# CONSIDERATIONS IN THE USE OF COMMERCIAL DATA FOR GENETIC EVALULATION IN BEEF AND SHEEP IN AUSTRALIA

**R. G. Banks and B. Tier**

Animal Genetics and Breeding Unit[*], University of New England, Armidale, NSW 2351

## SUMMARY

Using data collected on commercial animals for genetic evaluation is appealing due to the larger volumes of data potentially available, and the fact that such data may be on objective traits, or traits more closely correlated with consumer end-point traits. However, there are important considerations, the most significant being the degree of genetic relationship between commercial animals and current candidates for selection. There will continue to be advantages in using data collected in structured programs, rather than relying on commercial data.

## INTRODUCTION

There is growing interest in the use of commercial data in genetic evaluation of beef cattle and sheep, stimulated in part by the proliferation of means of capturing and transporting data and of interrogating databases at lower cost, coupled with broader discussion of “big data”. In general terms, big data refers to finding patterns or associations between factors in large datasets, with a lively debate continuing over what potential this offers and which traditional approaches to finding meaning will be rendered redundant.

One application is in livestock production, where databases built for different purposes are starting to be linked through the common identity provided by NLIS, (National Livestock Identification Scheme: - for example data collected in processing plants for meat eating quality assessment (Meat Standards Australia, or MSA)), and with genetic databases such as those containing pedigree, performance and genotype information for use in BREEDPLAN and Sheep Genetics evaluations. The appeal behind this area of application has several elements:

- Wanting to get the most out of the substantial investments in building these data systems
- Seeking to engage commercial producers more directly with genetic information, in part by providing a means for them to contribute to the evaluations, and
- Seeking to exploit data that is already captured on otherwise hard-to-measure (HTM) traits related to carcase and meat characteristics, and potentially animal health data.

This paper briefly explores this opportunity and highlights important considerations that impact the value of the enterprise of increasing connections between various data systems, to generate big data.

For this exploration, it is important to briefly summarise the genetic structure of the beef and sheep industry populations, but focussing mainly on beef.

## INDUSTRY GENETIC STRUCTURE

In both the beef cattle and sheep industries, there are well-defined and separate sire breeding, or seedstock, and commercial production sectors, with much larger numbers of animals in the commercial sector. For example, the national commercial cow herd is approximately 12m head and includes approximately 100,000 enterprises. The seedstock sector comprises approximately 250-300,000 animals in 1,500-2,000 enterprises.

The sires of commercial animals – herd bulls in the beef industry (and flock rams in sheep) –

---

[*] AGBU is a joint venture of NSW Department of Primary Industries and University of New England

are in simple terms bulls born in seedstock herds that are not retained as sires of sires, and which are marketable. These herd bulls will be bred to approximately 100 cows in a working life. Genetic evaluation is based predominantly on records of performance taken in seedstock herds on pedigree-recorded young bulls and heifers, and similarly in sheep.

Given that the majority of males born will be used as herd bulls – to sire commercial progeny – the closest commercial relatives of young candidate males will be the progeny of half-brothers. If any of these are themselves used to breed herd bulls, this adds 2 more steps in the relationship path between young candidates and commercial relatives. Because AI is essentially only used at the seedstock level, cases where young male candidates have commercial half-sibs that can be recorded are very rare in “normal” commercial practice.

This structure contrasts with that in dairy cattle, where commercial cows being herd-recorded are half-sisters of young bulls, and in pigs, where full- and half-sibs of young candidates can be recorded for slaughter/carcase traits. What does this mean for the value of commercial data in beef cattle and sheep?

## ACCURACY OF ESTIMATED BREEDING VALUES USING COMMERCIAL DATA

It is straightforward to calculate potential accuracies of EBVs for various data combinations. Table 1 shows the results for a sample of scenarios.

**Table 1. EBV accuracy for candidate animals, with varying heritability and combinations of effective records on relatives.**

| Heritability of trait | Records available | Accuracy of EBV for candidate animals in seedstock herd/flock |
|---|---|---|
| 0.4 | Own record | 0.63 |
| 0.4 | 29,850 grand-progeny of candidate’s paternal grand-sire (PGS). PGS has 100 sons each with 30 progeny | 0.25 |
| 0.4 | 930 progeny of 1 son of the same paternal grand-sire as the candidate | 0.125 |
| 0.3 | Own record | 0.55 |
| 0.3 | Own record plus 3 half-sibs plus 3 animals with same paternal grand-sire, each with 30 progeny | 0.60 |
| 0.3 | 60 progeny of 2 animals that share the same paternal grand-sire | 0.15 |

The overall pattern is simple and clear:

- Information from animals that share the same grand-parents is of limited value for genetic evaluation of young candidates in the seedstock sector
- Recorded and/or progeny-tested half-sibs can add accuracy

This simple example is for the situation where the trait recorded is the same for all animals. Data from relatives is potentially more useful when data can be collected on a trait more closely correlated with an objective trait, or that cannot be recorded on the candidate at all. This situation applies for carcase marbling (or other eating quality traits). Animals in seedstock herds can be indirectly assessed for marbling using live scanning or other correlated traits, but not for the objective trait itself.

In this situation, direct carcase measures on very small numbers of progeny generate more accuracy than even very large numbers of recorded half-sibs, for example. However, the contribution to accuracy of EBV on the candidates depends on the genetic relationship with the animals recorded for the objective trait – the value of the data declines as the square of the number of steps in the relationship path.

## MAKING COMMERCIAL DATA WORTH HAVING

From the perspective of genetic evaluation of young candidate animals in seedstock herds/flocks, the most likely situation to be practical is to collect data on commercial progeny of half-sibs, or on half-sibs themselves. The former requires that each crop of young sires generates commercial progeny which are recorded, the latter that seedstock sires are mated to produce both seedstock and commercial progeny routinely.

Examples of these 2 scenarios exist in beef cattle in Australia:

1. In the Team Te Mania program commercial herds use semen from stud sires of sires, or current young bulls, and capture slaughter data. Accordingly, young bulls are evaluated with data from animals with either the same sire (relationship = 0.5) or grand-sire (0.25)
2. In the Wagyu breed to date, a high proportion of commercial animals are AI progeny of widely used sires. This means that young bulls have commercial half-sibs with data. Zhang (2015, in press) details the data currently available in this breed for genetic evaluation, but in simple terms it is much more like dairy data than beef in terms of the relationship x data pattern.

In both these cases, higher accuracy of genetic evaluation is achieved for breeding objective traits of young seedstock animals than is usual in beef cattle in this country.

The reference populations (or information nucleus herds/flocks) established in beef and sheep combine some aspects of these 2 examples, but with potentially wider reach. In each, elite young sires are being progeny tested including for direct objective traits. The impact on accuracy of young animals in the seedstock population at large then depends on the animals' relationship with the animals being progeny tested. Here the intention is more to generate reference data for genomic selection, which partly overcomes the variable impact on accuracy of evaluation caused by the variation in relationships. At the same time, animals that are progeny-tested in this way will inevitably achieve higher accuracy of EBV for the traits recorded which in turn will increase their likelihood of being selected as parents in the seedstock sector.

## CAVEATS

There are two obvious concerns regarding use of commercial data:

1. unless data is collected in identifiable management groups, the heritability of the data will be compromised, and may in fact reach zero
2. such groups must have more than one sire represented, or if genomic pedigree is used, be sufficiently diverse to support some statistical contrast.

These aspects of data have been examined in the Wagyu case, which is to date simply a semi-random sample of commercial datasets, analysed by Zhang *et al.* (2015, pers. comm.). Within the data:

- 5,270 recorded were recorded in 1,161 management groups) for an average group size of 4.5 animals
- 692 management groups (60% of groups) contained 0 effective progeny, but 1,197 recorded animals (or 23% of the data)
- The overall average effectiveness of data (the ratio of effective number to actual) was 24%, with the value for groups with at least 1 effective progeny being 60%.

In general, utility of commercial data will be maximised when:

- Management groups are accurately recorded, and effective and actual progeny numbers are as similar as possible
- The commercial animals are as closely related to selection candidates as possible

- The animals are recorded for objective traits, or traits highly correlated with objectives.

## IMPLICATIONS FOR USE OF COMMERCIAL DATA

Even in the Wagyu situation, at least currently inherently favourable for use of commercial data, 60% of data collected (as measured by number of groups or datasets) provides no information that can be used for genetic evaluation.

More generally in the beef and sheep industries, it is not easy to predict the distribution of management group effective size, but there is no obvious reason to expect it to be dramatically different from in this example, and just as importantly, the genetic relationship between records and current candidates is likely to be lower. This last reflects the fact that widespread AI is very unlikely to be commercially practical – certainly use of AI in bull or ram multiplication makes growing sense as high merit, high accuracy sires are identified, but bulls and rams are extremely efficient AI technicians for extensive operations.

If these surmises are correct, it will be important to proceed carefully in harvesting commercial data, or more precisely, in what benefits are promoted from that harvesting. Even if the capture were free, costs will be generated in data storage and analysis. If the capture is not free, it will be imperative to develop, and communicate very clearly, ways of valuing data in advance of its collection so that informed investment decisions can be made.

The discussion to this point presupposes that the only purpose of capturing commercial data is for genetic evaluation. This may not be the case – management decision tools may be developed around real-time commercial data, essentially akin to herd recording in dairy cattle. In this scenario, as long as simple ways of screening in the data that is useful for genetic evaluation can be applied, some benefit is possible.

More generally, the multiplication and dissemination structure of the beef and sheep breeding and production industries inherently favour development of structured data collection. Harvesting data on close relatives of current selection candidates is genetically and hence economically more efficient, with the caveat that the “commercial” conditions must be commercial, otherwise there is a GxE to contend with and possible loss of confidence in the EBVs.

Livestock genetic improvement has been a “quite big data” enterprise ever since the introduction of BLUP methods, and is automatically becoming genuinely big as volumes of genotypic information grow. Because the field has always been focussed on extracting maximum value from precious (ie expensive) data, basic principles for valuing data and for designing efficient recording structures are well developed.

## CONCLUSIONS

Increasing use of large volumes of commercial data is very appealing in principle, and becoming more feasible through developments in data transfer and storage, and greater willingness to link databases.

However, “data ain’t data” – commercial data will vary enormously in its value for genetic evaluation, and simply assuming that incorporating large amounts of commercial data will lead to dramatic increases in genetic progress is misguided. It is almost certain that the greatest value will come from carefully structured and managed data collection, and those breeding enterprises that can incorporate such activity into their business plans will always be at an advantage.

## ACKNOWLEDGMENTS

This research is partly supported by Meat and Livestock Australia.

## REFERENCES

Zhang, Y., Tier, B. and Banks, R.G. (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21:** (in press)