

1 **Testing different genomic selection scenarios in a small cattle population by**  
2 **simulation**

3  
4 *J. Obšteter\**, *J. Jenko†*, *J. M. Hickey†* & *G. Gorjanc†*

5  
6 *\*Department of Animal Science, Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000*  
7 *Ljubljana, Slovenia*

8 *† The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh,*  
9 *Easter Bush, Midlothian, EH259RG, United Kingdom*

10 *cd*

11 *Jana Obšteter, Agricultural Institute of Slovenia, Department of Animal Science, Hacquetova ulica*  
12 *17, 1000 Ljubljana, Slovenia*

13 *+386 1 280 51 34*

14 [jana.obsteter@kis.si](mailto:jana.obsteter@kis.si) (Corresponding Author)

15 **Introduction (4000 characters words and spaces)**

16 Genetic gain depends on four factors, namely accuracy, intensity, genetic variance for the trait and  
17 generation interval. With the exception of genetic variance, genomic selection (GS) can influence  
18 remaining three parameters. Choosing the parents (animals) based on their genomic breeding values  
19 (gEBVs) allows for an earlier selection compared to the conventional selection, where the parents  
20 (animals) are chosen based on conventional breeding values (EBVs) obtained after progeny testing  
21 (PT). GS hence reduced the generation interval that along with a higher accuracy of early selection  
22 decisions increases the genetic gain (Schaeffer, 2006). However, the genomic prediction requires a  
23 large reference population. Consequently small cattle populations encounter difficulties with  
24 implementation of GS due to lacking financial resources (support?) and a small number of PT  
25 reference bulls. The latter diminishes the gain in accuracy of genomic prediction compared to  
26 parent average (PA) estimates what in turn decreases the advantage of using genomic information  
27 (the advantage of genomic over conventional selection) (Thomasen et al., 2014). Alternatively  
28 small cattle populations can participate in international associations for a joint genomic prediction  
29 with international reference population or they can include cows in the reference population.  
30 However, an information of a cow in the reference is not equal to the information of a PT bull,  
31 therefore a larger number of cows has to be included to achieve desired accuracies (de Roos, 2011).

32  
33 As mentioned, GS can also reduce the generation interval. The degree of reduction depends on the  
34 strategy of using the genomic information in a breeding program: in which selection paths to use the  
35 genomic information and to what extent. Previous studies mainly tested the scenarios in which the  
36 genomic information is used either for a pre-selection of young bulls for progeny testing (so called  
37 GS-PS) or for the selection of genomically tested young bulls directly (so called turbo scheme)  
38 against the conventional scenario with progeny tested sires without the use of genomic information  
39 (Pryce et al., 2010; Lillehammer et al., 2011; de Roos et al., 2011). The studies observed that we can

40 increase the genetic gain up to 30% by using genomic information for a pre-selection step and up to  
41 195% by using genomic information directly for selection of young bulls and dams as parents  
42 (M=100%!!!, realistic – 40% → up to 108%).

43

44 Further on, Thomasen et al., 2014, deterministically explored hybrid schemes that simultaneously  
45 use both progeny tested and young genomically tested bulls which changes the generation interval  
46 to an intermediate degree. The study varied the proportion of bull dams and cows mated with young  
47 / progeny tested sires and inspected the resulting annual monetary genetic gain (AMGG). They  
48 concluded that even in small populations, the genomic scenario was economically and genetically  
49 superior to the conventional scenario that uses progeny testes sires. However, with low reliabilities  
50 the hybrid scheme brought the higher AMGG whereas with high reliabilities the turbo scheme was  
51 the most superior (Thomasen et al., 2014).

52

53 In addition to maximising the genetic gain, breeding schemes have to manage genetic variation to  
54 assure sustainability of selection. The aim is therefore to balance short- and long-term success of  
55 selection. While short-term success depends only on the genetic gain in the next (few) generation,  
56 long-term success enables increase of genetic gain after many generations of selection.

57

58 Some studies also focused onto how the use of genomic information affects inbreeding and obtained  
59 contradictory results. While some studies observed a decrease in inbreeding rate per year with the  
60 use of genomic information (Pryce et al., 2010; Lillehammer et al., 2011), the others observed and  
61 increase (de Roos et al., 2011). Although reducing the generation interval may increase annual  
62 inbreeding rate, managing inbreeding rate per generation is more important for avoiding inbreeding  
63 depression (Daetwyler et al., 2007). The reason for a lower rate of inbreeding with GS lies in a

64 more accurate estimation of Mendelian sampling term – this results in lower co selection of siblings  
65 that reduces inbreeding rate per generation.

66  
67 However, a combine measure of genetic gain and genetic variability is needed to assure long-term  
68 competitiveness of selection. Efficiency of converting genetic variance into genetic gain measures  
69 how successful a specific scheme is in achieving both goals (Gorjanc et al., 2017).

70  
71 Although GS is a well established technology globally, small population still struggle with the  
72 adoption of the technology. Contrary to other parameters of GS, less work was done exploring how  
73 the use of genomic information in a breeding program affects genetic gain. Although the existed  
74 studies explored the genetic gain and / or inbreeding, questions remain about the optimal use of  
75 genomic information in small realistic cattle populations with overlapping generation that allow for  
76 a long-term success of selection. (LONG TERM, DETERMIISTIC; DISCRETE GENERATION,  
77 LARGE POPULATION). Also, studies mainly explored the use of genomic information for a  
78 pre-selection step or in a turbo scheme, but less work was done regarding the intermediate  
79 strategies. This study aimed to explore different strategies of the use of genomic information in  
80 small realistic cattle populations in order to determine the best one in terms of efficiency (the most  
81 efficient one).

82

### 83 **Material and Methods**

84 First we developed a simulator of a realistic cattle population under selection which included all  
85 selection steps and allowed for the user to define all the selection parameters. We compared five sire  
86 selection scenarios in which we varied the criterion for the selection of sires for insemination of bull  
87 dams and cows. Additionally we tested three sire use strategies with varying number of bulls chosen

88 each selection cycle and years kept in use. All tested scenarios were compared based on the  
89 achieved genetic gain, change (loss?) of genetic variance and efficiency of selection.

90

### 91 **Simulator of a cattle breeding program**

92 We developed a simulator of a realistic cattle populations under selection. The simulator was  
93 written as a Python wrapper around AlphaSim (Faux et al., 2016) and blupf90 (Miszta et al., 2002)  
94 software. The simulator enabled for the user to set all the selection parameters: the percentage of  
95 animals selected at each stage in each selection path, age at selection, selection criterion for each  
96 selection path, number of offspring per parent, years in use and number of selection cycles. The  
97 simulation was a continuous process resulting in overlapping generation. In each selection cycle  
98 three steps were performed: estimation of EBVs, selection of parents and mating. Breeding values  
99 were estimated as described below. Selection was performed based on single trait BVs.

100

### 101 **Simulated population**

102 Genome consisted of ten chromosomes that included 20,000 SNPs from which 10,000 SNPs were  
103 chosen at QTNs for a polygenic trait with heritability of 0.25. The effects of the QTNs for each  
104 chromosome were drawn from a normal distribution. The simulated population mirrored Slovenian  
105 Brown-Swiss population participating in milk recording consisting of ~30,000 active individuals,  
106 ~10,000 of which cows. In total 60 generations were simulated with 8640 animals born each  
107 selection cycle. First 20 generations represented the burn-in population in which random mating  
108 was implemented. This was followed by 20 generations of conventional selection using PT bulls to  
109 achieve a population structure resembling a cattle population under selection. The last 20  
110 generations implemented testing scenarios that differed in the use of genomic information in the  
111 male selection paths only.

112

113 **Selection of females**

114 Out of newborn females 1% was removed in their first year reflecting stillbirths and deaths  
115 occurring soon after birth. Out of these female calves 90% were inseminated in the second year and  
116 became cows in the third year. In each subsequent lactation we removed 20% of the cows and after  
117 the fourth lactation all cows were culled. This totaled to 10,653 active cows that were all screened  
118 for the trait and had their EBV estimated. In each generation 8550 cows with highest EBVs were  
119 chosen for mating. After the second completed lactation and estimation of EBVs, 43 cows with  
120 highest EBVs were chosen as bull dams. Bull dams were kept in use for three lactations,  
121 completing five lactations in total. From 129 bull dams 90 were chosen each year for contracted  
122 mated with sires of sires to produce new generation of selection candidates.

123

124 **Selection of sires**

125 Each year we obtained 45 male offspring from contracted matings. Sires were selected based on  
126 three overall strategies: a) progeny testing with pre-selection based on PAs (PT); b) progeny testing  
127 with pre-selection based on gEBVs (GS-PS); c) young genomically tested sires selected based on  
128 gEBVs (GS).

129

130 In the PT strategy, 27 out of 45 male candidates were randomly chosen for performance testing in  
131 their first year. In their second year 8/27 were chosen for progeny testing based on their PAs. In the  
132 sixth year 4/8 bulls were selected after obtaining progeny testing results. In the GS-PS strategy all  
133 45 male candidates were genomically tested in their first year and 8/45 were chosen for progeny  
134 testing in their second year based on their gEBVs. In the sixth year 4/8 bulls were selected after  
135 obtaining progeny testing results. In the GS strategy all 45 male candidates were genomically tested  
136 in their first year and in their second year 5/45 were chosen as young sires based on their gEBVs.

137

138 The chosen candidates entered either progeny testing or genomic testing. Young bulls for progeny  
139 testing were pre-selected either based on the performance test and EBVs (which equal PA since they  
140 have no information) (PT) or based on their gEBVs (GS-PS).

141 a) performance testing - progeny testing: 60% (n=27) of male calves from contracted mating were  
142 randomly chosen for performance testing. The random choice reflects the realistic situation where  
143 the availability of male calves depends also on unpredictable human factors (where not all the best  
144 male calves are available for performance testing). In the second year 8/27 bulls were chosen for  
145 progeny testing. The non-chosen bulls were used in natural service as cow sires. After five years 4/8  
146 progeny tested bulls were chosen as proven bulls.

147 b) genomic testing - progeny testing: all male calves from contracted mating were genomically  
148 tested in their first year. In the second year 8/45 bulls were chosen for progeny testing. The non-  
149 chosen bulls were used in natural service as cow sires. After five years 4/8 progeny tested bulls  
150 were chosen as proven bulls.

151 c) genomic testing: all male calves from contracted mating were genomically tested in their first  
152 year. In the second year 5/45 bulls were chosen as young sires. The non-chosen bulls were used in  
153 natural service as cow sires.

154

#### 155 **Tested scenarios**

156 We created five testing scenarios by using different categories of sires (PT, GS-PS, GS) for  
157 insemination of cows and bull dams. The tested scenario included one conventional scenario  
158 without the use of genomic information and four genomic scenarios. PT scenario used PT sires  
159 exclusively for insemination of all cows and bull dams. Similarly, GS-PS scenarios used GS-PS  
160 sires exclusively for insemination of all females. Next we created two hybrid scenarios that used  
161 GS-PS and GS sires simultaneously. GS-C used young genomically tested sires for insemination of  
162 cow population and GS-PS sires for insemination of bull dams. Contrary, in GS-BD we used GS

163 sires for the insemination of bull dams and GS-PS for insemination of cows. GS scenarios, also  
 164 referred to as “turbo”, used exclusively young genomically tested bulls for insemination of all  
 165 females.

166

167 Table 1: The five tested scenarios differing in sire category used for insemination of cows and bull  
 168 dams.

PT	All females are inseminated with progeny tested bulls (preselected for progeny testing based on their EBVs (PAs)).
GS-PS	All females are inseminated with progeny tested bulls (preselected for progeny testing based on their gEBVs).
GS-C	Cows are inseminated with young genomically tested bulls and bull dams with progeny tested sires pre-selected based on their gEBVs.
GS-BD	Bull dams are inseminated with young genomically tested bulls and cows with progeny tested sires pre-selected based on their gEBVs.
GS	All females are inseminated with young genomically tested sires.

169 EBV = estimated breeding value, gEBV = genomic breeding value, PA = parent average

170

171 **Sire use strategy**

172 All five scenarios were tested within three sire use strategies. The original strategy reflected the  
 173 situation in the simulated Slovenian Brown Swiss population where every year they select five bulls  
 174 and keep them in use for five years (SU 5/5). In the next strategy we increased the intensity of  
 175 selection for which we selected only one bull per year and kept him in use for five years (SU1/5).  
 176 We also created a strategy in which we reduced the generation interval even further by selecting  
 177 five bulls each year and replacing them all in the next year, i.e. kept them in use for only one year  
 178 (SU5/1).

179

180 **Estimation of breeding values**



Breeding values were estimated using pedigree or single-step BLUP depending on whether the genomic information was available. In genomic scenarios we assumed a initial reference population of ~11,000 cows and 100 progeny tested sires for the genomic prediction. The reference population was updated each selection cycle by removing the oldest generation of cows and adding an equivalently large sample of cows from the currently active cow population to the reference. Each year all the male offspring of contracted mating were genotyped and added to the reference population.

### Metrics of comparison

The scenarios were compared in terms of genetic gain per year (expressed with mean zero and in units of genetic standard deviation in the first generation of comparison) and efficiency averaged across 20 replicates. We defined the efficiency of selection as a regression of the achieved genetic gain on the amount of reduction of the genic standard deviation. We used the genic rather than the genetic standard deviation due to large fluctuations in the latter. When computing efficiency, both genetic gain and the genic standard deviation were standardised by the genic standard deviation in the first generation of comparison. Therefore, this efficiency metric indicates the potential genetic gain in units of genic standard deviation when all variation is converted into gain or lost due to drift. Results are presented as the mean of twenty replicates for each scenario on a per generation or cumulative basis.

## Results and Discussion

### Genetic Gain

The results for the genetic gain in the last generation of simulation / comparison (generation 60) of all scenarios within all three sire use strategies are shown in Table 2. The values for genetic gain are expressed relative to the genetic gain of PT scenario of the SU 5/5 strategy. First we inspected the

206 differences between the scenarios within each of the three strategies. As expected, the use of  
207 genomic information increases the genetic gain in all sire use strategies. This is due to higher  
208 accuracy for selection of young unphenotyped animals and reduction in generation interval. It is  
209 interesting that in this particular population we can substantially increase the genetic gain (up to  
210 44%) by using genomic information for pre-selection of male calves for progeny testing only  
211 (GS-PS) – although the generation interval is not reduced. The reason for this lies in higher  
212 accuracy of the pre-selection step, since gEBV (mean  $r = 0.76$ ) are more accurate than PA (mean  
213  $r = 0.27$ ). When we used GS sires for insemination of cow population (GS-C), we observed a  
214 reduction in generation interval in the sires of dams path up to 56% and between 53 and 65%  
215 increase in genetic gain. Similarly, when we used GS sires for insemination of bull dam population  
216 (GS-BD) – while cow population was inseminated with GS-PS bulls – we reduced generation  
217 interval up to 56% in the sires of sires selection path (56% in SU5/5, 46% in SU1/5 and 53% in  
218 SU5/1) and increased the genetic gain between 61 and 69%. This reveals that using young  
219 genomically tested bulls for insemination of bull dams and consequently as sires of sires results in a  
220 higher genetic gain than using them for insemination of cows (sires of dams). Lastly, the scenario  
221 that used GS sires exclusively resulted in the highest reduction in generation interval - up to 67%  
222 and 60% in the sire of sires and sires of dams selection path, respectively - and highest increase in  
223 genetic gain (between 88 and 120%). These results are in concordance with previous studies that  
224 observed a higher genetic gain with the use of genomic information for a pre-selection for PT or for  
225 selection of young sires (“turbo” scheme) (Pryce et al., 2010; Lillehammer et al., 2011; de Roos et  
226 al., 2011; Thomasen et al., 2014).

227

228 Changing the strategy of sire use affected genetic gain as well. Increasing the intensity of selection  
229 (SU 1/5) and reducing the generation interval even further (SU 5/1) both increased genetic gain.  
230 However, although we implemented a severely intensive selection by selecting one bull only in

231 SU 1/5 strategy (increased the intensity of selection by 57%) (intensity  $5 / 27 = 1.2664$ ;  $1/27 =$   
232  $1.9962$ ), we only observe up to 22% increase in genetic gain. In the SU 5/1 strategy, where the  
233 generation interval was reduced up to 41% compared to the corresponding scenarios in SU 5/5, the  
234 genetic gain was increased up to 32%.

235  
236 Table 2: Genetic gain of tested sire use strategies and sire selection scenarios averaged across 20  
237 replicates. The genetic gain is expressed relative to the genetic gain of the PT scenario in SU 5/5  
238 strategy.

		Sire use strategies		
		SU5/5	SU1/5	SU5/1
Sire selection scenario	PT	100%	122%	101%
	GS-PS	136%	144%	133%
	GS-C	153%	162%	165%
	GS-BD	160%	161%	169%
	GS	188%	208%	220%

239 PT = conventional scenario without use of genomic information; GS-PS = genomic information used for pre-selection  
240 of young bulls for progeny testing, GS-C / GS-BD / GS = genomically tested bulls used respectively for insemination of  
241 cows / bull dams / cows and bull dams. Sire use strategies: SU 5/5 = five bulls selected each year and used for five  
242 years, SU 1/5 = one bull selected each year and used for five years, SU 5/1 = five bulls selected each year and used for  
243 one year.

244  
245 **Genic variance**

246 Here we report the results for genic variance due to large fluctuations of the genetic variation over  
247 the generations. The evolution of the genic variance for the tested scenarios is shown in Figure X.  
248 We notice that the strategy SU1/5 loses in general loses the most genic variance which is a  
249 consequence of the high selection pressure for sire selection in this strategy. Also, completely  
250 genomic scenario (Genomic D) loses the most genic variance in all sire use strategies. We  
251 attribute this to a faster turn-over of the generations.

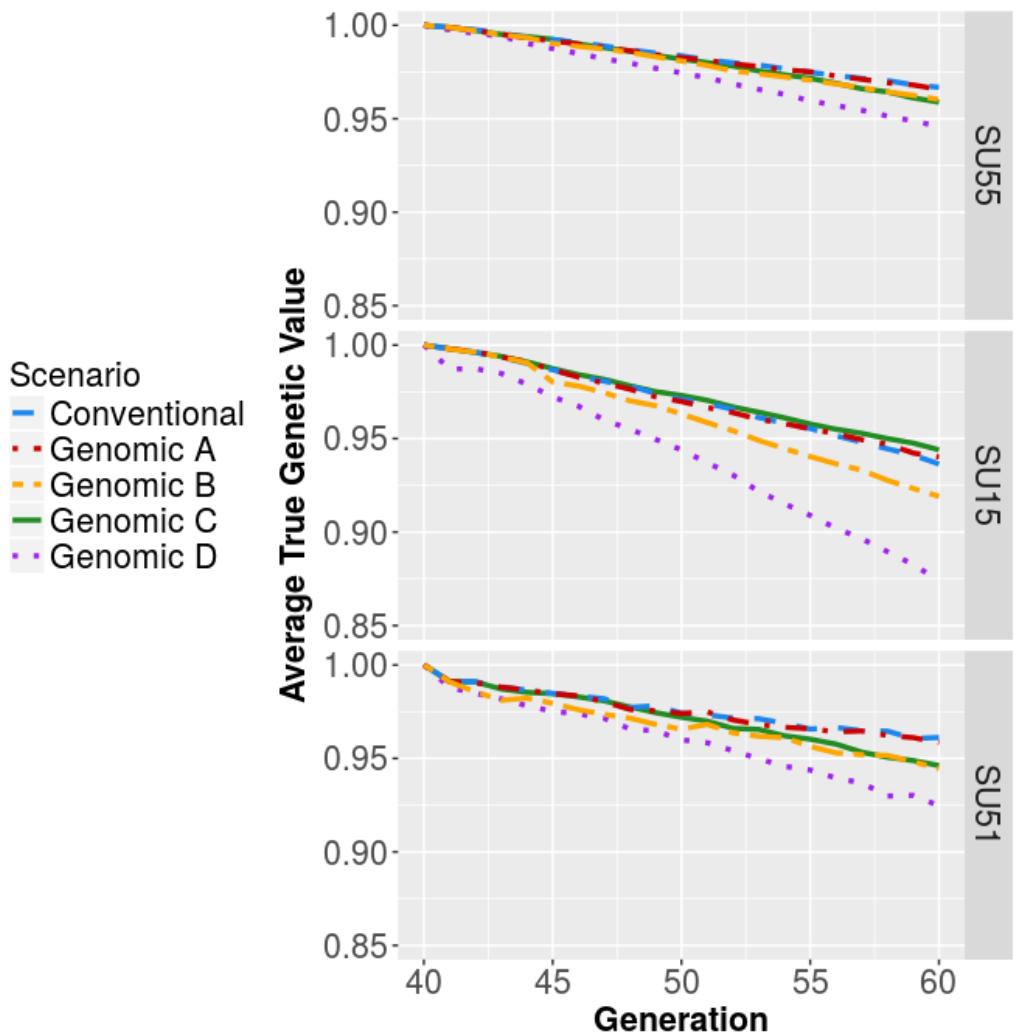


Figure: Genic variance for the five tested sire selection scenarios within the three sire use strategies averages across 20 replicates. Genomic A = genomic information used for selection of young bulls for progeny testing, genomic B / C / D = genomically tested bulls used respectively as sires of dams / as sires of sires / as sires of dams and sires of sires. Sire use strategies: SU 5/5 = five bulls selected each year and used for five years, SU 1/5 = one bull selected each year and used for five years, SU 5/1 = five bulls selected each year and used for one year.

**(Selection) efficiency**

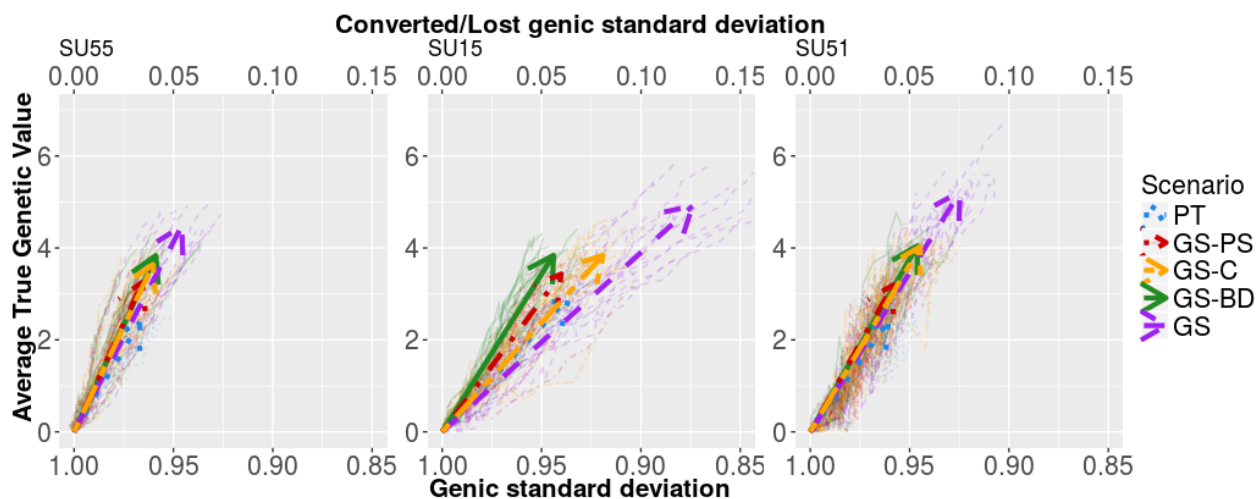
Selection efficiencies for five scenarios tested within three sire use strategies are shown in Figure 1. First we compared the efficiencies of the scenarios in the SU 5/5 strategy (Figure 1a). We observed that efficiencies do not follow the pattern of genetic gain. The efficiencies range between 71 and 100 with small differences between the scenarios (efficiency of e.g. 71 means that when we burn all

263 the genic variance we will reach genetic gain of 71 units). The PT scenario was the least efficient  
264 one and the most comprehensive genomic scenario (GS) was not the most efficient one. Only some  
265 of the differences were significant, i.e. between GS-PS, GS-C and GS-BD with PT scenario with  
266 p-values of 0.00, 0.008 and 0.000, respectively. Additional significant differences were between  
267 GS-BD and GS scenario (p-value = 0.026) and between GS-PS and GS (p-value = 0.048). Although  
268 not all differences were significant, the PT scenario was the least efficient in 10 / 20 replicates and  
269 in other 10 / 20 it was the second least efficient. GS scenario was the most efficient only in 2 / 20  
270 replicates. The most efficient in SU 5/5 were genomic scenarios GS-S and GS-BD, which were the  
271 most efficient in 7 / 20 and 6 / 20 replicates, respectively. We attribute that to a greater spread of  
272 sires EBV

273  
274 The results suggest that schemes that use progeny tested and young genomically tested bulls  
275 simultaneously are the most efficient one and therefore most appropriate for a long-term  
276 sustainability of selection (genetic gain?) – although they do not bring the highest genetic gain. We  
277 explain this by a greater spread of the true BVs of the PT sires due to a higher accuracy of their  
278 EBVs. Hence this hybrid schemes loses less genetic (genic) variance resulting in higher  
279 efficiencies.

280  
281 Next we compared selection efficiencies between the sire use strategies. Increasing the intensity of  
282 selection in the SU 5/1 strategy increased the differences between efficiencies of the scenarios  
283 which ranged from 39 to 68 (Figure 1b). The efficiencies were overall reduced compared to the  
284 SU 5/5 strategy which is due to a larger loss in genetic variance caused by selecting only bull per  
285 year (up to 8% larger loss in genic SD, SU 5/5 – 0.95, SU 1/5 – 0.87). The genomic scenario  
286 became the least efficient due to the largest loss in genic variance. We fixed this problem by  
287 selecting five bulls again in the SU 5/1 strategy. The plot of efficiencies in the SU 5/1 strategy

288 (Figure 1c) resembles more the one of the original SU 5/5 strategy (Figure 1a) since the differences  
289 between the scenarios are again decreased. However, the efficiencies are still reduced – range from  
290 60 to 72 – due to larger loss in genetic variance (up to ). We attribute this to a 0.1 unit drop in  
291 accuracy compared to SU 5/5 strategy, which we believe is due to a larger genetic distance between  
292 the training and the testing population. This together suggests that increasing the intensity of  
293 selection and reducing the generation interval even further has different consequences on short and  
294 long-term success of selection (genetic gain). Although both these strategies increase genetic gain  
295 they also loose more genic variance and decrease the efficiencies.



296 Figure 1: Change of genetic mean and genic standard deviation over the 20 years of selection  
297 by sire selection scenario within sire use strategies averaged across 20 replicates.

298 Arrows point the direction of change. Sire selection strategies: conventional = only progeny tested bulls (no genomic  
299 information), PT = conventional scenario without use of genomic information; GS-PS = genomic information used for  
300 pre-selection of young bulls for progeny testing, GS-C / GS-BD / GS = genomically tested bulls used respectively for  
301 insemination of cows / bull dams / cows and bull dams. Sire use strategies: SU 5/5 = five bulls selected each year and  
302 used for five years, SU 1/5 = one bull selected each year and used for five years, SU 5/1 = five bulls selected each year  
303 and used for one year.

304 **Conclusion**  
305

306 In summary, applying different strategies of the use of genomic information and sires has different  
307 effect on short- and long-term selection success. While a more comprehensive use of genomic  
308 information, increased intensity of selection and an additional reduction in generation interval  
309 increase genetic gain, all these strategies also result in a higher loss of genic variation and  
310 consequently reduced efficiencies. The optimal strategy for a specific small cattle population  
311 therefore depend on their risk preferences and selection goals.