Basic Principles of Response to Selection

Jack Dekkers

Response to selection = (mean BV of progeny) – (mean BV parental generation)

Predicting Mean BV of Progeny Generation

Guiding principle of genetic improvement: parents with high BV tend to have progeny with high BV

Single progeny: $g_o = \frac{1}{2}g_s + \frac{1}{2}g_d + g_m$

$E(g_o) = \frac{1}{2}g_s + \frac{1}{2}g_d$ because $E(g_m) = 0$

Progeny generation: $E(g_o) = \frac{1}{2} \text{ave. all males} + \frac{1}{2}(S_s + S_d)$

$S = \text{genetic superiority of the selected parents}$

$S = \text{average genetic superiority of selected parents}$

Response $= R = \Delta G = (\text{mean BV progeny generation}) – (\text{mean BV parental generation})$

$E(R) = g_o - g_p$ with $g_p = \frac{1}{2}(g_s + g_d)$

$E(R) = \frac{1}{2}(g_s + g_d) + \frac{1}{2}(S_s + S_d) - \frac{1}{2}(g_s + g_d)$

$\Rightarrow$ expected response to selection = average genetic superiority of selected parents

For current generation, genetic superiority of selected parents can be estimated based on their EBV:

If EBV are unbiased: $E(g) = \hat{g}$

Then – for an individual: $E(g_o) = \frac{1}{2} \hat{g}_s + \frac{1}{2} \hat{g}_d$

– for progeny generation: $E(g_o) = \frac{1}{2} \text{ave. all males} + \frac{1}{2}(S_s + S_d)$

$\hat{S} = \text{estimated genetic superiority of the selected parents}$

$\Rightarrow$ Expected response $= E(R) = \frac{1}{2}(S_s + S_d)$
Predicting Response per Generation into the Future

\( \hat{g} \) are not available for future generations or for alternative designs

\( \Rightarrow \) need to predict genetic superiority of selected parents by other means

- do this based on the linear relationship between the selection criterion, \( I \), and \( g \)

Predict \( g_i \) given \( I_i \):

\[
\hat{g}_i = \bar{g} + b_{gl} (I_i - \bar{I})
\]

Predict the mean BV of a group of selected animals:

\[
\bar{g}^* = \bar{g} + b_{gl} (\bar{I}^* - \bar{I})
\]

Predict the genetic superiority of selected parents:

\[
\hat{S} = \bar{g}^* - \bar{g} = b_{gl} (\bar{I}^* - \bar{I})
\]

Define **intensity of selection**: \( i = (\bar{I}^* - \bar{I}) / \sigma_I \)

How many SD (for the selection criterion) is the selected group better than all candidates

Using

\[
(\bar{I}^* - \bar{I}) = i\sigma_I
\]

\[
\hat{S} = b_{gl} i \sigma_I
\]

From regression:

\[
b_{gl} = r_{gl} \frac{\sigma_g}{\sigma_I}
\]

\[
\hat{S} = r_{gl} \frac{\sigma_g}{\sigma_I} (i \sigma_I) = i r_{gl} \sigma_g
\]

\( r_{gl} = \text{accuracy of selection} \)

\( \hat{S} = i r_{gl} \sigma_g \) is general equation to predict genetic superiority of selected parents

applies whenever selection criterion \( (I) \) is linearly related to the BV \( (g) \)

Use \( \hat{S} = i r_{gl} \sigma_g \) to model the genetic mean of the next generation recursively:

\[
\bar{g}_{o} = \frac{1}{2}(\bar{g}_s + S_s) + \frac{1}{2}(\bar{g}_d + S_d)
\]

or predict response per generation:

\[
R = \frac{1}{2}(S_s + S_d) = \frac{1}{2}(i_s \sigma_g + i_d \sigma_g)
\]

**Selection Intensity with Truncation Selection**
Adjustments to standard (large unrelated population) selection intensities:

- **small populations:** $i < z/p$
  - order statistics
  - approximation: $p = \frac{s + z^2}{2n + 2s}$
    
    $s = \#$ selected $n =$ population size
    
    $i^* = \frac{z^*}{p^*}$
    
    use $p^*$ rather than $p$ to derive intensity

- **Correct intensities for correlated EBV** among selection candidates due to:
  1) genetic relationships between candidates of selection
  2) use of the same information in calculating the EBV

  *e.g.* $n_{fs}$ full sib families with $n_w$ individuals per family and selection on pedigree

  $\Rightarrow \text{Correlation between EBV of full sibs} = 1$

  $\Rightarrow$ select $n_c/n_w$ families out of $n_{fs}$
  
  instead of: select $n_c$ individuals out of $n_{fs}n_w$

Approximations to account for effect of correlated EBV on intensity:

Rawlings (1976)

$t^* = \sqrt{1 - t_{av}^2} i$

$t_{av}$ = average correlation between the selection criterion across all possible pairs of selection candidates

Unrelated full sib families

$t_{av} = t_{fs} \frac{n_w - 1}{n_w n_{fs} - 1}$

$t_{fs}$ = correlation between selection criterion of full sibs

Full sib families nested within half sib families (Meuwissen 1991):

$n_{hs}$ sires mated with $n_{fs}$ dams each producing $n_w$ offspring

Correlation between half sibs $= t_{hs}$

Correlation between full sibs $= t_{fs}$

Correlation between non-relatives $= 0$

$t_{av} = \frac{t_{fs}(n_w - 1) + t_{hs}n_w(n_{fs} - 1)}{n_w n_{fs} n_{hs} - 1}$

Correlations between EBV of relatives based on selection index theory:

$t_{i,j} = \text{corr}(I_i, I_j) = \text{corr}(b'x_i, b'x_j) = \frac{b'\text{cov}(x_i, x_j)b}{b'Rb} = \frac{b'Rb}{b'Pb}$

$R = m \times m$ matrix with covariances between information sources on the relatives
Asymptotic Response per Unit Time

Equal selection in males and females: Response per generation: \( R = S = \frac{i \sigma_g}{r_{gl}} \)

Response per year: \( R_{yr} = \frac{S}{L} \)

Generation interval = \( L \) = average age of the parents when their progeny are born

= average time between birth of the parents and birth of progeny

Unequal selection in males and females:
Response per generation: \( R = \frac{1}{2}(S_s + S_d) \)

Response per year:
\[ R_{yr} = \frac{S_s + S_d}{L_s + L_d} \]

= ‘steady state’ or ‘asymptotic’ response

Responses from year to year obtained recursively based on mean of selected parents:
\[ \bar{g}_o = \frac{1}{2}(\bar{g}_s + S_s) + \frac{1}{2}(\bar{g}_d + S_d) \]

Derivation of asymptotic response:
\[ R_{yr} = \frac{S_s + S_d}{L_s + L_d} \]

When response constant at \( R_{yr}/yr \):
\[ \bar{g}_s = \bar{g}_o - L_sR_{yr} \]
\[ \bar{g}_d = \bar{g}_o - L_dR_{yr} \]
\[ \bar{g}_o = \frac{1}{2}(\bar{g}_s + S_s) + \frac{1}{2}(\bar{g}_d + S_d) \]

Substituting we get:
\[ \bar{g}_o = \frac{1}{2}(\bar{g}_o - L_sR_{yr} + S_s) + \frac{1}{2}(\bar{g}_o - L_dR_{yr} + S_d) \]
\[ = \bar{g}_o - \frac{1}{2}R_{yr}(L_s + L_d) + \frac{1}{2}(S_s + S_d) \]

Rearranging \( R_{yr} = \frac{S_s + S_d}{L_s + L_d} \)
Four pathways of genetic improvement (Rendel and Robertson, 1950):

- sires of males \( sm \)
- dams of males \( dm \)
- sires of females \( sf \)
- dams of females \( df \)

Asymptotic response/yr = 
\[
R_{yr} = \frac{S_{sm} + S_{dm} + S_{sf} + S_{df}}{\sum L_i} \\
= \frac{L_{sm} + L_{dm} + L_{sf} + L_{df}}{\sum i_i \cdot S_i} \quad S_i = i_l \cdot r_l \cdot \sigma_g
\]

Table 1 Intensity and accuracy of selection and generation interval in a highly efficient hypothetical progeny-testing program for improving milk yield in dairy cattle.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Proportion Selected ( (p_i) )</th>
<th>Intensity ( (i_l) )</th>
<th>Accuracy ( (r_l) )</th>
<th>Genetic Superiority ( (S_i = i_l \cdot r_l \cdot \sigma_g) )</th>
<th>Generation Interval (yr) ( (L_l) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires of males</td>
<td>2 %</td>
<td>2.42</td>
<td>0.90</td>
<td>2.178 ( \sigma_g )</td>
<td>6</td>
</tr>
<tr>
<td>Sires of females</td>
<td>10 %</td>
<td>1.75</td>
<td>0.90</td>
<td>1.575 ( \sigma_g )</td>
<td>7</td>
</tr>
<tr>
<td>Dams of males</td>
<td>0.5 %</td>
<td>2.89</td>
<td>0.60</td>
<td>1.743 ( \sigma_g )</td>
<td>5</td>
</tr>
<tr>
<td>Dams of females</td>
<td>90 %</td>
<td>0.19</td>
<td>0.60</td>
<td>0.114 ( \sigma_g )</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td>( \Sigma S = 5.601 \sigma_g )</td>
<td>( \Sigma L = 24 )</td>
</tr>
</tbody>
</table>

\[
R_{yr} = \frac{(2.42 \times 0.9 + 1.75 \times 0.9 + 2.89 \times 0.6 + 0.19 \times 0.6)}{6 + 7 + 5 + 6} \cdot \sigma_g = 0.233 \sigma_g \text{ per yr}
\]
Selection Across Multiple Age Groups

E.g. three age groups:

<table>
<thead>
<tr>
<th>Age group</th>
<th>Relative number of candidates</th>
<th>Generation interval</th>
<th>Mean within group</th>
<th>Accuracy of selection within group</th>
<th>Fraction selected within group</th>
<th>Genetic superiority within group</th>
<th>Genetic mean selected individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$w_1$</td>
<td>$L_1$</td>
<td>$g_1$</td>
<td>$r_1$</td>
<td>$p_1$</td>
<td>$S_1$</td>
<td>$g_1 + S_1$</td>
</tr>
<tr>
<td>2</td>
<td>$w_2$</td>
<td>$L_2$</td>
<td>$g_2$</td>
<td>$r_2$</td>
<td>$p_2$</td>
<td>$S_2$</td>
<td>$g_2 + S_2$</td>
</tr>
<tr>
<td>3</td>
<td>$w_3$</td>
<td>$L_3$</td>
<td>$g_3$</td>
<td>$r_3$</td>
<td>$p_3$</td>
<td>$S_3$</td>
<td>$g_3 + S_3$</td>
</tr>
</tbody>
</table>

$\sum w_i = 1$

Total proportion selected is: $P = p_1 w_1 + p_2 w_2 + p_3 w_3$

Genetic mean of selected individuals across groups is:

$g^* = \frac{1}{P} \left\{ p_1 w_1 g_1^* + p_2 w_2 g_2^* + p_3 w_3 g_3^* \right\}$

$g_i^* = g_i + S_i = i_i r_i \sigma_g$

Applied to separate selection of males and females:

Genetic mean of selected sires:

$g_s^* = \frac{1}{P_s} \sum p_{si} w_{si} (g_{si}^* + S_{si})$

Genetic mean of selected dams:

$g_d^* = \frac{1}{P_d} \sum p_{di} w_{di} (g_{di}^* + S_{di})$

Genetic mean of progeny:

$E(g_o) = \frac{1}{2} g_s^* + \frac{1}{2} g_d^*$

For asymptotic gain/yr: Compute average within-group genetic superiority and generation interval across age classes:

For sires: $S_s = \frac{1}{P_s} \sum p_{si} w_{si} S_{si}$

$L_s = \frac{1}{P_s} \sum p_{si} w_{si} L_{si}$

$R_{yr} = \frac{S_s + S_d}{L_s + L_d}$

For dams $S_d = \frac{1}{P_d} \sum p_{di} w_{di} S_{di}$

$L_d = \frac{1}{P_d} \sum p_{di} w_{di} L_{di}$

Selection Across Age Groups

e.g. Selection of Bull Dams $\sigma_g = 550$ kg

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Age at Birth of Progeny</th>
<th>% of Bull Dams</th>
<th>% Selected</th>
<th>$i$</th>
<th>$r$</th>
<th>Genetic Superiority in $\sigma_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heifers</td>
<td>2 yr</td>
<td>50%</td>
<td>2.5%</td>
<td>2.34</td>
<td>.55</td>
<td>707.9</td>
</tr>
<tr>
<td>1st Lact.</td>
<td>3 yr</td>
<td>30%</td>
<td>1.5%</td>
<td>2.53</td>
<td>.68</td>
<td>946.2</td>
</tr>
<tr>
<td>2nd Lact.</td>
<td>4 yr</td>
<td>20%</td>
<td>1.5%</td>
<td>2.53</td>
<td>.72</td>
<td>1001.9</td>
</tr>
</tbody>
</table>

Selection Across Age Groups

- Pooled Generation Interval

$L_{dm} = 50\% \ast 2$

$+ 30\% \ast 3$

$+ 20\% \ast 4$

= 2.7 yr

- Pooled Genetic Superiority

$S_{dm} = 50\% \ast 707.9$

$+ 30\% \ast 946.2$

$+ 20\% \ast 1001.9$

= 838.2 kg
Accounting for the use of young bulls

\( y = \text{proportion of females produced from young bulls} \)

**Dairy Cattle Progeny Testing Program**

- **Active Bulls**
- **Sires of Males**
- **Dams of Males**
- **Sires of Females**
- **Dams of Females**
- **(Inter)National Cow Population**

**Predicting Genetic Progress**

**Accounting for use of young bulls**

- **Progeny Test**
  - **Sires of Males**
  - **Sires of Females**
  - **Dams of Males**
  - **Dams of Females**

**Farmer Decisions**

- **Cows in commercial herds**

**Herd Replacements**

**Genetic superiority sires of females**

\( S_{sf} = yS_{yb,f} + (1-y)S_{pb,f} \)

Without MAS/GS

\( S_{yb,f} = 0 \) because \( p_{yb,f} = 1 \Rightarrow i_{yb,f} = 0 \) or \( r_{sf} = 0 \)

**Generation interval of sires of females**

\( L_{sf} = yL_{yb,f} + (1-y)L_{pb,f} \)

**Table 2.** Intensity and accuracy of selection and generation interval in a highly efficient hypothetical progeny-testing program for improving milk yield in dairy cattle with accounting for 20% use of young bulls to breed female replacements.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Proportion Selected (( p_i ))</th>
<th>Intensity (( i_i ))</th>
<th>Accuracy (( r_i ))</th>
<th>Genetic Superiority ( (S_i = i \cdot r_i \cdot \sigma_g) )</th>
<th>Generation Interval (yr) ( (L_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires of males</td>
<td>2 %</td>
<td>2.42</td>
<td>0.90</td>
<td>2.178( \sigma_g )</td>
<td>6</td>
</tr>
<tr>
<td>Sires of females</td>
<td>- Young - Proven</td>
<td>100 %</td>
<td>0</td>
<td>0</td>
<td>1.575 ( 1.260 \sigma_g ) ( 2 ) ( 7 ) ( 6 )</td>
</tr>
<tr>
<td>Dams of males</td>
<td>0.5 %</td>
<td>2.89</td>
<td>0.60</td>
<td>1.734( \sigma_g )</td>
<td>5</td>
</tr>
<tr>
<td>Dams of females</td>
<td>90 %</td>
<td>0.19</td>
<td>0.60</td>
<td>0.114( \sigma_g )</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td>( \Sigma S = 5.268 \sigma_g )</td>
<td>( \Sigma L = 23 )</td>
</tr>
</tbody>
</table>

Now response per year becomes:

\[ R = \frac{5.268}{23} \sigma_g = 0.230 \sigma_g \text{ per yr} \]

\( \Rightarrow \) Genetic_gain.xls
Optimizing selection across (age) groups

Optimize proportions to select from each group to maximize the average genetic value of the selected group (for total fraction selected \( P \)).

Assume that selection criterion is unbiased for each age group: \( E(g_i) = I_i \)

- truncation selection across distributions of \( I \)

- find the truncation point where selection across all distributions yields a total proportion selected of \( P \)

Various algorithms can be used
- Bisection
- gradient algorithms
- Genetic algorithms

\[ P = p_1w_1 + p_2w_2 + p_3w_3 \]

**Figure 3.2 Schematic representation of truncation selection of a total proportion \( P \) across multiple overlapping distributions**

\[ \sigma_1 \]

\[ p_1 \]

\[ \sigma_2 \]

\[ p_2 \]

\[ \sigma_3 \]

\[ p_3 \]

\[ g_1 \]

\[ g_2 \]

\[ g_3 \]

\[ T \]

**Multrunc.xls**

**truncselDE.xls**
Balancing Accuracy vs Generation Interval

Selection on EBV Across Age Groups

Young Bulls
\( r = .57 \)

Sib-tested Bulls
\( r = .65 \)

Proven Bulls
\( r = .90 \)