MAXIMISING GENETIC GAINS WITH DATA QUANTITY AND QUALITY IN MERINO FLOCKS

S.Z.Y. Guy[1] and D.J. Brown[1]

[1] Animal Genetics Breeding Unit[*], University of New England, Armidale, NSW, 2350 Australia

SUMMARY

Genetic gain can be maximised when selection is based on the most accurate breeding values and selection indices. To more explicitly take into account aspects pertaining to the quality of information used to estimate breeding values, metrics to characterise the quantity and quality of genetic evaluation data were previously proposed. This paper examines the relationships between these data quantity and quality metrics and genetic gains for Merino flocks. Stepwise regression analysis was used to analyse 3 metrics describing genetic gains: index accuracy, average index value and index trend. Index accuracy had the most number of significant predictors, with 4 quantity and 3 quality predictors explaining 85% of the observed variation. The most important metrics explaining index accuracy were level of genetic linkage for wool traits, average proportion of pedigree known in the last 3 years, and the level of wool and reproduction trait recording ($p <$ 0.0005). Data characteristic metrics were also associated with average index and index trend, although to a lesser level (~24% variation explained). This study demonstrates that both data quantity and quality are associated with index accuracy and genetic gains in Merino flocks. This decomposition provides a basis for informing ram breeders on improvements in their data recording. Used in conjunction with optimum selection decisions, this will enable higher rates of genetic progress.

INTRODUCTION

Genetic gain can be maximised when selection is based on the most accurate breeding values and selection indices. While the accuracy of estimated breeding values is calculated using the amount and structure of information utilised (i.e. quantity), some aspects pertaining to the quality of information can not explicitly taken into account in this calculation. Aspects of data quality, such as management group structure and accurate dates of birth, have been shown to affect the accuracy of estimation of genetic merit (Brown *et al*. 2001; Swan and Brown 2007). However, it is important to more specifically quantify the impact of data quality on the estimation of genetic merit due to the varying quality of data submitted by seedstock producers. Characterising both the quantity and quality of data will allow breeders to identify where their recording programs can be improved.

Sheep Genetics reports data quantity and quality metrics in their 'RAMping Up Genetic gains' (RUGG) report. Variation in these metrics has been shown to be associated with variation in rates of genetic progress (Stephen *et al.* 2018). Guy and Brown (these proceedings) reported considerable variation in key data metrics for Merino flocks, and proposed additional data quantity and quality metrics. This paper aims to demonstrate the value proposition of these metrics by examining their associations with genetic gains.

MATERIALS AND METHODS

Data quantity and quality metrics. The metrics examined in this paper were available for the 265 Merino flocks in the 12th December 2020 analysis that had a minimum of 100 animals per year

---

[*] A joint venture of NSW Department of Primary Industries and the University of New England

and data available for the last 5 years. These metrics (defined in Guy and Brown, these proceedings) were calculated as an average of the last 5 years and across contemporary groups.

Briefly, the data quantity metrics were:
- **fullped** (%): overall proportion of animals in the analysis with full pedigree.
- **avpedknown** (%): completeness of pedigree known from last 3 generations.
- **count**: average flock size per year.
- **ngeno** (%): average proportion of animals genotyped.
- Average proportion of year drop recorded for weight (**rec_weights**), reproduction (**rec_repro**) and wool traits (**rec_wool**).

The data quality metrics were:
- **synped** (%): proportion of animals with syndicate pedigree.
- **ages** (%): proportion of animals in contemporary groups with variation in age recorded.
- **bt** (%): proportion of animals in contemporary groups that have variation in birth type recorded.
- **eff** (%): the average proportion of effective progeny (Brown *et al.* 2001).
- Average proportion of animals directly genetically linked to other flocks for weight (**link_weights**), reproduction (**link_repro**) and wool traits (**link_wool**).
- **Dayinweek:** mean square error of birth date for days in week from an expected uniform distribution.

Data metrics and genetic gains. Index accuracy (indexacc), average index value (avindex) and index trend (indextrend) were calculated for each Merino flocks for the Merino Production + index, and averaged over the last 5 years. A series of linear regression models was used to examine the relationships between each genetic gain metric and data quantity and quality metrics:

$$\text{GeneticGains\_perc}_i = \mu + \beta_1 x_i + \varepsilon$$

where GeneticGains_perc is the percentile of indexacc, avindex or indextrend of flock i (quintiles 1 to 5, with 5 being the highest index), and x_i the flock's corresponding data quantity and quality metric refined above. Outliers, defined by $1.5 \times$ Inter Quartile Range below the 1st quartile and above the 3rd quartile, were removed from data metrics due to potential leverage and influential points affecting results of this analysis.

Multiple linear regression was conducted using all data quantity and quality metrics as predictor variables. Identification of the strongest associations with index accuracy and genetic gains was via stepwise regression (combining both backward elimination and forward selection), based on Akaike Information Criterion (AIC). The final model only included significant data characteristic metrics:

$$\text{GeneticGains} = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i + \varepsilon$$

where β_i is the partial regression coefficient of metric i. Flocks were included in the analysis if all data quantity and quality metrics were available ($n = 243$).

RESULTS AND DISCUSSION

There were significant differences in metrics across percentile groups, although there was considerable variation in each metric across the genetic gain percentile groups, with distributions overlapping (strongest relationships shown in Figure 1. Flocks with higher indexacc, avindex and indextrend ($P < 0.005$) had more reproduction traits recorded, more genotyped animals and a higher degree of average pedigree known in last 3 years (with the exception of avindex). These flocks also had greater linkage with other flocks for reproduction, weight and wool traits, actual birth dates and birth types recorded (not for avindex) and birth dates recorded evenly across days of the week. Flocks with higher indexacc also had greater average effective progeny numbers for sires and less syndicate recording (not shown in Figure 1). Therefore, flocks with more records and better quality data were associated with higher index accuracies and greater index gains.
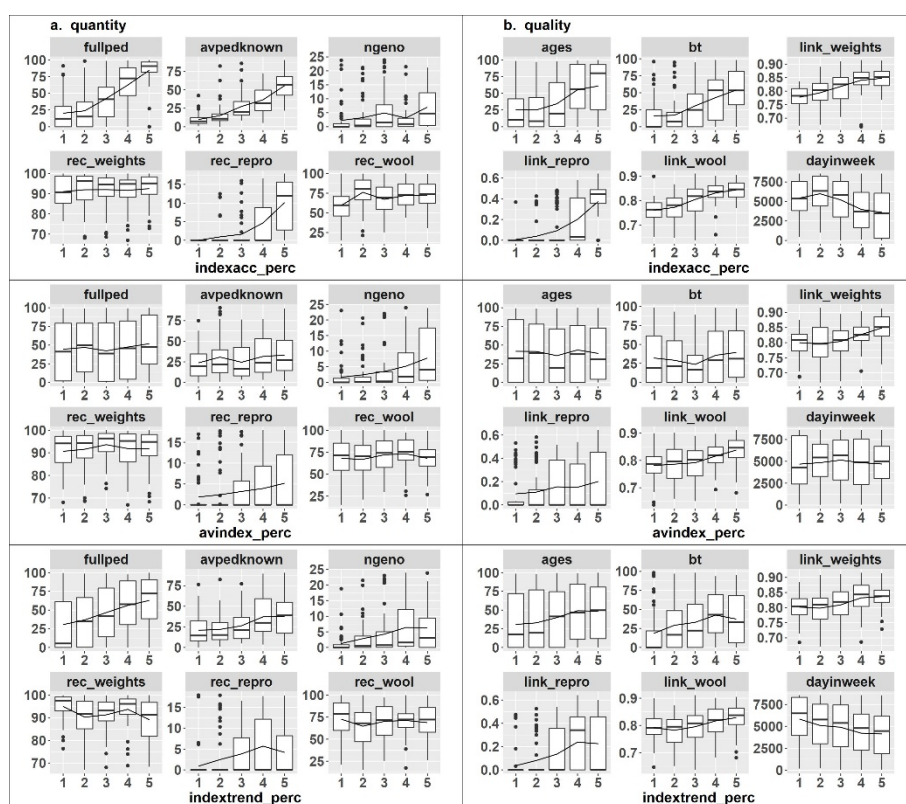
Figure 1. Univariate relationships between a) data quantity and b) data quality metrics (*y* axes) and percentiles of index accuracy (indexacc_perc), average index value (avindex_perc) and annual index trend (indextrend_perc) (*x* axes) for the Merino Production Plus index, from 265 Merino flocks. Lines are the average metric values for each percentile group, which reflect the strength of each relationship

Stepwise regression was used to analyse the indexacc and genetic gains, with all data quantity and quality metrics tested (Table 1). Indexacc had the most number of significant predictors, with 4 quantity and 3 quality in the final model. The most important descriptors of indexacc were level of linkage by wool traits, average pedigree known in the last 3 years, and level of wool and reproduction trait recording (P < 0.0005). This reflects the key priority areas of increasing complete pedigree, wool trait and reproduction trait recording identified by Sheep Genetics (Collison *et al.* 2018). The proportion of variation explained, taking into account number of predictor variables (R^2_{adj}), was also highest for indexacc (R^2_{adj} = 0.85). Avindex had 3 significant quantity and 2 quality metrics (R^2_{adj} = 0.25) and indextrend had 4 significant quantity and 1 quality metrics (R^2_{adj} = 0.23). The metric that was a significant predictor in all models was avpedknown, while synped, bt and dayinweek were not significant predictors for all models.

These results have implications for the development of an overall 'data quality score'. Indexacc more appropriately reflects data characteristics than avindex and indextrend, which are influenced by breeder selection decisions. Along with a large proportion of variance explained by data quantity and quality metrics, indexacc may be the most appropriate measure to 'train' an overall score.

Table 1. Stepwise analysis of index accuracy, average index and index trends, with significant data quantity and quality metrics as predictor variables ($p < 0.05$), for 243 Merino flocks

	Metrics[1]	Index accuracy[+] ($R^2_{adj} = 0.85$)	Average index[+] ($R^2_{adj} = 0.25$)	Index trends[+] ($R^2_{adj} = 0.23$)
		Coefficient estimates (SE)[*]		
Quantity	fullped	0.03 (0.01)	-0.10 (0.05)[*]	-
	avpedknown	0.11 (0.02)	0.16 (0.08)	0.03 (0.01)
	count	-	0.007 (0.003)	-
	ngeno	-	0.72 (0.17)	0.07 (0.03)
	rec_weights	0.06 (0.03)[*]	-	-0.09 (0.02)
	rec_repro	0.18 (0.08)	-	-0.14 (0.06)
	rec_wool	0.05 (0.01)	-	-
Quality	ages	-	-0.08 (0.04)	-
	eff	0.04 (0.02)	-	-
	link_weights	-	60.21 (23.98)	-
	link_repro	8.02 (2.20)	-	4.60 (1.70)
	link_wool	40.27 (4.56)	-	-

[+]using the Merino Production Plus index; [1] Description of metric acronyms provided in materials and methods section above; [*] P < 0.10

It is important to note that the most powerful predictors of the measures of genetic gain used in this study (indexacc, avindex and indextrend) are specific to this dataset and the index examined, and separately, that they may change over time. The effectiveness of providing feedback on data characteristics can be monitored by trends over time, and the cost-benefit of improved recording can be assessed. Future investigations may consider how genetic gains are also influenced by ram breeder selection decisions. This includes selection for traits not included in the index or use of outside genetics or selection differential, which has been identified as a key performance indicators of index gains across multiple beef cattle breeds (Johnston and Moore 2005).

CONCLUSIONS

This study demonstrates some key components of data quantity and quality which are associated with index accuracy and metrics describing genetic gains in Merino flocks. This decomposition provides a basis for informing ram breeders on improvements in their data recording. Used in conjunction with optimum selection decisions, this will enable higher rates of genetic progress.

REFERENCES

Brown D.J., Tier B. and Banks R. (2001) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **14**: 441.

Collison C.E., Brown D.J., Gill J.S., Chandler H.R., Apps R., Swan A.A. and Banks R.G. (2018) *Proc World Congreess on Genetics Applied to Livestock Production*. **11**: 661.

Guy S.Z.Y. and Brown D.J. (2021) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **24**: In Press.

Johnston D.J. and Moore K.L. (2005) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **16**: 161.

Stephen L.M., Brown D.J., Jones C.E. and Collison C.E. (2018) *Proc World Congreess on Genetics Applied to Livestock Production*. **11**: 433.

Swan A.A. and Brown D.J. (2007). *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 395.