

Balancing long-term and short-term selection response

J. W. James

Reprogen, Faculty of Veterinary Science, University of Sydney.

SHORT TERM RESPONSE

When we learn about designing animal breeding programs we usually begin with the classic equation of Rendel and Robertson (1950):

$$\Delta G = \frac{\sum i r s_A}{\sum l}$$

where i = standardised selection differential;
 r = correlation between true and estimated breeding values;
 s_A = standard deviation of true breeding value;
 l = generation interval;
 G = gain in mean breeding value;

and the summations are over all relevant pathways of improvement such as males to breed males, males to breed females, and so on.

If r is constant over all pathways, this equation reduces to

$$\Delta G = r s_A \frac{\sum i}{\sum l}$$

and maximising response to selection involves maximising the ratio $\sum i / \sum l$. There is normally some intermediate generation length which maximises genetic gain, though with very high reproductive rates it may be best to turn over generations as quickly as possible. A good general discussion of these matters was given by Ollivier (1974). There is a module in GENUP for doing these calculations.

If the amount of information on animals changes with age, as it often does, the optimum age structure will be affected, and in fact we can avoid having to specify the age structure which will maximise gain by noting that the greatest gain is achieved if the required number of parents are selected from among all those available on the basis of their estimated breeding values. This procedure will automatically give the best age structure (James, 1987). In fact, more generally, if the accuracy of estimated breeding values varies within age groups in any way at all, selection on EBVs will deal with all difficulties. In this sense, BLUP solves many of the problems which arise in the design of breeding programs. But the problems which it solves are those arising from making use of available data to select a specified number of animals. What it does not do is tell us what data to obtain, nor how many animals to select. Some aspects of these two questions will be dealt with in these notes.

First we will deal with the question of how many animals should be selected from among those available. To some extent this is governed by rates of reproduction: we need to select enough parents to produce the required number of offspring. This may fix the number of breeding females needed, at least within limits, but the number of males needed will generally be considerably smaller. With AI, the male reproductive rate may be very high indeed. Of course, the number of females required will be reduced if MOET is used, but the number of males needed will still usually be smaller, so that the opportunity arises for a stringent selection of males.

INBREEDING AND GENETIC DRIFT

If we take the basic equation and seek to maximise gain we see at once that i should be as large as possible, but in practice this is never done, and the number selected is larger than necessary for breeding the necessary number of progeny. The reason for this is that with very few parents selected the rate of inbreeding will be increased to an undesirable level. So it is common to set a minimum number of parents to limit inbreeding to an “acceptable” level.

There are three main effects of inbreeding, or limited population size, on response to selection.

1. Most traits of economic importance show some degree of inbreeding depression
2. The available genetic variance is reduced by inbreeding, so that future response is less than if there were no inbreeding.
3. Variation due to new mutations accrues more slowly, so that in later generations this source of variation contributes less to genetic gain.

LONG TERM RESPONSE

If we assume that selection is for a single trait in a monoecious population with discrete generations we can write simple approximate expressions for the effects of these three factors on genetic gain up to the t th generation.

Let the effective population size be N and let F_t be the population average inbreeding coefficient in generation t . Then

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

The inbreeding depression is $-DF_t$, where D is the regression of performance on inbreeding, assumed linear, so that D is the loss of performance in a completely inbred population.

The gain from exploitation of the original genetic variance in the base population is approximately

$$2Ni \frac{V_A}{s_p} F_t$$

where s_p is the phenotypic standard deviation. This expression must be multiplied by $(1 - \frac{1}{2n})$ if the selection program is started with a sample of n founders from the base population.

The gain from mutational variance is

$$2Ni \frac{V_M}{s_p} [t - 2NF_t]$$

where V_M is the new mutational variance arising each generation.

Putting all these factors together, we have for the response in generation t

$$2Ni \frac{V_A}{s_p} (1 - \frac{1}{2n}) F_t + 2Ni \frac{V_M}{s_p} (t - 2NF_t) - DF_t.$$

As t becomes large, F_t approaches unity and we find that the mean M_t becomes

$$M_t = 2Ni \frac{V_A}{s_p} [(1 - \frac{1}{2n}) + \frac{V_M}{V_A} (t - 2N)] - D.$$

This approximation ignores inbreeding depression arising from new mutational variance and so will overestimate the genetic gain from mutation. This aspect of the model could be improved, but it will not be pursued here. It can be seen that when t is large the only continued improvement comes from mutation, and that the gain per generation from this

source is $2Ni \frac{V_M}{s_p}$.

In some circumstances, such as when terminal sire breeds are bred so that inbreeding which accumulates in the pure-bred population is removed in the cross-bred progeny, or when the trait(s) selected show negligible depression, the term in D becomes unimportant. The value of V_M/V_A is likely to be small, less than 0.01 in most cases, so this factor is likely to be of importance only in the long term, where it is the sole source of progress.

In any case, we see that in the long term the greatest progress is made by maximising Ni . For a normally distributed selection criterion it is well known that the standardised selection differential is given by $i = z/p$, where p is the proportion selected, and z is the ordinate of the normal curve at the abscissa x truncating a fraction p in the upper tail of the distribution. If T is the total number of animals available for selection, the number of parents selected is $N = Tp$, so that $Ni = Tz$ and has a maximum at the mode of the normal curve, when $x = 0$ and $p = 0.5$. This result has been known for a long time, having been pointed out by Dempster (1955) and Robertson (1960).

BALANCING LONG AND SHORT TERM GAINS

However, although this gives the greatest long-term response the gain in the short term will be significantly less than would be achieved by selecting more intensely, and a breeder who adopted such a policy might well be out of business long before his strategy could pay off. So we need to consider a compromise between short-term and long-term gains. How can we make a sensible choice?

Most recommendations are made on the basis of a more or less arbitrary choice of an “acceptable” rate of inbreeding which will determine the number of parents to be used for breeding. While this probably works reasonably well, it does leave the decision as to what is “acceptable” to individual taste, which is in some ways an unsatisfactory state of affairs. Surely there should be a better way of reaching a compromise. I am going to propose just such a method – in fact I proposed it in 1970 (James, 1972). It was put forward independently by Dempfle (1973).

The proposal is based on the idea of discounting future gains to present value. This is a standard procedure in investment appraisal and cost/benefit analysis and was introduced to animal breeding by Poutous and Vissac (1962). The method was more widely used after it formed the basis of an appraisal of investment in animal breeding by Hill (1971). The principle of discounting is simple. If the rate of interest available is $100r\%$, one dollar can be invested at compound interest and in y years will be worth $(1 + r)^y$ dollars so it follows that a dollar obtained y years in the future is worth $1/(1 + r)^y$ dollars now. Therefore, returns achieved y years later can be discounted to their present value and compared with returns achieved at present. This provides a simple mechanism by which the conflict between short-term and long-term goals can be resolved.

If we have a well-defined breeding program we can predict the responses expected year by year or generation by generation. Let us suppose we have a monoecious population with discrete generations, and that the mean breeding value in generation t , measured from the current mean as origin, is M_t , while the economic value of a unit of breeding value is B dollars (or other currency unit). Then the gain in economic value in generation t is $W_t = BM_t$. However, in present value this is worth $W_t/(1 + R)^t$, where R is the per-generation discount rate. Therefore the present value of the entire breeding program, from now into the indefinite future, is

$$W = \sum_{t=1}^{\infty} BM_t / (1 + R)^t .$$

The discount rate is chosen to be appropriate for an investment with a degree of risk (which is always present) but should not include any allowance for inflation. Inflation should usually affect returns and it is assumed that all returns are evaluated at present-day currency values. Of course there may be cases where the returns from the breeding program are affected differently from the rest of the economy, but this is probably not the normal case. Therefore I suggest that for a suitably chosen discount rate the worth of a

breeding program be evaluated as $B \sum_{t=1}^{\infty} M_t / (1+R)^t$ and this measure of worth is what should be maximised.

Often such investment appraisals are made over a fixed time period, such as 20 years and Hill (1971) used a fixed time period in his influential paper. But in my opinion this is seldom justified. It may be so if a company is planning to build a toll road which will revert to public ownership after a set number of years, but presumably a genetically improved animal population will not be disposed of after a given time. Even if the breeder plans to sell the breeding operation and retire, the purchase price paid should reflect the value of the population as a going concern, which will depend on the prospects of future profits. If future profits are ignored, it is clearly best to select at maximum intensity in the last generation taken into account, since there is no need to allow for future improvement. In fact, Robertson (1970) showed that to maximise gain achieved by a given time a dynamic policy was required, with progressively more intense selection. The same would apply with discounted gains up to a specified time. In what follows I shall assume that the summation continues to infinity, though this is not necessary in principle, and the method can be used with any desired length of time. In fact, in many cases the distinction may be of relatively little importance, since gains made far into the future will be heavily discounted.

OPTIMISATION

To illustrate the method let us suppose we have selection in a monoecious population with discrete generations, and that mutational variance and inbreeding depression can be ignored. We then have approximately

$$M_t = 2Ni \frac{V_A}{s_p} \left[1 - \left(1 - \frac{1}{2N} \right)^t \right]$$

and on inserting this into the equation for W and simplifying we find

$$W = 2B \frac{V_A}{s_p} \left(\frac{1+R}{R} \right) \left(\frac{Ni}{2NR+1} \right).$$

Only the last term is affected by the breeder's decision on selection intensity, so W is maximised by finding the maximum of $Ni/(2NR+1)$. If T is the number of candidates for selection the criterion can be written as $Tz/(2TpR+1)$. It turns out that this is a maximum when

$$2TR = \frac{x}{z - px}.$$

When $R=0$ and future gains are not discounted we have $x=0$, corresponding to $p=0.5$ in agreement with the results of Dempster and Robertson.

For any given value of TR the corresponding value of p can be readily found from tables of the normal distribution or from a computer program. It is probably simplest to start

with a value of p from which the values of z and x are obtainable, and the value of TR which corresponds to the given value of p can be computed. By trying a few values of p the value which gives the required value of TR can be found by trial and error. The value of p does not usually need to be very precisely known. Alternatively, a computer program or subroutine which takes TR as input and iterates to the required value of p could be written.

It should be noted that for this model the rate of decline of genetic variance is constant so that the pattern does not change with generation. This means that the future looks the same with respect to the present at all times, so a fixed rather than a dynamic strategy is called for, in contrast to a fixed time horizon.

The above analysis is a very simple one, but it can be easily extended to the case of selection primarily in males. If S sires are selected each generation from T_S available candidates the effective population size is $4S$ and so the above analysis can be carried over with $4T_S$ being used instead of T , on the assumption that the number of females is much larger than that of males. Similarly, if the effective size is kN the replacement of T by kT allows the equations to be used. In practice the value of k is likely to depend on p as shown by Robertson (1961) and Wray and Thompson (1990) so that this simple device will probably not work very well.

If we look at the equation connecting p with TR , or plot a graph, we can see that p declines as TR increases. This is obvious enough for increasing values of R , since when future gains are more heavily discounted it will pay to put more emphasis on short-term gains rather than the relatively low-valued long-term gains. Similarly, as T increases the effective size increases and the time-scale over which response occurs is stretched, so that for a given selection intensity more gain is made at later times in large than in small populations and is more heavily discounted. Though the analysis above is of only a simple case, these conclusions about the effects of R and T on the optimum selection intensity are clearly more general.

EXAMPLE

Let us now apply the theory as developed above to the breeding program examined by Hill (1971) as an example. He proposed the establishment of a national breeding program in a herd of Hereford cattle which would be selected to improve slaughter weight when the bulls were crossed to Friesian cows to produce calves to be grown for beef. His analysis concluded that this would be a profitable investment.

The scheme he proposed involved selection in bulls only, cows being used for as short a period as possible in order to keep the generation interval as low as possible. Cows would be mated only three times, to calve at 2, 3 and 4 years of age, giving an average dam age of 3 years. Bulls would be mated once only, being 2 years old when their progeny were born. This would give a generation length of $l = 2.5$ years. It was assumed that selection would be for yearling weight, which had a phenotypic standard deviation of 40 kg and a heritability of 0.4. The economic value of one kg was then £0.15. It was

assumed that bulls from the herd would produce 200 000 crossbred slaughter calves per year, or 500 000 per generation, and these crossbred calves would show half the genetic gain made in the purebred herd. Thus the economic return from a gain of one kg in breeding value in the purebred herd would be $B = £0.15 \times \frac{1}{2} \times 500\,000 = £37\,500$.

If the number of cows in the breeding herd is 400 and the number of calves reaching mating age per cow mated is 0.8, the number of young bulls available for selection each year is 160, or 400 per generation. Assuming that the effective population size is determined very largely by the number of sires, the effective value of T is 1600.

The interest rate taken for our purposes will be 5% per annum. This can be converted to a per-generation interest rate by noting that for a generation of 2.5 years $(1 + R)$ is equal to $1.05^{2.5}$ or 1.1297 or $R = 0.1297$. We then need to find the value of p corresponding to $TR = 207.52$. The appropriate value of p is 0.015 to three decimal places. Then the number of bulls selected each year would be $160 \times 0.015 = 2.4$. We would therefore select 3 young bulls each year, to allow for possible wastage. (This example illustrates why it is not essential to solve the equation with great accuracy). The effective population size would be about $4 \times 3 \times 2.5 = 30$ per generation and the annual rate of inbreeding would be about 0.67%. Selecting 3 out of 160 bulls would give $i = 2.44$ so that averaged over sexes the selection differential would be 1.22. The initial rate of genetic gain would then be $1.22 \times 0.4 \times 40 / 2.5 = 7.8$ kg / yr. The overall worth of the program would be

$$£[2 \times 37500 \times \frac{0.4 \times 40^2}{40} (\frac{1.1297}{0.1297}) (\frac{30 \times 1.22}{2 \times 30 \times 0.1297 + 1})] = £43\,560\,419$$

which is a very considerable amount.

The effect of a higher interest rate can be seen by taking an annual rate of 8%, from which $R = 1.08^{2.5} - 1 = 0.2121$, so that $TR = 339.36$. The corresponding value of p is 0.010 and the number of bulls to be selected is 1.6. To allow for possible wastage we would select the best 2 bulls. The effective population size would be 20 and the annual rate of inbreeding about 1%. The standardised selection differential averaged over sexes would be 1.30 and the initial rate of gain would be 8.3 kg / yr. and the total worth of the breeding program would be £18 800 370. Because future gains are much more heavily discounted, the higher discount rate makes a very large difference to the worth of the breeding program.

In the mathematical analysis several simplifications have been made to make the mathematics easier and to avoid complications which would make the main points more difficult to illustrate. For a numerical analysis of a specific problem these simplifications are not necessary. The model appropriate for the problem can be programmed and the computations carried out for a range of options to locate the optimum.

OPTIMUM SCALE

The approach taken here can be extended to optimising the scale of a breeding program. Assume that the costs of setting up and operating a breeding program have some fixed

and some variable costs. Suppose the capital costs of setting up the breeding program are $C + DT$, while the recurrent costs are $U + VT$ per generation. Then the present value of the recurrent costs is $(U + VT)(1 + R)/R$ and the present value of all costs is

$$[C + U(1 + R)/R] + [D + V(1 + R)]T = F + KT.$$

If there is a lag of g generations before improvement is realised, the present value of all future improvement is $W_g = W/(1 + R)^g$ and so the present value of future profits is

$$P = W_g - F - KT.$$

To find the optimum size of the breeding program we must maximise P with respect to variation in both T and p , since we have seen that the optimum value of p changes with T . The solution depends on the ratio Q , the present value of the initial response per unit standardised selection differential divided by the present value of all costs per animal. That is,

$$Q = \frac{BV_A(1 + R)}{s_p R} / K(1 + R)^g.$$

On doing the required mathematics we find that the solution is given by taking

$$\frac{z}{2(z - px)^2} = Q$$

in conjunction with the previous equation involving TR . For a given value of Q we can find the optimum value of p and this gives us the optimum value of T .

In my 1972 paper I made such an analysis for the breeding program proposed by Hill using an 8% annual discount rate and found a Q value of 3500. This led to a calculation of about 600 bulls tested per year, of which the best 3 would be selected. This was a much larger operation than proposed by Hill. As you may perhaps have guessed, the UK government did not fund the program.

CONCLUSION

I would stress again that the important point here is one of principle. The equations I have given are primarily for illustration. In practice one would model the breeding program based on all available information, and use the predictions from this model to measure the total worth of the breeding program. Analyses would be based on year to year responses rather than gains per generation. The idea of choosing a selection intensity by this method of balancing short-term and long-term gain is the main thing I would like you to consider.

REFERENCES

- Dempfle, L. (1973) Bestimmung der optimalen Selektionsintensitat bei Selektion uber mehrere Generationen. *Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, **90**:160 – 168.
- Dempster, E.R. (1955) Genetic models in relation to animal breeding problems. *Biometrics*, **11**:535 – 536
- Hill, W.G. (1971) Investment appraisal for national breeding programs. *Animal Production*, **13**:37 – 50.

- James, J.W. (1972). Optimum selection intensity in breeding programmes. *Animal Production*, **14**:1 – 9.
- James, J.W. (1987). Determination of optimal selection policies. *Journal of Animal Breeding and Genetics*, **104**:23 – 27.
- Ollivier, L. (1974) Optimum replacement rates in animal breeding. *Animal Production*, **19**:257 – 271.
- Poutous, M. and Vissac, B. (1962) Recherche théorique des conditions de rentabilité maximum de l'épreuve de descendance des taureaux d'insémination artificielle. *Annales de Zootechnie*, **11**:233 – 256.
- Rendel, J.M. and Robertson, A. (1950) Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *Journal of Genetics*, **50**:1 – 8.
- Robertson, A. (1960) A theory of limits in artificial selection. *Proceedings of the Royal Society, Series B*, **153**:234 – 249.
- Robertson, A. (1961) Inbreeding in artificial selection programmes. *Genetical Research*, **2**:189 – 194.
- Robertson, A. (1970) Some optimum problems in individual selection. *Theoretical Population Biology*, **1**:120 – 127.
- Wray, N.R. and Thompson, R. (1990) Prediction of rates of inbreeding in selected populations. *Genetical Research*, **55**:41 – 54.

Sire Purchase Policies

J.W.James

Reprogen, Faculty of Veterinary Science, University of Sydney.

THE BREEDING PYRAMID

Although there are some vertically integrated animal production systems where the breeding program is controlled by the commercial producer, or where commercial producers work under contract to suppliers of breeding stock, for the most part genetic improvement of livestock is carried out in studs or breeding nucleus herds or flocks. Genetic improvement is then disseminated to commercial stock by the transfer of animals (mainly sires) or gametes.

This pattern was described by Robertson and Asker (1951) in British Friesian cattle using pedigree records from herdbooks. They showed that the structure was pyramidal, with a small group of elite studs at the apex of the hierarchy with one or two layers of “multiplier” studs between the elite nucleus and commercial herds. The elite studs were identified as those who exchanged animals among themselves, transferred animals to other herds, but did not introduce stock from these other herds. Other layers can be defined in a similar way. This work has been confirmed in many other studies. With changes in animal breeding over the last 50 years there have been changes, but it is still true that genetic improvement is made in a special nucleus and transferred to commercial herds and flocks through animals or gametes.

The result of this hierarchical structure is that, after an initial settling in period, the whole population improves at the same rate, which is set by the nucleus. Each tier in the pyramid lags behind the one above it. In a very simple case, where all sires in any level are introduced from the tier above, there is no selection of females and the sires bought are of average breeding value in the group in which they were born, each tier lags behind the one above it by two generations of genetic improvement. More complex situations require more complex expressions for lag, and have been considered by Bichard (1971) and James (1977) among others. It should be noted that any selection in the lower tiers does not affect the steady-state rate of genetic gain in the population, unless there is upward transfer of genes in the hierarchy (an open nucleus). But it does affect the lag, as does selection of better-than-average sires from the upper tiers. It will also affect lag if producers buy some breeding stock from two tiers above rather than one.

In principle it is possible for the lag to be negative. This can happen if (say) the nucleus is making no genetic progress, but selection of superior females in the commercial population raises its mean breeding value above that of the nucleus.

Another point to note is that in a population with a pyramidal structure the rate of inbreeding is controlled entirely by what happens in the elite nucleus, and since this is typically a small fraction of the whole population, the effective size of the population will

be much smaller than might be expected from the population size. The dominance by the nucleus of the population inbreeding is for the same reason as for its effect on genetic improvement. Genes flow from the nucleus to the rest of the population but not vice versa, so that eventually the whole population consists of genes from the nucleus, the genes at different levels originating in the nucleus at various times in the past.

VALUING GENETIC SUPERIORITY

One might expect that a nucleus breeder would like to limit expenses in running a genetic improvement program to produce animals for sale. If stock can be sold for satisfactory prices, avoiding spending on an effective breeding program may be a good option. If the nucleus begins at a higher genetic level than a commercial flock the commercial breeder can achieve genetic improvement by grading up to the nucleus, but the difference is halved each generation and would soon disappear without genetic gain in the nucleus. For the nucleus to remain significantly superior to commercial stock it must run an effective improvement program. This raises production costs in the nucleus and will be justified only if a suitable premium can be charged for breeding stock. A commercial producer, on the other hand, will want to reduce sire purchase costs, which can be done by buying sires less often or by buying cheaper ones. Thus a sire buyer will want to have a strategy covering frequency of purchase and cost (presumably related to breeding value) of sires.

An approach to devising such a strategy can be based on discounted gene expressions. This concept was introduced by McClintock and Cunningham (1974) to provide a rational method of giving economic weights to meat and milk traits in dual-purpose cattle, and was based on the idea of a farmer who was considering buying a dose of semen to inseminate a cow. Sires whose semen is available for purchase have EBVs for meat and milk traits. If the calf born is male it will be slaughtered for meat. If it is female it will be kept for milking over a number of lactations. Thus the meat traits in a male are expressed once, while milk traits will be expressed several times. When the cow is culled its sale value may be influenced by its meat traits. So the milk traits will contribute more often than meat traits, but returns from females will occur later than from males, and so will be discounted more heavily. However, there is a further complication. Female calves may be used for breeding and pass on half their breeding value for both meat and milk to their progeny, which will then express them after another delay, so even more heavily discounted. So in order to give appropriate weights to meat and milk traits it is necessary to take account not only of the number of expressions of the traits but the pattern of expression. Thus to choose a dose of semen from different sires we must weight their EBVs for meat and milk in a suitable way. We also need to compare expected returns with costs.

The approach of McClintock and Cunningham was used by Napier and Jones (1976) to evaluate purchase of a ram rather than a dose of semen. Essentially all we need to do is consider that when the ram is bought he can generate many progeny rather than the single one from a dose of semen. If he is to be used for only one mating season, his value will be that of M doses of semen, where M is the number of ewes to which he is mated. If he

is to be used for several seasons, then successive seasons can be added, each being discounted to present value while making allowance for possible death or disability. In this way the value of a purchased ram can be calculated and compared with the purchase price.

One problem with this method of evaluation is that each additional season of use adds more value to the sire, and an uncritical use of this criterion could lead a producer to use a sire for as long as possible. This would be the correct procedure if better sires are not available at a competitive price. But what should the producer do if better sires are continually becoming available? The buyer needs a policy which can give guidance on when to replace a sire he already owns with a new one. Of course if the sire can no longer produce the required progeny he has to go. But assuming he is still fertile, should he be retained or replaced?

REPLACEMENT POLICY

The basic principle is the same for terminal sires and sires which will breed replacements, but some details are different, and another approach is possible for replacement-breeding flocks. Let us therefore begin with the simpler case of buying a terminal sire. The females may be bred on-site as part of a breeding program or purchased from another breeder. The question of how to improve the females is a separate one, and here we concentrate on the purchase of terminal sires.

A method of assessing the value of terminal sires was given by Ollivier and James (1986) and James (1994). Here we follow the presentation of James (1994). Let P be the number of progeny produced by a sire in a given period, and let W be the probability that the sire survives to the next mating period, P and W being assumed constant (a convenient but not necessary assumption). The maximum number of mating periods for which a sire will be used is T , possibly determined by declining fertility. If d denotes the discount rate per period, the discount factor is defined as $r = 1/(1 + d)$, and we denote by Y the number of periods between sire purchase and first returns. Let B be the monetary value of a unit of overall breeding value. Then the present value to the producer of a unit of breeding value in the sire is

$$v = 0.5r^y BP(1 - r^T W^T)/(1 - rW).$$

Suppose the producer has the choice among several potential sires, the j th having breeding value A_j and cost C_j . Then the profitability of the j th sire is $vA_j - C_j$, and this provides a criterion for choosing the best sire to buy from among those available, on the assumption that a purchase is to be made.

Next we need to consider the profitability of a sire already owned. This sire has breeding value A_k but his purchase cost is now zero. However, if he was bought t time periods ago he will be used for only another $T - t$ time periods. Thus his profitability will be

$$0.5A_k B Pr^Y (1 - r^{T-t} W^{T-t})/(1 - rW).$$

Thus the difference in profitability between the new and old sires is

$$\frac{0.5r^y BP}{1-rW} [(A_j - A_k) - (rW)^T \{A_j - \frac{A_k}{(rW)^t}\}] - C_j$$

and sire k should be replaced by sire j if this is positive. In practice we would not use this criterion, but simply evaluate all potential sires and choose the most profitable. But this expression allows us to gain some insights.

Let us suppose sire j has the same breeding value relative to his nucleus contemporaries as sire k had when he was bought, and that the breeding nucleus has been making genetic progress at the rate of G per time period. Then $A_j = A_k + tG$ and the criterion for replacement becomes

$$\frac{0.5r^y BP}{1-rW} [tG\{1 - (rW)^T\} + A_k\{(rW)^{T-t} - (rW)^T\}] - C_j.$$

This has an especially simple form when sires will not be replaced at any fixed age but will be kept indefinitely until a more profitable sire is found. Then $T \rightarrow \infty$ and the criterion becomes

$$\frac{0.5r^y BP}{1-rW} tG - C_j.$$

This shows clearly what is obvious anyway, namely that, when $G > 0$, as t increases the more likely it is that a sire should be replaced, and that the time before replacement should be shorter when more rapid genetic progress is being made. The more expensive potential sires are, the less frequently they should be bought, other things being equal.

ADDITIONAL EVALUATION

For New Zealand sheep, Amer (1999) has discussed the evaluation of purchased rams and ewes by computing the expected numbers of discounted genetic expressions of slaughter traits and repeated traits such as amount and quality of wool and number of lambs born. He did this separately for terminal sires and for rams breeding self-replacing ewes. For a basic situation the numbers of discounted expressions of lamb slaughter traits were 145 and 196 respectively, while for ewe traits the number was 243, greater but not remarkably so. The reason is that though the ewe traits are expressed more frequently they are expressed later than slaughter traits, and so are more heavily discounted. The discount rate used was 7% and the planning horizon used was 10 years. A planning horizon of 10 years seems rather short, but Amer claimed that that it gave close agreement with a much longer period. With a lower discount rate the agreement might not be so good. In this work the number of years for which a sire would be used is taken as given. The paper gives a detailed account of how the calculations are done.

A modification of this approach can be used to value a sire with a known genotype for an important locus. For example, Wood, van der Werf and Parnell (2004) have shown how to value a bull known to be homozygous for a desirable recessive allele. Clearly the bull's progeny will show no advantage if the cows are all homozygous for the "bad" allele, since then they will all be heterozygotes. However, in the following generation there should be some "good" homozygotes. What is required is to trace genotype frequencies across generations to find the discounted number of homozygous recessive

genotypes produced. If there are some “good” genes in the cows, there will be some immediate production of homozygous recessives, so that the value of the homozygosity of the bull would be higher. Obviously the evaluation needs to be done in relation to the current gene frequency in the cow herd. This process was modelled and applied to the case of marbling. The case of an additive gene would show a different pattern of expression, but would also depend on gene frequency.

REPLACEMENT-BREEDING FLOCKS

We saw earlier that the value of a terminal sire could be evaluated simply because only its immediate progeny produce expressions of his breeding value. With a sire of self-replacing females, his grandchildren, great-grandchildren and so on also contribute expressions of his breeding value, albeit halved each generation and progressively more discounted. It was shown by James (1980) that the effect of these extra expressions could be accounted for by the factor J , which depends on the breeding age structure and fertility of females and the discount factor. If f_j is the genetic contribution of females aged j to each progeny crop, so that $\sum_j f_j = 0.5$, then

$$J = 1 / (1 - \sum_j f_j r^j).$$

If a sire for a female replacement-breeding flock has a breeding value A and a cost C then his profitability can be found by multiplying the value of v for a terminal sire by J and is

$$0.5r^Y JBPA[1 - (rW)^T] / [1 - rW] - C.$$

Thus the same type of analysis may be made as for terminal sires except for the inclusion of the factor J .

As an example of the calculation of J , let us consider the following example. Suppose the chosen discount rate is 5%, and that the reproductive contributions are:

Age	1	2	3	4	5
Contribution	0	0.15	0.15	0.12	0.08

Then we have

$$1 - 1/J = \frac{0}{1.05} + \frac{0.15}{1.05^2} + \frac{0.15}{1.05^3} + \frac{0.12}{1.05^4} + \frac{0.08}{1.05^5} = 0.4270$$

so that $J = 1.7453$. If $r = 1$ and there is no discounting we have $J = 2$ so that discounting has reduced the effective contributions by about 13%.

Although James (1980) presented this result, he devoted more attention to a different approach to determining sire buying policies. He introduced the concept of a lag cost, based on the idea that the profitability of a commercial flock or herd will depend on its productivity in relation to its competitors. Therefore the profitability can be regarded as dependent on the mean breeding value as well as other environmental and managerial factors. Ignoring these other factors as not involved in sire buying decisions, the measure of the effects of sire purchase will depend on the mean breeding value, which at any particular time can be measured relative to the sire breeding nucleus, or the improvement lag.

The lag in a particular herd or flock will depend on the selection practised on females, on the rate of turnover of breeding stock and the mean breeding value of purchased sires. As shown by James (1977) the lag can be written as $2(l_B G - S_B)$ where l_B is the generation interval in the commercial population, G is the rate of overall genetic gain and S_B is the genetic selection differential in the commercial population. Then $S_B = \frac{1}{2}(D_B + d_B)$, where D_B is the mean breeding value of purchased sires relative to their contemporaries in the nucleus and d_B is the genetic superiority of selected female replacements. If l_{MB} and l_{FB} are the average ages of male and female parents in the commercial population then we have $l_B = \frac{1}{2}(l_{MB} + l_{FB})$, and we have

$$\text{lag} = (l_{MB}G - D_B) + (l_{FB}G - d_B).$$

If N is the total number of sires used in any mating period, there are P progeny per sire and B is the value per animal of a unit of breeding value, then if the lag were zero the extra value of production per period would be $NPB[(l_{MB}G - D_B) + (l_{FB}G - d_B)]$. This can thus be defined as the lag cost. Letting n be the number of sires replaced in each period at an average cost of C , the "total cost" can be written as

$$NPB\{(l_{MB}G - D_B) + (l_{FB}G - d_B)\} + nC.$$

The second part of the expression in $\{ \}$ does not depend on sire buying policy, so if we ignore it we can write the cost relevant to sire purchase as

$$NPB(l_{MB}G - D_B) + nC.$$

Now if we again assume that the probability that a sire survives to the next mating period has a constant value W and that a sire will be used at most T times, we have

$$l_{MB} = a + \frac{1}{1-W} - \frac{TW^T}{1-W^T}$$

where a is the age of a sire at birth of its first progeny, and

$$n = \frac{N(1-W)}{1-W^T}.$$

If there is no wastage ($W = 1$) the corresponding expressions are $a + (T + 1)/2$ and NC/T . Therefore the cost relevant to sire purchase is

$$NPB[G(a + \frac{1}{1-W} - \frac{TW^T}{1-W^T}) - D_B] + \frac{N(1-W)}{1-W^T} C,$$

or, when there is no wastage

$$NPB[G(a + \frac{T+1}{2}) - D_B] + \frac{NC}{T}.$$

Although the analysis can be done without assuming there is no wastage, we will illustrate with the simpler form.

If we divide by $NPBG$, the value of one period's genetic gain, we obtain the criterion

$$U = a + \frac{T+1}{2} - \frac{D_B}{G} + \frac{C}{PBG}.$$

Writing C/PBG as Q and D_B/G as S we have

$$U = a + \frac{T+1}{2} - S + \frac{Q}{T}.$$

We wish to minimise this cost and find that

$$\frac{\partial U}{\partial T} = \frac{1}{2} - \frac{Q}{T^2}$$

and U is a minimum when $T = \sqrt{2Q}$. The more expensive sires are with respect to the value of genetic gain, the longer they should be used for breeding. Thus if the price of a sire is twice the value of one season's genetic gain for a mating of that sire, he should be mated twice, while if his price is 8 times the value of genetic gain he should be used four times.

In the above analysis we have calculated the optimum usage for sires of a given genetic superiority S , assuming that this has been chosen as an appropriate class of sires for purchase. But how do we decide which class of sire to buy? Clearly it will depend on the relation between price of sire and breeding value. Obviously from what we have seen above the optimum replacement rate would change as sire prices and breeding values altered. We need to consider their simultaneous optimisation. Therefore let us suppose that $Q = Q_0 k(S)$ where Q_0 is the value of C/PBG for a sire of average breeding value among his contemporaries and $k(S)$ is the function relating sire price to EBV. Then

$$U = a + \frac{T+1}{2} - S + Q_0 k(S) / T.$$

Then the task is to minimise U with respect to variation in both S and T .

Upon differentiation we find

$$\begin{aligned} \frac{\partial U}{\partial T} &= \frac{1}{2} - Q_0 k(S) / T^2 \quad \text{as before, and} \\ \frac{\partial U}{\partial S} &= -1 + \frac{Q_0}{T} \frac{\partial k}{\partial S}. \end{aligned}$$

Thus the shape of the curve relating C to S determines the optimum. Essentially what one has to do is find the price – breeding value relationship and then find how to minimise cost.

James (1980) argued that if there is a class of sires, as judged by EBV and price, which is more profitable than other classes, competition for these sires should drive up their price, while the price obtainable for less profitable sires should decline. In the long run the price – breeding value relationship should shift so that all classes of sire are equally profitable if used for the correct number of mating seasons. He showed that this would lead to

$$k(S) = (1 + S / \sqrt{2Q_0})^2$$

for which

$$T = S + \sqrt{2Q_0}$$

and

$$U = a + 0.5 + \sqrt{2Q_0}.$$

This relationship is not what is observed. James obtained sale prices of boars sold from French testing stations with known selection index values in 1978 and found that the relation between price (francs) and index values was roughly

$$\text{Price} = 1500 + 5(I - 100) + 0.5(I - 100)^2.$$

G was about 2.5 index points, each point being worth 1.5 francs per pig, and with 200 progeny per boar per period, $PBG = 750$ francs. Since index values had a mean of 100, the value of Q_0 was 2, so that since $S = (I - 100)/2.5$ the price function can be written

$$\text{Price} = 1500 + 12.5 S + 3.125 S^2$$

from which

$$k(S) = 1 + S/120 + S^2/480.$$

With such a price function it would be best to buy the very best available boars, say with index values of about 140. Then the optimum value of T would be 2.58. So the optimum with this price function would be to buy relatively expensive boars and use them for 2 or 3 time periods (12 or 18 months) before replacing them. The analysis indicates that the best boars were not achieving the market premium they deserve, on the basis of the assumptions made.

CONCLUSION

Perhaps because I introduced it, I have found this lag approach sensible, but it has not found any favour with others. The discounted expressions approach has met with reasonable acceptance for valuation, but again the optimisation of replacement rates in the formal fashion suggested here has not been widely used. For example, Amer (1999) has assumed that the number of times a sire is used is determined externally.

REFERENCES

- Amer, P.R. (1999) Economic accounting of numbers of expressions and delays in sheep genetic improvement. *New Zealand Journal of Agricultural Research*, **42**:325 – 336.
- Bichard, M. (1971) Dissemination of genetic improvement through a livestock industry. *Animal Production*, **13**:401 – 411.
- James, J.W. (1977) Open nucleus breeding systems. *Animal Production*, **24**:287 – 305.
- James, J.W. (1980) The analysis of sire buying policies. *Annales de Génétique et de Sélection Animale*, **12**:33 – 47.
- James, J.W. (1994) Terminal sire purchase policies. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, **18**:248 – 250.
- McClintock, A.E. and Cunningham, E.P. (1974) Selection in dual-purpose cattle populations: defining the breeding objective. *Animal Production*, **18**:237 – 247.
- Napier, K.M. and Jones, L.P. (1976) The value of genetic improvement to commercial sheep producers. *Proceedings of the Australian Society of Animal Production*, **11**:17 – 20.
- Ollivier, L. and James, J.W. (1986) Analyse des politiques d'achat de verrats dans un élevage de production. *Journées de Recherche Porcine en France*, **18**:253 – 260.
- Robertson, A. and Asker, A.A. (1951) The genetic history and breed-structure of British Friesian cattle. *Empire Journal of Experimental Agriculture*, **19**:113 – 130.
- Wood, B.J., van der Werf, J.H.J. and Parnell, P.F. (2004) Valuing DNA marker tested bulls for commercial beef production. *Australian Journal of agricultural Research*, **55**:825 – 831.

Optimising expenditure on trait measurement

J.W.James

Reprogen, Faculty of Veterinary Science, University of Sydney

In the past thirty years there has been a great development of the use of computers in genetic evaluation, due in part to the rapidly increasing power of these machines and in part on the development of methods such as BLUP and REML which rely on computing power for their practical implementation. We know very well now how to include all available information in our genetic evaluations, and how to make selection decisions based on these evaluations. This is all on a very sound theoretical basis. Of course, we often do not have the information we would like, but we can use what we have. We are still faced with the question of what information we should gather. How can we decide on such matters?

PROGENY TESTING

It was nearly 50 years ago that Robertson (1957) sought to rationalise progeny testing programs in dairy cattle. This is now so well established that it is hard to realise how original it was at the time. Essentially he asked the question: How can a limited amount of test material, the cows to be mated to young bulls for testing, best be used? He pointed out that the greater the number of young bulls tested, the greater would be the selection intensity, but the smaller would be the accuracy of selection, and that the product of the standardised selection differential and the correlation between true and estimated breeding value was the quantity to be optimised. He found a simple approximation to the optimum family size. This well-known theory is not usually thought of as involving optimising the distribution of expenditure on measurement over candidates for selection, but it is an example. Some young bulls have progeny test measurements, while others do not. In the theory it is assumed that the breeder has no useful information to choose among the young bulls, so all selection is based on the progeny test results. This would have been a reasonable assumption in the early days of dairy cattle progeny testing, even if less so now. Others soon extended Robertson's idea to other situations, e.g. Rendel (1958) and Smith (1959).

INCOMPLETE TESTING

In discussing the design of pig testing schemes, Smith (1959) pointed out that if the number of animals tested was no greater than that required for use, progress was possible only if part of the required breeding stock were selected from the tested animals, with the remainder being made up of randomly chosen untested animals. He remarked that it was always preferable to choose at random from among untested stock than to choose a tested animal known to be below average, since the EBV of an animal with no records is zero, and hence higher than that of a below-average tested animal. (This assumes that the mean EBVs of tested and untested animals are equal). James (1966) applied this idea to selecting a fixed number of animals from a population on the basis of a measurement that

is relatively expensive compared to the expected gain from selection. He found the optimal fractions to measure for given proportions culled and ratios of benefit to cost. For instance, if the cost of a measurement exceeded 0.3989 of the value of a standard deviation superiority in an animal it was best to do no measurement. Although a formal solution was given, in practice it would be straightforward to program a computation to find the most profitable measurement fraction. It was also pointed out that if comparison of a cheap and an expensive method was made assuming complete testing there might be a bias against the expensive method, because complete testing is less likely to be optimal for it than it is for the cheap one.

PREVIOUS INFORMATION

The analysis by Robertson (1957) of optimal family size in progeny testing was based on the assumption that nothing was known about the breeding values of the young bulls, or at least that there was no difference among their EBVs prior to the progeny test. If there is previous information, this ought to be combined with the progeny test data in the final evaluation. Another limitation placed on the progeny testing is that each tested sire has the same number of progeny. Thus the distribution of progeny among the available bulls is such that only two family sizes are possible: zero or n . But the general situation is that each available bull could be given an arbitrary number of progeny, subject to the restriction that the total number of progeny matches the testing facilities. James (1979) considered the optimum distribution of varying numbers of progeny to available sires subject to selection of a fixed number and that the total number of progeny is fixed. We present his approach here.

Suppose the prior information is summarised in the form of an EBV, \hat{G}_1 , scaled so that the regression of true breeding value on this EBV is unity, and that the correlation between G , the true breeding value, and \hat{G}_1 is r_1 . A second set of data is then obtained such that if the prior information is ignored breeding value can be estimated as \hat{G}_2 with a correlation r_2 with G . Let r be the correlation between these two EBVs. Then the best combination of these two EBVs is \hat{G} , where

$$\hat{G} = \frac{1 - rr_2 / r_1}{1 - r^2} \hat{G}_1 + \frac{1 - rr_1 / r_2}{1 - r^2} \hat{G}_2.$$

Then each sire is evaluated on its value of \hat{G} and the required number is selected by truncation (Henderson, 1963). This truncation point X can be taken as a deviation from the prior EBV as $(X - \hat{G}_1) / \mathbf{s}_{\hat{G}-\hat{G}_1}$ which we denote as x . The corresponding probability of selection is p and z is the ordinate of a normal curve at x . Thus the expected gain following selection on \hat{G} is

$$\frac{\sum p(\hat{G}_1 + i\mathbf{s})}{\sum p} = \frac{\sum p\hat{G}_1 + \sum z\mathbf{s}}{\sum p}$$

where we have dropped the subscript on $\mathbf{s}_{\hat{G}-\hat{G}_1}$, and summation is over all prospective sires with Sp chosen to give the required number of selected sires. The problem is then

to choose the appropriate value of $s_{\hat{G}-\hat{G}_1}$ conditional on the prior EBV. James gave a formal solution, but it was so complex as to be of little practical use.

However, a few general points can be brought out. The equations show that the response to selection can be regarded as having two components: selection among groups with the same prior EBV, and selection within these groups on the difference between final and initial EBV. Sires with small numbers of progeny have small $s_{\hat{G}-\hat{G}_1}$ values, and in the limit with no progeny will have a final EBV equal to the initial one. Animals without progeny will be selected on \hat{G}_1 so that if $\hat{G}_1 > X$, an animal which is not progeny tested is certain to be selected, while if $\hat{G}_1 < X$ it is certain to be culled. Independent culling levels can be regarded as corresponding to the second case. There is then reason to suggest that animals with very high initial EBVs need not be progeny tested, nor should those with very low values, attention being concentrated on testing those for which further information will be critical for making selection decisions. Of course, such a situation can arise only if the accuracy of the prior EBV is rather high, which would probably occur only rarely when progeny testing is being considered. Although the theory was developed with progeny testing in mind, it applies to any further gathering of information.

It would be usual to obtain further data on the animals judged best from initial EBVs when the amount of further data collection was limited, which is a form of independent culling level selection. The suggestion that attention might be concentrated on animals with intermediate EBVs is similar to the concept of selection of extremes due to Abplanalp (1972). Independent culling levels involves a series of tests, all of which a candidate must pass in order to be selected. Selection of extremes also involves a series of tests, all of which an animal must fail in order to be culled. Abplanalp showed that in some situations it was better than independent culling levels in terms of genetic gain, though it is more expensive, since each animal must be tested for all traits, while with independent culling levels those that fail the first test will not be given the second test.

TESTING MALES AND FEMALES

Suppose a breeder has facilities for testing 1000 animals for a quantitative trait and will select 10 males and 500 females from among those tested. If he tests 500 of each sex he will have no selection differential in females unless he replaces the worst tested females with untested ones. Suppose this is done. Then he selects 10 out of 500 males with a selection differential of 2.42s and 250 out of 500 females with a selection differential of 0.80s together with 250 untested females so that the overall female selection differential in females is 0.40s and averaged over both sexes it is 1.41s. But if the number of males tested was to be 400, the males would have a selection differential of 2.34s, and there would be 300 selected tested females with a selection differential of 0.80s and 200 with a value of zero, so that the average would be 0.48s, and averaged over sexes it would be 1.41s also. Essentially these are the same.

Now suppose we want to select 10 males and 250 females. In this case with equal testing we can select all animals required from among those tested. The selection differential in males would again be 2.42s, while that in females would be 0.80s with an average of 1.61s. If 400 males and 600 females are tested the selection differentials are 2.34s and 0.93s respectively, giving an average of 1.64s. There is thus in this case a small advantage to testing more females than males.

Smith (1969) made a theoretical analysis of this situation using his approximation to the standardised selection differential

$$i = 0.8 + \ln[(1 - p) / p] = 0.8 + \ln(q - 1)$$

where i is the standardised selection differential, p is the proportion selected, and $q = 1/p$. Here I will give a brief idea of what is involved.

Suppose we have T testing places and we need to select S sires and D dams from among N available animals of each sex, while $T = 2N$, so that not all can be tested. The breeder decides to allocate a fraction f of testing places to males, and $1 - f$ to females. To avoid complications we assume that all animals required may be selected from those tested, as in the second example above. The required numbers as fractions of the numbers tested will then be S/Tf and $D/T(1 - f)$ respectively. Then if we write q_M and q_F as the reciprocals of the selection intensities the selection response is proportional to $i_M + i_F$ which can be written as, approximately,

$$0.8 + \ln(q_M - 1) + 0.8 + \ln(q_F - 1).$$

To find the optimal value of f we differentiate this expression and set the derivative to zero. The derivative can be written

$$\frac{T/S}{q_M - 1} - \frac{T/D}{q_F - 1} = \frac{T}{fT - S} - \frac{T}{(1 - f)T - D},$$

and setting this to zero we find

$$f = 0.5 - \frac{D - S}{2T} \quad \text{and} \quad 1 - f = 0.5 + \frac{D - S}{2T}.$$

Clearly there will be more complicated considerations when there are not enough testing places to meet the requirement that all selected animals are tested. The condition will be met if the number of females tested is at least twice the number required, or $3S + D = T$. For the numerical example above with $S = 10$, $D = 250$, $T = 1000$, the theory above would give optimal numbers tested as 380 males and 620 females, quite close to the values used in the example. The advantage in this case is fairly small, but the principle is clear; it is better to use extra facilities to increase the selection differential in the sex in which more animals are to be selected.

In many cases, the two sexes will be selected for different traits, with different correlations with overall breeding value, and with different measurement costs. The analysis above can be easily modified to deal with these complications.

Let r_M and r_F be the correlations between male and female traits and overall breeding value, and let c_M and c_F be the respective costs of measuring a male and a female, while C_T is the total cost of measurement in the budget. If we set $T_M = C_T/c_M$ with T_F defined similarly for females, these are the total numbers of males and females which can be

measured if the whole budget is devoted to measuring one sex. Let f be the fraction of the testing budget devoted to testing males. Then the number of males tested is fC_T/c_M or fT_M while the number of females tested is $(1 - f)T_F$. The selection intensities in the two sexes are S/fT_M and $D/(1 - f)T_F$.

The response to selection is proportional to $i_M r_M + i_F r_F$ and we again use Smith's approximation to i . Here $?_M = fT_M/S$ and $?_F = (1 - f)T_F/D$ and we need to maximise

$$r_M[0.8 + \ln(?_M - 1)] + r_F[0.8 + \ln(?_F - 1)]$$

with respect to variation in f . On differentiating and setting the derivative to zero we have

$$\frac{r_M T_M}{S(q_M - 1)} = \frac{r_F T_F}{D(q_F - 1)}$$

from which the solution is

$$f = \frac{r_M}{r_M + r_F} \left[1 - \frac{D}{T_F} + \frac{r_F}{r_M} \frac{S}{T_M} \right].$$

This shows that, other things being equal, more should be expended on the more accurate evaluation, which comes as no surprise. Again, if the achievable proportion selected is greater in females than in males, other things being equal, more effort should be devoted to increasing the measurement of females. This conclusion is sometimes regarded as wrong because of a prejudice that since male selection is more powerful than female selection more effort should be put into evaluating males. This is not necessarily true. The optimal value above can easily be seen to reduce to the previous one if evaluation is equally accurate and equally costly in the two sexes, as it should.

A somewhat different approach was initiated by Jackson, Lax and Wilson (1986), who assumed that a range of measurements could be made on males and females, and that the accuracy of selection was dependent on measurement expenditure with diminishing returns. Specifically, they assumed that the accuracy of selection $r_{G\hat{G}}$ is related to the cost of measurement c by the relation $r_{G\hat{G}} = r_{\max}(1 - k^c)$ so that the accuracy is zero when $c = 0$ and is r_{\max} as c becomes very large. Of course, costs will not increase smoothly as implied by this equation, but will rise in a stepwise manner as more traits are added, but the equation allows some consideration to be given in a general manner. This approach was also used by Wade and James (1990), also with reference to Merino sheep breeding. They followed Jackson et al. in taking $k = 0.7$ and presented some results to indicate the type of outcome which might be observed. Assuming 50% of ewes and 1% of rams are required to be kept among the 500 candidates of each sex, a unit of cost = \$0.50 and a total measurement budget of \$1250 is allowed, they found that the optimum allocation was :

Test 170 rams at \$4.05 per head	\$687.65
Test 500 ewes at \$1.13 per head	\$562.50

The selected proportions would be 3% in rams and 50% in ewes, with accuracies of $0.94r_{\max}$ in males and $0.55r_{\max}$ in females. They also found that with this relation between cost and accuracy it always paid to measure at least twice as many animals as were required if that was possible, so that there was no need to select untested animals. Of course if more than half of the available females must be kept, this is impossible. In this example the proportion of the budget spent on testing males is $687/1250 = 0.55$. If we

apply the formula previously derived we have $r_M/(r_M + r_F) = 0.63$, $T_M = 1250 / 4.05 = 309$, $T_F = 1250 / 1.23 = 1106$, $r_F/r_M = 0.58$, so that

$$f = 0.63 \left[1 - \frac{250}{1106} + 0.58 \frac{5}{309} \right] = 0.49.$$

Thus the approximation does not give the same result, though both methods suggest that about the same amount should be spent on evaluating each sex, but that fewer males should be tested with a higher accuracy than females.

In the treatment given so far it has been assumed that the generation interval is fixed. However, it was pointed out by Ollivier (1990) that if generation interval is allowed to vary, the generation length will be affected by the amount of testing done, since the testing facilities will be on a per year basis, and by changing the numbers required for testing the breeder could alter the time needed for breeding replacements. He showed that when testing facilities are very limited, only males should be tested, and the maximum proportion to be selected is 27%, any other males required being taken at random from untested males. This is based on the following development.

Assume that the proportions of males and females required for breeding are p_M and p_F while the remainder $(1 - p_M)$ and $(1 - p_F)$ are chosen at random from untested animals. The standardised selection differentials are i_M and i_F respectively. The generation intervals for the selected males and those taken at random are t_{M1} and t_{M0} with similar definitions for females. Then the annual selection response is proportional to

$$\frac{p_M i_M + p_F i_F}{p_M t_{M1} + (1 - p_M) t_{M0} + p_F t_{F1} + (1 - p_F) t_{F0}}$$

Assuming no wastage of breeding animals and fertility the same at all ages, the generation lengths can be written

$$\begin{aligned} t_{F1} &= a + p_F n_F / 2q_F & t_{F0} &= a + (1 - p_F) / 2(? - q_F) \\ t_{M1} &= a + p_M n_M / 2f q_M & t_{M0} &= a + (1 - p_M) / 2f(? - q_M) \end{aligned}$$

where ? is the number of progeny of each sex per dam per year, f is the number of females mated to each male, n_M being the number of males tested per year per male selected, with n_F defined similarly, and q_M is the number of males tested per breeding female per year, with q_F defined similarly. If k is the proportion of the available animals of both sexes which can be tested annually, and f is the fraction of testing places given to males, then $q_F = 2k?(1 - f)$ and $q_M = 2k?f$. Using these relations and the relation between proportion selected and selection differential gave the result quoted above.

Ollivier considered a number of cases other than this simplest one, and identified conditions which led to testing of males only, either with or without some use of untested males, and to testing of some females. He let k_1 be the value below which some untested males should be used and k_2 denote the value above which some females should be tested. A small sample of the results is given in the table below.

Females per male (f)	10			100		
Value of a?	0.5	1.0	5.0	0.5	1.0	5.0
k_1	0.061	0.037	0.009	0.006	0.004	0.001
k_2	0.77	0.65	0.28	0.87	0.75	0.35

For higher female reproduction rate the value of k at which all selected males should come from those tested is lower, and in any case it is only for very restricted testing facilities that untested males should be used. Of course, untested females will be used in many situations. Ollivier's paper should be consulted for further details.

TWO-STAGE SELECTION

All of the above analyses were carried out assuming selection in a single stage. But it is often pointed out that if sequential culling is used, only those animals which survive the first culling need to be measured at the second stage, and therefore the survivors can be measured for more traits than would be possible for the same cost if there were no culling. Wade and James (1996) dealt with the case of two-stage selection in conjunction with the approach taken above for single-stage selection.

There was no simple solution in this case (at least none was found) and results were obtained by using computer programs which are described in the paper. First they estimated the relationship between cost and accuracy of EBVs using a series of measurements appropriate for selection of Merino sheep (about 10 years ago). The value of k in the relation $1 - k^c$ with c measured in dollars was 0.8, which fitted quite well, though there were a few measurement combinations for which the fit was not so good. A numerical example will illustrate what can be done.

A breeder has a flock of 1200 ewes which produce 1000 lambs annually, and a budget of \$2000 for measurement to allocate among the 500 progeny of each sex. The proportions required for breeding were 5% of rams and 50% of ewes. The optimum was to allocate 0.75 of the total to males and 0.25 to females. Thus the expenditure should be \$3 per ram and \$1 per ewe. Further analysis gave the following allocation of testing funds.

Test 250 ram hoggets at \$3.90 per head in stage 1, then choose the best 75 and test them in stage 2 at \$7.00 per head. This gives a total male testing cost of

$$250 \times \$3.90 + 75 \times \$7.00 = \$1500.$$

For females the best option was to test 500 in stage 1 at a cost of \$0.65 per head, select the best 400 and test them at stage 2 at a cost of \$0.44 per head, giving a cost of

$$500 \times \$0.65 + 400 \times \$0.44 = \$501,$$

and the total cost was \$2001 as budgeted.

In practice, these solutions would not be achievable. An appropriate stage 1 criterion for males might be clean fleece weight at \$4.00 per head, with perhaps mean fibre diameter and coefficient of variation of diameter at \$6.00 per head in stage 2. A visual classing criterion at \$1.00 per head might be added. For females, no measurements meeting the suggested costs are obvious. It might then be best to measure all ewes at a single stage for a trait such as hogget body weight or a visual wool trait. It would be necessary to consider the costs and accuracies of available traits in order to find a feasible plan.

There were often only small differences between various options and similar gains could be achieved using quite different allocations of funds, so not too much attention should

be paid to the actual optima reported. What is important is that in planning a breeding program a variety of possible strategies needs to be considered.

CONCLUSION

With the ready availability of quick and cheap computing, the planning of a breeding program should involve the modelling of a wide range of possible schemes with attention given to the way that funds are spent on obtaining data which provide efficient means of making genetic improvement.

REFERENCES

- Abplanalp, H. (1972) Selection of extremes. *Animal Production*, **14**:11 – 15.
- Henderson, C.R. (1963) Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* (W.D. Hanson & H.F. Robinson eds) NAS-NRC Publication **982**:141 – 163.
- Jackson, N., Lax, J. and Wilson, R.L. (1986) Sex and selection for fleece weight in Merino sheep. *Journal of Animal Breeding and Genetics*, **103**:97 – 115.
- James, J.W. (1966) Incomplete sampling for selection of a fixed proportion. *Australian Journal of Statistics*, **8**:127 – 137.
- James, J.W. (1979) Optimum family size in progeny testing when prior information is available. *Zeitschrift für Tierzucht und Züchtungsbiologie*, **95**:194 – 203.
- Ollivier, L. (1990) Mass selection in livestock using limited testing facilities. *Genetics Selection Evolution*, **22**:109 – 117.
- Rendel, J.M. (1958) Optimum family size in half sib family selection. *Biometrics*, **15**:376 – 381.
- Robertson, A. (1957) Optimum group size in progeny testing and family selection. *Biometrics*, **13**:442 – 450.
- Smith, C. (1959) A comparison of testing schemes for pigs. *Animal Production*, **1**:113 – 121.
- Smith, C. (1969) Optimum selection procedures in animal breeding. *Animal Production*, **11**:433 – 442.
- Wade, C.M. and James, J.W. (1990) Optimal allocation of test funds in selection for genetic gain. *Proceedings of the Australian Association of Animal Breeding and Genetics*, **8**:115 – 118.

Wade, C.M. and James, J.W. (1996) Optimum allocation of resources considering two sexes and selection in

Calculation of discounted genetic expressions

Calculations will follow the procedures given by P.R.Amer (1999) *N.Z.J.Agric.Res.*, **42**:325 – 336.

Amer defines the following numbers of genetic expressions discounted to a ewe age of one year.

Terminal Ewes.

X_{FEA} : ewe lambing trait expressed by a ewe over her lifetime.

X_{FEC} : ewe end of life trait.

X_{FEL} : ewe's progeny trait at birth.

X_{FES} : ewe's progeny trait at slaughter.

X_{FEH} : ewe hogget trait. $X_{FEH} = 1$ by definition.

Ewes Breeding Replacements

X_{FRA} The H subscript refers to hoggets. Otherwise the numbers correspond to those

X_{FRC} for terminal ewes, the R subscript denoting that the ewes breed replacements.

X_{FRH} A is for adult, C for cull, L for lamb, S for slaughter. F is for female, M for male.

X_{FRL}

X_{FRS}

Terminal Sire With All Progeny Slaughtered.

X_{MTL}

X_{MTS}

Sire Breeding Terminal Daughters.

X_{MEA}

X_{MEH}

X_{MEC}

X_{MEL}

X_{MES}

Sire Breeding Self-replacing Daughters.

X_{MRA}

X_{MRH}

X_{MRC}

X_{MRL}

X_{MRS}

To calculate these values, flock structure data, survival rates, ages at culling, etc. need to be specified. Amer gave the following definitions.

s vector of n elements with s_i = probability of survival from age $i - 1$ to age i . n is the greatest ewe age considered possible. We can set $s_1 = 0$, since it is not used.

p vector of n elements with p_i = average number of progeny per ewe reaching slaughter or reproductive age.

a vector of probabilities that a ewe survives and lambs at age i , given she was present at age 1.

$$a_i = \prod_{j=2}^i s_j \quad i = 2, \dots, c, \quad = 0 \text{ otherwise.}$$

where c is the age, $=n$, at which a ewe will be culled. For example, $a_4 = s_2 s_3 s_4$.

d vector of probabilities a ewe dies or is culled at age i .

$$d_i = 1 - a_i \text{ for } i = 2; \quad = a_{i-1} - a_i \text{ for } i = 3 \text{ to } c - 1; \quad = a_{i-1} \text{ for } i = c; \quad = 0 \text{ otherwise.}$$

Q discounting matrix.

$$Q_{ii} = [1/(1+r)]^{i-1} \text{ for } i = 1, \dots, n. \quad Q_{ij} = 0 \text{ if } i \neq j.$$

1 a vector of n ones.

$$X_{FEA} = \mathbf{1}'\mathbf{Q}\mathbf{a}$$

$$X_{FEC} = \mathbf{1}'\mathbf{Q}\mathbf{d}$$

l_s is probability of surviving from birth to one year of age.

$$X_{FEL} = \mathbf{p}'\mathbf{Q}\mathbf{a}/2l_s$$

$$X_{FES} = \mathbf{p}'\mathbf{Q}\mathbf{a}/2$$

h is number of years in planning horizon from birth of replacement ewe..

D is a transition matrix (h by h) of survival probabilities lagged by one row for each birth year.

$$D_{ij} = a_{i-j} \text{ for } j < i - 1 \text{ and } i - j < c; \quad = 0 \text{ otherwise.}$$

E is a matrix (h by h) of terminal lambs produced per ewe age group adjusted for ewe survival and lagged for birth year.

$$E_{i,j} = a_{i-j}(p_{i-j} - f) \text{ for } j < i - 1 \text{ and } i - j < c; \quad = 0 \text{ otherwise.}$$

f is the number of lambs required as replacements at first reproductive age per ewe lambing per year. With a constant age structure,

$$f = 1 / \sum_{i=1}^c a_i.$$

F, **G** and **H** are similar matrices for lambs, culled ewes and hoggets respectively.

$$F_{ij} = a_{i-j}p_{i-j} \text{ for } j < i - 1 \text{ and } i - j < c; \quad = 0 \text{ otherwise.}$$

$$G_{ij} = d_{i-j} \text{ for } j < i - 1 \text{ and } i - j < c + 1; \quad = 0 \text{ otherwise}$$

$$H_{ij} = 1 \text{ for } i - 1 = j; \quad = 0 \text{ otherwise.}$$

\mathbf{g}_k is a vector of h elements for increments of gene flow for each generation k from 1 to m where m is the maximum number of generations considered.

$$\mathbf{g}'_1 = (1 \ 0 \ 0 \ \dots \ 0)$$

$$\mathbf{g}_k = \frac{1}{2} \mathbf{f} \mathbf{D} \mathbf{g}_{k-1}$$

\mathbf{g}_{sum} is the aggregate yearly genetic expressions accumulated over generations.

$$g_{\text{sum}} = \sum_{k=1}^m g_k \cdot$$

\mathbf{q} is a vector of h elements of discount factors.

$$q_i = [1/(1+r)]^{i-2} \text{ so the discount factor is 1 at a ewe age of one year.}$$

Then

$$X_{\text{FRA}} = \mathbf{g}_{\text{sum}}' \mathbf{D}' \mathbf{q}$$

$$X_{\text{FRC}} = \mathbf{g}_{\text{sum}}' \mathbf{G}' \mathbf{q}$$

$$X_{\text{FRH}} = \mathbf{g}_{\text{sum}}' \mathbf{H}' \mathbf{q}$$

$$X_{\text{FRL}} = \mathbf{g}_{\text{sum}}' \mathbf{F}' \mathbf{q} / 2l_s$$

$$X_{\text{FRS}} = \mathbf{g}_{\text{sum}}' \mathbf{E}' \mathbf{q} / 2$$

Average number of lambs / ewe lambing either sold or breeding a replacement is

$$v = \mathbf{a}' \mathbf{p} / (\mathbf{1}' \mathbf{a}) \quad (\text{vectors are of length } n)$$

Average number of ewes lambing per sire used in any one year is e .

Average number of years a sire is used is y .

Proportion of sire's daughters selected as terminal ewes is u .

\mathbf{w} is a vector of y elements such that $w_i = [1/(1+r)]^i$.

\mathbf{z} is a vector of y elements such that $z_i = [1/(1+r)]^{i+1}$

Then

$$X_{\text{MTL}} = \frac{1}{2} \mathbf{1}' \mathbf{w} e v / l_s$$

$$X_{\text{MTS}} = \frac{1}{2} \mathbf{1}' \mathbf{w} e v$$

$$X_{\text{MEA}} = \frac{1}{4} \mathbf{1}' \mathbf{z} e v u X_{\text{FEA}}$$

$$X_{\text{MEH}} = \frac{1}{4} \mathbf{1}' \mathbf{z} e v u X_{\text{FEH}}$$

$$X_{\text{MEC}} = \frac{1}{4} \mathbf{1}' \mathbf{z} e v u X_{\text{FEC}}$$

$$X_{\text{MEL}} = \frac{1}{2} \mathbf{1}' \mathbf{z} e v (\frac{1}{2} u X_{\text{FEL}} + [1+r]/l_s)$$

$$X_{\text{MES}} = \frac{1}{2} \mathbf{1}' \mathbf{z} e v (\frac{1}{2} u X_{\text{FES}} + [1 - \frac{1}{2} u][1+r])$$

In Amer's paper the equations for X_{MEL} and X_{MES} give \mathbf{w} instead of \mathbf{z} . Dr. Amer has pointed out the error in a personal communication.

$$X_{\text{MRA}} = \frac{1}{4} \mathbf{1}' \mathbf{z} e v u X_{\text{FRA}}$$

$$X_{\text{MRH}} = \frac{1}{4} \mathbf{1}' \mathbf{z} e v u X_{\text{FRH}}$$

$$X_{\text{MRC}} = \frac{1}{4} \mathbf{1}' \mathbf{z} e v u X_{\text{FRC}}$$

$$X_{\text{MRL}} = \frac{1}{2} \mathbf{1}' \mathbf{z} e v (\frac{1}{2} u X_{\text{FRL}} + [1+r]/l_s)$$

$$X_{\text{MRS}} = \frac{1}{2} \mathbf{1}' \mathbf{z} e v (\frac{1}{2} u X_{\text{FRS}} + [1 - \frac{1}{2} u][1+r])$$

The basic structure used by Amer was as follows.

Oldest accepted age of ewe: $n = 7$.

The vector $\mathbf{s}' = (0 \ 0.95 \ 0.94 \ 0.93 \ 0.88 \ 0.83 \ 0.70)$

The vector $\mathbf{p}' = (0 \ 1.1 \ 1.25 \ 1.25 \ 1.35 \ 1.30 \ 1.10)$

Culling age: $c = 5$.

Planning horizon: $h = 10$ years

Discount rate: $r = 0.07$

Average ram working life: $y = 3$ years
Ewes lambing per sire joined: $e = 90$
Proportion of ewe lambs retained: $u = 0.85$.
Lamb survival to mating: $l_s = 0.8$.

The object of this exercise is to use Microsoft Excel to calculate these numbers of discounted expressions from the base data.

This will involve setting up the various matrices, with some matrix transposition and matrix multiplication, and vector additions.

The matrix (vector) operations required can be done as shown on the following page provided by Julius van der Werf.

Matrix calculations using Excel

You can do some basic matrix calculations with MS Excel.

First put in the values of your matrices

To multiply two matrices:

- select an area of the size of the resulting matrix
- type: =**MMULT**(
- select the area of the first matrix
- type a comma (,)
- select area of the second matrix
- type a close bracket)
- press: Ctrl_Shift_Enter

To add or subtract a matrix (vector):

- select an area of the size of the resulting matrix
- type: = (
- select the area of the first matrix
- type a + or -
- select area of the second matrix
- type a close bracket)
- press: Ctrl_Shift_Enter

To invert a matrix:

- select an area of the size of the resulting matrix
- type: =**MINVERSE**(
- select the area of the first matrix
- type a close bracket)
- press: Ctrl_Shift_Enter

To transpose a matrix (vector):

- select an area of the size of the resulting matrix
- type: =**TRANSPOSE**(
- select the area of the first matrix
- type a close bracket)
- press: Ctrl_Shift_Enter

A more specialized matrix calculation program is MATLAB. It contains many more matrix functions and mathematical function than excel. MATLAB allows you to make and run programs, draw graphs, and run simulation). A MATLAB student version is very well suitable for animal breeding problems and quite easy to use.