WCGALP, Rotterdam, The Netherlands.

Brown D.J., Swan A.A. and Gill J.S. (2010) *Anim. Prod. Sci.* **50:** 1060.

Bird-Gardiner T.L., Brown D.J., Smith J. and Mortimer S.I. (2013) *Proc. Assoc. Advmnt. Anim. Breed. Genet*. **20:** 183.

Dehnavi E., Swan A.A., Ramsay A.M.M., Bird-Gardiner T.L. and Brown D.J. (2023) *Proc. Assoc. Advmnt. Anim. Breed. Genet*. **25:** *These Proceedings*.

Gilmour A. R., Gogel B. J., Cullis B. R., Welham S.J. and Thompson R. (2015) ASReml User Guide Release 4.2 structural specifications. In: *VSN International Ltd: Hemel Hempsted*, UK.

Greeff J.C., Karlsson L.J.E. and Schlink A.C. (2014) *Anim. Prod. Sci.* **54:** 125.

Habier D., Tetens J., Seefried F.R., Lichtner P. and Thaller G. (2010) *Genet. Sel. Evol.* **42:** 5.

Lane J., Jubb, T., Shephard R., Webb-Ware J. and Fordyce G. (2015) Final report of project B.AHE.0010, Meat and Livestock Australia Ltd.

Legarra A and Reverter (2018) *Genet. Sel. Evol.* **50:** 53.

Smith J.L., Brewer H.G. and Dyall T. (2009) *Proc. Assoc. Advmnt. Anim. Breed. Genet.* **18:** 334.

Swan A.A., Brown D.J. and van der Werf J.H.J. (2016) *Anim. Prod. Sci.* **56:** 87.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci.* **50:** 998.