# Discussion

Our results show that any dairy breeding programme can implement genomic selection without extra costs by optimizing the investment into phenotyping and genotyping, which could more than double genetic gain. The estimation of breeding values requires continuous investment in data collection. While breeding programmes usually have stable funding for phenotyping, the funding for genotyping is not yet well established. We show that by reallocating a part of phenotyping resources to genotyping, conventional breeding programmes can implement genomic selection and substantially increase genetic gain regardless of the amount and cost of genotyping, and availability of initial training population. With genomic selection, genetic gain further increases with increasing the investment into genotyping, despite simultaneously decreasing phenotyping. Although in genomic scenarios we reduced the number of phenotype records per animal, we increased the selection accuracy for non‑phenotyped candidates. These results raise four discussion points: 1) how optimizing the investment in phenotyping and genotyping affects genetic gain; 2) how optimizing the investment in phenotyping and genotyping affects accuracy ; 3) implications for breeding programmes; and 4) limitations of the study. For genetic gain and accuracy we first discuss the results of scenarios with equal price of phenotyping and genotyping, and initial training population available. We then remark on what changed at a different relative price or without an initial training population.

## Genetic gain

### Genomic vs conventional selection

Implementing genomic selection by optimizing the investment in phenotyping and genotyping increased genetic gain compared to the conventional selection.With an initial training population of 10,000 cows, all genomic scenarios outperformed the conventional scenario, mainly due to reduced generation interval in sire selection. This is in agreement with previous modelling studies and real data. Modelling studies showed that genomic selection increases genetic gain due to reduced generation interval, despite reduced selection accuracy in comparison to progeny test (Schaeffer, 2006; Pryce et al., 2010; Obšteter et al., 2019). Analysis of real data confirmed that the main driver of genetic gain with genomic selection is the reduced generation interval in the sire of bulls and sire of dam’s paths. In the US Holstein population it decreased between 25% and 50% compared to the conventional selection (Garcia-Ruiz et al., 2016). The amount of reduction in generation interval impacts the benefit of genomic over conventional selection. Van Grevenhof et al. computed a break‑even size of a training population to achieve a comparable response between genomic and conventional selection. They showed, that if the generation interval is not reduced and the number of phenotypes is limited, genomic selection cannot compete with conventional selection. But when generation interval is halved, a training population with ~2,000 or ~3,500 individuals gives comparable response as traditional selection based on own performance or 10 progeny per sire at an intermediate heritability. While the assumption of an available initial training population might not be realistic, it could be achieved through participating in international consortia. An example of such if InterGenomics for Brown Swiss in Central Europe (Jorjani, 2012).

Genomic scenarios were better also because the reduced number of phenotype records did not proportionally translate into reduced **accuracy**. While genomic scenarios only slightly decreased the selection accuracy for male candidates, they actually increased the selection accuracy for female candidates. We discuss the reasons for this in more details below.

Another major advantage of the genomic scenarios wasincreased intensity of sire selection. A costly and lengthy progeny-testing limits the number of tested sires in conventional selection. Genomic selection significantly reduces the cost of testing (Schaeffer, 2006) and thus increases the number of tested sires. In US Holstein population, genomic selection improved the selection differential for all traits, particularly for traits with low heritability, such as health and fertility (Garcia-Ruiz et al., 2016).

### Increasing the investment into genotyping

Genetic gain increased by increasing the investment into genotyping. This was mainly due to **increased intensity** of sire selection. Increasing the investment into genotyping allowed us to increase the number oft tested male candidates, but select the same number. We can see this as increasing investment into genotyping did not affect the generation interval nor accuracy of sire selection (discussed in the next section). Increasing the investment into genotyping also allowed for increasing theupdate and total size of the training population, which assisted in increasing genetic gain. This is in agreement with Thomasen et al., 2020, who showed that adding more cows yearly to the training population increases genetic gain. In our simulation a larger training population in turn increased selection accuracy of female candidates. The benefit of this was however not large, since the intensity of selection in females was very low.

The increase in genetic gain had diminishing relationship with increasing investment into genotyping. This has important implications for breeding programmes, since they use phenotypes also for management (discussed below). Results showed that investing resources of more than six phenotype records into genotyping did not significantly improve the genetic gain. The first reason for this is, that the accuracy of sire selection in genomic scenario was high regardless of the amount of genotyping when there were at least 10,000 animals in the training population. We discuss the reasons for this in detail below. Second, the intensity of sire selection had diminishing relationship with increasing genotyping. This agrees with Reiner-Benaim et al. showing that genetic gain increases with the number of tested male candidates, but with a diminishing return. While with four sires selected they achieved the maximum profit with 1721 tested candidates, they achieved 99% or 90% of the maximum profit with 740 or 119 tested candidates. Third, increasing investment into genotyping did not proportionally increase the size of the training population due to limited number of animals in the population and limited size of the training population. Once the investment sufficed to genotype all the females or when the size of the training population hit 25,000, investing more into genotyping did not increase the size of the training populationg. Due to the same three reasons we also achieved a comparable maximum genetic gain regardless the relative price of phenotyping to genotyping. In general, selecting less than 2% of the tested males and updating the training population with more than 35% of first parity cows resulted in the maximum genetic gain.

Our results agree with previous studies showing that adding females to the training population has diminishing return relationship with accuracy and genetic gain (Van Grevenhof et al., 2012; Gonzalez-Recio et al., 2014). Consequently, when the number of females in a training population is large, an additional record has a smaller additional value than when a training population is small. Since our scenarios with initial training population started ~10,000 genotyped and phenotyped cows, enlarging the training population had small effect. Increasing the training population beyond that decreased the value of additional record even further .

While genetic gain increases with the number of females in training population, adding repeated records does not have the same effect. As we increased the number of females in the training population, the number of repeated records decreased (Figure S1). The scenarios with largest genetic gain therefore had a training population with many cows and few repeated records. However, since we ran single-step genomic prediction, the phenotypes of the non-genotyped animals contributed to the estimation as well. Effectively, all scenarios thus operated with the same number of phenotyped animals.

We should emphasize, that some of the high‑genotyping scenarios achieved the observed genetic gain at a lower total cost, since they could not use all the available resources for genotyping females. The saved resources could be invested back into phenotyping females for milk production or novelty traits, genotyping more male candidates, or some other breeding action. Buch et al., 2011, showed that for new functional traits, it is possible to achieve adequate accuracy of genomic prediction within three years of recording a new trait.

### Scenarios without an initial training population

We also considered that some small populations do not have access to a training population and have to initialize one themselves. These genomic scenarios still increased genetic gain compared to the conventional scenario, but achieved lower genetic gain than corresponding scenarios with an initial training population available. This was mainly due to delay in implementing genomic selection and smaller training population. Consequently, increasing the investment into genotyping compensated for starting without a training population in two ways. Firstly, it shortened the time to obtain the 2000 genotypes required to implement genomic selection down to one year in high‑genotyping scenarios. Secondly, it shortened the time to build a training population in which an additional record had negligible effect on accuracy (OR accuracy comparable to when we had an initial TP). Gonzales-Recio et al. showed, that for most traits the additional gain from increasing the number of females above 10,000 is negligible.

We should note, that when implementing genomic selection with a delay, we did not observe any decrease compared to the conventional scenario prior to the implementation, despite reduced phenotyping. This suggests that breeding programmes could run a conventional breeding programme with reduced phenotyping until they accumulate genotypes to initiate a training population, without harming the genetic gain in the accumulation or transition period.



**Figure S1:**The number of animals and phenotypes in the training population.

## Accuracy

Despite reduced phenotyping, genomic scenarios increased the accuracy for young non‑phenotyped calves and cows.In general, genomic prediction increases the accuracy of the Mendelian sampling term. This is the main reason for increase in accuracy with genomic prediction when the accuracy of parent average is high. But when the accuracy of parent average is low, such as for non‑phenotyped parents or parents with little own or progeny information, genomic information increases accuracy both for the parent average and the Mendelian sampling term (Daetwyler, 2007; Wolc, 2011).

### Accuracy for males with initial training population

For male candidates, genomic prediction more than doubled the accuracy compared to the parent average in conventional scenario. This is partly in agreement with Wolc et al., 2011, and Schaeffer, 2006 who showed that genomic prediction can increase the accuracy of early selection up to two‑fold. However, in our study, this increase was even higher, since genomic prediction also increased the accuracy of parent average.

Within the genomic scenarios, the accuracy for male candidates was high regardless of the amount of genotyping and phenotyping for two reasons. First, due tothehigh accuracy of their parent average, since we tested all the offspring of elite matings. Second, starting with a 10K training population gave an adequate starting point for accurate prediction. The accuracy was additionally boosted by using all available information jointly through single‑step genomic prediction.

In contrast, reducing phenotyping decreased the accuracy for sires, despite increased genotyping. This was due to two reasons. First, since we used truncation selection to select the sires, their breeding values lie in the tail of distribution. Each additional phenotypic record increased the precision of individuals breeding values, although only marginally, and helped to distinguish the sires. Second, with increasing investment into genotyping the training population grew quicker and reached the limit of 25,000 at which the sires genotypes were removed. However, since this is the accuracy after the selection has been made, it is not of great interest for breeding.

Although sires already had phenotyped progeny, their accuracy was lower than for male candidates and had a larger standard deviation. First, this was due to a small number of sires, since each year we selected only five. Second, both male candidates and sires came from a truncated distribution with reduced variance, but the variance for the sires was even smaller This in turn reduced the empirical accuracy computed as the Pearson’s correlation coefficient between the true and estimated values.

***Accuracy for females with initial training population***

Genomic scenarios increased the accuracy for cows compared to the conventional scenario. Besides increasing the accuracy of Mendelian sampling term, using genomic information increases genetic connectedness between individuals from different management units (Yu et al., 2017, Powell et al., 2019). This in turn increases the accuracy of prediction regardless of the heritability, number of QTLs, and number of markers (Yu et al., 2018).

The accuracy for dams further increased with increasing investment into genotyping,despite reduced phenotyping. Increasing investment into genotyping translated into growing training population and a larger training population update. As shown by previous studies (Bijma, Gonzales-Recio), the accuracy of genomic prediction increases with increasing size of a female training population, even up to 100,000 females. Same studies also shown that the accuracy of 0.70 is achieved with ~20,000 animals, which agrees with our results. However, these studies did not account for varying degree of genetic distance between the training and the evaluation population. As shown by previous studies, we can increase the accuracy in the evaluation population with a higher relationship to the training population (Pszczola et al., 2012, Habier et al., 2010; Clark et al, 2011). Increasing the investment into genotyping allowed us to genotype more females and include more females from the most recent cow generation in the training set. This decreased the genetic distance between training and evaluation population and in turn increased the accuracy. Genotyping more females had two additional benefits. Firstly, more cows had both genomic and phenotypic information available, which increased the accuracy of their breeding values. Secondly, as shown by Yu et al., 2018, increasing the number of genotyped animals increases genetic connectedness. As with genetic gain, increasing the size of the training population had a diminishing return relationship with accuracy(Bijma, Gonzales-Recio)**.** Correspondingly, investing resources of more than six phenotype records into genotyping plateaued.

Accuracy for female candidates followed the accuracy trend for the dams, but at lower values. Female candidates were not genotyped nor phenotyped, hence their accuracy mainly reflected parent average accuracy. Increasing genotyping increased the accuracy for dams and in turn increased the accuracy of the parent average for female candidates.

### Accuracy without an initial training population

Accuracy in scenarios without an initial training population closely followed the trends of the corresponding scenarios with initial training population available. We observed minor differences in the low genotyping scenarios that had reduced accuracy for male candidates and sires. We attribute this to a smaller training population. Buch et al, 2011, showed that for new traits and with large scale recording, we can achieve 75% of the maximum genomic accuracy within first two to three years of recording. In our study we shortened this period even more by including the historical data through single-step genomic prediction.

## Implications

* 1. We show that any dairy breeding programme can implement genomic selection without increase in costs but only by optimizing the investment into breeding actions. Here we propose funding the genotyping with a part of resources for milk recording, since it can manipulate with number of repeated records. Breeding programmes could reduce phenotyping for a different trait that they record repeatedly and is perhaps less crucial for management. They could also reallocate the funds from another breeding action, if it does not result in cancelling a crucial activity.
     1. When breeding programmes have limiting resources, they could optimize which individuals to genotype and phenotype, which we did not consider in this study. We expect this would further increase the genetic gain for the same level of investment or require less investment for the same genetic gain. Selective phenotyping can increase the accuracy of genomic selection up to 20% with a larger increase observed with small sample sizes (Heslot et al.., 2017; Akdemir and Isidro-Sanchez, 2019). Researchers also suggested the use of phenotyping farms, which could be contracted and paid to provide records (ICAR 2011 Coffey presentation, no abstract). Similarly, Jenko et al., 2017, showed, that selective genotyping of cows from the distribution tails increases the accuracy of genomic prediction by 15% compared to random selection.
  2. When breeding programmes do not have access to high performance computers necessary for genomic evaluation of big genotyped populations, they could optimize the computational cost. As shown in our study, we can achieve large genetic gain with a relatively small training population of recent genotypes. This implies that breeding programmes do not have to use all available genotypes for prediction. The problem of a large number of genotypes can be alternatively solved by using methods with reduced computational costs, such as algorithm for proven and young (Misztal et al., 2014) or singular value decomposition of the genotype matrix (Ødegård et al., 2018).

### Target population

1. The economic efficiency of the programmes strongly depends on who pays for which action. The scenarios presented in this paper are of little value for programmes, where phenotyping and genotyping funding is disconnected. But different programmes have different investment schemes, often intricate, which could benefit from suggested solutions. Similarly, optimizing the investment into phenotyping is not of interest for breeding programmes with abundant use of automated milking systems. With automated systems the cost of phenotyping does not depend on the number of records. But in populations with small herds the use of automated system is still limited, since its benefits do not make up for the high initial cost. Further on, the genomic selection could be more beneficial for some settings than the others. Powell et al., 2019, showed, that genomic information is especially important for generating genetic connectedness in systems with small herd sizes, geographically dispersed farms, and limited use of artificial insemination, often found in low to mid income countries. Kasap et al., 2018, showed the same benefit for sheep breeding, where herds do not actively exchange of sires between herds. In such settings, we can additionally increase the prediction accuracy with spatial modelling of the data. This establishes environmental connectedness, which helps to separate the genetic and environmental effects (Selle et al., 2020).

### The use of genotypes

In our study we used genotypes only for the prediction of genomic breeding values and achieving genetic gain. In breeding programmes, the genomic information has additional value for the breeders. Firstly, animals genomic information could be used for parentage verification or parentage discovery (ICAR Guidelines for Parentage Verification and Parentage Discovery Based on SNP Genotypes). This eliminates the cost of an alternative method, such as obtaining animal’s microsatellite information. Secondly, genotypes provide information on causative loci for some monogenic diseases and traits (included for free or for a small royalty). This information can prevent large economic loss caused by spreading the lethal alleles. It can also create economic gain by adding value to the product, such as branding A2 β-casein milk or producing B κ-casein milk with better coagulation properties. Thirdly, the genomic information could be used for a better monitoring and control of inbreeding (Woolliams et al., 2012), and optimization of matings (Obšteter et al., 2019). These additional uses of genotypes increase the return on investment of genomic selection, also in long-term. Further on, maintaining the system is more economically efficient in genomic than in conventional selection (König et al., 2009) for at least three reasons. Firstly, genomic selection removes the need for costly and lengthy progeny testing. Secondly, to maintain high accuracy of prediction across the generations, genomic selection requires only a minor update of the training population, while conventional selection requires another round of progeny-testing (Gonzales-Recio). And thirdly, genomic selection increases the value of the phenotype, since it prolongs its usefulness to many generations (compared to few in conventional selection). In order for this to hold, breeders and breeding organizations should genotype the phenotyped animals (Bijma reference).

## Limitations of the study

### Reduced number of phenotype records

* 1. In this study, we optimized the number of repeated test-day records with the aim to estimate individual’s breeding value and achieve genetic gain. In reality, breeding programmes have to balance the number of records for achieving genetic gain and managing the herd, which we did not consider in this study. Farmers use phenotype records to manage animals’ health and feed composition, which affect milk yield and composition. Besides managing production, milk recording is also important from an environmental perspective. By managing the milk urea concentration, herds can decrease the nitrogen footprint per kg of milk (Verbič et al., 2019). Breeding programmes do also not use records directly for predicting the breeding values. Instead, they use test day records to estimate the 305-day milk yield according to standard lactation curves using various regression methods (reviewed in ICAR Guidelines: Computing of Accumulated Lactation Yield, 2020; Jeretina et al., 2013). This additional prediction step decreases the accuracies below the ones observed in our study. While previous studies quantified the accuracy of this prediction, determining the value of phenotype for management is more complex.
  2. The shortest ICAR standard recording interval that we tested in a genomic setting was five weeks, corresponding to nine records per lactation. In some settings, this sufficed to achieve the maximum genetic gain while in others we achieved only 68% of the maximum genetic gain. Studies suggest, that using nine instead of eleven records would not greatly affect the accuracy of predicting the 305‑day milk yield. They observed a high correlation (between 0.96 and 0.98) of prediction based on 5‑weekly and weekly records (Pool and Meuwissen, 1999). Gartner et al., 2008, similarly observed high correlation of 0.96 between predicting the 305-day milk yield from 11 (ICAR A4 standard) or eight (ICAR A6 standard) test day records. They however showed that using eight records yields a high bias and significantly underestimates the 305-day milk yield by 500-1000 kg.
  3. The longest sampling interval tested in our study and still approved by ICAR was nine weeks, which yielded five records per lactation and invested the resources of six records into genotyping. In most settings this sufficed to achieve the maximum genetic gain. Previous studies also showed a good predictive ability of such scheme for estimating the 305-day milk yield. Pool and Meuwissen, 1999, showed that the correlation of prediction based on weekly and 9‑weekly records was between 0.92 and 0.96. Berry et al., 2005, showed that the mean error of 305-day yield estimated from five test day records was 6.8kg with 0.99 correlation with 305-day yield estimated from 11 records. Studies also showed that choice of the model affects the prediction outcome, hence the prediction could be optimized (Pool and Meuwissen, 1999; Lidauer et al., 2003).
  4. Investing more than the resources of six records into genotyping did not prove as necessary in our study, since it did not increase the genetic gain, accuracy for selection candidates, nor used all the available resources. Also, collecting only one record does not allow to estimate animal’s permanent environment effect and in turn decreases the accuracy of breeding values. However, breeding programmes could want to invest more than the resources of six records into genotyping when initializing genomic selection and aiming to quickly build a training population. Kong et al., 2017, explored using three records to predict the 305‑day milk yield. They achieved the accuracy between 0.67 and 0.99 in the first lactation, between 0.92 and 1.00 in the second, and between 0.91 and 1.00 in the third lactation, depending on the statistical model.
  5. The effect of reduced records on herd management is much more intricate and less measurable. The number of records required for efficient herd management highly depends on management practices. Studies confirm this by showing that herd-test day variance, which reflects the variance due to management, can greatly exceed the genetic variance for milk yield (Caccamo et al., 2008) or be less than it (e.g. Špehar et al., EAAP 2008 poster).

### Limited size of the training population

In our simulation the upper limit for a training population was 25K. Although we achieved high accuracies, increasing the size of the training population could increase them even further. However, as already mentioned, the value of additional female decreases with the size of the training population. Studies also showed that increasing the training population reduced the economic efficiency of genomic selection (Azizian et al., 2016). Since we included the most recent animals in the 25K set, increasing the size would also result in adding older animals to the training population. These animals are genetically more distant from the evaluation population and of lesser value.

### Single additive trait

We simulated milk yield as a single polygenic trait with additive genetic as well as permanent, common and random environmental effect. In reality, non‑additive genetic effects also affect the trait. According to previous studies, the dominance effect can amount to between 12% and 45% of the additive effect for milk yield (Fuerst and Sölkner, 1994; Ertl et al., 2014; Aliloo et al., 2016; Jiang et al., 2017). In our simulation, we did not directly simulate nor account for them, but individual’s permanent effect includes non-additive genetic effects or individual specific environmental effects. Studies showed, that around 25% of the permanent environment variance is due to dominance effects (Aliloo et al., 2016). We also simulated milk yield in different lactations as a single trait, whereas genetic correlation between different lactations is not unity. Studies observed correlation between 0.82 and 0.97 for milk yield in different lactations (Meyer, 1984; Dong and Van Vleck, 1989; Swalve and Van Vleck; 1987).

### Genomic selection of females

In this study we did not implement genomic selection in the female path nor did we assume the use of female reproductive technologies, such as embryo transfer or sexing semen. This would further decrease the generation interval and increase selection intensity on female side, which would increase genetic gain of genomic scenarios even more (Pryce et al., 2010; Garcia-Ruiz et al., 2016). Implementing genomic selection of females would require only a minor modification, that is, genotyping heifers instead of first-parity cows. Implementing female reproductive technologies would require a larger modification and larger investment. However, some of the tested scenarios saved some of the available resource and could invest in embryo transfer or some other technology.

# Conclusion

This study shows that any dairy breeding programme can implement genomic selection and maximize the return on investment with no extra costs by optimizing the investment into milk phenotyping . We showed, that genomic scenarios increased both genetic gain and selection accuracy for non-phenotyped candidates, despite reduced phenotyping. The increase was observed regardless the amount and cost of genotyping, and availability of initial training population, which indicates the advantage of proposed solutions for a range of breeding programmes. However, increasing investment in genotyping had diminishing returns, which suggests that breeding programmes should balance the investment into phenotyping and genotyping to maximise return on investment.