Data investment strategy - the core issue for genetics organisations in the genomics era

R.G. Banks[1], S. Ring[2] and R. Evans[2]

[1]Animal Genetics and Breeding Unit, Armidale, Australia
[2]ICBF, Cork, Ireland
Corresponding author: rbanks@une.edu.au

Abstract

Genetic evaluation and improvement in the genomics era is underpinned by genomic reference populations. Traditionally, evaluation systems servicing multi-stakeholder situations (such as national or breed evaluation systems) have acted as recipients of data – essentially dealing with what is submitted by breeders and others, usually with some protocols around data collection and encouragement in relation to traits seen as most important. Typically, this stance results in a preponderance of data relating to easy-to-measure traits, such as milk volume or early life weight records. This may be despite such traits only comprising a moderate proportion of the value defined by the breeding objective(s), and accordingly, genetic improvement must be limited both in direction (alignment with the breeding objective) and speed (rate of progress for the objective). This limitation can become worse under genomic selection, as differences in accuracies among traits with differing numbers of records will be increased.

Addressing this potential problem requires an active strategy for phenotypes – either via direct investment and/or by contracting to obtain more desirable (economically valuable).

The extent of potential problems is here explored via a case study using ICBF data for several beef and dairy breeds, and elements of an active phenotype investment strategy are then discussed.

Failure to implement such strategies will increase risks of fragmentation, leading to possible reductions in rate and value of genetic improvement. Data investment strategy also interacts with progeny testing and approaches to finding new elite animals, and reflect the importance for evaluation organisations of clarifying their role(s)

Introduction

Genomic prediction is now central to all major livestock genetic improvement applications, including evaluation systems at both the breed and national level.

Fundamental to useful genomic prediction is the genomic reference population (or information, or library), which consists of animals with pedigree information of some form, along with phenotype and genotype records.

The size, sampling and trait coverage determines the accuracy of breeding values estimated for animals without phenotypes, and can make a significant contribution to the accuracy of breeding values for animals with phenotypes. In broad terms, the reference population thereby determines the attraction to submit animals for evaluation, whether they are potential candidates for selection or simply to assess their value for commercial production.

To date, genetic evaluation systems have in the main conducted analyses for the traits that are present ie that have been recorded at some reasonable scale, in the population of interest. This pragmatic approach may be despite the breeding objective(s) developed for the population: for example, fertility may be an objective trait (ie one with an economic weight) but if recording for the trait is limited or absent, then any genetic change in fertility will result almost solely from correlated responses to selection for well-recorded traits.

Extending this point, where a novel trait is developed via research and introduced, volumes of phenotypes for that trait may be limited for some time, particularly if it is costly or otherwise difficult to record. In this situation, the accuracy of genomic BVs for that trait will likely remain low despite its economic importance.

This situation generates risk for an evaluation system in two ways:

- The risk that selection for the trait will be limited, generating an opportunity cost
- The risk that alternative providers of genetic information may offer BVs for that trait, with whatever level of technical rigour, and attract participants away

This perspective points to a view of evaluation systems, whether at breed or national level, as platforms – infrastructure which can support a range of information services. As a general principle, the more such services can be obtained via a platform, the more attractive that platform is to existing and potential new users. In the context of traits, that means having sufficient data available to the platform for all economically valuable traits – including those that are assessed as having potential future value.

Attractiveness of the platform is critical because as a general principle, platforms whether at national or breed level will (or should) seek to maximise participation, in order to:

- Spread fix costs over a broader user audience
- Maximise scope for genetic improvement by maximising numbers of animals evaluated
- And maximise relationship to the active population, enhancing the accuracy and hence utility of genetic parameters and information

Given that data collected underpins the utility of the platform, this in turn points to the importance for platform managers of having an informed and active strategy for data – which given the fact that data collection costs money, means an investment strategy: what traits will be recorded, by whom, and on what terms made available to the platform.

This paper focussed on a first step in considering strategies for investment in data to support genetic evaluation and improvement: the assessment of the strength or balance of a data portfolio.

Materials and methods

The value of a data "portfolio" can be summarised simply as the average or expected accuracy of genomic BVs for young animals evaluated via the platform. The balance of that portfolio can be assessed by the relationship between the array of trait accuracies and the corresponding array of trait economic weights.

Here, trait reliabilities and economic weights for the Irish national beef and dairy evaluations are considered.

ICBF data was provided by Dr Siobhan Ring (ICBF) for the following:

- Trait reliabilities (REL) for the last 6 years for the traits in the beef and dairy evaluations. Individual trait reliabilities were averaged across the 6 years (NB: while not comprehensively analysed at this stage, in general trait reliabilities were relatively constant across years).
- Trait economic weights (EW) for breeding objective traits were collated for the beef terminal and maternal indexes, and their corresponding reliabilities logged. In the dairy traits, where a trait was associated with economic weights for more than one index, the total absolute value of the economic weight was summed. Absolute values were used for all economic weights (reflecting the fact that the size of the economic weight is an indication of how much we want to change the trait, or in simple terms, its relative importance).
- These were calculated for Angus and Limousin of the beef breeds, and Holstein (HOL) of the dairy breeds.

The correlation between trait average reliability and economic weight was calculated.

Plots have been made of the average reliability against the absolute economic weight, shown as proportion of the highest absolute economic weight of the breeding objective traits (PropEW). The x-axis for these plots is sorted from highest to lowest absolute economic weight (and PropEW).

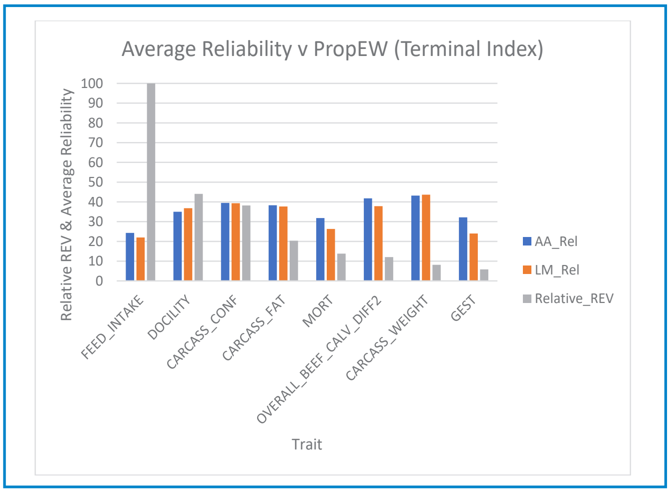
Results

Results shown here are:

- The correlation between average reliability and absolute economic weight for breeding objective traits for each of the breeds in the categories beef terminal, beef maternal and dairy
- The respective plots of average reliability against absolute economic weight.

Beef Terminal

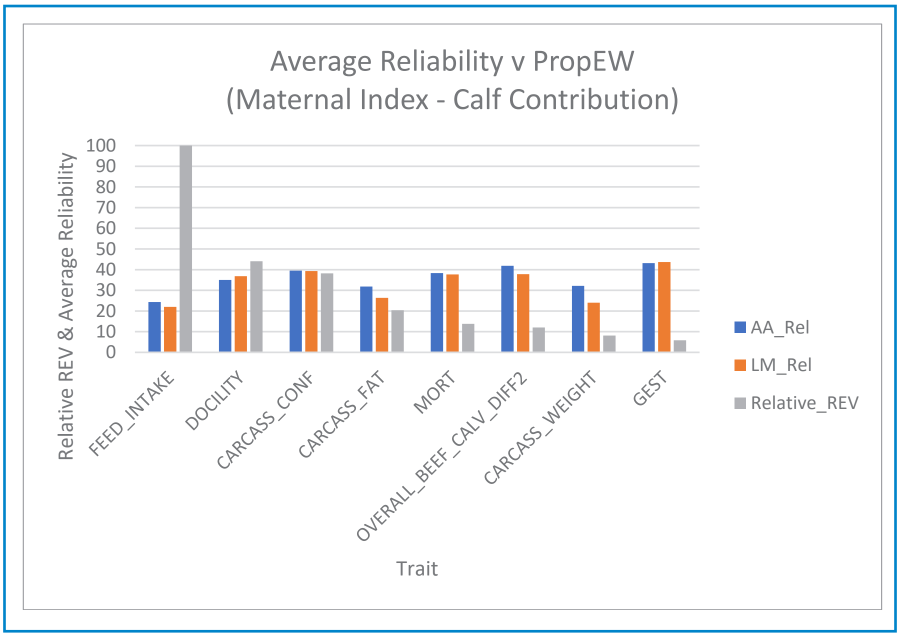
The correlations are -0.68 (ANG) and -0.40 (LIM) as shown in the below picture.

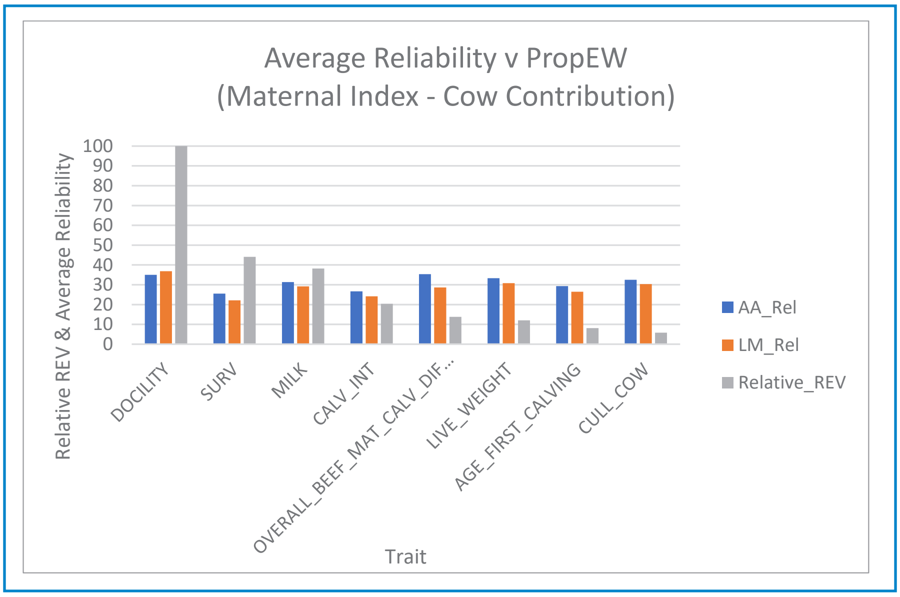


Beef – Maternal

The correlations, calculated as calf traits and dam traits separately, and then together, are:

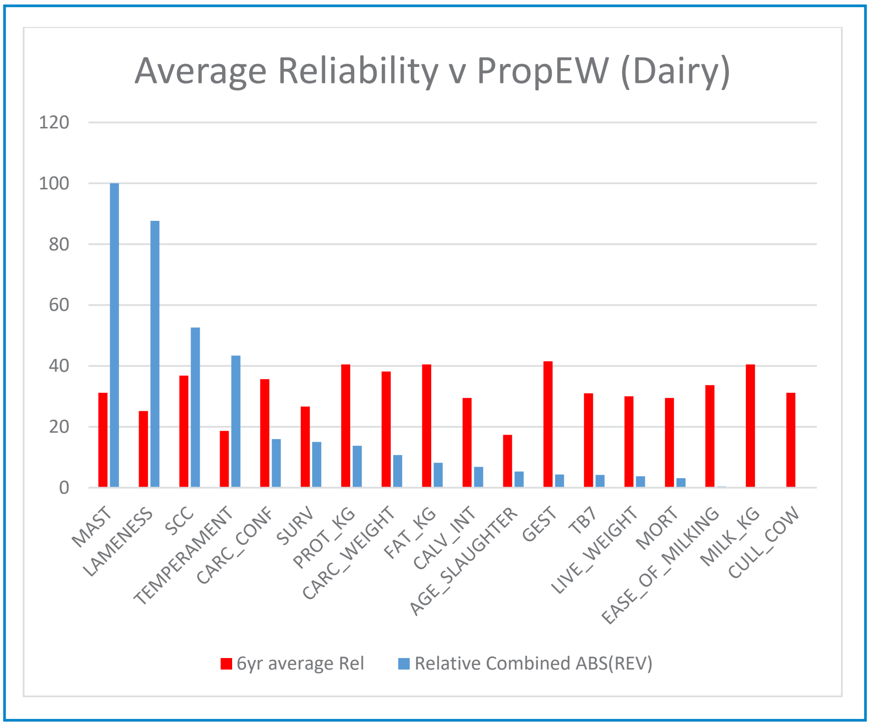
	ANG	LIM
Calf correlation	-0.69	-0.41
Cow correlation	0.38	0.69
Overall correlation	-0.05	0.15





Dairy

The correlation is -0.40.



Discussion

The simple approach used here is based on the fact that accuracy (of EBVs) is one of the determinants of the rate of genetic progress (R = i. rIT.▯T, where rIT = accuracy).

Naively, the accuracies of traits in the breeding objective should correspond to their relative importance – the more valuable it is to change a trait, the more accurate the BVs would ideally be.

In principle, improved alignment of accuracy with trait importance provides the maximum opportunity to steer the genetic improvement in the trait direction implied by the EWs, with equal accuracy for all traits enabling best alignment with that direction.

Where there is variation in trait accuracy, that will likely result in genetic change being less than perfectly aligned with the optimal direction of change in trait space. (NB: this does not mean that the genetic change will not be valuable, simply that it could be more valuable).

This perspective should be modified to take account of the cost of recording different traits, but where capture of returns from genetic improvement is over a sufficiently large scale, this consideration becomes insignificant (Banks, 2022).

Given this perspective, what do the simple results show?

- In beef (terminal) for ANG and LIM, alignment between reliabilities and EWs is poor, largely reflecting the low accuracy for feed intake.
- At the same time, there is in a sense an “excess of accuracy” for Carcase Fat, Mortality, Overall Calving Difficulty, Carcase Weight and Gestation.

The correlation is poorer in ANG than LIM, reflecting greater excess accuracy in the lower EW traits.

In beef (maternal), there is a marked difference between the calf traits and the cow traits: the correlation REL:EW is negative in the calf traits, reflecting low reliability for Feed Intake in both breeds. For the cow traits, overall the alignment is closer, but reliability for Docility is markedly lower than its relative importance.

In dairy, the overall alignment is relatively poor (correlation = -0.4) with a “deficit” of reliability for MAST, LAMENESS, SCC and Temperament, and “excess” for the remaining traits.

How significant are these findings, and what might be done in response?

The traits for which there are “deficits” of reliability include one obviously hard-to-measure trait in beef (feed intake), and several disease and ease of handling traits in dairy.

Assuming that the trait EBVs are from multi-trait analysis, then the extent of potential change in these traits is implicit in their reliabilities, and accordingly, potential responses in traits that have been defined as economically important will be constrained.

The obvious response is to invest in capturing more data for these traits. In the case of the disease and ease of handling traits in dairy, this does not seem to imply any or large investment in recording costs – simply getting the traits recorded by farmers. In the case of feed intake, off-farm or R&D investments seem the likely route in the short term, and are underway (e.g. Tully Station, Andrew Cromie, pers. comm.) – the question is whether enough data is being collected.

In the case of feed intake, it is possible that analysis of the cost-benefit of increased investment in recording should be done, although given the scale of the industry, it seems likely to be profitable (and see comment below re methane).

Some other observations prompted by these results include:

- The reliabilities seem very stable, and are not changing year to year (ie increasing) despite assumed accumulation of more data (phenotypes and genotypes).
- Reliabilities for some sets of traits are very similar – does this reflect assumed genetic parameters?
- Is Feed intake not important in dairy?
- Feed intake is a proxy for methane (Barwick *et al.*, 2019), and one would expect that the importance of feed intake would be increased if its use as a genetic correlate of methane were incorporated into the models – or more precisely, methane had its own economic value.

Conclusions

A simple analysis of trait reliabilities for the Irish national beef and dairy evaluations shows weak relationship between trait economic importance and the reliability of the corresponding trait genomic BVs. Based on informal discussions, this observation is likely to be repeated in most breed and country evaluations.

This is not surprising, in that to date, recording has been mainly for traits that can readily be recorded on-farm. However, this limits the accuracy (reliability) that can be achieved for hard-to-measure traits, and so limits the attractiveness of participation in the evaluation system.

Increasingly, attention will need to be paid to obtaining (or encouraging) appropriate volumes of data for hard-to-measure traits, and strategies to achieve this will become central to organisations (Banks, 2022). The simple approach used here can quickly highlight deficiencies in data, and help to focus recording efforts and investment.

References

Banks, R.G. (2022) Proc. WCGALP 12: 1804-1807 10.3920/978-90-8686-940-4

Barwick, S.A., Henzell, A.A., Herd, R.M., Walmsley, B.J., and Arthur, P.F. (2019) Genet. Sel. Evol. 51:18 https://doi.org/10.1186/s12711-019-0459-5