**Discussion**

The estimation of breeding values requires stable investment for continuous collection of data. While breeding programmes usually have a secure funding for phenotyping, the funding for genotyping is not yet well established to initiate and regularly update the training population for genomic prediction. In this paper we propose implementing genomic selection by optimizing the investment into phenotyping of milk production traits and genotyping. We show that by reallocating a part of phenotyping resources to genotyping, we can substantially increase genetic gain regardless of the cost and amount of genotyping, and availability of initial training population. We also show that we can increase the genetic gain even further by increasing the investment into genotyping, despite simultaneously decreasing the amount of phenotyping. Similarly, although reduced phenotyping decreased the phenotype accuracy, genomic prediction increased the accuracy for non-phenotyped selection candidates. These results raised four discussion points 1) how optimizing investment in phenotyping and genotyping affects genetic gain with or without an initial training population; 2) how optimizing investment in phenotyping and genotyping affects accuracy with and without an initial training population; 3) limitations and remarks of the study; 4) implications for breeding programmes. For all measures, w**e discuss in details the trends of scenarios with equal price of phenotype and genotype and initial training population available and then comment of what changed when we changed the relative price of phenotyping to genotyping or when we started without an initial training population.**

**1 Genetic gain**

**Implementing genomic selection by optimizing the investment in phenotyping and genotyping increased genetic gain compared to the conventional scenario despite reduced phenotyping, even at minimum investment into genotyping.** In this study we assumed that some small populations have access to an international training population, such as InterGenomics for Brown Swiss in Central Europe (Jorjani, 2012). With an initial 10K training population available, all genomic scenarios outperformed the conventional scenario, mainly due to reduced generation interval. This is in agreement with real data and previous simulation studies. Garcia-Ruiz et al. analyzed US Holstein data and showed that the main driver of genetic gain in genomic selection is the reduced generation interval in the sire of bulls and sire of dams paths, that decreased between 25% and 50% compared to the conventional selection. Simulation studies also confirmed that the genomic selection increases genetic gain due to reduced generation interval, despite reduced selection accuracy (Pryce et al., 2010; Obšteter et al., 2019). Van Grevenhof et al. computed the break-even size of the training population required to achieve a response comparable with conventional selection. They showed, that if the generation is not reduced and the number of phenotypes is limited, genomic selection cannot compete with conventional selection. But as the generation decreases, the break-even size decreases rapidly. When generation interval is halved, only ~2000 or ~3500 individuals are needed to achieve the response of selection on traditional BLUP-EBV based on own performance or 10 progeny per sire.

Another major advantage of the genomic scenarios was increased intensity of sire selection. A costly and lengthy procedure of progeny-testing limits the number of tested sires in the conventional schemes. Genomic selection significantly reduced the cost of testing male candidates (Schaeffer, 2006) and thus allowed for increased intensity of male selection. In US Holstein population, genomic selection increased (improved) the selection differential for all traits, even more for the low heritable ones, such as health and fertility traits (Garcia-Ruiz et al:).

Assisting the superiority of genomic scenarios was also the fact, that although reduced phenotyping decreased the phenotype accuracy, it did not affect the selection accuracy in the same way. While the genomic prediction only slightly decreased the selection accuracy for sires, it actually increased the selection accuracy for females. We discuss the reasons for this in more details below.

**We further increased the genetic gain with increasing the investment into genotyping**, mainly due to increased intensity of sire selection. Investing more into genotyping resulted in more male candidates tested, but same number selected, which intensified selection and drove the genetic gain. This can be seen as increasing investment into genotyping did not further reduce generation interval nor increase the accuracy of sire selection candidates (discussed in the next section). A minor drive of increasing genetic gain was also enlarging the update and total size of the training population. This is in agreement with Thomasen et al., 2020, showing 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 diminished, since the intensity of selection in females was very low. It is also worth mentioning, that some of the high-genotyping scenarios achieved the observed genetic gain at a lower total cost, since they did not use all the 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.

Although increasing investment into genotyping increased genetic gain, the increase was not proportional. Instead, **increasing genotyping had a diminishing return relationship with genetic gain,** **which reached a plateau**. Results showed that when phenotyping and genotyping had equal cost, investing more than the resources of 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 did not increase with increasing genotyping, but was high regardless. We discuss the reasons for this in more details below. Secondly, the intensity of male selection was high in all top performing scenarios. This agrees with Reiner-Benaim et al., 2017, showing that genetic gain increases with the number of tested candidates, but with a diminishing return. They showed that with 4 sires selected, the optimal number of tested calved yielding maximum profit is 1721. They also showed, that 99% or 90% of the profit is achieved with 740 or 119 calves tested. However, they assumed that the price of genomic evaluation is $95, which has reduced since then. And lastly, the top performing scenarios had a large percentage of training population update and a large training population. Due to similar reasons, **we 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 a female to the training population has diminishing return relationship with accuracy and economic genetic gain (Van Grevenhof et al., 2012; Gonzalez-Recio et al., 2014). Gonzales-Recio et al. showed that when the number of females in a training population is small, an additional record has a larger value for the genetic gain than when the female training population is large. Since our scenarios in question all started with a ~10K training population, additional revenue from enlarging the training population was small to begin with, but still decreased with increasing genotyping.

In this study we additionally showed, that while genetic gain does increase with the number of females in training population (although with diminishing return), adding repeated phenotypes does not have the same effect. As the scenarios increased the number of females in the training population, they also decreased the number of (repeated) phenotypes (Figure S1). The top performing scenarios therefore had a training population with the most females but also the least (repeated) phenotypes. However, since we ran single-step genomic prediction, the phenotypes of the non-genotyped animals contributed to the estimation as well.

In this paper we also considered a situation, when small populations **do not have access to a training population and have to initialize one themselves**. **These genomic scenarios still increased the genetic gain between 31% and 134% compared to the conventional scenario**. However, compared to the corresponding scenarios with an initial training population available, these scenarios achieved lower genetic gain. This was mainly due to smaller training population and delay in implementing genomic selection. Increasing the investment into genotyping compensated for starting without a training population due to two reasons. Firstly, investing more into genotyping shortened the delay in implementation of genomic selection (down to one year in high‑genotyping scenarios). And secondly, a smaller initial training population (until reaching 25K) did not proportionally translate into smaller genetic gain, since increasing the number of females in the training population has diminishing return. For example, Gonzales-Recio et al. showed, that for most traits the additional gain from increasing the number of females above 10,000 is negligible.

When implementing genomic selection with a delay, we did not observe any increase in genetic gain above the conventional scenario prior to implementing genomic selection of sires. On the other hand, we also did not observe a decreased genetic gain compared to the conventional scenario prior to the implementation, despite reduced phenotyping. This suggests that breeding programmes can 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 (transition?) period.

However, 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. This would further decrease the generation interval and increase genetic gain of genomic scenarios (Pryce et al., 2010; Garcia-Ruiz et al., 2016). Implementing genomic selection of females would require a minor modification of the scheme used in this paper, i. e. genotyping heifers instead in first-parity cows. Regarding female reproductive technologies, some of the tested scenarios saved some of the available resource and could invest in embryo transfer or some other technology.



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

**2 Accuracy**

Despite reduced phenotyping, genomic scenarios increased the accuracy for young non‑phenotyped animals and dams.For male candidates, genomic scenarios increased the accuracy of parent average in the conventional scenario (first stage of selection) by more than two fold. This is partly in agreement with Wolc et al., 2011, which showed that prediction based on genomic relationship 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. **The accuracy for male candidates also persisted high regardless of the amount of genotyping and phenotyping.** This was firstly due tohigh accuracy of parent average, since we tested the offspring of elite and other high-performing matings. And secondly, starting with a 10K training population gave a good base for genomic prediction / single-step!!! This might be higher then reported elsewhere!

In contrast, reducing phenotyping decreased the accuracy for sires, despite increased genotyping. Since we used truncation selection to select the sires, their breeding values all lie very closely at the top of the distribution. To distinguish between them and rank them correctly, their breeding values therefore have to be very precise. Hence, each additional phenotypic record helps to correctly distinguish between sires. However, since this is the accuracy after the selection has been made, it is not of great interest for the breeders.

The accuracy for the sires was also below the accuracy for male selection candidates and and has a larger standard deviation. This was due to a smaller number of sires selected than male candidates tested. Also, although both male candidates and sires come from a truncated distribution with reduced variance, the variance for the sires is even more reduced. This is turn reduced the accuracy, computed as Pearson’s correlation coefficient.

**Genomic scenarios increased the accuracy for dams** compared to the conventional scenario. The first reason for this is that genomic prediction estimates Mendelian sampling term with higher accuracy. As shown by Daetwyler et al., 2007, when the accuracy of parent average is low, genomic information helps to predict both parent average and Mendelian sampling term. And when the accuracy of parent average is high, the increase in accuracy with genomic prediction relates to the Mendelian sampling term. Secondly, using genomic information increases genetic connectedness between individual (from different management units) (Yu et al., 2017). This in turn increases the accuracy of prediction regardless the heritability, number of QTLs, and number of marker (Yu et al., 2018). However, the number of markers have to be large enough to capture the QTL signal. **The accuracy for dams also increased with increasing investment into genotyping,** despite reduced phenotyping. This is perhaps surprising, since own-performance contributes most to the accuracy of prediction. Increasing investment into genotyping translated into growing training population and a larger size of the update. As shown by previous studies (Bijma, Gonzales-Recio), the accuracy of genomic prediction does increase with increasing size of a female training population, even up to 100,000 females included. Same studies also shown that the accuracy of 0.70 is achieved at ~20,000 animals, which agrees with the result of this paper. However, these studies did not account for varying degree of genetic distance between the training and the evaluation population. As shown by Pszczola et al., 2012, we can increase the accuracy in the evaluation population with a higher relationship and less generations to the training population. Increasing investment into genotyping allowed for include more females from the most recent cow population into the training set, which decreased the genetic distance between training and evaluation population and in turn increased the accuracy. Further on, as the investment into genotyping increased, we were able to obtain genomic information for more females and use it in conjunction with own phenotypic data to predict their breeding values with added value. Also, as already mentioned, genomic information increases genetic connectedness and as shown by Yu et al., 2018, increasing the number of connecting animals increases the connectedness eve further. As with genetic gain, **increasing the size of the training population had a diminishing return relationship with accuracy. Correspondingly,** investing more than the resources of six phenotypic records into genotyping did not further increase accuracy.

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

Accuracy in scenarios without an initial training population closely followed the trends of the corresponding scenarios with an initial training population available. We observed minor differences in the low‑genotyping scenarios, that had a reduced accuracy for male candidates and sires. For male candidates this stemmed from a smaller training population, and for sires from an interplay of the number of phenoype records and number of genotyped daughters.????

In our simulation we did not account for the lactation curve of milk yield. In practice, test day records are used to compute the 305-day milk yield according to standard lactation curves and using different regression methods (reviewed in ICAR Guidelines: Computing of Accumulated Lactation Yield, 2020). Previous studies explored, how removing test day records affects the accuracy of prediction the full lactation yield. Berry et al., 2005, concluded that with five records (or more), which corresponds to ICAR 8-weeks sampling scheme, we can predict the 305-day milk yield with high accuracy. They showed that the mean error of estimating 305-day yield from five test day records was 6.8kg with 0.99 correlation with 305-day yield estimated from 11 records. Kong et al., 2017, explored the accuracy of estimating 305-day milk yield from three vs. six test day records. They showed, that while in the first lactation using six vs three records increased the accuracy between 0.01 and 0.31, in the second and third lactation the difference decreased to 0.00-0.07 and 0.01-(-0.12), depending on the breed. On the other hand, Gartner et al., 2008, observed significant differences between 305-day milk yield predicted from eleven (ICAR A4 standard) or eight (ICAR A6 standard) test day records. Although they observed a correlation of 0.96 between the predictions, they observed that prediction on eight records yield a high bias and underestimates the 305-day milk yield by 500 – 1000 kg.