

Accounting for preselection in data

Thorvaldur Árnason
IHBC, AUI

Background

- Pre-selection or culling on an unobservable criterion before testing is a common problem in horse breeding
- Culling on criterium correlated with tested traits causes bias in estimation of genetic parameters and genetic evaluations
(Robertson,1966,Meyer & Thompson,1984)

Background cont.

- Survival = Racing-status = **Test-status** = 0/1 binary trait
- **If** record available on continuous trait **then** test-status=1 **else** test-status=0
- Requires pedigree information on culled animals
- Estimated heritability of test-status moderate to high in many horse populations
- EBVs for test-status often important

Background cont.

- Árnason(1999) showed by simulations that bivariate genetic evaluations involving racing-status and a continuous trait reduced selection bias, increased accuracy, and increased genetic response when **true** genetic parameters were available
- The procedure was validated by method **R** in Swedish standardbred trotters (SST)
- Racing-status has been included in MT-AM-BLUP for genetic evaluations of SST since 1995

Background cont.

- Genetic and environmental correlations between test-status and the continuous performance traits are not readily estimable
- Genetic covariance estimable if environmental covariance is constrained to predefined value
- Expected environmental covariance ?

Objective

- To study the consequences of assuming **zero environmental covariances** between a binary test-status and a continuous trait on:
 - A) the estimates of genetic parameters in **culled data**, when the true covariances deviates from zero, using REML and Gibbs sampler methods
 - B) genetic progress from selection based on BLUP and Gibbs sampling estimates of EBVs

Simulation procedure

- True genetic parameters: $h^2_1 = h^2_2 = 0.4$;
 $r_A = 0.5$; $r_E = 0.0$; 0.5 ; -0.5
- $BV = \mathbf{a} = 0.5\mathbf{a}_s + 0.5\mathbf{a}_d + \mathbf{m}$ (Mendelian sampling term)
- $\mathbf{m} = \sqrt{0.5c} \mathbf{C}_G \mathbf{z}$; $c = (1 - 0.5(F_s + F_D))$; $\mathbf{V}_G = \mathbf{C}_G \mathbf{C}'_G$;
 \mathbf{z} = bivariate random sampled vector $(\mathbf{0}, \mathbf{I})$
- $\mathbf{P} = \mathbf{X}\mathbf{b} + \mathbf{a} + \mathbf{C}_e \mathbf{z}$ (underlying phenotypes)
- Fixed effects = generation effects
- Culling frequency 0.5 and 0.8
- 10 - 100 replicates

Simulated data for estimation of genetic parameters

- Base population: random 25 males 500 females = 525 animals
- 3 generations random mating 500 females and 25 males per generation = 1575 animals. No phenotypes
- 5 generations: Selection on phenotypic records (mass selection across generations): 500 females and 25 males selected per generation producing 1500 offspring per generation = 7500 records
- Pedigree list=9600 animals

Simulated data for genetic selection

- Base population: random 25 males 500 females = 525 animals
- 3 generations random mating 500 females and 25 males per generation = 1575 animals. No phenotypes
- 5 generations: Selection on EBVs (selection across generations): 500 females and 25 males selected per generation producing 1500 offspring per generation = 7500 records
- Pedigree list=9600 animals

Software and methods for estimation of genetic parameters

- DMU (Jensen & Madsen, 2008)
- Animal Model Average Information Restricted Maximum Likelihood (AM-AI-REML)
- Multiple Trait Animal Model Linear Threshold Markov Chain Monte Carlo Gibbs Sampler (MT-AM-LT-MCMC-GS)

Results



Estimated variance components for the continuous trait ($\sigma^2_A = 0.4$; $\sigma^2_E = 0.6$; $r_A = 0.5$). *ST-AM-AI-REML* analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	\hat{h}^2 (s.e)
50% culling			
0.0	0.383 (0.016)	0.588 (0.014)	0.394 (0.015)
0.5	0.338 (0.012)	0.509 (0.010)	0.399 (0.013)
-0.5	0.498 (0.020)	0.520 (0.016)	0.489 (0.018)
80% culling			
0.0	0.403 (0.014)	0.565 (0.011)	0.416 (0.012)
0.5	0.334 (0.022)	0.469 (0.015)	0.414 (0.036)
-0.5	0.534 (0.029)	0.485 (0.018)	0.522 (0.021)

Estimated variance components for the continuous trait ($\sigma^2_A = 0.4$; $\sigma^2_E = 0.6$; $r_A = 0.5$). **ST-AM-AI-REML** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	\hat{h}^2 (s.e)
50% culling			
0.0	0.383 (0.016)	0.588 (0.014)	0.394 (0.015)
0.5	0.338 (0.012)	0.509 (0.010)	0.399 (0.013)
-0.5	0.498 (0.020)	0.520 (0.016)	0.489 (0.018)
80% culling			
0.0	0.403 (0.014)	0.565 (0.011)	0.416 (0.012)
0.5	0.334 (0.022)	0.469 (0.015)	0.414 (0.036)
-0.5	0.534 (0.029)	0.485 (0.018)	0.522 (0.021)

Estimated variance components for the Gaussian trait ($\sigma^2_A = 0.4$; $\sigma^2_E = 0.6$; $r_A = 0.5$). **MT-AM-LT-GS** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	\hat{h}^2 (s.e)	\hat{r}_A (s.e)
50% culling				
0.0	0.397 (0.016)	0.606 (0.013)	0.395 (0.014)	0.478 (0.021)
0.5	0.349 (0.012)	0.519 (0.009)	0.402 (0.013)	0.278 (0.040)
-0.5	0.534 (0.017)	0.540 (0.014)	0.497 (0.018)	0.621 (0.018)
80% culling				
0.0	0.402 (0.019)	0.598 (0.010)	0.401 (0.014)	0.466 (0.048)
0.5	0.325 (0.019)	0.491 (0.013)	0.396 (0.020)	0.100 (0.041)
-0.5	0.610 (0.035)	0.505 (0.019)	0.544 (0.022)	0.715 (0.027)

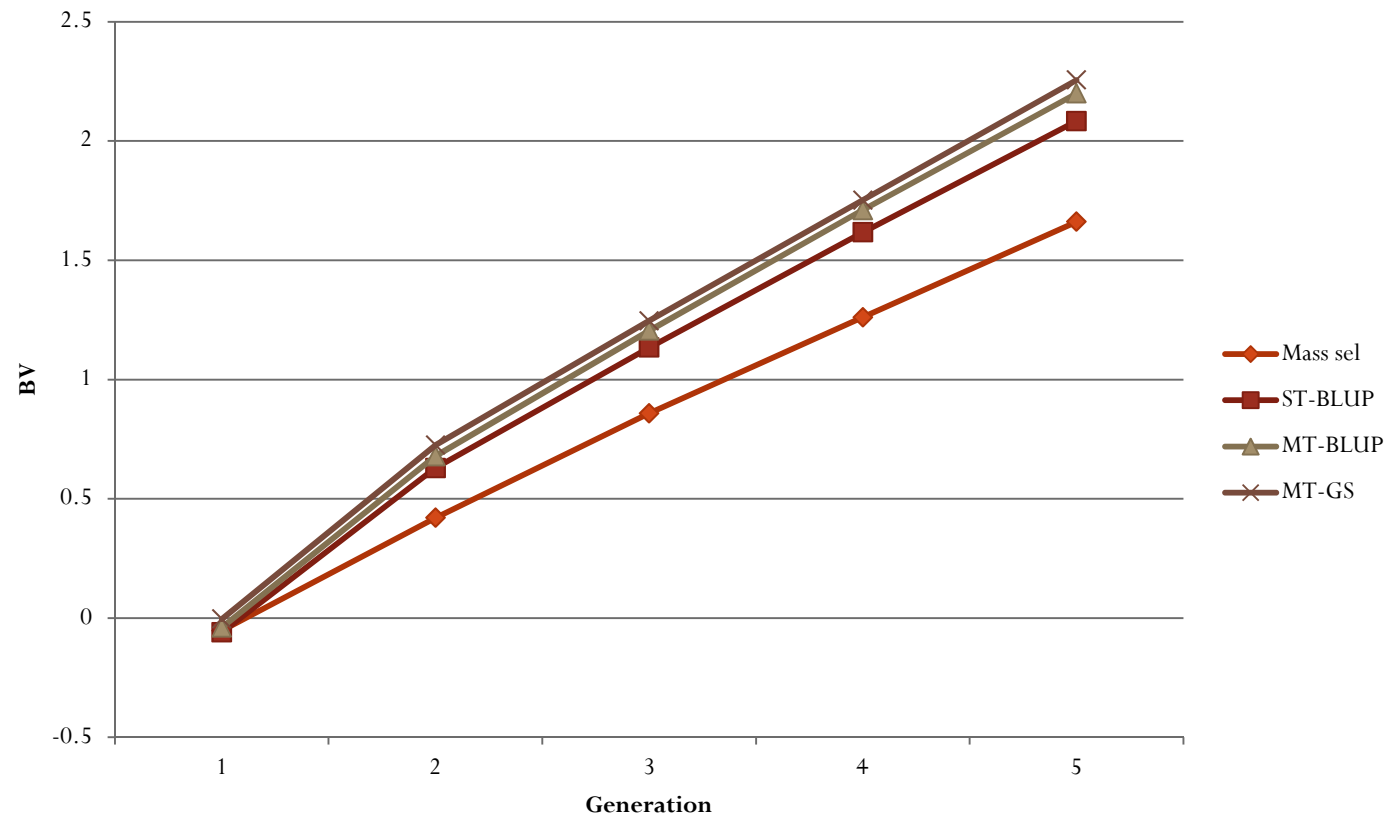
Estimated variance components for the Gaussian trait ($\sigma^2_A = 0.4$; $\sigma^2_E = 0.6$; $r_A = 0.5$). **MT-AM-LT-GS** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	\hat{h}^2 (s.e)	\hat{r}_A (s.e)
50% culling				
0.0	0.397 (0.016)	0.606 (0.013)	0.395 (0.014)	0.478 (0.021)
0.5	0.349 (0.012)	0.519 (0.009)	0.402 (0.013)	0.278 (0.040)
-0.5	0.534 (0.017)	0.540 (0.014)	0.497 (0.018)	0.621 (0.018)
80% culling				
0.0	0.402 (0.019)	0.598 (0.010)	0.401 (0.014)	0.466 (0.048)
0.5	0.325 (0.019)	0.491 (0.013)	0.396 (0.020)	0.100 (0.041)
-0.5	0.610 (0.035)	0.505 (0.019)	0.544 (0.022)	0.715 (0.027)

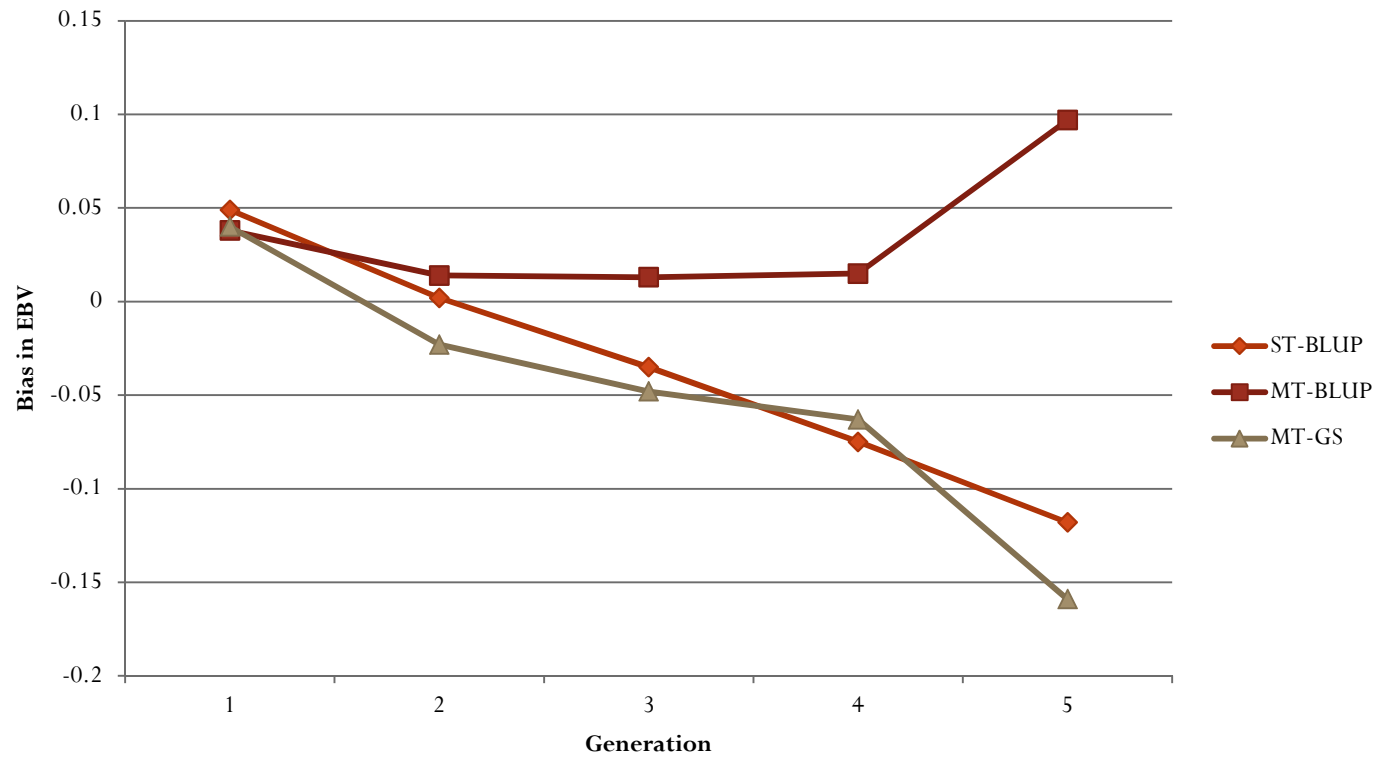
Genetic progress (ΔG) from selection on three different selection criteria

r_E	Culling rate	Mass selection	STL-BLUP	MTLL-BLUP
	0%	0.463	0.568	
0.0	50%	0.430	0.524	0.560
	80%	0.341	0.489	0.505
0.5	50%	0.429	0.530	0.551
	80%	0.341	0.466	0.507
-0.5	50%	0.459	0.557	0.566
	80%	0.366	0.491	0.509

Real genetic trend from four selection criteria for $r_E = 0.0$ and culling rate of 50%



Bias in EBVs from four selection criteria for $r_E = 0.0$ and culling rate of 50%



Genetic analysis of racing status and racing performance traits in Swedish standardbred trotters

Trait	h^2	r_A
Racing status	0.31	
		0.59
Racing performance (av)	0.31	

Results from Árnason, 1999; 40% culling

Genetic analysis of test status and riding-ability traits in Icelandic horses

Trait	h^2	r_A
Test status	0.63	
		0.68
Riding ability (av)	0.52	

Results from Albertsdóttir et al., 2011; 80% culling

What are the main reasons for preselection of data for genetic evaluation in horse breeding ?

- Preselection based on criterion strongly correlated with pedigree?
- Preselection based on criterion strongly correlated with within family deviation?
- $BV = a = (0.5a_s + 0.5a_d)$ (Pedigree) + m (Mendelian sampling term)

Results from simulation of 50%
 culling on parent's BV average.
 Mass-selection, 5 generations,
 $h^2=0.4$

Analysis/ data type	$h^2_{\text{Test status}}$	$h^2_{\text{Performance}}$	r_A	r_E
ST-underlying P		0.42		
ST- observed P		0.17		
MT- underlying P	0.68	0.17	0.99	-0.20
MT- observed P	0.69	0.34	0.97	-

Results from simulation of 50%
culling on Mendelian deviation.
Mass-selection, 5 generations,
 $h^2=0.4$

Analysis/ data type	$h^2_{\text{Test status}}$	$h^2_{\text{Performance}}$	r_A	r_E
ST-underlying P		0.43		
ST- observed P		0.49		
MT- underlying P	0.00	0.43	- (?)	0.45
MT- observed P	0.00	0.49	- (?)	-

Jump to the Conclusions



Conclusions

- For the parameter combination tested, heritability estimates for the continuous trait were significantly biased (upwards) when r_E was negative
- Genetic correlations between test-status and the continuous trait were systematically underestimated when r_E was positive and overestimated when r_E was negative

Conclusions

- Genetic progress from selection is always reduced when the data are subject to culling
- Genetic progress from ST-AM-BLUP selection is much less affected by culling than it is for mass selection
- The inclusion of test-status in a multivariate framework together with the continuous traits reduces the loss in genetic progress further still!
- The MT genetic evaluations are robust to moderate errors in estimated genetic parameters

Conclusions

- High heritability estimates of test(racing) status in horses and high estimated genetic correlations between test status and the continuous performance traits point to high impact of preselection based on pedigree, possibly in combination with some within family selection which varies between sires.
- True underlying environmental correlation between test status and performance traits in horses is probably close to zero or low negative

Dank u - Thank you

