



An approach to coordinating and encouraging investment in phenotypes

R.G. Banks

*Animal Genetics and Breeding Unit, University of New England, Armidale, NSW,
Australia 2350
Corresponding author: rbanks@une.edu.au*

Performance records have always been the critical “raw material” for genetic improvement. This is even more the case in the genomics era, when genomic relationships enable leveraging of phenotypic data across large numbers of animals that may not be recorded at all. In extensive industries such as beef cattle and sheep, performance recording has traditionally been conducted by individual seedstock enterprises, with the volume and quality (in terms of trait coverage and identification of fixed effects) of data varying widely within (and between breeds).

Genomics enables leveraging of the recording effort, but if that effort is not rewarded, there is a risk that breeders will withdraw that effort, leading to decline in accuracy of breeding values. Addressing this coordination problem requires a mechanism(s) for cost-sharing, which needs to address diversity in quality of data (which is linked to contribution to accuracy) and cost of recording. In addition, recording effort may be augmented by government or industry contributions, particularly for hard-to-measure or “future” traits.

A simple model is presented that equitably and efficiently shares costs and rewards of recording across multiple players. The model accounts for diversity in recording cost and data quality, and provides a framework for optimising pooled investment in novel traits, or traits for which market signals are imperfect or absent.

Keywords: Genomic reference, recording, costs, coordination.

The reference population is the critical core or foundation for implementation of genomic selection (Goddard and Hayes, 2009). This can be considered as an extension of the fundamental requirement for performance records for genetic evaluation and subsequent selection, but with the added dimension that a genomic reference population can be leveraged to a greater extent than performance records can be under simple BLUP evaluation.

By leverage, we mean the ability to use information contained in the reference in evaluation of other animals, in particular those without their own performance records. Within a single breeding unit or organisation such leveraging is effectively an extension of the standard principle of focussing recording effort in a nucleus, from which genetic improvement is disseminated throughout the tiers of multiplication and commercial production (Bichard, 1971). In such cases the transition from BLUP-based to genomic

Abstract

Introduction

evaluation may not necessitate any significant organisational or investment modification: relevant traits will be recorded in a small component of the overall population, and genotyping may be used primarily to achieve increases in accuracy of estimated breeding values and to extract more value from recording hard-to-measure traits.

In the situation typical of beef cattle and sheep breeding and production, and especially so in countries with larger populations of these species, where there are large numbers of breeding units and (typically) larger numbers of commercial production enterprises, this relative simplicity may not apply. Here, the recording efforts of what may be a small number of enterprises can be “harvested” by many others, both breeders (enterprises selling bulls or rams) and commercial producers. If this harvesting erodes competitive advantage for the enterprises that do record, incentive to continue recording will likely decline, leading to loss of accuracy in genomic prediction. Such harvesting is a form of “free-riding”, and the risk it poses suggests that some mechanism(s) for sharing the cost of the reference should be developed.

Cost and accuracy of the reference

The simplest model of a reference population would be a single herd (or flock) where recording is practiced for the objective traits, and costs can be clearly defined, and some number of enterprises “draw on” the reference, by genotyping. Here, the accuracy generated and costs incurred would be simple to determine, and the costs could be shared simply by a levy on genotyping, equal to the reference cost divided by the known or estimated number of genotypes sampled outside the reference.

A more typical situation would be where a number of herds (flocks) collect performance data and genotypes, potentially including specific resource herds (flocks) – who could be termed contributors. Almost invariably, there would be variation in both the accuracy generated reflecting different combinations of traits recorded by different contributors and different costs, reflecting variation in herd (flock) size, business arrangements etc.

The contributor herds (flocks) could include industry- or government-funded or assisted units, likely particularly for hard-to-measure traits. Costs for farmer-owned units may need to be estimated by survey, or in more sophisticated models by some form of tender system.

This variation in accuracy generated and cost incurred across some number of contributors makes no difference to how costs of the reference can be shared across genotyping activity, but adds a dimension of complexity in allocating any levy collected amongst those contributors to the reference.

In this situation, we have:

- A range of contributors, varying in the accuracy their records generate when genotyping is added, and the cost of the accuracy they contribute, and
- Some number of “drawers”, who will also vary in the accuracy they obtain via genotyping, but likely little or no variation in genotyping cost

For each agent (contributor or drawer) we can in principle estimate (ex ante) or determine (ex post) the combination accuracy obtained and cost incurred, and hence for each case, the return on investment, as accuracy/cost. This can form the basis of determining equitable levies (charged to drawers) and rebates (distributed to contributors).

Equalising return on investment via levies and rebates

Costs can be equitably shared across contributors and drawers provided the following are known:

- Cost of reference ie amount to be collected [$cost_{ref}$]
- Accuracy of genomic breeding values for reference animals [acc_{ref}]
- Accuracy of genomic breeding values for genotype-only animals [acc_{go}]
- Costs of genotyping at each level (ie can be different) [$cost_{ref}$ & $cost_{go}$]
- Expected numbers of animals genotyped in the two categories [n_{ref} & n_{go}]

The levy to be imposed on genotypes is calculated as:

$$Levy = [(acc_{go} \times cost_{ref}) - (acc_{ref} \times cost_{go})] / [acc_{ref} + ((n_{go}/n_{ref}) \times acc_{go})] \quad (1)$$

and the rebate applied to reference data contributions is calculated as:

$$Rebate = Levy \times (n_{go}/n_{ref}) \quad (2)$$

The application of this simple approach can be illustrated with an example. This is based on beef cattle in Australia, and assuming three categories of contributors and one of drawers (Table 1). The 3 categories of contributors reflect “research-level” recording – which might involve some industry or government assistance to record a trait such as individual feed intake, farms where all traits that can be readily recorded are, and farms recording only minimal records (such as weaning weight).

Table 1: A simple example of the cost-sharing approach

Parameter	Contributors			Drawers
	A	B	C	
Number of animals	150	450	500	12,500
Accuracy generated or obtained	0.55	0.38	0.20	0.25
Cost of phenotyping	\$143	\$55	\$20	
Cost of genotyping	\$30	\$30	\$30	\$30
Levy				\$3.36
Rebate	\$99.60	\$34.29	\$28	
Final cost after levy and rebate	\$73.40	\$42	\$29.69	\$33.36
Return on investment*	0.75	0.75	0.75	0.75

*: calculated as Accuracy x 100/final cost.

The simple example shows how costs of the reference can be shared equitably, and that under realistic conditions, the required levy can be modest. This is desirable, as the core of the approach is to share the reference costs over large numbers of animals that are only genotyped, but in practice, that number cannot be known in advance. This points to challenges for practical implementation.

The first is the number of animals that will be genotyped can either be estimated allowing the levy to be estimated for an upcoming year or other time period, or observed over a period and the retrospective number used to calculate the levy. Either way, there is likely to be some error, generating some difference in return on investment for contributors and drawers, and/or potentially some loss (or benefit) for whatever organisation is managing the levy and rebating system.

The second challenge is that the calculations involved depend on a single basis for calculating accuracy. This implies that the individuals comprising contributors and drawers can agree on some common breeding objective – such that accuracy is the

Challenges for implementation

accuracy for the objective. This may not be readily achieved where individuals vary widely in their goals for the breed.

The third challenge is simply to establish some coordinating mechanism: some group or process that determines the levy and rebates, communicates those, and actually collects the levy and distributes the rebates. This might be a breed association, or in situations where there is industry and/or government involvement, a genetic evaluation unit or service. No matter the situation, communication must be clear and effective – the mechanism needs to be understood and accepted.

The fourth challenge is around inclusion of traits for which there is no current market value. An example is methane output – in some countries, absence of any market price means that consideration of inclusion of methane emission in the objective involves more uncertainty than inclusion of a trait like sale weight. This uncertainty, or more precisely the resulting under-investment in recording, may be alleviated by some form of industry and/or government investment. If this is quantified, the approach used here can be extended to include such traits, with their cost of recording adjusted to reflect the investment, but the decision on the economic weight to be applied to such traits must be made in such a way that it is accepted by the contributors and drawers.

Two final practical challenges for implementation can be identified. The first is that the value of performance records and genotypes collected on reference animals varies through time: basically, there are diminishing returns to additional records (Goddard and Hayes, 2009). Strictly, this should lead to rebates declining as more reference animals are recorded – this can be handled mathematically in real-time, but in practice, the rebating mechanism may need to be simplified such that the first n records receive rebate x , the next n records some fraction of x , the next n records some smaller fraction of x , and so on. This would require very clear communication. The second is that the mechanism requires that the coordinating group – for example, a breed association – acts as a bank, particularly in making payments to the contributors. It may be simpler to avoid such payments by adjusting all levies and rebates to only addressing the recording costs not considered as “normal costs of business” – for example, focussing on the recording of hard to measure traits, but this does not address variation in the level of recording of traits that can be normally done on farm without expensive equipment. It is hard to avoid concluding that all costs, and hence all contributors, need to be eligible for rebates.

Implications

The underlying message of this paper is that phenotyping, and the accuracy that it generates, is the fundamental resource of genomics (“phenotyping is king!”). Whereas in the BLUP era, the costs and benefits of phenotyping were largely respectively borne and captured by the individual or business investing in the recording, with genomics recording generates a shared or public good (strictly, a club good). Such goods typically require some coordinating or governance mechanism, at least more than simply leaving decisions about investment in recording to the market (Ostrom, 1990). Such mechanisms can be effective over long periods and across a wide range of shared and/or co-created resources (Ostrom, 1990), but require active engagement of the participants. For breed associations, or industry- or government-supported organisations, this implies effective and ongoing communication – primarily around the breeding objective(s) and how recording generates accuracy, and around understanding the shared or common pool resource that is being managed. Organisations that cannot deliver this will almost inevitably suffer from sub-optimal investment in recording, and resulting loss of competitiveness.



Bichard, M. 1971. Dissemination of genetic improvement through a livestock industry. *Animal Production* 13: 401-411

Goddard, M.E., and B.J. Hayes, 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* 10: 381-391

Ostrom, E., 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK: Cambridge University Press.

List of references