PERSISTENCE OF GENOMIC ESTIMATED BREEDING VALUES AND IMPACT OF MISSING RECORDS IN COMMERCIAL PIG EVALUATIONS

by

MARY KATE HOLLIFIELD

(Under the Direction of Ignacy Misztal)

ABSTRACT

Increasing the accuracy of estimated breeding values (EBV) improves the rate of genetic gain, resulting in superior animals and greater profitability. Accuracies of genomic EBV can be maximized if enough information is available in the reference population, and more importantly, the effects of the independent chromosome segments (M_e) are explained. Commercial farm populations are typically highly related; hence, inheritance includes large chromosome segments. If the effects of the independent chromosome segments are well estimated, genetic predictions will have high accuracy. The persistence of accuracy for commercial populations will remain more stable over time as the available data increases. Livability and retained tag information can be included in growth trait evaluations; however, accuracy does not improve when these causes for missing measurements are included. The objective of this thesis was to analyze genetic parameters and accuracies of (G)EBV using varying traits, quantities of data, and effects of including the reasons for missing records on pig genetic evaluations.

INDEX WORDS: genetic selection, growth traits, LR method, predictive ability, swine

PERSISTENCE OF GENOMIC ESTIMATED BREEDING VALUES AND IMPACT OF MISSING RECORDS IN COMMERCIAL PIG EVALUATIONS

by

MARY KATE HOLLIFIELD

B.S., North Carolina State University, 2019

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

PERSISTENCE OF GENOMIC ESTIMATED BREEDING VALUES AND IMPACT OF MISSING RECORDS IN COMMERCIAL PIG EVALUATIONS

by

MARY KATE HOLLIFIELD

Major Professor:	Ignacy Misztal
Committee:	Daniela Lourenco Romdhane Rekaya Jeremy Howard

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2021

DEDICATION

To my grandmother, in loving memory.

ACKNOWLEDGEMENTS

I would like to think Dr. Ignacy Misztal for the opportunity to be a part of this incredible group of researchers and mentors. I am so grateful for all the knowledge and skills I have gained these past two years. The class discussions, research guidance, and specifically, the lunch conversations will forever be memorable and cherished. Thank you to all my other committee members, Dr. Daniela Lourenco, Dr. Romdhane Rekaya, and Dr. Jeremy Howard, for the valuable knowledge and advice, both academically and personally.

I am so thankful for the internship opportunity at Smithfield Premium Genetics and Dr. Jeremy Howard's guidance. I now have so much appreciation for the hard-working farm technicians who put in blood, sweat, and tears collecting phenotypes.

This accomplishment would not have been possible without the immense support from each and every colleague, visitor, postdoc, researcher, and professor that I connected with along the way. The UGA Animal Breeding and Genetics group is more like a family, and I am so excited to spend a few more years with this special group as a Ph.D. student.

Lastly, I would like to thank my family, friends, and especially my parents for your unconditional support and encouragement throughout my academic career.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

BREEDING PROGRAMS FOR COMMERCIAL PIG PRODUCTION

The goals of swine breeding are dependent on consumer demand and profitability, both of which vary widely around the world due to culture and available resources. The desired traits can be divided into meat quality characteristics, resource efficiency, and durability. All of which focus on maximizing economic gain. Geneticists optimize the economically valued traits by implementing a three-way crossbred pig as the commercial terminal product (Lutaaya et al., 2001).

Generally, commercial producers use an integrated production system to update genetics by generating estimated breeding values (EBV) to decide which animals to keep as breeders based on information collected from the nucleus and commercial herds. The integrated system involves each commercial terminal animal to transport from a farrowing farm, a nursery farm, and a finishing farm before harvest. A few reasons for this production method are more profitability for more farmers, efficient use of the barns (more space needed for finishing pigs than nursery pigs), and optimizing the animal's living conditions and needs in each stage of life. While this method has been successful for decades, it is difficult for breeders to model the effects of each new environment, pen grouping changes, and transportation on production traits. Unlike the dairy industry, pork production is centered around optimizing crossbred performance and has

less emphasis on purebreds (Knol et al., 2016). Closed nucleus farms are used for breeding purebred sows and boars to be the parents and grandparents of the terminal animals. The best performing purebred animals remain in the nucleus herd as parents, and the others are culled. The animals selected from the most updated models are in the nucleus herd. The genetic effects from the selection decisions made in the nucleus herd take many generations to reach the commercial terminal animals.

The most economically important trait of the breeding sow is the number of piglets weaned per litter (Serenius et al., 2004). Culling a sow due to nonsufficient reproductive performance is a high cost for the producers (Serenius and Stalder, 2006). In the early 2000s, piglet mortality in a commercial setting was approximately 20% (Grandinson et al., 2002). Piglet mortality is often due to crushing, low birth weight, or lack of essential nutrients. Mortality is a heritable trait that can be selected. Roehe (1999) found that lighter piglets have a higher probability of mortality. For maximum production, survivability and birthweight models need to be improved to have more pigs surviving to harvest.

It is hypothesized that piglet mortality increased as breeders increased selection on litter size, consequently lowering birth weight per pig (Quiniou et al., 2002). Fix et al. (2010) found that animals with a heavier birth weight had a faster daily gain resulting in heavier body weight at harvest. The genetic correlation between birth weight and hot carcass weight was calculated as 0.55 ± 0.15 in a crossbred commercial pig population (Dufrasne et al., 2013). With this positive correlation, it seems it would be of interest to directly select for birth weight to increase hot carcass weight and decrease piglet mortality; however, birth weight and litter size have a negative relationship, so it is difficult to select for more pigs and heavier pigs simultaneously.

Previous studies have shown that it is possible to considerably improve several economically essential traits by incorporating an associative social interaction effect in the breeding program (Muir and Schinckel, 2002; Bijma et al., 2007a). The associative effects from group mates that alter an animal's performance or phenotype are considered indirect genetic effects (IGE). Linear and moderately heritable traits (i.e., hot carcass weight) can also be affected by IGE and improved through social interaction models (Bergsma et al., 2008; Chen et al., 2009). Bijma et al. (2007b) found that two-thirds of heritable variation in chicken morality is social interaction. Since, by definition, a phenotype is equal to the sum of the direct effect and all associative effects from group members, the heritable variation due to social effects is often hidden in classical analyses (Bijma et al., 2007a). Including IGE in the model improves response to selection for growth traits, and as a result, increases heritability (Bergsma et al., 2008). Selecting pigs that are not as aggressive or interactive will decrease injuries, which will improve animal welfare. Accounting for social interaction requires individual identification and group information. It is not common for producers to collect this data in a commercial setting. Moreover, to improve growth traits and consequential profitability, social interaction models should be implemented in commercial breeding programs.

GENOMIC SELECTION

The genomic estimated breeding values (GEBV) of animals are estimated by incorporating single nucleotide polymorphism (SNP) marker information in genomic models and are used widely across commercial livestock breeding programs (Meuwissen et al., 2001; de Roos et al., 2011). The initial excitement of genotyping began with the first draft of the human genome project in 2001 (Sachidanandam et al., 2001). Since then, the amount of genomic

information available for livestock populations is rapidly growing as the cost of genotyping is constantly reduced. Genotyping pigs began around 2009 when SNPs became commercially available. The most commonly used SNP chip in the pork industry is the Illumina PorcineSNP60 v2 BeadChip (http://www.illumina.com).

The most extensively adopted technique to incorporate genomic information in evaluations is single-step GBLUP (ssGBLUP), which allows incorporating information from SNP, pedigree relationships, and phenotypic data into a single model to obtain GEBV (Aguilar et al., 2010; Christensen and Lund, 2010). For ssGBLUP implementation, the inverse of **H** ($\mathbf{H}^{-1}$) is used instead of the inverse of **A** ($\mathbf{A}^{-1}$) in the mixed model equations, where **A** is the pedigree-based relationship matrix, and $\mathbf{H}^{-1}$ is:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (1)$$

where subscript 22 refers to genotyped animals, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix (VanRaden, 2008). Since its development in 2009, many livestock breeding programs have implemented ssGBLUP into their routine evaluations (Lourenco et al., 2020).

Genomic information is used in genomic evaluation for quantifying relationships between animals and estimating marker effects. The resulting GEBV from genomic evaluations explains each animal's merit by the accumulated effects from SNP marker information, pedigree relationships, and phenotypic data (Schaeffer, 2006). Animal breeders use GEBV to select the next generations' parents and increase the rate of genetic change over time. The incorporation of GEBV has increased accuracy, decreased generation intervals, and dramatically reduced cost and time for progeny testing (Schaeffer, 2006).

Pedigree tracking is necessary for the use of ssGBLUP and is usually from recording visual or electronic animal identification information (such as ear tags or microchips), which is

prone to substantial human error. Genotyping can more accurately track parentage and save farmers time and labor costs on administering and reading the traditional identification methods, and in return, they will have more genotyped animals to include in genomic evaluations. Using pooled semen for swine breeding is commonly practiced and increases reproduction rates (Maiorano et al., 2019); however, the sires are unknown when this method is used. Through genotyping for parentage, producers will continue to have optimal reproduction rates and identify which boar sired each animal. As more animals are being genotyped and genomic prediction methodology is becoming more advanced, the rate of genetic gain is increasing (Misztal et al., 2020).

ACCURACY OF PREDICITONS

Accuracy is one of the essential metrics in genetic evaluations. Validation metrics for models test how well the performance of future animals is predicted. The magnitude of accuracy is most often discussed when comparing methods. Traditionally the accuracy of (G)EBV is defined as the correlation between true and estimated BV or the variance of prediction error; however, true BV are not available for real data. As accuracy approaches one, associations between true BV and EBV become more substantial, and similarly, associations become weaker as accuracy values approach zero.

The response to selection and rate of genetic change are functions of accuracy. Breeders want to use the model with the highest accuracy to observe selection results faster. Several methods of computing accuracy exist. Breeders choose the appropriate method based on the phenotype of interest, the heritability of the trait, or the dynamics of the breeding program for a specific species. Dairy cattle breeders typically use daughter yield deviations (DYD) or

deregressed proofs (DRP) (VanRaden et al., 2009). This method is necessary to obtain (G)EBV from bulls to produce females with high milk yield. Since bulls cannot obtain this phenotype, this method is necessary to select bulls and obtain an estimator of accuracy. This method is not necessary or useful for phenotypes that are not sex-limited. A commonly used measure of accuracy is predictive ability, which is computed by adjusting the phenotype by the fixed effects' estimates and correlating with (G)EBV with phenotypes removed for the validation animals (Legarra et al., 2008). This method is challenging to use and does not output logical accuracies for models with multiple random effects or traits of low heritabilities.

Legarra and Reverter (2018) have developed a method, the LR method, to calculate validation statistics for complex models and traits. This method uses a full and reduced dataset to estimate bias, dispersion, and accuracy. Bermann et al. (2021) tested the LR method on a lowly heritable binary trait and found consistent and logical accuracy estimates. The LR accuracy is computed by: $\hat{\rho}_{\mathrm{cov(whole,partial)}} = \sqrt{\frac{\mathrm{cov}(\widehat{\boldsymbol{u}}_{whole},\widehat{\boldsymbol{u}}_{partial})}{(1-\bar{\mathrm{F}})\hat{\sigma}_u^2}}$; where $\widehat{\boldsymbol{u}}_{whole}$ and $\widehat{\boldsymbol{u}}_{partial}$ are the (G)EBV for the whole dataset and dataset with phenotypes removed from validation animals, $\bar{\mathrm{F}}$ is the average inbreeding for the validation animals, and $\hat{\sigma}_u^2$ is the additive genetic variance for the whole population. The LR method is promising for increasing the genetic gain of many economically important traits in the pig industry. With the elaborate breeding systems and objectives in the commercial swine industry, breeders hope to maximize accuracy and, consequently, genetic gain.

The persistence of accuracy depends on the decay of linkage disequilibrium and the genetic relatedness among the animals (Habier et al., 2007). Due to the decay of linkage disequilibrium over time, the recent generations are more genetically different than the most ancestral generations. Previous studies conducted in mice show that distant relatives' information

results in lower accuracies than closely related individuals (Legarra et al., 2008). Each parent can explain 50% of the genetic variation in their progeny, but this is continuously reduced by half as generations proceed.

It is common to calculate accuracies by dividing the dataset into a training and validation set. The training set fits the model on the data, and the validation set has information removed to test the model. It is practicable to have old animals in the training set and young animals in the validation set to mimic population structures by resembling successive genetic evaluations. These two datasets are referred to as the "whole" and "partial" datasets in Legarra and Reverter (2018). This strategy resembles the real population scenarios in which young animals have pedigree information or are genotyped and do not have phenotypes.

The additive genetic relationship between the reference and validation population affects the accuracy of GEBVs in the validation population (Habier et al., 2007). A limited number of independent chromosome segments explains a population's additive genetic information (Pocrnic et al., 2016a). As the amount of additive information explained increases, accuracy increases. To accurately estimate the effects of the independent chromosome segments (M_e), linkage disequilibrium (LD) must exist between single nucleotide polymorphisms (SNP) and quantitative trait loci (QTL), which is the basis of genomic selection (Meuwissen et al., 2001).

Maximizing accuracy for genomic predictions is a function of optimizing the reference population's size, which depends on the dimensionality of the genomic relationship matrix and the effective population size (N_e) (Pocrnic et al., 2016a). Therefore, if enough information exists to explain the effects of the independent chromosome segments, the additive genetic variance can be explained, and accuracies will be adequate (Misztal, 2016). The number of independent

chromosome segments can be estimated by finding the number of eigenvalues in the genomic relationship matrix that explain most of the variation (Pocrnic et al., 2016a). The number of eigenvalues that obtain most of the variation in the population insinuates if the population is too small or diverse or if there is redundant information in the genomic relationship matrix. Regarding commercial pig populations, Pocrnic et al. (2016b) found that approximately 5,000 independent chromosome segments are a sufficient amount of information to obtain 98% of variance explained in the genomic relationship matrix. With more records and more independent chromosome segments explained, accuracy increases.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. Journal of Dairy Science 93(2):743-752. doi: https://doi.org/10.3168/jds.2009-2730

Bergsma, R., E. Kanis, E. F. Knol, and P. Bijma. 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics 178(3):1559-1570. doi: 10.1534/genetics.107.084236

Bermann, M., A. Legarra, M. K. Hollifield, Y. Masuda, D. Lourenco, and I. Misztal. 2021. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. J Anim Breed Genet 138(1):4-13. doi: 10.1111/jbg.12507

Bijma, P., W. M. Muir, E. D. Ellen, J. B. Wolf, and J. A. Van Arendonk. 2007a. Multilevel selection 2: Estimating the genetic parameters determining inheritance and response to selection. Genetics 175(1):289-299. doi: 10.1534/genetics.106.062729

Bijma, P., W. M. Muir, and J. A. Van Arendonk. 2007b. Multilevel selection 1: Quantitative genetics of inheritance and response to selection. Genetics 175(1):277-288. doi: 10.1534/genetics.106.062711

Chen, C. Y., R. K. Johnson, S. Newman, S. D. Kachman, and L. D. Van Vleck. 2009. Effects of social interactions on empirical responses to selection for average daily gain of boars. J Anim Sci 87(3):844-849. doi: 10.2527/jas.2008-0937

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genetics Selection Evolution 42(1):2. doi: 10.1186/1297-9686-42-2

de Roos, A. P., C. Schrooten, R. F. Veerkamp, and J. A. van Arendonk. 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J Dairy Sci 94(3):1559-1567. doi: 10.3168/jds.2010-3354

Dufrasne, M., I. Misztal, S. Tsuruta, J. Holl, K. Gray, and N. Gengler. 2013. Estimation of genetic parameters for birth weight, preweaning mortality, and hot carcass weight of crossbred pigs. Journal of animal science 91doi: 10.2527/jas.2013-6684

Fix, J. S., J. P. Cassady, W. O. Herring, J. W. Holl, M. S. Culbertson, and M. T. See. 2010. Effect of piglet birth weight on body weight, growth, backfat, and longissimus muscle area of commercial market swine. Livestock Science 127(1):51-59. doi: https://doi.org/10.1016/j.livsci.2009.08.007

Grandinson, K., M. S. Lund, L. Rydhmer, and E. Strandberg. 2002. Genetic Parameters for the Piglet Mortality Traits Crushing, Stillbirth and Total Mortality, and their Relation to Birth

Weight. Acta Agriculturae Scandinavica, Section A — Animal Science 52(4):167-173. doi: 10.1080/090647002762381041

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. p 2389. GENETICS SOCIETY OF AMERICA, United States.

Knol, E. F., B. Nielsen, and P. W. Knap. 2016. Genomic selection in commercial pig breeding. Animal Frontiers 6(1):15-22. doi: 10.2527/af.2016-0003

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genetics Selection Evolution 50(1):1-18. (article) doi: 10.1186/s12711-018-0426-6

Legarra, A., C. Robert-Granié, E. Manfredi, and J.-M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180(1):611-618. doi: 10.1534/genetics.108.088575

Lourenco, D., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguilar, and I. Misztal. 2020. Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. Genes (Basel) 11(7)doi: 10.3390/genes11070790

Lutaaya, E., I. Misztal, J. W. Mabry, T. Short, H. H. Timm, and R. Holzbauer. 2001. Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model. Journal of Animal Science 79(12):3002-3007. doi: 10.2527/2001.79123002x

Maiorano, A. M., A. Assen, P. Bijma, C.-Y. Chen, J. A. I. V. Silva, W. O. Herring, S. Tsuruta, I. Misztal, and D. A. L. Lourenco. 2019. Improving accuracy of direct and maternal genetic effects in genomic evaluations using pooled boar semen: a simulation study1. Journal of Animal Science 97(8):3237-3245. doi: 10.1093/jas/skz207

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. p 1819. WAVERLY PRESS INC, United States.

Misztal, I. 2016. Inexpensive Computation of the Inverse of the Genomic Relationship Matrix in Populations with Small Effective Population Size. Genetics 202(2):401-409. doi: 10.1534/genetics.115.182089

Misztal, I., D. Lourenco, and A. Legarra. 2020. Current status of genomic evaluation. Journal of Animal Science 98(4)doi: 10.1093/jas/skaa101

Muir, W., and A. Schinckel. 2002. Incorporation of competitive effects in breeding programs to improve productivity and animal well being

Pocrnic, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The Dimensionality of Genomic Information and Its Effect on Genomic Prediction. Genetics 203(1):573-581. doi: 10.1534/genetics.116.187013

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. Genetics Selection Evolution 48(1):82. doi: 10.1186/s12711-016-0261-6

Quiniou, N., J. Dagorn, and D. Gaudré. 2002. Variation of piglets' birth weight and consequences on subsequent performance. Livestock Production Science 78(1):63-70. doi: https://doi.org/10.1016/S0301-6226(02)00181-1

Roehe, R. 1999. Genetic determination of individual birth weight and its association with sow productivity traits using Bayesian analyses. Journal of Animal Science 77(2):330-343. doi: 10.2527/1999.772330x

Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P.-Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, D. Altshuler, S. N. P. M. W. G. The International, L. Cold Spring Harbor, I. National Center for Biotechnology, C. The Sanger, L. Washington University in St, and M. I. T. C. f. G. R. Whitehead. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409(6822):928-933. doi: 10.1038/35057149

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123(4):218-223. doi: 10.1111/j.1439-0388.2006.00595.x

Serenius, T., M. l. Sevón-aimonen, A. Kause, E. A. Mäntysaari, and A. Mäki-tanila. 2004. Selection potential of different prolificacy traits in the finnish landrace and large white populations. Acta Agriculturae Scandinavica, Section A — Animal Science 54(1):36-43. doi: 10.1080/09064700310019082

Serenius, T., and K. J. Stalder. 2006. Selection for sow longevity1,2. Journal of Animal Science 84(suppl_13):E166-E171. doi: 10.2527/2006.8413_supplE166x

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414-4423. doi: 10.3168/jds.2007-0980

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for

North American Holstein bulls. Journal of Dairy Science 92(1):16-24. doi: https://doi.org/10.3168/jds.2008-1514

CHAPTER 2

DETERMINING THE STABILITY OF ACCURACY OF GENOMIC ESTIMATED BREEDING VALUES IN FUTURE GENERATIONS IN COMMERCIAL PIG POPULATIONS [1]

[1] Hollifield, M.K., D. Lourenco, M. Bermann, J. T. Howard, and I. Misztal. 2021. *Journal of Animal Science*. 99(4). Reprinted here with permission of publisher.

ABSTRACT

Genomic information has a limited dimensionality (number of independent chromosome segments [M_e]) related to the effective population size. Under the additive model, the persistence of genomic accuracies over generations should be high when the nongenomic information (pedigree and phenotypes) is equivalent to M_e animals with high accuracy. The objective of this study was to evaluate the decay in accuracy over time and to compare the magnitude of decay with varying quantities of data, and with traits of low and moderate heritability. The dataset included 161,897 phenotypic records for a growth trait (GT) and 27,669 phenotypic records for a fitness trait related to prolificacy (FT) in a population with dimensionality around 5,000. The pedigree included 404,979 animals from 2008 to 2020, of which 55,118 were genotyped. Two single-trait models were used with all ancestral data and sliding subsets of 3-, 2-, and 1-generation intervals. Single-step genomic best linear unbiased prediction (ssGBLUP) was used to compute genomic estimated breeding values (GEBV). Estimated accuracies were calculated by the linear regression (LR) method. The validation population consisted of single generations succeeding the training population and continued forward for all generations available. The average accuracy for the first generation after training with all ancestral data was 0.69 and 0.46 for GT and FT, respectively. The average decay in accuracy from the first generation after training to generation 9 was -0.13, and -0.19 for GT and FT, respectively. The persistence of accuracy improves with more data. Old data has a limited impact on predictions for young animals for a trait with a large amount of information but a bigger impact for a trait with less information.

INTRODUCTION

The addition of genomic information to routine genetic evaluations reduced generation interval and increased the accuracy of genomic estimated breeding value (GEBV), defined as the correlation between true and estimated breeding values (VanRaden, 2009). These factors are the main forces driving the increase in the rate of genetic gain over time (VanRaden, 2008; García-Ruiz et al., 2016). Genomic information helps to identify the best young animals accurately even before phenotypes are recorded; therefore, it is of interest to determine the accuracy of GEBV for generations without new data recording and the magnitude of decay of accuracy over time. The selection of novel traits and traits difficult to measure is mainly dependent on the accuracies of GEBV. For example, milking speed and temperament have shown promising genetic progress due to genomics (Chen et al., 2020). Initial studies in genomic selection showed great persistence in the accuracy of genomic predictions over time. Results from Meuwissen et al. (2001) showed marginal decay in accuracy with a decrease from 0.84 to 0.72 over five new generations without phenotypes. This created initial excitement for the potential of selection with genomic information; however, the parameters of the simulated population cannot be compared with present-day commercial livestock populations. In the simulation, there was no selection, and only a few major genes explained the additive genetic variance of the trait. Under strong selection, steep decay in accuracy occurs (Muir, 2007). In small, simulated populations, Muir (2007) found that the accuracy of GEBV decays more rapidly than expected when under strong selection compared to random selection.

We hypothesize that the decay will be minimized even under selection if enough phenotypes and genotypes are available to represent the population structure. The reason is that a

limited number of independent chromosome segments (M_e) theoretically explains the additive genetic variance in a population (Pocrnic et al., 2016a). Therefore, if enough information exists to precisely estimate the effects of M_e, the additive genetic variance can be explained, and accuracies will be adequate and stable over time. The number of M_e is dependent on the effective population size (N_e) and genome length (L) (Stam, 1980). Pocrnic et al. (2016a) showed that the optimal amount of M_e can be estimated by computing the number of eigenvalues that explain a certain proportion of variation in the genomic relationship matrix (GRM), which is used in GBLUP (VanRaden, 2008) and ssGBLUP (Aguilar et al., 2010). This creates a threshold for the amount of information that is nonredundant, that is, information that can increase accuracy, and the amount of which new data no longer increases accuracy. Hence, the GRM has a limited dimension. Whereas N_eL eigenvalues explain most information, no new information is added after $4N_eL$ (Stam, 1980; Pocrnic et al., 2016a). Goddard (2009) showed that accuracy is inversely related to N_e. As N_e increases, accuracy decreases. It is estimated that genome lengths for pigs range from 18 – 23 Morgan (Rohrer et al., 1994; Archibald et al., 1995; Marklund et al., 1996; Tortereau et al., 2012), and N_e range from 55 – 113 (Welsh CS, 2009; Uimari and Tapio, 2011; Pocrnic et al., 2016b). Pocrnic et al. (2016b) found that 5000 segments explain approximately 98% of the variation in commercial pig populations. With enough data relative to the independent chromosome segments, high accuracy could be achieved. Additionally, if the segments are well estimated, there should be less decay of predictivity under the additive model even under selection.

The inverse of the genomic relationship matrix can be obtained by recursion on a group of animals (Faux et al., 2012; Misztal, 2014), with the optimal group size equal to the dimensionality of the genomic information (Misztal, 2016). The recursion means that the

breeding value of any animal can be estimated with near-perfect accuracy from exact breeding values of $4N_eL$ other animals. Bradford et al. (2017) showed by simulation that the accuracy of GEBV was the same whether the recursion was based on animals from the last generation or a distant generation. Their results suggest that, under the additive model, the persistence of genomic evaluations is very high if the reference population includes $4N_eL$ animals with high accuracy or equivalent.

Although accuracy is dependent on the proportion of variance explained by the eigenvalues of the GRM, the distribution of eigenvalues is not consistent, and a small percentage of the largest eigenvalues explain the majority of the genetic variation (Pocrnic et al., 2019). Additionally, the animals necessary to explain the largest eigenvalues carry almost the same genomic information. Hence, selection by GBLUP-based models occurs on clusters of independent chromosome segments, not individual chromosome segments (Pocrnic et al., 2019). In pig populations, the segments can be well estimated if there are around 5000 animals available with very high accuracy (e.g., theoretical EBV accuracy based on prediction error variance) or an equivalent number of animals with less accuracy. Despite a large amount of data available, the decay will be more dramatic if genomic selection induces faster epistatic changes (Huang and Mackay, 2016). Epistatic interactions between genes may reduce the value of old data, and epistatic effects may be unstable across populations because of the fluctuation in allele frequencies (Varona et al., 2018).

With the commercial pig production systems and population structure, the N_e and the M_e are small. The purpose of this study is to determine how accuracy and the decay in accuracy are affected by the quantity of data available, the heritability of the trait, and removing data from

ancestral generations. With genotypes now available for many generations in pigs, reliable predictions for generations without new phenotype recordings may be possible.

MATERIALS AND METHODS

DATA

Data for animals born between 2008 to 2020 were provided by Smithfield Premium Genetics (Rose Hill, NC). The population consisted of 273,382 animals, of which 55,118 were genotyped or imputed to the 50k SNP panel for autosomal markers only. Quality control removed SNP with minor allele frequency lower than 0.05, SNP and animals with call rates lower than 0.9, SNP with the difference between expected and observed frequency of heterozygous greater than 0.15 (departure from the Hardy-Weinberg equilibrium), and animals with parent-progeny Mendelian conflicts. After quality control, 39,263 SNP remained for 53,147 genotyped animals.

The dataset consisted of 27,669 records for a repeated fitness trait related to prolificacy (FT) from 13,883 animals and 161,495 records for a single growth trait (GT). The population consisted of 11 generations. Generations were constructed by tracing the population back to the oldest animals with no recorded parents. These animals were considered generation 1, and their progeny, grand-progeny, great-grand-progeny, were placed in generations 2, 3, and 4, respectively, and continued until generation 11. The birth year of the animals without parent records was considered when joining the successions to be more precise and to account for the age variation of animals without parent records. Table 2.1 shows the number of animals with genotypes, phenotypes, and pedigree per generation.

MODEL AND ANALYSIS

Variance components were estimated using AIREMLF90 (Misztal, 2014) without genomic information. The heritabilities were 0.21 and 0.06 for GT and FT, respectively, with standard errors less than 0.01. GEBV were computed using single-step genomic BLUP (ssGBLUP) (Aguilar et al., 2010). Two single-trait models were used in the analyses:

$$\mathbf{y}_{GT} = \mathbf{X}_{GT}\mathbf{b}_{GT} + \mathbf{Z}\boldsymbol{u}_{GT} + \mathbf{W}_1\boldsymbol{cl}_{GT} + \boldsymbol{e}_{GT} \quad (1)$$

$$\mathbf{y}_{FT} = \mathbf{X}_{FT}\mathbf{b}_{FT} + \mathbf{Z}\boldsymbol{u}_{FT} + \mathbf{W}_2\boldsymbol{pe}_{FT} + \boldsymbol{e}_{FT}\,, \quad (2)$$

where $\mathbf{y}_{GT}$ is a vector of GT observations; $\mathbf{b}_{GT}$ is a fixed vector of systematic effects including contemporary group (farm, year, and week of birth), sex, and age in days at recording; $\boldsymbol{u}_{GT}$ and $\boldsymbol{cl}_{GT}$ are random vectors of direct additive genetic and common litter effects, respectively. Elements of $\mathbf{y}_{GT}$ are related to elements of $\boldsymbol{cl}_{GT}$ by the incidence matrix $\mathbf{W}_1$. The $\mathbf{y}_{FT}$ is a vector of FT observations; $\mathbf{b}_{FT}$ is a fixed vector of systematic effects including contemporary group (farm, year, and month of birth) and parity; $\boldsymbol{u}_{FT}$ and $\boldsymbol{pe}_{FT}$ are random vectors of direct additive genetic and permanent environmental effects, respectively. Elements of $\mathbf{y}_{FT}$ are related to elements of $\boldsymbol{pe}_{FT}$ by the incidence matrix $\mathbf{W}_2$. In both models, $\mathbf{X}$, $\mathbf{Z}$, are incidence matrices relating elements of $\mathbf{y}$ to $\mathbf{b}$, and $\boldsymbol{u}$, respectively, and $\boldsymbol{e}$ is a vector of random residuals. The covariance matrices were assumed to be:

$$Var\begin{bmatrix}\boldsymbol{u}_{GT}\\ \boldsymbol{cl}_{GT}\\ \boldsymbol{e}_{GT}\end{bmatrix} = \begin{bmatrix}\mathbf{H}\sigma^2_{uGT} & 0 & 0\\ 0 & \mathbf{I}\sigma^2_{cl} & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{eGT}\end{bmatrix} \quad (3)$$

$$Var\begin{bmatrix}\boldsymbol{u}_{FT}\\ \boldsymbol{pe}_{FT}\\ \boldsymbol{e}_{FT}\end{bmatrix} = \begin{bmatrix}\mathbf{H}\sigma^2_{uFT} & 0 & 0\\ 0 & \mathbf{I}\sigma^2_{pe} & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{eFT}\end{bmatrix}, \quad (4)$$

where σ^2_{uGT} and σ^2_{uFT} are variances for additive genetic effects for GT and FT, respectively; σ^2_{cl} is the variance for the common litter effect; σ^2_{pe} is the variance for the permanent environmental effect; σ^2_{eGT} and σ^2_{eFT} are the variances for the residual effects for GT and FT, respectively; **I** is the identity matrix; **H** is a matrix combining pedigree and genomic relationships among animals as applied in ssGBLUP (Aguilar et al., 2010). The inverse of the pedigree-based relationship matrix ($\mathbf{A}^{-1}$) is replaced by the inverse of **H** ($\mathbf{H}^{-1}$) in the ssGBLUP mixed model equations, which is written as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (5)$$

where **G** was constructed using the first method of VanRanden (2008), then 95% of **G** was blended with 5% of the pedigree relationship matrix for genotyped animals ($\mathbf{A}_{22}$), and finally tuned so the means of the diagonal and off-diagonal elements were similar to those of $\mathbf{A}_{22}$ (Chen et al., 2011). The allele frequencies used to compute **G** were calculated based on all genotyped animals in the dataset.

In this study, the accuracy and dispersion of GEBV were estimated with the linear regression (LR) method (Legarra and Reverter, 2018). This method uses two datasets, namely the *whole* dataset and the *partial* dataset, hereinafter denoted with the subscripts *w* and *p*, respectively. The former contains all the available phenotypes up to a certain time *t*, whereas the latter contains phenotypes up to a time period before *t*. The focal individuals, that is, the individuals for whom the accuracy of GEBV will be estimated, are defined as the genotyped animals with phenotypes in the whole dataset but without in the partial dataset.

To investigate the impact of the amount of data on the accuracy of GEBV for focal individuals, GEBV were sequentially estimated by changing the definition of focal individuals

and partial datasets using a sliding approach based on generation. Figure 2.1 shows four definitions of focal groups that included generations 5 to 9, 6 to 9, 7 to 9, and 8 to 9. Accuracy and dispersion were then calculated separately for each generation of focal individuals. Additionally, to investigate the impact of ancestral data, four partial datasets were created for each focal group: (i) the *ancestral group*: contained all the ancestors of the focal individuals, (ii) the *3-generation group*: consisted of the ancestors up to the great-grandparents of the focal individuals, (iii) the *2-generation group*: included the grandparents and parents of the focal individuals, and (iv) the *1-generation group*: contained only the parents of the focal individuals. A total of 16 different combination of groups of focal individuals and partial datasets were created (Figure 2.1).

The benchmark for each validation, i.e., $GEBV_w$, remained unchanged, whereas $GEBV_p$ were updated as the partial datasets were modified. Due to the lack of phenotypes and genotypes in generations 10 and 11, these animals were removed from all analyses as they were incomparable with the other validation generations. Accuracies were estimated for each generation in each set of focal individuals using: $\hat{\rho}_{cov(w,p)} = \sqrt{\frac{cov(\hat{\boldsymbol{u}}_{\boldsymbol{w}},\hat{\boldsymbol{u}}_{\boldsymbol{p}})}{(1-\bar{F})\hat{\sigma}_u^2}}$ (Legarra and Reverter, 2018; Macedo et al., 2020b), where $\bar{F}$ is the average inbreeding coefficient among focal individuals in a specific generation and $\hat{\sigma}_u^2$ is the estimated additive genetic variance of the population. Inbreeding coefficients for each animal were calculated with a recursive method based on pedigree using INBUPGF90 (Aguilar and Misztal, 2008). The slope of the regression of $\hat{\boldsymbol{u}}_{\boldsymbol{w}}$ on $\hat{\boldsymbol{u}}_{\boldsymbol{p}}$, is used to assess the dispersion of partial GEBV and is equal to $b_{w,p} = \frac{cov(\hat{\boldsymbol{u}}_{\boldsymbol{w}},\hat{\boldsymbol{u}}_{\boldsymbol{p}})}{var(\hat{\boldsymbol{u}}_{\boldsymbol{p}})}$. The primary purpose of this research was to compare accuracies over time with varying amounts of ancestral data for two traits of differing heritabilities; therefore, other statistical parameters were

not used. Accuracy and dispersion are well-researched and logical to use as a function over time (Macedo et al., 2020a). Additional statistics proposed by the LR method have not been widely tested as a function of time. Including those values would output uninterpretable comparisons and should be further researched.

RESULTS AND DISCUSSION

Figures 2.2 and 2.3 show the accuracy for GT and FT over time using the partial datasets belonging to each group. When comparing traits, GT had higher accuracy and less decay in accuracy over time compared with FT. For example, when considering the partial dataset composed of generations 1 to 4 from the ancestral group, the accuracy decreased from 0.55 in generation 5 to 0.42 in generation 9 for GT (Figure 2.2A), and from 0.46 to 0.22 for FT (Figure 2.3A), respectively. These results are expected and agree with those from Muir (2007) since GT has higher heritability than FT, and low heritability traits require a large number of records to achieve high accuracy; FT had about 1/6th of the records compared to GT.

Persistence for both traits can be inferred by observing the initial and final accuracy for each line in Figures 2.2 and 2.3. The slopes for FT are greater in magnitude than the slopes for GT, meaning that the latter showed more persistence. The differences in persistence between the two traits may be explained by the heritability and the amount of phenotypic information. Roughly, the amount of information in this study can be approximated as accuracies of hypothetical 5000 ($4N_eL$) sires with as many progeny as the number of animals with records, and with progeny equally distributed per sire. For a trait with 32 progeny per sire and heritability of 0.21, the accuracy per sire would be approximately 0.80. For a trait with 5 progeny per sire and heritability of 0.06, the equivalent accuracy would be only 0.25.

The distance between different lines in Figures 2.2 and 2.3 show the impact that the different sources of information, namely parents, grandparents, etc. have in the estimation of the accuracy of GEBV. This fact can be observed for the focal individuals in generation 8 (Figures 2.2 and 2.3). In this case, the purple line includes the parents of the named focal individuals, whereas, for the blue line, the closest generation used to estimate their accuracies was that of their grandparents. When comparing the difference between both lines, it can be deduced that removing the parents drops the accuracy about 0.11, on average for GT, whereas the average drop for FT was about 0.04. To compare the two traits across time, the average decreases in accuracy for GT (FT) were 16.0% (10.1%) after removing parents and 79.3% (34.4%) after removing three generations (parents, grandparents, and great-grandparents).

The magnitude and slope of the regression of $\widehat{\boldsymbol{u}}_{\boldsymbol{w}}$ on $\widehat{\boldsymbol{u}}_{\boldsymbol{p}}$ overtime for both traits explains the effect of heritability and quantity of data on GEBV prediction. Regression coefficient less than one indicates the GEBV of the focal animals are over-dispersed (overestimated) compared to GEBV from the whole dataset. In Figure 2.4, the partial datasets include generations 1 through 4 for both traits. The partial datasets are not updated over time; therefore, the focal animals become less related to the partial datasets as generations proceed. In relation to animals in generation 4, the GEBV for focal animals were overestimated for progeny, grand-progeny, great-grand-progeny, great-great-grand-progeny, and great-great-great-grand-progeny, which are generations 5, 6, 7, 8, and 9, respectively. Analogously to accuracy, $b_{w,p}$ remained greater and more persistent over time for GT than FT. The $b_{w,p}$ decreased from 0.84 to 0.66 for GT from generations 5 and 9, respectively. Similarly, it decreased from 0.63 to 0.21 for FT. A steep negative trend for $b_{w,p}$ over time indicates there was not enough information available to predict the amount of dispersion in further generations. The differences in the persistence of accuracy

and dispersion confirm that for traits with low heritability, the impact of information from closely related individuals is less than traits with high heritability.

Apparently, this is subject to the fact that all chromosome segments are represented in the population (Pocrnic et al., 2016a). Thus, with sufficient genotyped animals, it is expected that chromosome segments would be well represented in the population. Consequently, the gain in accuracy when adding information from individuals more closely related will be minimal if the corresponding trait has low heritability. It is important to highlight that in this study, the accumulation of ancestors was considered a new source of information, not the addition of progeny of the focal individuals. Logically, the accuracy of GEBV for focal individuals will largely depend on the incorporation of their progeny in the genetic evaluation, regardless of the heritability of the trait and the representation of the chromosome segments in the population.

To maximize the accuracy of genomic predictions, an optimal size of the training population is necessary to capture most of the variation in the population. This optimal subset is theoretically related to a limited dimension of the genomic information. This limited dimension is a function of N_e and L. If ~ $4N_eL$ largest eigenvalues are contained in the GRM, the M_e is likely obtained, and ample information is provided to achieve high accuracies (Pocrnic et al., 2016a). According to Misztal (2016), each independent chromosome segment has an additive effect, and the sum of the effects of the existing chromosome segments in individual animals composes the breeding values. If enough chromosome segment effects are captured in the population, more variation is explained in the population, and thus, it is expected that accuracies will also show more persistence over time.

As explained in a study conducted by Hayes et al. (2009), the accuracy of genomic selection is crucially dependent on the number of phenotypic records available, and the

heritability of a trait. In their study, approximately 5000 phenotypes were required to achieve an accuracy of GEBV equal to 0.6 for a trait with a heritability of 0.2 in a population with an N_e of 1000. In our study, for FT, generation 6 and 7 contained 4278 and 3348 records, respectively. Compared to GT that had 26,474 records for generation 6 and 28,260 for generation 7, it can be concluded that FT does not have enough information to achieve an accuracy as high as GT. This can explain the lack of persistency and low accuracy over time when analyzing FT with 1-generation partial datasets. In every analysis for FT and GT, 2 or 3 generations of data seem sufficient enough to reach a comparable maximum accuracy to all ancestral data. As heritability decreases, the number of required phenotypic records to achieve the desired accuracy of GEBV increases (Hayes et al., 2009).

The selection pressure and complexity of a trait significantly affect the accuracy of GEBV over time (Muir, 2007; Gorjanc et al., 2015). In this study, different intensities and types of selection pressure were placed on the two separate traits. GT was heavily selected upon over time, and this trait was directly selected across all generations. FT, however, was only indirectly selected, meaning that the selection pressure on FT depended on the selection pressure of a different trait with a more favorable relationship with pre-weaning mortality. These differences in selection for both traits can be observed in Figure 2.5, where the genetic trends of GEBV across generations for both GT and FT are shown. To make both traits comparable, GEBV were standardized. As seen in the trends over time, GT increased at a steadier rate, whereas FT increased less directly, implying less selection. Also, FT is more challenging to select upon and predict its performance since it is a categorical trait, compared to the continuity of GT.

One important limitation of this is that the accuracy for generations that were distant from the reference populations was computed for preselected animals, and preselection decreases

realized accuracies (Bijma, 2012; Lourenco et al., 2015). Therefore, the future accuracies may be underestimated, although the LR method may partially account for the preselection.

The issue of persistence of GEBV is also important in the dairy industry where young bulls are selected from other young bulls only based on the genomic information. For Holsteins with a large amount of information and the genomic dimensionality around 15,000 (Pocrnic et al., 2016b), the reliability for production traits two generations ahead of the reference population was 90% of that of one generation ahead (VanRaden et al., 2010). If the persistence of the evaluations is high, the importance of phenotyping may be reduced. However, the persistence is likely to be lower for lower heritability traits, especially with fewer records, keeping phenotyping relevant. Additionally, in the long run, very strong selection and epistatic interactions may possibly reduce the persistence, keeping the need for phenotype recording.

CONCLUSIONS

When the reference population is large enough to accurately estimate the effects of the independent chromosome segments, GEBV can be persistent, with minimal decay of accuracy over generations. In such a case, the impact of old data is minimal. The decay is larger with less information, particularly for lower heritability traits, and with necessarily lower selection pressure, the impact of old data is likely larger. It would be desirable to estimate the decay as a function of many parameters analytically, however, the complexity of selection and side effects of faster selection (e.g., Bulmer effect and epistasis) are likely to make such a theory complex.

REFERENCES

Aguilar, I., and I. Misztal. 2008. Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. J Dairy Sci 91(4):1669-1672. doi: 10.3168/jds.2007-0575

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93(2):743-752. doi: https://doi.org/10.3168/jds.2009-2730

Archibald, A. L., C. S. Haley, J. F. Brown, S. Couperwhite, H. A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, A. K. Winterø, M. Fredholm, N. J. Larsen, V. H. Nielsen, D. Milan, N. Woloszyn, A. Robic, M. Dalens, J. Riquet, J. Gellin, J. C. Caritez, G. Burgaud, L. Ollivier, J. P. Bidanel, M. Vaiman, C. Renard, H. Geldermann, R. Davoli, D. Ruyter, E. J. M. Verstege, M. A. M. Groenen, W. Davies, B. Høyheim, A. Keiserud, L. Andersson, H. Ellegren, M. Johansson, L. Marklund, J. R. Miller, D. V. Anderson Dear, E. Signer, A. J. Jeffreys, C. Moran, P. Le Tissier, Muladno, M. F. Rothschild, C. K. Tuggle, D. Vaske, J. Helm, H. C. Liu, A. Rahman, T. P. Yu, R. G. Larson, and C. B. Schmitz. 1995. The PiGMaP consortium linkage map of the pig (*Sus scrofa*). Mammalian Genome 6(3):157-175. doi: 10.1007/BF00293008

Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. Journal of Animal Breeding and Genetics 129:345-358. doi: 10.1111/j.1439-0388.2012.00991.x

Bradford, H. L., I. Pocrnić, B. O. Fragomeni, D. A. L. Lourenco, and I. Misztal. 2017. Selection of core animals in the Algorithm for Proven and Young using a simulation model. Journal of Animal Breeding and Genetics 134(6):545-552. doi: https://doi.org/10.1111/jbg.12276

Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale1. Journal of Animal Science 89(9):2673-2679. doi: 10.2527/jas.2010-3555

Chen, S.-Y., H. R. Oliveira, F. S. Schenkel, V. B. Pedrosa, M. G. Melka, and L. F. Brito. 2020. Using imputed whole-genome sequence variants to uncover candidate mutations and genes affecting milking speed and temperament in Holstein cattle. Journal of Dairy Science 103(11):10383-10398. doi: https://doi.org/10.3168/jds.2020-18897

Faux, P., N. Gengler, and I. Misztal. 2012. A recursive algorithm for decomposition and creation of the inverse of the genomic relationship matrix. Journal of Dairy Science 95(10):6093-6102. doi: https://doi.org/10.3168/jds.2011-5249

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proceedings of the National Academy of Sciences 113(28):E3995-E4004. doi: 10.1073/pnas.1519061113

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136(2):245-257. doi: 10.1007/s10709-008-9308-0

Gorjanc, G., P. Bijma, and J. M. Hickey. 2015. Reliability of pedigree-based and genomic evaluations in selected populations. Genetics Selection Evolution 47(1):65. doi: 10.1186/s12711-015-0145-1

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92(2):433-443. doi: 10.3168/jds.2008-1646

Huang, W., and T. F. C. Mackay. 2016. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. PLoS genetics 12(11):e1006421. doi: 10.1371/journal.pgen.1006421

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genetics Selection Evolution 50(1):1-18. (article) doi: 10.1186/s12711-018-0426-6

Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A. Legarra, and I. Misztal. 2015. Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. Genetics Selection Evolution 47(1):56. doi: 10.1186/s12711-015-0137-1

Macedo, F. L., O. F. Christensen, J. M. Astruc, I. Aguilar, Y. Masuda, and A. Legarra. 2020a. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. Genet Sel Evol 52(1):47. doi: 10.1186/s12711-020-00567-1

Macedo, F. L., A. Reverter, and A. Legarra. 2020b. Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. Journal of Dairy Science 103(1):529-544. doi: https://doi.org/10.3168/jds.2019-16603

Marklund, L., M. J. Moller, R. K. Juneja, P. Mariani, H. Ellegren, L. Andersson, B. Høyheim, W. Davies, M. Fredholm, and W. Coppieters. 1996. A comprehensive linkage map of the

pig based on a wild pig - Large White intercross. Animal Genetics 27(4):255-269. doi: 10.1111/j.1365-2052.1996.tb00487.x

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819-1829.

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Aguilar, I., Legarra, A., & Vitezica, Z. . 2014. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.

Misztal, I. 2016. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics 202(2):401-409. doi: 10.1534/genetics.115.182089

Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. p 342. Blackwell Publishing Ltd, Germany.

Pocrnic, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203(1):573-581. doi: 10.1534/genetics.116.187013

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. Genetics Selection Evolution 48(1):82. doi: 10.1186/s12711-016-0261-6

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, and I. Misztal. 2019. Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest

eigenvalues: a simulation study. Genetics Selection Evolution 51(1):75. doi: 10.1186/s12711-019-0516-0

Rohrer, G. A., L. J. Alexander, J. W. Keele, T. P. Smith, and C. W. Beattie. 1994. A microsatellite linkage map of the porcine genome. Genetics 136(1):231.

Stam, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genetical Research 35(2):131-155. doi: 10.1017/S0016672300014002

Tortereau, F., B. Servin, L. Frantz, H.-J. Megens, D. Milan, G. Rohrer, R. Wiedmann, J. Beever, A. L. Archibald, L. B. Schook, and M. A. M. Groenen. 2012. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics 13(1):586. doi: 10.1186/1471-2164-13-586

Uimari, P., and M. Tapio. 2011. Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. Journal of Animal Science 89(3):609-614. doi: 10.2527/jas.2010-3249

VanRaden, P., J. O'Connell, G. R. Wiggans, and K. Weigel. 2010. Combining different marker densities in genomic evaluation. Interbull Bullet. 42

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414-4423. doi: 10.3168/jds.2007-0980

VanRaden, P. M., Van Tassel, C.P., Wiggians, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Schenkel, F.S. 2009. Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92(1):16-24.

Varona, L., A. Legarra, M. A. Toro, and Z. G. Vitezica. 2018. Non-additive Effects in Genomic Selection. Frontiers in Genetics 9(78)(Review) doi: 10.3389/fgene.2018.00078

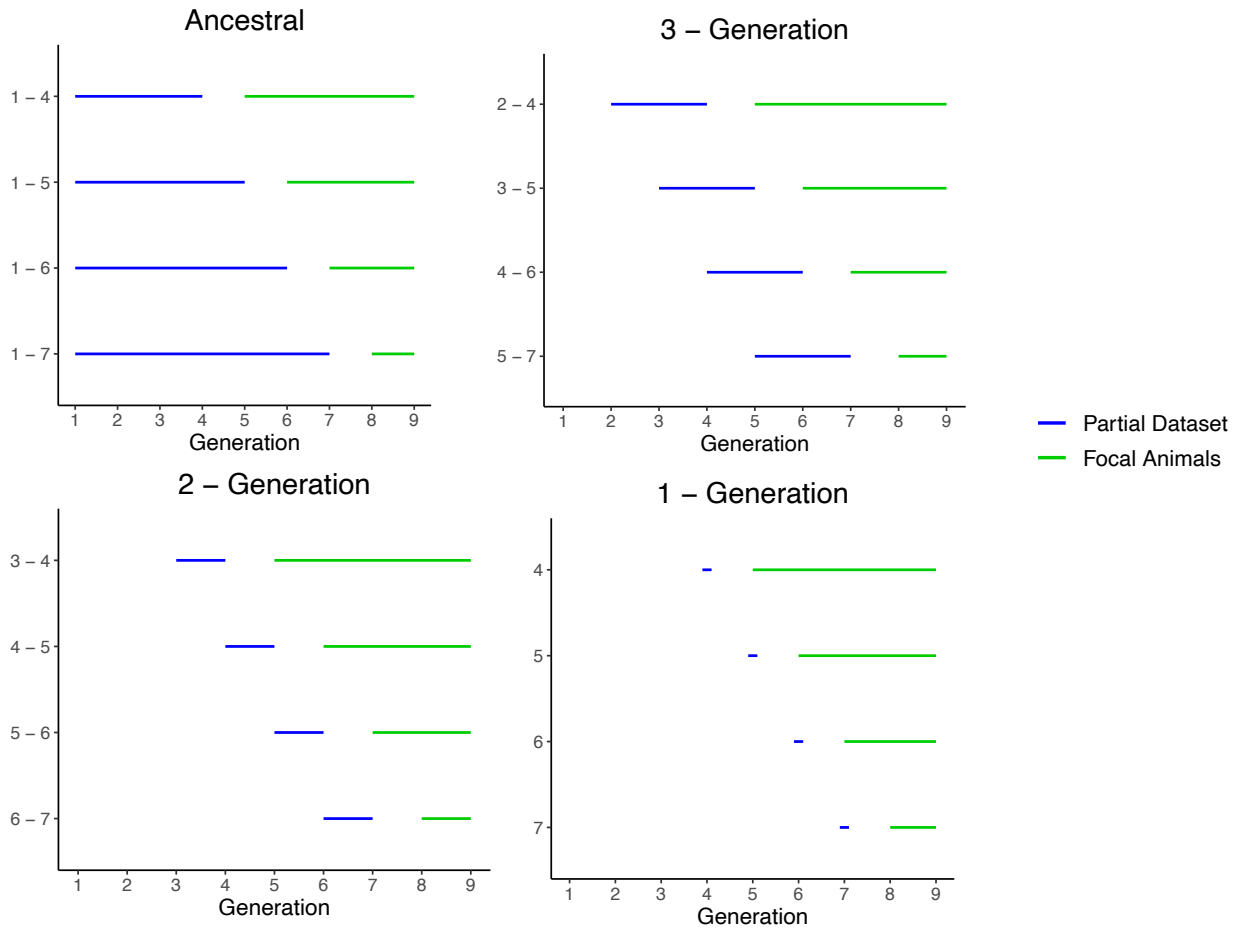
Welsh, C.S., H. D. Blackburn, and C. Schwab. 2009. Population status of major U.S. swine breeds. In: Proceedings of American Society of Animal Science Western Section, Fort Collins, CO

TABLES

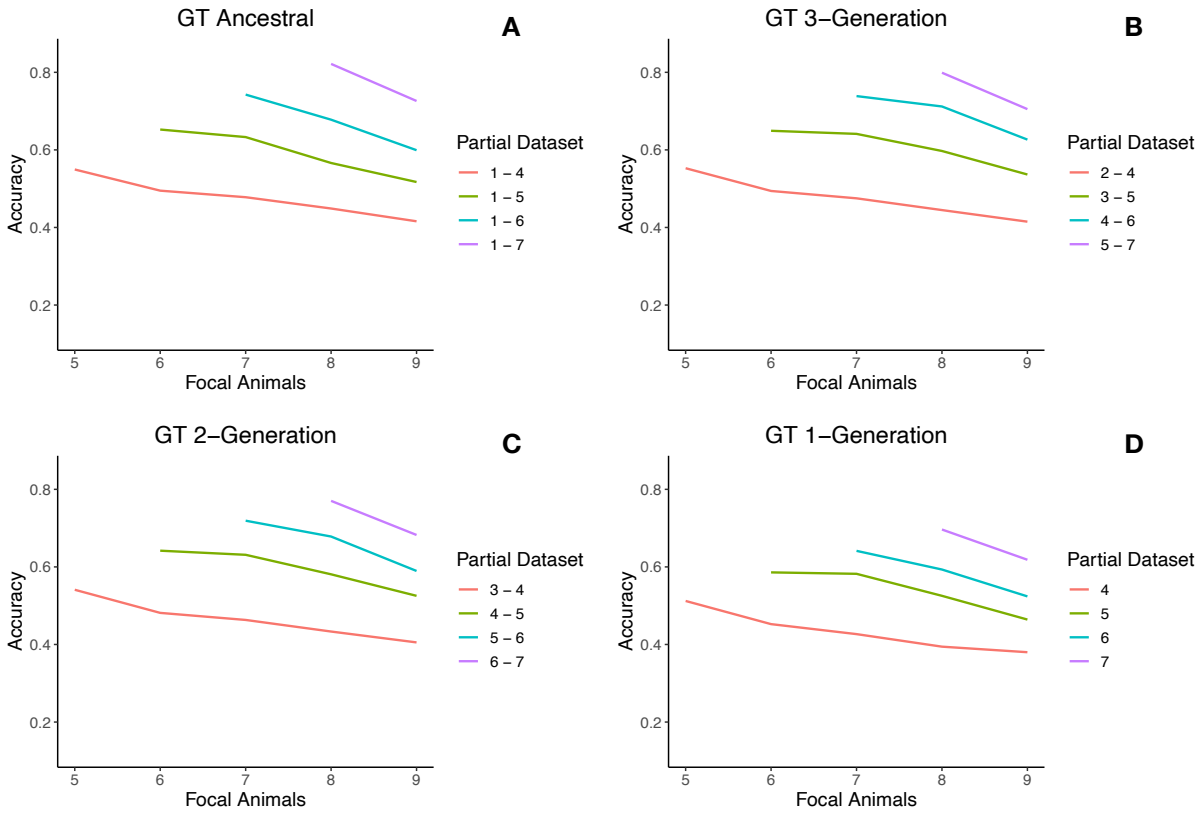
Table 2.1 Number of animals in the pedigree, genotyped animals, and records for GT and FT per generation.

Generation	Pedigree	Genotypes	GT	FT
1	758	214	658	1,991
2	12,513	384	4,767	2,098
3	15,190	831	7,697	3,447
4	29,017	1,929	16,491	3,753
5	38,316	2,775	23,211	4,302
6	42,476	6,158	26,474	4,278
7	44,363	10,769	28,260	3,348
8	39,082	11,345	25,002	2,290
9	27,445	8,636	16,989	1,435
10	17,084	6,149	8,762	570
11	7,138	3,957	3,184	157

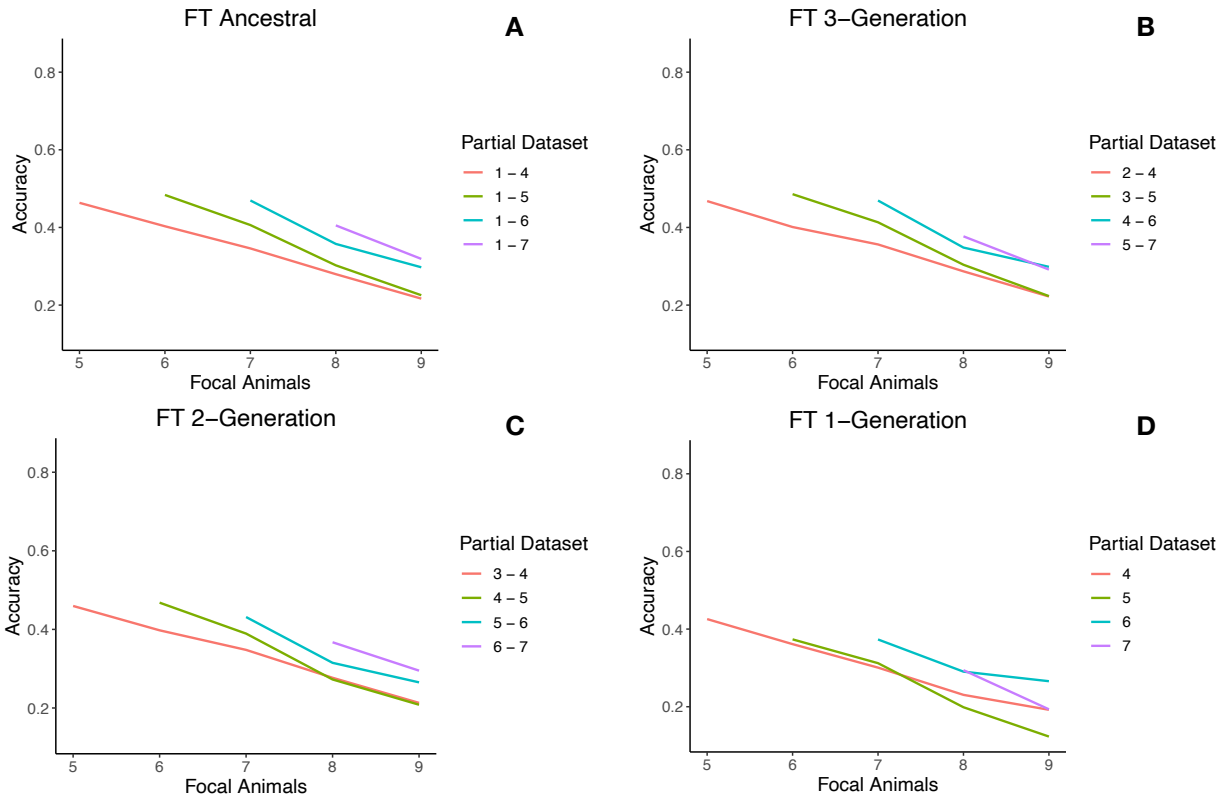
FIGURES



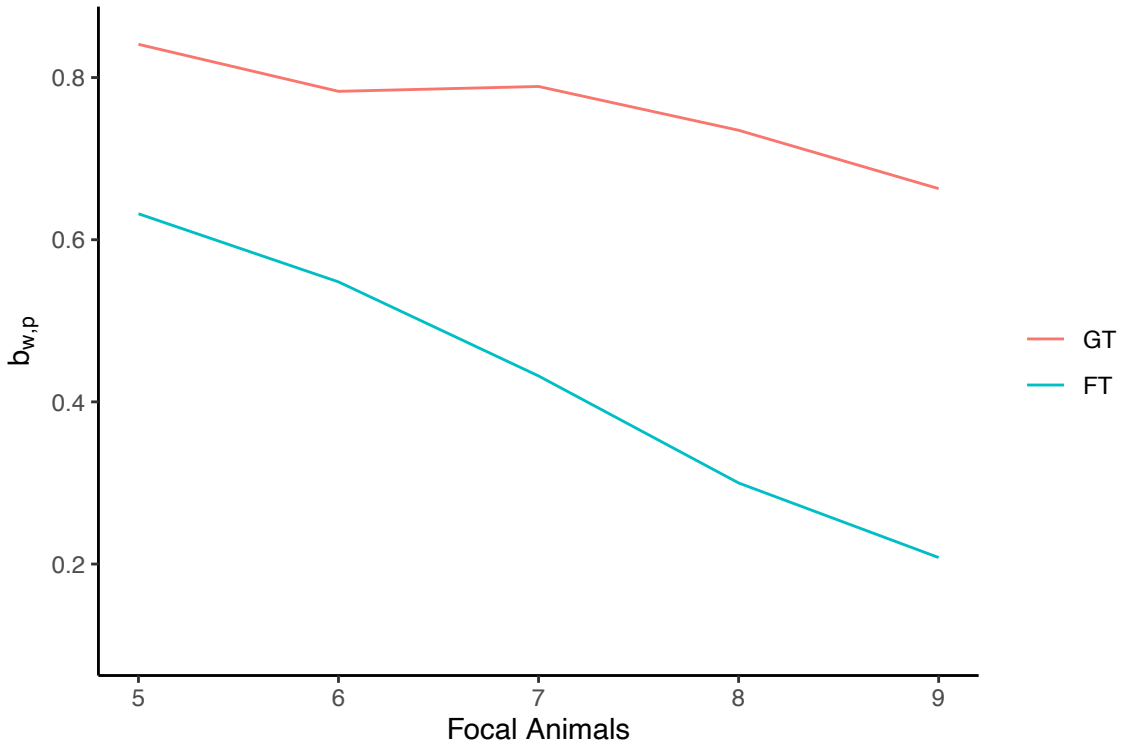
2.1 Scheme for partial datasets and focal animals. The four partial dataset groups include ancestral, 3-, 2-, and 1-generation subsets. In each scenario, the genomic and pedigree information is included for all animals and remain unchanged, but only phenotypes exist for animals in the partial dataset. Generations are not grouped for the focal animals, and accuracies are calculated for each generation separately.



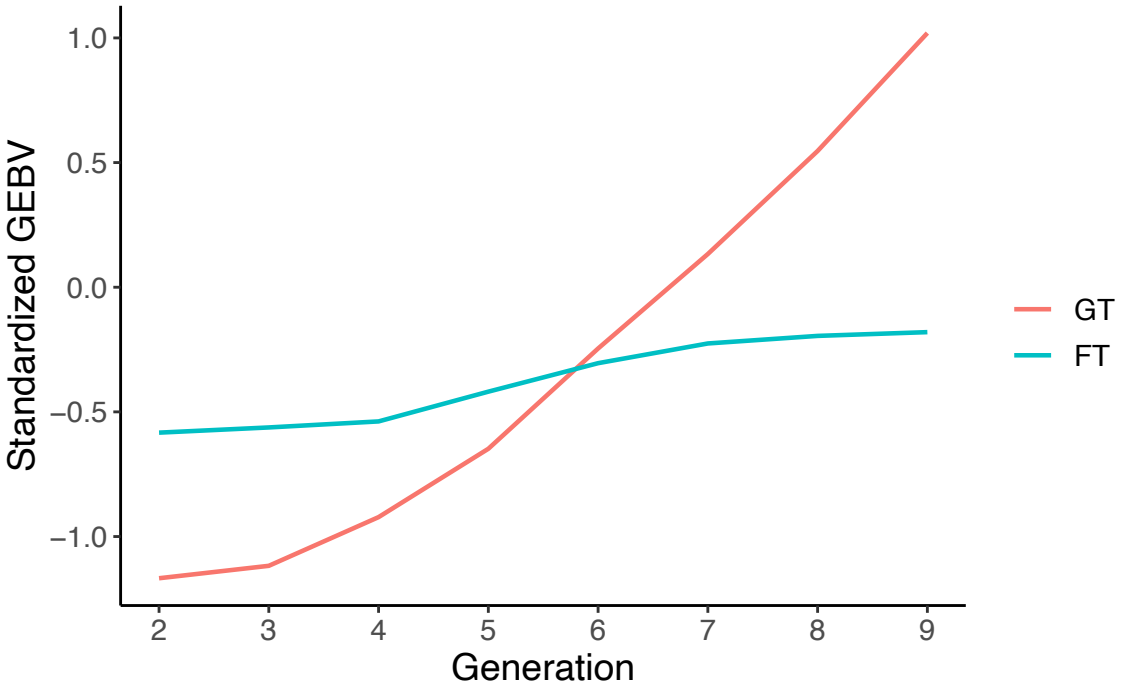
2.2 Accuracy over time with four partial dataset groups for GT. The partial datasets are updated over time, increasing a generation of data for the ancestral groups (A) and adding a recent generation of data while removing the oldest generation of data for 3-, 2-, and 1- generation subsets (B, C, and D, respectively). Accuracy is calculated for each generation separately, beginning with the first generation following the partial dataset and ending at generation 9.



2.3 Accuracy over time with four partial dataset groups for FT. The methods are the same as in Figure 2.2.



2.4 Dispersion trends over time for GT and FT. The partial datasets include ancestral data from generations 1-4 and are not updated over time. Each generation beyond generation 4 is a generation of focal animals becoming less related to the partial dataset animals. The slope of the regression of GEBV whole on GEBV partial ($b_{w,p}$) was used to estimate dispersion. Dispersion was calculated for each generation separately, beginning with generation 5 and ending at generation 9.



2.5 Genetic trends for GT and FT with average standardized GEBV. Generation 1 was excluded from the trend due to the lack of animals with phenotypic records.

CHAPTER 3

IMPACT OF INCLUDING THE CAUSE OF MISSING RECORDS ON GENETIC EVALUATIONS FOR GROWTH IN COMMERCIAL PIGS [1]

[1] Hollifield M. K., D. Lourenco, S. Tsuruta, M. Bermann, J. T. Howard, I. Misztal. Submitted to *Journal of Animal Science,* 04/23/2021.

ABSTRACT

It is of interest to evaluate crossbred pigs for hot carcass weight (HCW) and birth weight (BW); however, obtaining a HCW record is dependent on livability (LIV) and retained tag (RT). The purpose of this study is to analyze how HCW evaluations are affected when herd removal and missing identification are included in the model and examine if accounting for the reasons for missing traits improves the accuracy of predicting breeding values. Pedigree information was available for 1,965,077 purebred and crossbred animals. Records for 503,716 commercial three-way crossbred terminal animals from 2014 to 2019 were provided by Smithfield Premium Genetics. Two pedigree-based models were compared; model 1 (M1) was a threshold-linear model with all four traits (BW, HCW, RT, and LIV), and model 2 (M2) was a linear model including only BW and HCW. The fixed effects used in the model were contemporary group, sex, age at harvest (for HCW only), and dam parity. The random effects included direct additive genetic and random litter effects. Accuracy, dispersion, bias, and Pearson correlations were estimated using the linear regression method. The heritabilities were 0.11, 0.07, 0.02, and 0.04 for BW, HCW, RT, and LIV, respectively, with standard errors less than 0.01. No difference was observed in heritabilities or accuracies for BW and HCW between M1 and M2. Accuracies were 0.33, 0.37, 0.19, and 0.23 for BW, HCW, RT, and LIV respectively. The genetic correlation between BW and RT was 0.34 ± 0.03, and between BW and LIV was 0.56 ± 0.03. Similarly, the genetic correlation between HCW and RT was 0.26 ± 0.04, and between HCW and LIV was 0.09 ± 0.05, respectively. The positive and moderate genetic correlations between BW and other traits imply a heavier BW resulted in a higher probability of surviving to harvest. Genetic correlations

between HCW were lower due to the effects of the large quantity of missing records. Despite the heritable and correlated aspects of RT and LIV, results imply no major differences between M1 and M2; hence, it is unnecessary to include these traits in classical models for BW and HCW.

INTRODUCTION

Profitability for commercial pig breeding is contingent on optimizing all traits contributing to the economic value of the terminal line. Mortality and culling of animals are the most detrimental to financial gain. Many of the high economically valued traits, such as livability (LIV), have low heritabilities, resulting in a lengthy genetic progress. The occurrence of an animal not living to harvest can be accounted for in the evaluations by including a censored trait if death records are available (Arango et al., 2005b). If an animal dies or is removed from the herd, then its survivability record becomes uncensored. Active animals in the herd have censored survivability records (Schaeffer, 2019). Harvested animals that obtain a HCW measurement then have an uncensored record for survivability and HCW. To incorporate censored data in the analysis, the reason for death and the stage of life when the animal died must be recorded.

The growth and carcass traits are economically important, and breeders are continuously working to improve these rates of genetic gain. Because the rate of genetic process is slow at the commercial level in swine breeding, improving the model and individual identification methods will ultimately improve performance (Arango et al., 2005a). Selection for heavier birth weight (BW) is essential for commercial pig models as it leads to greater chances of LIV and faster growth rates (Grandinson et al., 2002; Arango et al., 2006). Previous studies have shown that it is possible to considerably improve several economically important traits by incorporating an

associative social interaction effect in the breeding program (Muir and Schinckel, 2002; Bijma et al., 2007; Bergsma et al., 2008).

Individual identification is essential for traceability, phenotype tracking, and advancing breeding programs. The identification device must be retained and readable throughout the entire process to record measurements from birth to slaughter. A feasible identification method would accommodate the systematic processes at commercial harvest and provide a logical cost-benefit return. Efforts for social interaction models require a reliable animal identification method and group information, so group mates and their indirect genetic effects can be identified. However, group information is not usually attained in most commercial pig operations, and the percentage of animals that lose the identification tag can be as high as 30%. Accounting for the reason animals were unable to obtain a HCW measurement may help overcome this issue and provide better estimates of hot carcass weight (HCW), given data were not available for some animals because of mortality and missing tags.

The objective of this study was to compare genetic parameters, correlations, and breeding values for BW and HCW in a two-trait model or a four-trait model that also accounted for retained tag (RT) and LIV records.

MATERIALS AND METHODS

DATA

Data were recorded from two farms for animals born between 2014 to 2019 and were provided by Smithfield Premium Genetics (Rose Hill, NC). The pedigree included 1,965,077 animals; however, phenotypes were only available for 503,716 commercial three-breed cross terminal animals. The phenotyped animal's dams were crossbred Landrace and Large White, and

sires were purebred Duroc. The traits included BW, HCW, and two binary traits, RT and LIV. All 503,716 animals used in the dataset had a BW record. There were 237,041 animals with a HCW measurement. Each farm brought their animals to a different harvest site, in which the instrumentation used to measure HCW may differ between sites. However, this potential difference is accounted for by including farms in the contemporary group.

RT and LIV traits consisted of reasons for the animals' inability to obtain a HCW record and were included in the model to analyze their effects on HCW evaluation. The RT categories were retained tag and non-retained tag and coded as 1 and 2, respectively. If an animal was missing its ear tag, the HCW trait was unobtainable, and death information was not recorded; thus, it is unknown if the animal was harvested. RT is treated as a success or failure based on if the animal retained its tag. Once an animal loses its tag, phenotypes can no longer be recorded for the remainder of its life. There were no data available indicating at which life stage an animal lost its tag.

LIV evaluates if the animal lived to be a full-value pig and was harvested or if the animal failed to live until harvest. A missing ear tag is considered a missing record for LIV since it is unknown if the animal made it to harvest or was removed from the herd before harvest, and the animal could not obtain a HCW record. A total of 11,013 animals survived to harvest, retained their ear tag, but did not have a HCW measurement. This could be due to scale malfunction, errors in pig identification after initial processing, etc. Table 1 includes the number and proportion of animals that have each trait and level.

The dataset included 471,360 animals after editing. Summary statistics for all continuous traits and effects after editing are in Table 2. Records were discarded for all animals born in 2014 due to the lack of LIV phenotypes. Animals in contemporary groups containing less than ten

animals were also excluded from the dataset. Contemporary groups were composed of farm, week, and year of birth. Group or pen information was not recorded and cannot be included in the contemporary groups. All animals were identified by a unique identification number on a plastic ear tag administered at birth.

MODEL AND ANALYSES

Pedigree-based analyses were performed using a four-trait threshold-linear model (BW-HCW-RT-LIV) and a two-trait linear model (BW-HCW) defined as M1 and M2, respectively. M1 was considered to be the full model and compared with the reduced M2. The equation for both models can be expressed as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{e}\,, \tag{1}$$

where **y** is the vector of phenotypes; **b** is the fixed vector of systematic effects; **u** is the vector of random additive genetic effects; **c** is the vector of random litter effects; **e** is the vector for random residual effects; **X**, **Z**, and **W** are incidence matrices relating elements of **y** to **b**, **u**, and **c**, respectively. The systematic effects included in vector **b** were contemporary group (farm, year, and week of birth), sex, age at harvest (only for HCW), and dam parity.

The (co)variance component analyses were run as a single Gibbs chain of 50,000 rounds, with 1 in every 10 samples stored. The prior distributions were assumed to be uniform for fixed effects. The vectors **u**, **c**, and **e** were assumed to be distributed as MVN with mean zero and the following covariance structure:

$$Var\begin{bmatrix}\mathbf{u}\\ \mathbf{c}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\otimes\mathbf{G_0} & 0 & 0\\ & \mathbf{I}\otimes\mathbf{L_0} & 0\\ symm & & \mathbf{I}\otimes\mathbf{R_0}\end{bmatrix}. \tag{2}$$

Estimates of (co)variance components and EBV were obtained using THRGIBBS1F90 for both models (Tsuruta and Misztal, 2006). After discarding the first 15,000 sampled as burn-in, 3,500 samples were kept to calculate the means and standard deviations of the posterior distributions of variance components. Posterior means and standard deviations were used as estimations of (co)variances and their errors for the remainder of the analyses. Breeding values were obtained based on a Gibbs chain of 50,000 rounds with a burn-in of 15,000.

Validation metrics were estimated with the linear regression (LR) method to compare both models (Legarra and Reverter, 2018). The validation dataset consisted of 73,617 animals born in 2019. EBV were calculated for the animals in the validation set with all data available ($\widehat{\boldsymbol{u}}_{whole}$) and with phenotypes removed for the validation animals ($\widehat{\boldsymbol{u}}_{partial}$). The validation measurements obtained were accuracy, dispersion, bias, and Pearson correlations. These measures were obtained to compare the estimability of HCW for both models. Accuracy was calculated for the focal animals using: $\hat{\rho}_{\mathrm{cov(whole,partial)}} = \sqrt{\frac{\mathrm{cov}(\widehat{\boldsymbol{u}}_{whole},\widehat{\boldsymbol{u}}_{partial})}{(1-\bar{F})\hat{\sigma}_u^2}}$ (Legarra and Reverter, 2018), where $\bar{F}$ is the average inbreeding coefficient for animals born in 2019, and $\hat{\sigma}_u^2$ is the estimated additive genetic variance of the whole dataset. INBUPGF90 was used to calculate inbreeding coefficients for each animal by a recursive method based on pedigree (Aguilar and Misztal, 2008). Dispersion (b_1) was measured as the regression coefficient of the regression of $\widehat{\boldsymbol{u}}_{whole}$ on $\widehat{\boldsymbol{u}}_{partial}$: $b_1 = \frac{cov(\widehat{\boldsymbol{u}}_{whole},\widehat{\boldsymbol{u}}_{partial})}{var(\widehat{\boldsymbol{u}}_{partial})}$. The bias is defined as the difference in the average EBV from partial and whole datasets. Lastly, Pearson correlations were calculated between $\widehat{\boldsymbol{u}}_{whole}$ and $\widehat{\boldsymbol{u}}_{partial}$.

RESULTS AND DISCUSSION

VARIANCE COMPONENTS

Variances for the direct additive genetic, litter and residual effects for both models are shown in Table 3. The estimated BW variances were the same for both models (M1 and M2) and were 0.09, 0.24, and 0.48 for the direct additive genetic, litter, and residual effects, respectively. The estimated HCW variances for M1 (M2) were 26.4 ± 1.32 (25.9 ± 1.24), 45.1 ± 0.85 (43.6 ± 0.83), and 285.6 ± 2.25 (273.8 ± 1.13) for additive genetic, litter, and residual effects, respectively. For the binary traits, the residual variances were set to 1.00. There was no difference in variance estimates for BW and HCW between M1 and M2 in agreement with the lowly heritable aspects of RT and LIV. The variance estimates for RT (LIV) were 0.02 (0.05) and 0.10 (0.18) for the additive genetic and litter effects, respectively.

Table 4 shows the heritability and genetic correlations for both models and between all traits. Genetic correlations between traits were either weak or moderate. The genetic correlation between BW and RT was 0.34 ± 0.03, and between BW and LIV was 0.56 ± 0.03. These positive, moderate genetic correlations are logical with the code used for RT and LIV (Table 1). Previous studies have shown that piglets with a heavier BW have greater survival chances (Arango et al., 2006). As BW increases, the probability of an animal to live to harvest increases. Similarly, piglets with a lighter BW have a higher probability of early death, culling, or not retaining their ear tag. HCW had similar genetic correlations between the two binary traits as BW but to a lesser degree, which can be explained by the impact of the inability of 53% of this population to obtain HCW records. The heritability of HCW was less than BW, which explains the more significant impact of HCW between RT and LIV.

It should be noted that no information was given on the number of animals in each pen or if animals were removed from the pen at different times. If the larger animals were removed from the pen first, and the smaller animals had more time and pen space available to grow, this could impact the predictions for HCW. The genetic correlations between BW and HCW also showed no significant difference between models and were 0.31 ± 0.03 and 0.32 ± 0.03 for M1 and M2.

The heritabilities for BW (0.11) and HCW (0.07) showed no difference between models. Heritability estimates for RT and LIV were 0.02 and 0.04. Currently, there is no published research in estimating the heritability of RT for any species. As the genetic correlations are moderate between RT and LIV and the weight traits, as well as h^2 for weight traits, the indirect selection for weight may take care of RT and LIV. Accounting for RT and LIV gives no additional benefits for variance component estimations of HCW and BW evaluations.

VALIDATION

The validation measures give a further justification of the insignificant differences between the models (Table 5). Bermann et al. (2021) showed that the LR method is suitable for binary traits and yields consistent accuracy measures (Legarra and Reverter, 2018). The accuracy, dispersion, and correlations for HCW were higher than BW (Table 5). The EBV for HCW and BW were more biased in M1 than M2. Bias was less than 0.01 for BW and -0.01 for HCW in M1. In M2, bias was 0.01 for BW and 0.06 for HCW. Biases were less than 0.01 for RT and LIV. The dispersion for HCW was less than for LIV and BW. The greatest dispersion was for RT (b_1 = 0.65). The binary traits had lower accuracy and correlations than both linear traits, indicating the difficulty modeling binary traits of low heritability.

We hypothesized that by including the reasons for missing records, RT, and LIV information, HCW evaluations would have better predictions. However, no performance distinctions were observed when this information was accounted for in the model. Bias was marginally less in M1 compared to M2 for both BW and HCW. The dispersion was 0.02 greater for HCW and 0.01 less for BW when missing record information was included in the analyses. It is logical that including the missing trait information does not benefit models for BW evaluations since RT and LIV are traits measured after BW is recorded and can cause extra noise in the model. Despite subtle differences between the models, the inconsistencies are negligible, and the prediction performance is the same for both models. As in Arango et al. (2005b), censoring models could not be implemented with this dataset since there were no records of in which life stage each animal lost its ear tag. An alternative would be to link animals with missing tags back to the data by using parentage tests based on SNP (Maiorano et al., 2019); however, this would require much cheaper genotyping platforms because the crossbreds are terminal animals that do not become breeders.

CONCLUSIONS

HCW and BW accuracies were unchanged when the causes of missing records were included in the model. Positive genetic correlations were observed between BW and HCW and the binary traits indicating relationships exist between these traits. Low genetic correlations between HCW can be attributed to this trait's high percentage of missing records. Results imply a higher survival probability with heavier BW, shown in the moderate and positive genetic correlations between BW and RT and LIV. The low heritabilities of RT and LIV potentially explain the small impact of including animal removal reasons on HCW evaluations. An

alternative option would be to implement a social interaction model; however, group information and a more reliable identification method are needed. A low-density, inexpensive parentage SNP panel could possibly help with the latter. This study shows no major differences in results when accounting for causes of missing records, and RT and LIV traits are not necessary to include in HCW evaluations.

REFERENCES

Aguilar, I., and I. Misztal. 2008. Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. J Dairy Sci 91(4):1669-1672. doi: 10.3168/jds.2007-0575

Arango, J., I. Misztal, S. Tsuruta, M. Culbertson, and W. Herring. 2005a. Estimation of variance components including competitive effects of Large White growing gilts1. Journal of Animal Science 83(6):1241-1246. doi: 10.2527/2005.8361241x

Arango, J., I. Misztal, S. Tsuruta, M. Culbertson, and W. Herring. 2005b. Study of codes of disposal at different parities of Large White sows using a linear censored model. Journal of Animal Science 83(9):2052-2057. doi: 10.2527/2005.8392052x

Arango, J., I. Misztal, S. Tsuruta, M. Culbertson, J. W. Holl, and W. Herring. 2006. Genetic study of individual preweaning mortality and birth weight in Large White piglets using threshold-linear models. Livestock Science 101(1-3):208-218. doi: 10.1016/j.livprodsci.2005.11.011

Bergsma, R., E. Kanis, E. F. Knol, and P. Bijma. 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics 178(3):1559-1570. doi: 10.1534/genetics.107.084236

Bermann, M., A. Legarra, M. K. Hollifield, Y. Masuda, D. Lourenco, and I. Misztal. 2021. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. J Anim Breed Genet 138(1):4-13. doi: 10.1111/jbg.12507

Bijma, P., W. M. Muir, E. D. Ellen, J. B. Wolf, and J. A. Van Arendonk. 2007. Multilevel selection 2: Estimating the genetic parameters determining inheritance and response to selection. Genetics 175(1):289-299. doi: 10.1534/genetics.106.062729

Grandinson, K., M. S. Lund, L. Rydhmer, and E. Strandberg. 2002. Genetic Parameters for the Piglet Mortality Traits Crushing, Stillbirth and Total Mortality, and their Relation to Birth Weight. Acta Agriculturae Scandinavica, Section A — Animal Science 52(4):167-173. doi: 10.1080/090647002762381041

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genetics Selection Evolution 50(1):1-18. (article) doi: 10.1186/s12711-018-0426-6

Maiorano, A. M., A. Assen, P. Bijma, C.-Y. Chen, J. A. I. V. Silva, W. O. Herring, S. Tsuruta, I. Misztal, and D. A. L. Lourenco. 2019. Improving accuracy of direct and maternal genetic effects in genomic evaluations using pooled boar semen: a simulation study1. Journal of Animal Science 97(8):3237-3245. doi: 10.1093/jas/skz207

Muir, W., and A. Schinckel. 2002. Incorporation of competitive effects in breeding programs to improve productivity and animal well being

Tsuruta, S., and I. Misztal. 2006. THRGIBBS1F90 for estimation of variance components with threshold-linear models. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production 89:27-31.

TABLES

Table 3.1: Number of animals with records for each trait and level.

Trait	Code	Level	N	%
BW[1]			471360	100.0
HCW[2]			221311	47.0
RT[3]	1	Missing Tag	134523	28.5
	2	Retained Tag	336837	71.5
LIV[4]	0	Missing Tag	134523	28.5
	1	Died/Culled	104513	22.2
	2	Harvested	232324	49.3

[1]Birth weight

[2]Hot carcass weight

[3]Retained tag

[4]Livability

Table 3.2: Summary statistics for continuous traits and effects.

Trait	Min	Max	Mean	SD
BW, kg	0.24	2.70	1.42	0.41
HCW, kg	51.3	153.8	100.2	9.9
Age at Harvest, d	150.0	210.0	182.4	12.5

Table 3.3: Variances for direct additive genetic, litter, and residual effects for both models and all traits. Standard deviations are shown for HCW. All standard deviations for BW, RT, and LIV were less than 0.01.

	σ_u^2		σ_c^2		σ_e^2	
	M1	M2	M1	M2	M1	M2
BW	0.09	0.09	0.24	0.24	0.48	0.48
HCW	26.4 ± 1.32	25.9 ± 1.24	45.1 ± 0.85	43.6 ± 0.83	285.6 ± 2.25	273.8 ± 1.13
RT	0.02		0.10		1.00	
LIV	0.05		0.18		1.00	

Table 3.4: Estimates of heritability (diagonal) and genetic correlations (off-diagonal) for both models.

	BW	HCW	RT	LIV
Model 1				
BW	0.11 ± 0.00	0.31 ± 0.03	0.34 ± 0.03	0.56 ± 0.03
HCW		0.07 ± 0.00	0.26 ± 0.04	0.09 ± 0.05
RT			0.02 ± 0.00	0.00 ± 0.06
LIV				0.04 ± 0.00
Model 2				
BW	0.11 ± 0.00	0.32 ± 0.03		
HCW		0.07 ± 0.00		

Table 3.5: Validation statistics for both models.

Trait	Model	acc_{LR}[1]	b_1[2]	Bias[3]	$cor(\hat{u}_{whole}, \hat{u}_{partial})$[4]
BW	M1	0.33	0.74	0.00	0.59
	M2	0.33	0.75	0.01	0.59
HCW	M1	0.37	0.93	-0.01	0.74
	M2	0.37	0.91	0.06	0.74
RT	M1	0.19	0.65	0.00	0.56
LIV	M1	0.23	0.78	0.00	0.56

[1]Accuracy as defined in the LR method

[2]Dispersion: the coefficient of the regression of $\widehat{\boldsymbol{u}}_{partial}$ on $\widehat{\boldsymbol{u}}_{whole}$

[3]The difference in the average of $\widehat{\boldsymbol{u}}_{whole}$ and $\widehat{\boldsymbol{u}}_{partial}$ in terms of genetic standard deviation

[4]Pearson correlation between $\widehat{\boldsymbol{u}}_{whole}$ and $\widehat{\boldsymbol{u}}_{partial}$

CHAPTER 4

CONCLUSIONS

The magnitude and persistence of accuracy are high for moderately heritable traits, populations with enough data available to estimate the effects of the independent chromosome segments, and less selection pressure. Genotyping more animals can improve accuracy by adding more information to the evaluation and generating accurate animal relationships and identification. As genotyping becomes more inexpensive, commercial producers will be able to reap more benefits from this technology.

The proportion of animals that do not survive to harvest is a concerning animal welfare issue and a major cost to producers. With more genotyping and improvement in accuracy, commercial pig breeding can be more efficient and sustainable. It will be possible to improve complex and costly traits easily with more accurate models. As genotype data becomes more available, commercial pig breeders will have the ability to select for more specific traits, continue to meet the consumer's demand, and create happier and healthier animals.