

## On the value of the phenotypes in the genomic era

O. González-Recio<sup>1,2</sup>, M.P. Coffey<sup>3</sup>, J.E. Pryce<sup>1,2,4</sup>.

<sup>1</sup> Department of Environment and Primary Industries, Bundoora, Australia, <sup>2</sup> Dairy Futures Cooperative Research Centre, Australia, <sup>3</sup> SRUC, Easter Bush Campus, Midlothian UK, EH25 9RG <sup>4</sup> La Trobe University, Bundoora, Australia

**ABSTRACT:** Genetic improvement programs around the world rely on the collection of accurate phenotypic data. These phenotypes have an inherent value that can be calculated as the contribution of an additional record to genetic gain. The contribution of phenotypes were calculated using traditional progeny testing and 2 genomic selection strategies that included either exclusively males or females in the reference population. The main factors affecting the value of a phenotype are: the economic value of the trait; the number of phenotypic records already available for the trait and its heritability. Further, the cost of achieving a reliability of 0.60 using genomic selection is lower using sires with small progeny sizes (e.g. 20 daughters), unless the heritability is large and the cost of phenotyping exceeds a few hundred dollars, then female reference populations are economically more efficient.

**Keywords:** dairy cattle; phenotype; phenomics

### Introduction

In recent years, there has been a decline in the number of farmers participating in official milk-testing programs in some countries because of improved affordability of fully-automated and self-contained milk recording systems. Consequently, a reduced number of phenotyped animals with records available to national evaluation centres may create a shortage of data for both progeny test proofs and genomic evaluations. The cost of phenotyping has usually been ignored or not estimated in animal breeding as estimating the value of a phenotype is not straightforward and, until now, most phenotypes have been made freely available through the national milk recording activity. Phenotypes are not only used for genetic evaluations and breeding, but also for management and financial purposes that accumulate value over time. It is possible that future breeding schemes may have to purchase phenotypes for expensive to measure traits and so a method is required to value each additional phenotype for payment purposes.

There is an even greater challenge for novel traits such as feed efficiency, methane emissions, levels of hormones or metabolites that are not routinely recorded, are expensive to measure and are therefore available for only some animals in the population. An example is individual cow feed intake data, which is currently only available from research herds, because the equipment required is too expensive for commercial farmers to invest in. However, large populations are required to get genomic breeding values that are accurate enough for genetic evaluation units to include DMI in routine genetic evaluations. The total value of increasing the accuracy of genomic selection for difficult to

measure traits at an industry wide scale may be large. The questions that arise are:

1. What is the incremental value of an additional phenotypic record for a trait?
2. For the same investment into phenotyping, is the best return achieved from progeny-testing or genomic selection?
3. Is a genomic reference population of males or females preferable?

With a finite budget to pay for phenotypes the opportunity costs of a range of choices in both phenotyping and genotyping must be considered.

In this study a procedure to estimate the economic contribution of a phenotype to a breeding program is described for both GS and PT breeding programs using examples of a single trait breeding goal and a multitrait selection index, as well as some guidelines for the most costly efficient strategy to establish a reference population with regard to the phenotyping cost.

### Materials and Methods

The expected economic contribution of phenotypes in the commercial population was calculated by deriving the expected reliability with respect to the number of phenotypes.

$$Rel' = \frac{\partial Rel}{\partial n}$$

The reliability in a progeny testing breeding program was approximated as described by Cameron (1997):

$$Rel = 0.5 * \frac{\frac{1}{2}nh^2}{1 + (n-1)\frac{1}{4}h^2}$$

In a genomic selection breeding program, the reliability from a given reference population was approximated following Daetwyler et al. (2008):

$$Rel = \frac{Nh^2}{Nh^2 + q}$$

where  $N$  is number of individuals genotyped and phenotyped in the reference population,  $h^2$  is the heritability of the trait or the reliability of the breeding values in a sire population as  $\frac{n}{\frac{4-h^2}{h^2} + n}$ , with  $n$  being the progeny size,  $q =$

number of independent chromosome segments where  $q = 2NeL$  where  $Ne$  is the effective population size and  $L$  is the length of the genome in Morgans (assumed to be 30). Daetwyler et al. (2008) have a final correction of the prediction of  $r$  which is to add  $(r+r^2q)/2N$ .

Reliability was calculated for female and male reference populations separately. It was assumed that only one phenotype was available per female for reference populations that varied in size between 1,000 and 100,000. The male genomic reference populations varied in size between 1,000 and 50,000.

Then, the contribution of the increase in reliability per phenotype was used to calculate the corresponding increase in economic genetic gain:

$$ev \cdot \frac{\partial \Delta G}{\partial \sqrt{ReI}} \cdot \frac{\partial \sqrt{ReI}}{\partial n} = ev \cdot \frac{i \cdot \sigma_a}{L} \frac{\partial \sqrt{ReI}}{\partial n}$$

Selection intensity ( $i$ ) and generation interval ( $L$ ) were assumed to be 2 and 3.5, respectively,  $ev$  was the economic value of the trait. Traits and parameters used are shown in table 1. The number of cows in the population was assumed to be 500,000.

**Table 1. Economic values ( $ev$ ), heritability ( $h^2$ ) and additive genetic standard deviations ( $\sigma_a$ ) of traits used in the analyses.**

Trait <sup>&amp;</sup>	$ev$	$h^2$	$\sigma_a$
CI	-3.02	0.05	7.97
RFI	-102.61	0.27	0.20
MY	-0.05	0.30	445.42
PY	5.99	0.30	12.47
FY	1.49	0.30	16.32

<sup>&</sup>CI: Calving interval, RFI: residual feed intake, MY: Milk yield, PY: protein yield, FY: fat yield.

**Relative cost of female vs. male reference populations.** The cost of a GS program was naively calculated as a function of phenotypes and genotypes needed in the reference population. The cost of a cows only reference population was

$$C_{Female} = N_F \cdot g + N_F \cdot p$$

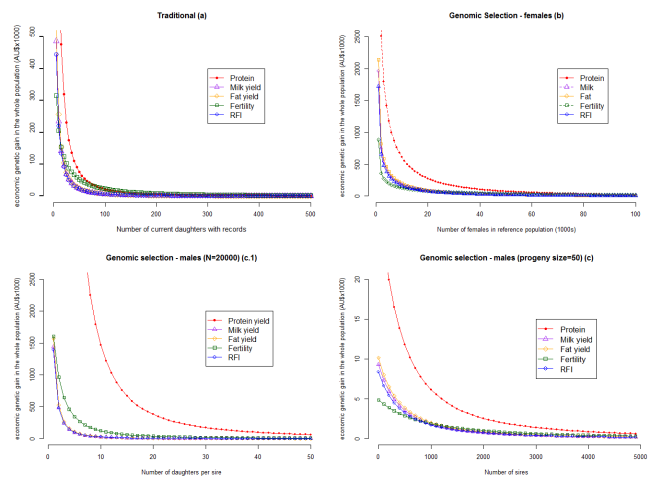
And the cost of a sire reference population was:

$$C_{sire} = N_s \cdot g + dtrs \cdot N_s \cdot p$$

where  $N_s$  and  $N_F$  are the number of sires or cows needed to get an accuracy of 0.60, respectively,  $g$  is the cost of genotyping set at AU\$60,  $p$  is the cost of phenotyping ranging between AU\$ 0.01 and 2000, and  $dtrs$  is the progeny size in the sire reference population (either 20 or 100). We evaluated the cost of three typical reference populations: one  $C_{Female}$ , and two  $C_{sire}$  scenarios when the number of progeny is 20 and 100 daughters in order to obtain reliabilities in young animals of 0.60.

## Results and Discussion

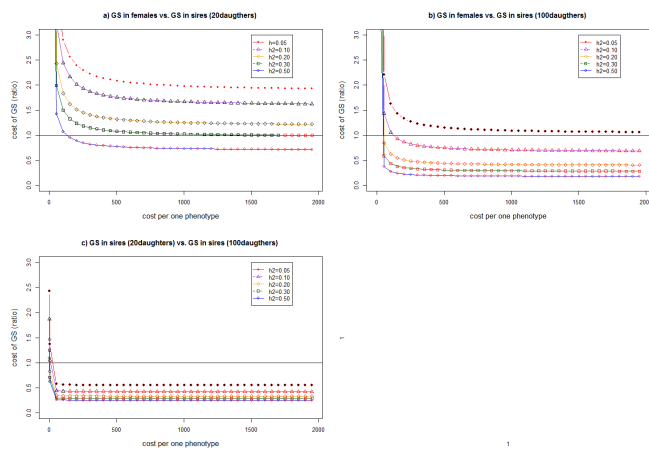
**Economic contribution of phenotypes to the whole commercial population.** The theoretical economic contribution to the Australian dairy industry of an additional phenotype for the traits included in the study is shown in Figure 1. In the early stages of a breeding program, when limited phenotypic data are available, it is better to obtain phenotypes for traits with larger economic weights in a selection index (e.g. CI or PY) for both traditional and genomic selection. However, their economic contribution rapidly decays as more daughters are phenotyped when the contribution of traits with lower heritabilities (e.g. fertility) becomes then more important. When more than 50 equivalent daughters are phenotyped for all traits, the economic contribution of an additional phenotype for calving interval is larger than an extra phenotype for milk yield. Also, when 100 equivalent daughters are phenotyped for protein yield, but no phenotypes are available for calving interval, then the economic contribution of one daughter with a fertility phenotype per sire is 81% larger than that of an additional daughter per sire with PY phenotype. The same trend is observed for the genomic reference population, in fact for the same given reliability Figure 1a is identical to Figure 1b. As expected the relative importance of liquid milk is less than that of protein yield, due to the larger economic value of the latter in Australia. Figure 1c shows that when a large sire reference population is already established (e.g. 20,000 sires) increasing the number of phenotyped progeny for fertility provides a larger economic benefit to industry than phenotypes for fat or milk yield until every sire has a progeny size larger than 30 daughters. However once the progeny size is equal to 50, increasing the number of sires in the reference population would produce larger economic benefit for production traits.



**Figure 1: Economic contribution in \$AU to the total Australian Holstein population of an additional phenotype in (a) traditional or (b) genomic selection breeding program.** <https://dl.dropboxusercontent.com/s/22381009/GonzalezRecioWCGALP2014.Figure1.tif>

**Relative cost of female vs. male reference populations.** Figure 2 shows the cost ratio of creating a reference population of females vs. sires. In general, when the phenotyping cost is low (<AU\$2) sire reference populations are

economically more efficient than female reference populations. Female reference populations cost less than reference populations of sires with 100 daughters each if the phenotyping cost exceeds AU\$50, unless the heritability of the trait is less than 0.10, in which case sire reference population is still economically preferable. When progeny groups are large (100 equivalent daughters), then sire reference populations are the lowest cost when the cost of phenotyping is < AU\$20. However, when the cost of phenotyping is large, smaller progeny group sizes (e.g. 20 equivalent daughters) are the most cost effective option for establishing a reference population. As the heritability of the trait and the cost of phenotyping increases, female reference populations can obtain a similar predictive ability but at lower cost.



**Figure 2: Relative cost of achieving a reliability of 0.60 in a genomic selection program applied in a female reference population versus a sire reference population<sup>2</sup>, regarding the heritability of the trait and the cost of phenotyping.** <https://doi.org/10.1009/GonzalezRecioWCGALP2014.Figure2.tif>

At the start of a new breeding program, or in adding a new trait to the breeding objective, compensating farmers for measuring phenotypes may be justified. For instance, feed efficiency traits are very expensive to measure but having breeding values for this trait could provide large financial benefits to all farmers when used in selection decisions despite only some farmers doing the recording. Phenotyping in experimental conditions is also expensive, and it is not worth collecting more phenotypes when no further increase in the economic response is expected. The cost of obtaining one phenotype is critical in the choice of the reference population in terms of sex and progeny group size. Populations of sires with large progeny group sizes (e.g. 100 equivalent daughters) are cheaper in obtaining a given reliability than smaller progeny group sizes (e.g. 20 equivalent daughters) and is also superior to female reference populations if the cost of obtaining one phenotype does not exceed a few dollars. These types of reference populations would be feasible for traits like conformation traits, where the cost of visual phenotyping is not too expensive and the heritability is large. When the cost of phenotyping increases, smaller progeny group sizes are more efficient, regardless the heritability of the trait. Fertility or milk composition would fit in this type of reference popula-

tion. Female reference populations should be considered when the heritability of the trait is high and the cost of phenotyping exceeds AU\$ 300. For instance, in a trait like feed efficiency or methane emission, where the cost of phenotyping can exceed AU\$1000 and the heritability is between 0.3 and 0.5, female reference populations are clearly both the most practical and economically best options to use.

Currently, most countries around the world have male only reference populations, or mixed reference populations, including females as well, ideally after removing the female's phenotype from the sire's breeding value, if both sire and daughter are in the reference population. Exclusively female reference populations are already being established, such as the Australian Ginfo project. Currently these populations are designed to augment reference populations that comprise bulls that have very reliable phenotypes from their progeny. However, with the increased popularity of young genomic over time, it is possible that genotyped females will eventually dominate reference populations.

There are some limitations to be considered. For instance, the genetic base of the reference population must be representative of the whole population. Further, some phenotypes are available regardless of the breeding program. For instance, milk recording schemes are maintained for purposes other than breeding programs, and the cost of phenotyping would be diluted among different partners and benefits to the industry. Finally, female reference populations allow exploiting dominance, epistasis and mating programs in a more comprehensive manner.

## Conclusion

The main contribution of this study is in quantifying the value of phenotypes in genomic selection breeding programs that have either male or female reference populations compared with a progeny-test breeding program.

Data recording is essential regardless of the scheme and can be used to evaluate the cost versus return of collecting phenotypes for a novel trait. When only few phenotypes are available, the value of each one to genetic gain is much larger than when a large number of phenotypes are available. However the accuracy of the trait is higher when more phenotypes are available.

Female reference populations must be considered when the heritability of the trait under selection is large, and the cost of phenotyping exceeds a few hundred dollars per animal. Finally, compensation, or payment for collecting phenotypes can be justified by the increase in profitability in the population resulting from genetics.

## Literature Cited

- Cameron, N.D. (1997). Selection indices and prediction of genetic merit in animal breeding. CAB International.  
 Daetwyler, H. D., Villanueva, B. and Woolliams, J. A. (2008). PLoS One, 3: e3395.  
 Ducrocq, V., and Wiggans, XX. (2013). Journal. #:pag-pag