Molecular Breeding Values for MAS or GS

Jack Dekkers

Accuracy of Genomic Predictions   After Goddard et al. (2011, JABG 128);
notation after Dekkers (2007, JABG 124)
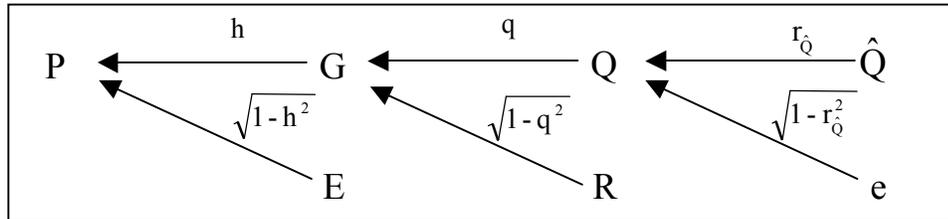
Accuracy of $\hat{Q}$ as a predictor of G = r_{MBV} = q $r_{\hat{Q}}$

Model for phenotype: P = G + E   Trait heritability = h^2

Model for BV: G = Q + R   G = total BV
Q = genetic effects captured by marker(s)
R = residual polygenic effects

$$\sigma_G^2 = \sigma_Q^2 + \sigma_R^2 \qquad \sigma_Q^2 = r^2_{M\text{-}QTL}*2pq\alpha^2 \quad \text{for single marker}$$

in LD (=$r^2_{M\text{-}QTL}$) with 1 QTL

$q^2 = \sigma_Q^2 / \sigma_G^2$ = proportion of genetic variance captured by markers

(i.e. with infinite training size)
- depends on marker-QTL
LD/linkage/relationships

$q^2 = M/(M_e + M)$ for GS with many QTL and markers and QTL
have similar properties (MAF). q^2 is often lower.

with $M_e = 2N_eLk/\ln(N_eL)$ or $= 2N_eLk/\ln(2N_e)$
M = number of markers
M_e = effective number of independent chromosome
segments segregating in the population
N_e = effective population size
L   = average length of a chromosome in Morgan
k   = number of chromosomes

$\hat{Q}$ = Molecular Breeding Value (MBV)
= estimate of Q based on "training data"

$$Q = \hat{Q} + e \qquad\qquad e = \text{prediction error}$$

$r_{\hat{Q}}$ = Corr($\hat{Q}$,Q) = Accuracy with which Q is estimated by markers

$r_{\hat{Q}}^2 = \theta/(1+\theta)$ for genomic selection with many small QTL

with $\theta = Tq^2h^2/M_e$ (Goddard et al. 2011 JABG)
Note: q^2h^2 = marker-based h^2

Accuracy of $\hat{Q}$ as a predictor of G = r_{MBV} = q $r_{\hat{Q}}$

1

Daetwyler *et al.* (2008) $\mathbf{r_{MBV}}^2 = \lambda h^2/(\lambda h^2 + 1)$ with $\lambda = T/M_e$ with $M_e = 2N_eLk$

Assumes complete genome coverage ($q^2 = 1$)

Correction when $T \gg M_e$ $r_{MBV\_corrected}^2 = r_{MBV}^2 (1 + r_{MBV}^4 M_e/2T)$
Accounts for reduction in residual variance with accurate prediction.

See spreadsheet GS_accuracy.xls

Additional issues (based in part on Goddard et al. 2011):

• Which N_e to use? Current or historical? Blend of current/historic N_e
(~340 for Holstein – Goddard personal comm.)

N_e at t=1/2c generations in the past can be estimated from average LD (r^2) between SNPs that are c cM apart (Hayes et al. 2003) based on:

$$E(r_c^2) = \frac{1}{1 + 4N_{e,t}c} \quad \Longrightarrow \quad \hat{N}_{e,t} = \frac{1}{4c}\left(\frac{1}{r_c^2} - 1\right)$$

De Roos et al. (Genetics 2009)
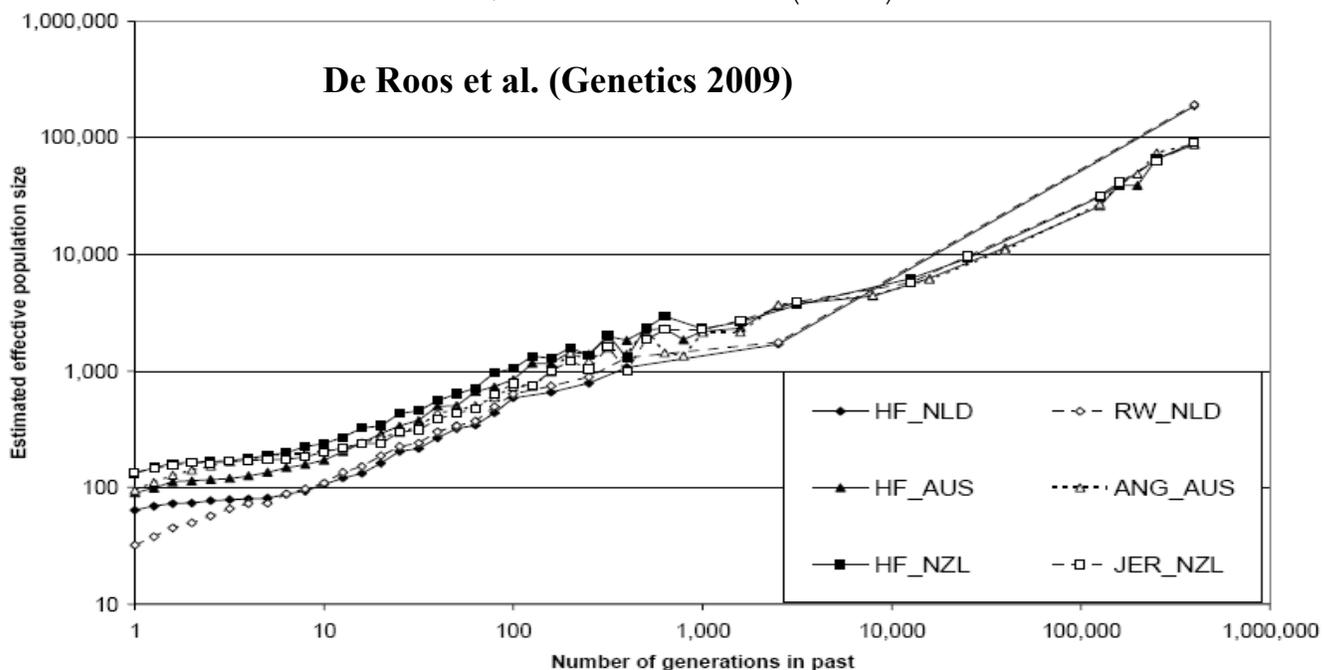
- Predictive equations using M_e assume many QTL with small effects.
  - May underestimate accuracy from variable selection method when QTL of large effect segregate
- Predicted accuracies apply to an average individual.
  - Accuracies will be higher for individuals that are more closely related to the training individuals.
- Presence of population structure affects accuracies (Daetwyler et al. 2012. JAS)

- M_e needs to reflect the structure of the population
  - M_e smaller if the population is more closely related
    - Goddard et al. (2011) provide equations to estimate M_e from the observed relationship matrix.
  - M_e needs to reflect the relationship of target population with training population
  - M_e larger when predicting across breeds

- Accuracies scale ~equally with M_e and T ➔ Combinations with constant T/M_e have approximately equal accuracy

  ➔ for stochastic simulation of smaller genomes, training population sizes need to be scaled accordingly (Meuwissen 2009).

  ➔ if for a simulated genome of 10 chromosomes of 1 Morgan each (instead of 30 chromosomes of 1 Morgan), accuracy = 0.7 with T=2,000
    ➔ T needs to be 3*2,000 = 6,000 for a full genome.

References:

Goddard M.E. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245-257.

Goddard M.E., Hayes B.J., Meuwissen T.H. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breed Genet. 128:409-421.

Hayes B.J., Visscher P.M., Goddard M.E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* 91: 47–60.

Meuwissen, T. H. E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35.

Meuwissen, Hayes and Goddard. 2013. Accelerating improvement of livestock with genomic selection. Annu. Rev. Anim. Biosci. 1:221-237.

Incorporation MBV in Index Calculations of Total EBV (GEBV)
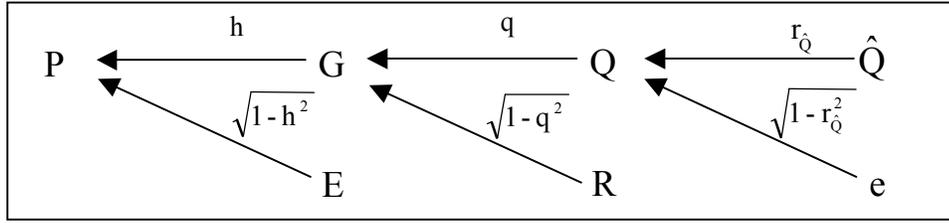
Lande and Thompson (1990, Genetics) index: $I_i = b_Q \hat{Q}_i + b_P P_i$

   $\hat{Q}_i$ = MBV for individual i, = individual's EBV based on markers alone

   P_i = individual's phenotype

   $r_{MBV} = q\, r_{\hat{Q}}$ = accuracy of MBV as a predictor of total BV G

   $q^2 = \sigma_Q^2 / \sigma_G^2$ = proportion of genetic variance captured by markers (with inf. train size)

Assuming unbiased MBV (regression of G (or P) on MBV = 1; if not then divide MBV by $b_{P,MBV}$
to make it = 1)

- Variance of MBV = Var($\hat{Q}$) = $r_{MBV}^2 \sigma_G^2$
- Cov($\hat{Q}$,P) = Cov($\hat{Q}$,G)   = $r_{MBV}^2 \sigma_G^2$

Table 1. Example calculation of MBV and index of phenotype and MBV with 3 SNPs with allele substitution effect estimates (allele A vs. B) of +10, +5, and −10 for for SNPs 1, 2, 3. The SNPs jointly explain 50% of the genetic variance for a trait with heritability 0.5. Resulting index weights on MBV and phenotype are $^2/_3$ and $^1/_3$, respectively.

Animal	QTL 1 Genotype	Value	QTL 2 Genotype	Value	QTL 3 Genotype	Value	MBV	Phenotype	Index value
1	AA	10	AA	5	AA	-10	5	35	15.0
2	AA	10	AA	5	BB	10	25	-10	13.3
3	AB	0	BB	-5	AB	0	-5	-15	-8.3
4	AB	0	BB	-5	AA	-10	-15	15	-5.0
5	BB	-10	AA	5	AB	0	-5	25	5.0

GEBV = $I_i = b_Q \hat{Q}_i + b_P P_i$ Derive optimal index weights by sel. index theory:

$$\begin{bmatrix} b_Q \\ b_P \end{bmatrix} = \mathbf{P}^{-1}\mathbf{G} = \begin{bmatrix} r_{MBV}^2 \sigma_G^2 & r_{MBV}^2 \sigma_G^2 \\ r_{MBV}^2 \sigma_G^2 & \sigma_P^2 \end{bmatrix}^{-1} \begin{bmatrix} r_{MBV}^2 \sigma_G^2 \\ \sigma_G^2 \end{bmatrix} = \begin{bmatrix} \dfrac{1-h^2}{1-r_{MBV}^2 h^2} \\[2em] h^2 \dfrac{(1-r_{MBV}^2)}{1-r_{MBV}^2 h^2} \end{bmatrix}$$

Thus, the relative weight on the MBV relative to phenotype is: $\dfrac{b_Q}{b_P} = \dfrac{1/h^2 - 1}{1 - r_{MBV}^2}$

Table 2. Index weight on molecular score relative to phenotype (b_Q/b_P) for different heritabilities and accuracy of MBV.

Heritability	Squared accuracy of MBV (r_{MBV}^2)				
(h^2)	0.10	0.25	0.50	0.75	1.00
0.10	10	12	18	36	Total weight
0.25	3.33	4	6	12	Total weight
0.50	1.11	1.33	2	4	Total weight
0.75	0.37	0.44	0.67	1.33	Total weight
1.00	0	0	0	0	Either

The Lande and Thompson (1990) formulation of the index can be extended to situations where phenotypes of relatives are used:

$$GEBV = I = \mathbf{b'X} = \begin{bmatrix} \mathbf{b'_Q}, \mathbf{b'_P} \end{bmatrix} \begin{bmatrix} \mathbf{X_Q} \\ \mathbf{X_P} \end{bmatrix}$$

$\mathbf{X_Q}$ = vector with marker-based EBV on the individual itself and/or its relatives

$\mathbf{X_P}$ = vector with phenotypic records on the individual itself and/or its relatives

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_Q \\ \mathbf{b}_P \end{bmatrix} = \mathbf{P}^{-1}\mathbf{G} = \begin{bmatrix} Var(\mathbf{X_Q}) & Cov(\mathbf{X_Q}, \mathbf{X'_P}) \\ Cov(\mathbf{X_P}, \mathbf{X'_Q}) & Var(\mathbf{X_P}) \end{bmatrix}^{-1} \begin{bmatrix} Cov(\mathbf{X_Q}, G) \\ Cov(\mathbf{X_P}, G) \end{bmatrix}$$

Elements ij of all matrices and vectors that involve $\mathbf{X_Q}$ = $a_{ij}\ r_{MBV}^2\ \sigma_G^2$ a_{ij} = genetic relationship

These methods can also be extended to include data on multiple traits and multiple trait breeding goals (see Lande and Thompson 1990).

$$\text{Accuracy of GEBV} = \mathbf{r_{GEBV}} = \sqrt{\frac{\mathbf{b'G}}{\sigma_G^2}}$$

It is useful to note that the index $GEBV = I_i = b_Q \hat{Q}_i + b_P P_i$ can be reparameterized into an equivalent index of MBV and the **phenotype adjusted for the MBV** as follows:

$$GEBV = I_i' = b_Q' \hat{Q}_i + b_P' P_i' \qquad \text{with} \qquad P_i' = P_i - \hat{Q}_i$$

Residual heritability = heritability of phenotype adjusted for the MBV:

$$h_{P'}^2 = \frac{\sigma_G^2 - r_{MBV}^2 \sigma_G^2}{\sigma_P^2 - r_{MBV}^2 \sigma_G^2} = \frac{h^2(1 - r_{MBV}^2)}{1 - r_{MBV}^2 h^2}$$

$$\begin{bmatrix} b_Q' \\ b_P' \end{bmatrix} = \mathbf{P}^{-1}\mathbf{G} = \begin{bmatrix} r_{MBV}^2 \sigma_G^2 & 0 \\ 0 & \sigma_P^2 - r_{MBV}^2 \sigma_G^2 \end{bmatrix}^{-1} \begin{bmatrix} r_{MBV}^2 \sigma_G^2 \\ \sigma_G^2 - r_{MBV}^2 \sigma_G^2 \end{bmatrix} = \begin{bmatrix} 1 \\ h_{P'}^2 \end{bmatrix}$$

Thus, the resulting index is: $I_i' = \hat{G}_i = \hat{Q}_i + h_{P'}^2 P_i'$

Advantage of index I' over index I is that its index weights remain constant over generations as the variance of MBV changes (with changing marker frequencies)

Note: $h_P^2.P_i'$ = individual's EBV for polygenes, $\hat{Q}_i$, based on own phenotype adjusted for the QTL.

This index can be expanded to BLUP EBV from a model that includes QTL or markers as a fixed or random effect.

Such models result in estimates of molecular scores, $\hat{Q}_i$, and EBV for residual genetic effects, $\hat{R}_i$, with accuracy $r_{\hat{R}}$.

Index weights for combining these two estimates, realizing that the variance of Residual EBV is equal to $r_{\hat{R}}^2\sigma_R^2$, where $\sigma_R^2 = h_{P'}^2(\sigma_P^2 - r_{MBV}^2\sigma_G^2)$ is the polygenic variance, can be derived as:

$$\begin{bmatrix} b_Q' \\ b_P' \end{bmatrix} = \mathbf{P^{-1}G} = \begin{bmatrix} r_{MBV}^2\sigma_G^2 & 0 \\ 0 & r_{\hat{R}}^2\sigma_R^2 \end{bmatrix}^{-1} \begin{bmatrix} r_{MBV}^2\sigma_G^2 \\ r_{\hat{R}}^2\sigma_R^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus the GEBV is: $\quad I_i' = \hat{G}_i = \hat{Q}_i + \hat{R}_i$

Exercise derivation of indexes and responses with own and sire genotypes.

Predicting Response to Marker-Assisted or Genomic Selection

Standard selection index theory can be used to predict responses to selection on GEBV, assuming multi-variate normality.

Consider the previously derived selection index of MBV and own phenotype:

$$\text{GEBV} = I_i = b_Q \hat{Q}_i + b_P P_i$$

$$\begin{bmatrix} b_Q \\ b_P \end{bmatrix} = \mathbf{P}^{-1}\mathbf{G} = \begin{bmatrix} r_{MBV}^2 \sigma_G^2 & r_{MBV}^2 \sigma_G^2 \\ r_{MBV}^2 \sigma_G^2 & \sigma_P^2 \end{bmatrix}^{-1} \begin{bmatrix} r_{MBV}^2 \sigma_G^2 \\ \sigma_G^2 \end{bmatrix} = \begin{bmatrix} \dfrac{1-h^2}{1-r_{MBV}^2 h^2} \\ h^2 \dfrac{(1-r_{MBV}^2)}{1-r_{MBV}^2 h^2} \end{bmatrix}$$

The accuracy of this index and response to selection can be derived by standard selection index theory as:

$$r_{GEBV} = \sqrt{\frac{\mathbf{b'G}}{\sigma_G^2}} = \sqrt{\begin{bmatrix} \dfrac{1-h^2}{1-r_{MBV}^2 h^2} & h^2 \dfrac{(1-r_{MBV}^2)}{1-r_{MBV}^2 h^2} \end{bmatrix} \begin{bmatrix} r_{MBV}^2 \\ 1 \end{bmatrix}}$$

$$= \sqrt{\frac{r_{MBV}^2 - 2r_{MBV}^2 h^2 + h^2}{1-r_{MBV}^2 h^2}} = \sqrt{r_{MBV}^2 + h^2 \frac{(1-r_{MBV}^2)^2}{1-r_{MBV}^2 h^2}}$$

Similarly for the alternate index parameterization:

$$\text{GEBV'} = I_i' = b_Q' \hat{Q}_i + b_P' P_i'$$

and

$$r_{GEBV'} = \sqrt{\frac{\mathbf{b'G}}{\sigma_G^2}} = \sqrt{\begin{bmatrix} 1 & h_{P'}^2 \end{bmatrix} \begin{bmatrix} r_{MBV}^2 \\ 1-r_{MBV}^2 \end{bmatrix}} = \sqrt{r_{MBV}^2 + h_{P'}^2 (1-r_{MBV}^2)}$$

Using $h_{P'}^2 = \dfrac{h^2(1-r_{MBV}^2)}{1-r_{MBV}^2 h^2}$ it can be shown that $r_{GEBV'} = r_{GEBV}$, i.e. the two indexes are equivalent

Assuming equal selection in males and females, with selection intensity i, response to selection can be predicted as:
$$R_{MAS} = i\, r_{GEBV}\, \sigma_g$$

Response to phenotypic selection without QTL information is:
$$R_P = i\, h\, \sigma_g$$

Thus, the efficiency of selection using marker information, defined as response to MAS relative to response without marker information, is given by:

$$E = \frac{R_{MAS}}{R_P} = \frac{r_{GEBV}}{h} = \sqrt{\frac{r_{MBV}^2}{h^2} + \frac{(1-r_{MBV}^2)^2}{1-r_{MBV}^2 h^2}}$$

An equivalent equation can be derived using the alternate index I':

$$E = \frac{R_{MAS}}{R_P} = \frac{r_{GEBV'}}{h} = \frac{1}{h}\sqrt{r_{MBV}^2 + h_{P'}^2(1-r_{MBV}^2)}$$

Efficiency of MAS vs phenotypic selection

$h^2=0.05$
$h^2=0.10$
$h^2=0.25$
$h^2=0.50$
$h^2=0.75$
$h^2=1.00$

Efficiency

Fraction of variance associated with MBV r_{MBV}^2
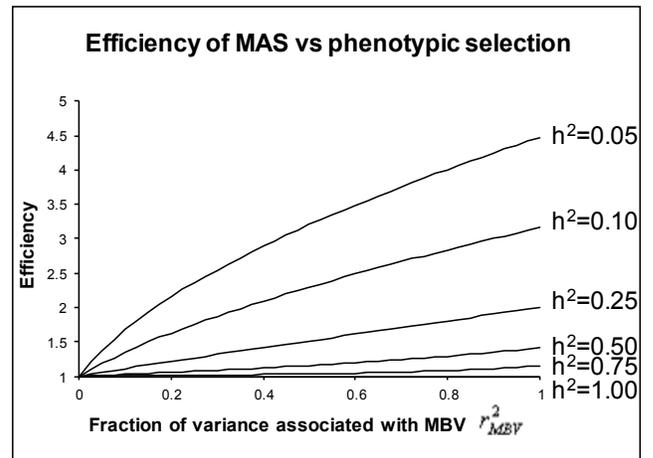
Modelling MBV as a Correlated Trait

After Dekkers (2007, JABG 124);
Schrooten et al. (2005, JABG 123)

When based on multiple regions/markers, marker-based EBV behave as a Mendelian inherited polygenic trait with heritability =1:

$$MBV = \hat{Q} = \sum_j (\hat{g}_j^{pat} + \hat{g}_j^{mat})$$

$\hat{g}_j^{pat}$ and $\hat{g}_j^{mat}$ = BLUP estimate of the effects of the paternal and maternal marker allele or haplotype for interval j.

$$\hat{Q}_{progeny} = \tfrac{1}{2}\hat{Q}_{sire} + \tfrac{1}{2}\hat{Q}_{dam} + (\sum_j \hat{g}_{ij}^{pat} - \tfrac{1}{2}\hat{Q}_{sire}) + (\sum_j \hat{g}_{ij}^{mat} - \tfrac{1}{2}\hat{Q}_{dam})$$

➔ marker-based EBV represent estimates that can be viewed and modeled as a genetic trait that can be observed on individuals without error ➔ heritability =1.

This allows standard selection index software to be used to model genomic selection (e.g. SelAction), including modelling the Bulmer effect and methods for prediction of inbreeding.
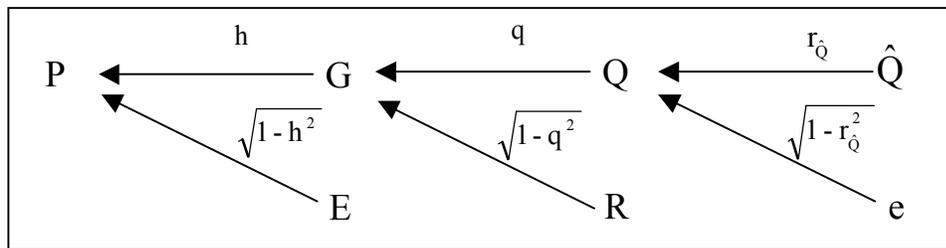
Marker-based EBV, $\hat{Q}$, are estimates of genetic effects Q.

$$Q = \hat{Q} + e \qquad e = \text{prediction error}$$

$r_{\hat{Q}}$ = accuracy of $\hat{Q}$ as a predictor Q, = correlation between Q and $\hat{Q}$.

Then, the correlation of $\hat{Q}$ with G, i.e. the accuracy of the MBV is equal to $r_{MBV} = q\,r_{\hat{Q}}$

= accuracy of the marker-based EBV as a predictor of the total genetic value G

Define marker-based EBV, $\hat{Q}$, as a correlated trait with $h^2 = 1$, along with the regular trait. Correlations required for selection index calculations (e.g., based on path diagram):

Genetic correlation between traits: $\qquad r_{G\hat{Q}} = q r_{\hat{Q}} = r_{MBV}$

Phenotypic correlation between traits: $\qquad r_{P\hat{Q}} = hq r_{\hat{Q}} = h\, r_{MBV}$

Phenotypic and genetic st.dev of MBV $\qquad = r_{MBV}\,\sigma_G$

Use of these parameters results in variances and covariances that are identical to the elements in matrix **P** and vector **G** of the Lande & Thompson (1990) index.

E.g.: $\quad Cov(\hat{Q}_i, P_j) = Cov(\hat{Q}_i, G_j) = a_{ij}Cov(\hat{Q}, G) = a_{ij}r_{G\hat{Q}}\sigma_G r_{MBV}\sigma_G = a_{ij} r_{MBV}^2 \sigma_G^2$

Extension to multiple traits

$\rho_{G_{12}}$, $\rho_{R_{12}}$, and $\rho_{Q_{12}}$ = genetic correlations traits 1 and 2 for the genetic components G, R, Q.

$$r_{G_i\hat{Q}_j} = q_i r_{\hat{Q}_i} r_{\hat{Q}_i\hat{Q}_j} + q_i\sqrt{1-r_{\hat{Q}_i}^2} \qquad r_{\hat{Q}_je_i} = q_i r_{\hat{Q}_i} r_{\hat{Q}_i} r_{\hat{Q}_j}\rho_{Q_{12}} + q_i\sqrt{1-r_{\hat{Q}_i}^2}\, r_{\hat{Q}_j}\sqrt{1-r_{\hat{Q}_i}^2}\rho_{Q_{12}} = q_i r_{\hat{Q}_j}\rho_{Q_{12}}$$

$$r_{P_i\hat{Q}_j} = h_i r_{G_i\hat{Q}_j} = h_i q_i r_{\hat{Q}_j}\rho_{Q_{12}}$$

Path diagram:

- $P_1 \xleftarrow{h_1} G_1 \xleftarrow{q_1} Q_1 \xleftarrow{r_{\hat{Q}_1}} \hat{Q}_1$, with $\sqrt{1-h_1^2}$ to E_1, $\sqrt{1-r_{\hat{Q}_1}^2}$ to e_1
- $G_1 \to E_1 \to R_1 \to e_1$
- $P_2 \xleftarrow{h_2} G_2 \xleftarrow{q_2} Q_2 \xleftarrow{r_{\hat{Q}_2}} \hat{Q}_2$, with $\sqrt{1-h_2^2}$, $\sqrt{1-q_2^2}$, $\sqrt{1-r_{\hat{Q}_2}^2}$ to E_2, R_2, e_2

Legend:

① $r_{E_1E_2} = \rho_{E_{12}}$

② $r_{R_1R_2} = \rho_{R_{12}}$

③ $r_{e_1e_2} = \sqrt{1-r_{\hat{Q}_1}^2}\sqrt{1-r_{\hat{Q}_2}^2}\rho_{Q_{12}}$

④ $r_{\hat{Q}_1e_2} = r_{\hat{Q}_1}\sqrt{1-r_{\hat{Q}_2}^2}\rho_{Q_{12}}$

⑤ $r_{\hat{Q}_2e_1} = r_{\hat{Q}_2}\sqrt{1-r_{\hat{Q}_1}^2}\rho_{Q_{12}}$
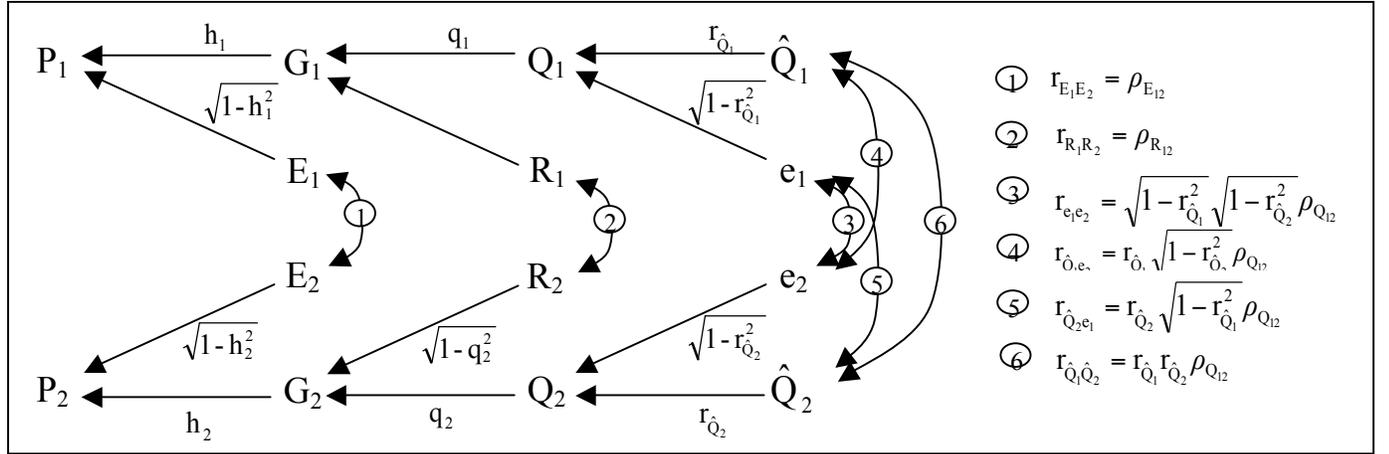
⑥ $r_{\hat{Q}_1\hat{Q}_2} = r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{Q_{12}}$

With random allocation of markers (GS) $E(q_1^2)=E(q_2^2)=q^2$ and $E(\rho_{R_{12}})=E(\rho_{Q_{12}})=\rho_{G_{12}}$

$$\rightarrow\ r_{\hat{Q}_1\hat{Q}_2} = \frac{Cov(\hat{Q}_1,\hat{Q}_2)}{\sqrt{Var(\hat{Q}_1)Var(\hat{Q}_2)}} = \frac{r_{\hat{Q}_1}^2 r_{\hat{Q}_2}^2\rho_{Q_{12}}}{r_{\hat{Q}_1}r_{\hat{Q}_2}} = r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{Q_{12}} = r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{G_{12}}$$

Table 1. Genetic parameters[1] for 4 traits considered for derivation of selection criteria: phenotype for trait (P_1) and trait 2 (P_2), and MBV for trait 1 ($\hat{Q}_1$) and trait 2 ($\hat{Q}_2$).

	P_1	P_2	$\hat{Q}_1$	$\hat{Q}_2$
P_1	h_1^2	$\rho_{P_{12}}$	$h_1q_1r_{\hat{Q}_1} =^2 h_1 r_{MBV_1}$	$h_1q_1r_{\hat{Q}_2}\rho_{Q_{12}} = h_1 r_{MBV_2}\rho_{G_{12}}$
P_2	$\rho_{G_{12}}$	h_2^2	$h_2q_2r_{\hat{Q}_1}\rho_{Q_{12}} = h_2 r_{MBV_1}\rho_{G_{12}}$	$h_2q_2r_{\hat{Q}_2} = h_2 r_{MBV2}$
$\hat{Q}_1$	$q_1r_{\hat{Q}_1} = r_{MBV_1}$	$q_2r_{\hat{Q}_1}\rho_{Q_{12}} = r_{MBV_1}\rho_{G_{12}}$	1	$r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{Q_{12}} = r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{G_{12}}$
$\hat{Q}_2$	$q_1r_{\hat{Q}_2}\rho_{Q_{12}} = r_{MBV_2}\rho_{G_{12}}$	$q_2r_{\hat{Q}_2} = r_{MBV_2}$	$r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{Q_{12}} = r_{\hat{Q}_1}r_{\hat{Q}_2}\rho_{G_{12}}$	1

[1] h_i^2 = heritability of phenotype for trait i

q_i^2 = proportion of genetic variance associated with markers for trait i

$r_{\hat{Q}_i}$ = accuracy of $\hat{Q}_i$ as a predictor of marker-associated genetic effects, Q_i.

r_{MBV_i} = accuracy of $\hat{Q}_i$ as a predictor of the total genetic value, G_i

$\rho_{G_{12}}$ = genetic correlation between traits 1 and 2

ρ_{P12} = phenotypic correlation between traits 1 and 2

$\rho_{Q_{12}}$ = correlation between Q_1 and Q_2

$\rho_{R_{12}}$ = correlation between residual genetic effects for traits 1 (R_1) and 2 (R_2)

[2] Results after the equality signs assume $q_1=q_2$ and $\rho_{G_{12}}=\rho_{Q_{12}}=\rho_{R_{12}}$, and use $q_i r_{\hat{Q}_i} = r_{MBV_i}$

$$\rho_{G_{12}} = q_1q_2\rho_{Q_{12}} + \sqrt{1-q_1^2}\sqrt{1-q_2^2}\rho_{R_{12}} \qquad \rho_{P_{12}} = h_1h_2\left(q_1q_2\rho_{Q_{12}} + \sqrt{1-q_1^2}\sqrt{1-q_2^2}\rho_{R_{12}}\right)$$

Examples of Modelling MAS/GS using SelAction by Modelling MBV as Correlated Traits
(Dekkers, 2007, JABG 124)

Each generation, 20 males were selected. Each male was mated to three selected females, which each producing eight offspring (four male, four female). Heritability of the trait was 0.1 or 0.4 and selection was on BLUP EBV based on phenotypic and/or marker data.
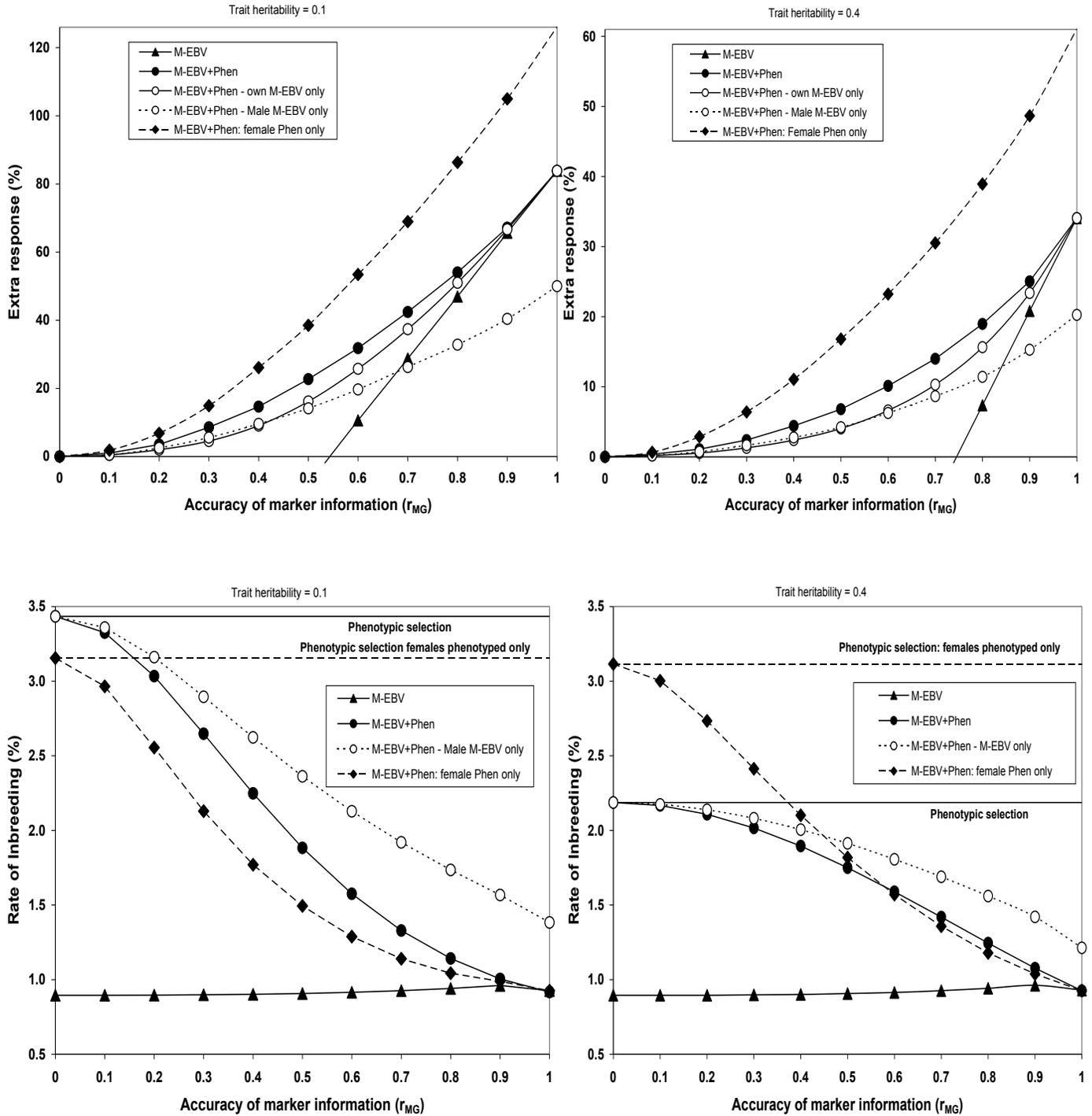
Table 2. Genetic parameters for selection on a breeding goal of two traits (P_1 and P_2) with and without marker information and resulting responses to selection in individual traits and the breeding goal (ΔH) and rates of inbreeding (ΔF). Marker-based EBV ($\hat{Q}_1$ and $\hat{Q}_2$) have accuracies of 0.8, based on markers explaining 62.4% of the genetic variance.

Correlations[1]	P_1	P_2	$\hat{Q}_1$	$\hat{Q}_2$		
P_1	--	-0.5	0.438	-0.131		
P_2	-0.3	--	-0.076	0.253		
$\hat{Q}_1$	0.8	-0.24	--	-0.243		
$\hat{Q}_2$	-0.24	0.8	-0.243	--		
Heritability	0.3	0.1	1	1		
Phenotypic SD	1	1	0.8	0.8		
Economic value	1	1	0	0		
Response to selection					ΔH	$\Delta F(\%)$
Phenotype only	0.408	0.041	0.394	0.052	0.448	2.36
Markers only	0.418	0.068	0.655	0.167	0.486	0.94
Combined	0.469	0.074	0.582	0.148	0.543	1.29

[1] Phenotypic correlations above the diagonal; genetic correlations below the diagonal

Impact of the Bulmer effect with Genomic Selection
(after Grevenhof, Bijma, van Arendonk GSE 2012, 44:26)

General equation for impact of selection on variable *w* on:

- covariance between variables *x* and *y*:
$$\sigma_{xy}^* = \sigma_{xy} - k\sigma_{xy}^* \frac{\sigma_{wx}\sigma_{wy}}{\sigma_w^2}$$

- variance of variable *w*:
$$\sigma_w^{*2} = (1-k)\sigma_w^2$$

- variance of correlated variable *x*:
$$\sigma_x^{*2} = (1-kr_{wx}^2)\sigma_x^2$$

Genomic Selection = select on *w* = MBV = trait with h² = 1 (equal male/female selection)

Variance MBV candidates generation t =
$$\sigma_{MBV,t}^2 = r_{MBV,t}^2 \sigma_{G,t}^2$$

Variance MBV selected parents gener t =
$$\sigma_{MBV,t}^{*2} = (1-k)\sigma_{MBV,t}^2$$

Variance MBV next generation t+1 =
$$\sigma_{MBV,t+1}^2 = .5(1-k)\sigma_{MBV,t}^2 + .5\sigma_{MBV,0}^2$$

Genetic variance selected parents gener t =
$$\sigma_{g,t}^{*2} = (1-r_{MBV,t}^2 k)\sigma_{g,t}^2$$

Genetic variance next generation t+1 =
$$\sigma_{g,t+1}^2 = 0.5(1-r_{MBV,t}^2 k)\sigma_{g,t}^2 + .5\sigma_{g,0}^2$$

At equilibrium/Limit:
$$\sigma_{MBV,L}^2 = \sigma_{MBV,0}^2/(1+k) \qquad \text{from } \sigma_{MBV,t+1}^2 = \sigma_{MBV,t}^2$$

Response to GS = $i\sigma_{MBV}$:
$$R_{(L)}/R_{(0)} = \sigma_{MBV,L}/\sigma_{MBV,0} = \frac{1}{\sqrt{1+k}}$$

Note: same as BLUP selection

Genetic variance
$$\sigma_{g,L}^2 = \sigma_{MBV,L}^2 + \text{PEV}$$
$$= [r_{MBV,0}^2 \sigma_{MBV,0}^2/(1+k)] + (1-r_{MBV,0}^2)\sigma_{g,0}^2$$
$$= \sigma_{MBV,0}^2(1-k\,r_{MBV,0}^2)/(1+k)$$

$\text{PEV}_L = \text{PEV}_0$
$$(1-r_{MBV,L}^2)\sigma_{g,L}^2 = (1-r_{MBV,0}^2)\sigma_{g,0}^2$$

Accuracy-squared
$$r_{MBV,L}^2 = 1 - [(1-r_{MBV,0}^2)\sigma_{g,0}^2/\sigma_{g,L}^2]$$

Table 1 Comparison of the Bulmer-effect for mass selection and genomic selection

Selection methoda	h^2	$r_{g\hat{g}_0}$	Equilibrium genetic variance	Equilibrium accuracy	Δ% b
Mass selection	0.25	0.5	0.21	0.47	−14%
Genomic selection	0.25	0.5	0.22	0.39	−27%
Mass selection	0.10	0.32	0.093	0.306	−7%
Mass selection	0.50	0.71	0.367	0.651	−21%
Mass selection	-	any value	-	-	−27%

[a]Comparison of the Bulmer-effect for mass selection and genomic selection with different heritabilities (h^2) and accuracies of EBV ($r_{g\hat{g}_0}$); phenotypic variance equals 1; selected proportion equals 5%; [b]Δ% is the relative difference between the initial response and the Bulmer-equilibrium response.

Note:

GS → greater Bulmer reduction in response than mass selection
       GS targets a proportion of the genetic variation with full accuracy
       Mass selection targets the full genetic variation with limited accuracy

GS → Bulmer reduction in response unaffected by accuracy
       - same as BLUP selection → no need to account for Bulmer effect
       when comparing GS to BLUP selection with equal intensity

Mixture distribution approach to predicting response to marker-assisted selection (after Dekkers and van Arendonk, 1998, Theor. Appl. Genet.)

Selection index approach to accuracy and response prediction assumes multivariate normality. Selection on major gene requires use of mixture distributions.

Select on: $\text{GEBV'} = I_i^{'} = \hat{Q}_i + \hat{R}_i$

$\hat{Q}_i$ takes on discrete values, depending on genotype at the major gene. Thus, the distribution of GEBV is a mixture of normal distributions.

$I_i^{'} \sim \text{Normal}(\hat{Q}_i, \text{Var}(\hat{R}_i))$

➔ Truncation selection across 3 distributions for biallelic major gene.

$\qquad I_i^{'} = b_Q^{'} \hat{Q}_i + \hat{R}_i$

$\hat{Q}_i$ = EBV for QTL with two alleles

➔ 4 genotypes (when distinguishing parental origin)
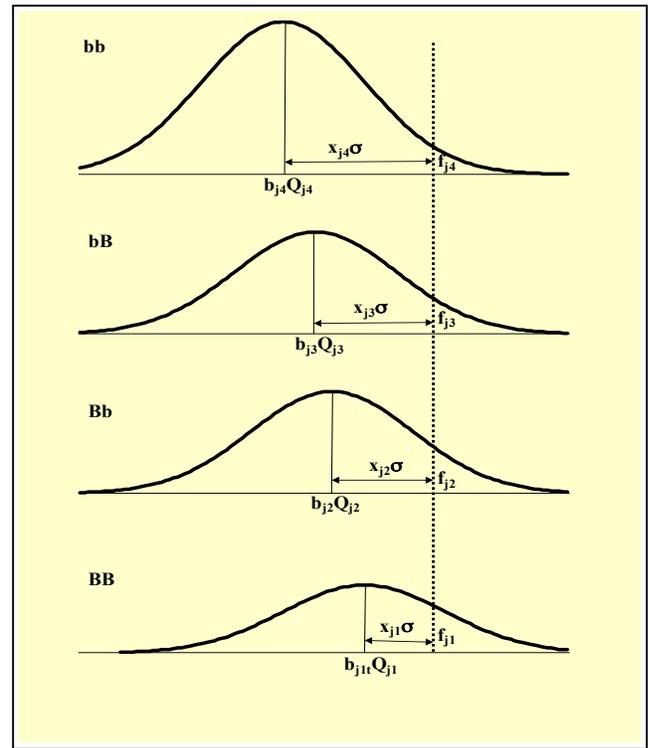
Summary of notation used for selection on a QTL with two alleles (B and b) in generation

Genotype	No	Genotype frequency[1]	Mean polygenic breeding value[2]	Mean genetic value	Mean BV, deviated from genotype Bb[3]	Prop. selected in sex j	Index wts	B gamete production, fraction	Selection differential[4]
BB	1	$p_s p_d$	$\bar{u}_1 = A_{sB} + A_{dB}$	$a + \bar{u}_1$	$\alpha + \bar{u}_1 - \bar{u}_2$	f_{j1}	b_{j1}	1	$i_{j1}\sigma_j$
Bb	2	$p_s(1-p_d)$	$\bar{u}_2 = A_{sB} + A_{db}$	$d + \bar{u}_2$	0	f_{j2}	0	½	$i_{j2}\sigma_j$
bB	3	$(1-p_s)p_d$	$\bar{u}_3 = A_{sb} + A_{dB}$	$d + \bar{u}_3$	$\bar{u}_3 - \bar{u}_2$	f_{j3}	b_{j3}	½	$i_{j3}\sigma_j$
bb	4	$(1-p_s)(1-p_d)$	$\bar{u}_4 = A_{sb} + A_{db}$	$-a + \bar{u}_4$	$-\alpha + \bar{u}_3 - \bar{u}_2$	f_{j4}	b_{j4}	0	$i_{j4}\sigma_j$

p_s and p_d = frequencies of allele B among selected sires and dams that are used to produce the next generation.

$\bar{u}_m$ = mean polygenic breeding value of individuals of genotype m in current generation,

A_{jB} and A_{jb} = mean polygenic values of gametes from sex j that carry allele B or b and were used to produce the current generation.

$\alpha = a + (1 - p_s - p_d)d$ = standard QTL allele substitution effect in current generation

σ_j is the SD of estimates of polygenic breeding values for sex j;

i_{ji} = selection intensity for individuals of sex j and genotype i.

Mean total genotypic value of the population in current generation =
$$\bar{g} = (p_s + p_d - 1)a + (p_s + p_d - 2p_s p_d)d + p_s A_{sB} + (1-p_s)A_{sb} + p_d A_{dB} + (1-p_d)A_{db}$$

Selection Model with QTL Information

Optimal truncation point across 4 distributions determined by multrunc

→ frequency in gametes of selected parents: $p_j^* = [p_s p_d f_{j1} + \frac{1}{2} p_s (1 - p_d) f_{j2} + \frac{1}{2} (1 - p_s) p_d f_{j3}]/F_j$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ F_j is the total proportion selected for sex j.

Expected mean polygenic breeding values of B and b gametes that form the next generation by sex of parent:

$A_{jB}^* = \frac{1}{2}[f_{j1} p_s p_d (\bar{u}_1 + i_{j1} \sigma_j) + \frac{1}{2} f_{j2} p_s (1 - p_d)(\bar{u}_2 + i_{j2} \sigma_j) + \frac{1}{2} f_{j3} (1 - p_s) p_d (\bar{u}_3 + i_{j3} \sigma_j)]/F_j p_j^*$

$A_{jb}^* = \frac{1}{2}[\frac{1}{2} f_{j2} p_s (1 - p_d)(\bar{u}_2 + i_{j2} \sigma_j) + \frac{1}{2} f_{j3} (1 - p_s) p_d (\bar{u}_3 + i_{j3} \sigma_j) + f_{j4} (1 - p_s)(1 - p_d)(\bar{u}_4 + i_{j4} \sigma_j)]/F_j (1 - p_j^*)$

Enables modeling changes in frequency and polygenes for any index of QTL and polygenic EBV

$$I_i' = b_Q' \hat{Q}_i + \hat{R}_i$$

- ***Standard index MAS*** : $\quad\quad\quad\quad\quad\quad\quad\quad$ $b_Q' = 1$
- ***Optimal index MAS*** (see later): $\quad\quad\quad$ b_Q' optimized
- ***Regular selection***: $\quad\quad\quad\quad\quad\quad\quad\quad$ $b_Q' = r_j^2$