DISSERTATION

# RANDOM REGRESSION MODELS AND THEIR IMPACT IN THE GENETIC EVALUATION OF BINARY FERTILITY TRAITS IN BEEF CATTLE

Submitted by

Miguel Angel Sánchez Castro

Department of Animal Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2021

Doctoral Committee:

Advisor: Scott E. Speidel
Co-advisor: Milton G. Thomas

R. Mark Enns
Stephen J. Coleman
W. Marshall Frasier

ABSTRACT

RANDOM REGRESSION MODELS AND THEIR IMPACT IN THE GENETIC EVALUATION OF BINARY FERTILITY TRAITS IN BEEF CATTLE

Female fertility is one of the most important economic drivers of cow-calf operations, however, the achievement of genetic improvement for female fertility traits is challenging due to the biological complexity of reproductive performance and the difficulties related to its statistical modeling. Among the traits relevant to beef cattle breeding practices, those related to key fertility events such as conception and calving are binary in nature. Traditional evaluations of binary traits involve the use of threshold models (TM) that convert categorical phenotypes to an underlying normally distributed range of genotypic values known as liabilities. Despite the successful influence that TM have had on genetic trends of categorically evaluated traits within livestock species, these models also have drawbacks. Among the most important weaknesses are their susceptibility to the extreme category problem (ECP) and their lack of flexibility to incorporate genomic information differently than using genomic relationship matrices whose inverse is difficult to obtain when the number of genotyped animals is high. These deficiencies of TM prevent them from comprehensively utilize all available phenotypic data and preclude their utilization in single-step genomic prediction methodologies based on marker effects models.

Contrastingly, random regression models (RRM) have emerged as an attractive alternative for the evaluation of binary fertility traits in cattle due to their ability to overcome ECP problems and utilize all available information to produce more accurate results in comparison to TM. Furthermore, these models are flexible enough to accommodate any of the single-step genomic

evaluation procedures that have been developed. Consequently, their extension to genomic evaluation procedures that avoid the need of inverting dense genomic relationship matrices such as the recently developed super-hybrid marker effects models, represents a novel approach to evaluate binary fertility traits in beef cattle. Traits like heifer pregnancy (HPG), first-service conception rate (FSCR) and stayability (STAY) constitute important elements of the breeding objective of beef cattle producers, therefore, they were selected as the traits to evaluate in this study. All the reproductive data utilized in this investigation was produced by the Angus cattle population of the John E. Rouse Colorado State University Beef Improvement Center (CSU-BIC). In general, this dissertation was divided in three different studies according to the physiological status of the females producing the phenotypic record (e.g., heifer *vs.* multiparous cows) and the number of instances that such phenotype can be recorded on the life of the animals (non-longitudinal *vs.* longitudinal).

The first study involved the comparison of expected progeny differences (EPD) and genetic parameters obtained with TM and RRM in genetic evaluations of singly-observed heifer dichotomous fertility traits such as HPG and FSCR. Breeding and pregnancy ultrasound records of 4,334 Angus heifers (progeny of 354 sires and 1,626 dams) collected between 1992 to 2019 at the CSU-BIC were utilized. Observations for HPG and FSCR (1, successful; 0, unsuccessful) were defined by fetal age at pregnancy diagnosis performed approximately 130 d post artificial insemination (AI). Traditional evaluations for both traits were performed using univariate TM, whereas alternative evaluations were performed by regressing HPG (or FSCR) on age at first exposure (AFE) using linear RRM with Legendre Polynomials as the base function. Heritability ($h^2$) estimates were 0.04 and 0.03 for HPG and FSCR using TM; whereas RRM derived $h^2$ estimates were 0.02 and 0.006 for the average AFE for HPG and FSCR, respectively. Pearson and

rank correlations between EPD obtained with each methodology were 0.97 and 0.96 for HPG, while for FSCR were 0.75 and 0.72, respectively. Regression coefficients from RRM predictions on those obtained with TM were 0.27 and 0.15 for HPG and FSCR, respectively. Differences in mean accuracies of prediction calculated at the average AFE were minimal between methodologies; however, RRM produced consistently higher accuracies than TM especially when considering young selection candidates. These results suggested that RRM genetic predictions for singly-observed fertility traits in beef heifers were feasible. More importantly, moderate to strong degrees of concordance were found between predictions obtained with both methodologies for both traits, implying that RRM could substitute for TM in genetic evaluations of heifer binary fertility traits.

The second study focused on the comparison of EPD and genetic parameters yielded by TM and RRM in genetic evaluations for longitudinal binary fertility traits such as STAY and FSCR in multiparous Angus cows. Calving performance data, as well as, breeding and reproductive ultrasound records of Angus cows collected between 1990 to 2019 at the CSU-BIC were used for the study. Ten STAY endpoints defined as whether a cow calved at age 3, 4, and up to 12 yr given she calved as a 2-yr-old were assigned observations (1, successful; 0, unsuccessful). Similarly, ten FSCR age specific observations were assigned depending on the age of exposure of the females (ages ranged from 2 to 11 yr) and were defined by fetal age at pregnancy inspections performed approximately 130 d post-AI. Traditional evaluation for STAY was performed using a TM that only considered the success/failure of females reaching the age of 6 (STAY06), since this age is considered as the financial break-even point for cows within the beef industry. Conversely, given there is no specific age of interest for a multiparous cow to conceive in response to her first AI, the traditional evaluation for FSCR was performed using a repeatability TM. Alternative

evaluations for both traits were performed by regressing each trait on its corresponding age specific endpoints using univariate linear RRM with Legendre Polynomials as the base function. Heritability ($h^2$) estimates obtained for STAY were 0.10 and 0.04 for the TM and the RRM, respectively. In the case of FSCR, age was not a significant longitudinal descriptor for the trait; however, only with documentation purposes, $h^2$ estimates were reported. For the TM the $h^2$ estimate was 0.03 whereas for the RRM, heritabilities ranged between 0.02 to 0.05 for all the ages at exposure considered in the model. Pearson ($r_p$) and Spearman's ($r_s$) correlations between EPD obtained with each method for STAY were 0.84 and 0.86. For FSCR, correlations were calculated between the EPD obtained with the repeatability TM and each one of the age-specific EPD obtained with the RRM; therefore, results for the $r_p$ ranged between 0.70 to 0.99; whereas results for $r_s$ ranged between 0.69 to 0.99, depending on the age of exposure considered in the RRM. Although mean accuracies of prediction were higher using RRM than using TM for both traits, increments were much more relevant for STAY than for FSCR. The strong degrees of concordance found between predictions obtained with both methodologies for STAY, suggests that RRM could effectively substitute TM in genetic evaluations of this trait. For FSCR, no improvements were achieved by evaluating the trait using RRM, mainly due to the lack of influence that age had on the ability of cows to conceive in response to their first AI at any age.

Finally, the third study had as objectives 1) to explore the feasibility of implementing single-step random regression super-hybrid models (ssRR-SHM) for the genomic evaluation of HPG, FSCR and STAY; 2) to assess the impact of differing data structures in the resulting genomic predictions of ssRR-SHM for all traits; 3) to identify quantitative trait loci (QTL) associated with the binary fertility traits contemplated in this dissertation. Two types of genetic evaluations were implemented for each trait, the first type was a pedigree-based RRM that utilized Legendre

polynomials as the base function in where the phenotype of interest was regressed on an appropriate age covariate. The second evaluation type was a ssRR-SHM that also used Legendre polynomials as the base function and regressed observations of the trait of interest on its appropriate age covariate, but that included random effects of marker and extra polygenic effects. Within each trait, four different data structure scenarios were created depending on the phenotypic performance of the genotyped and non-genotyped subsets of animals. The behavior of the genomic predictions was assessed through the calculation of Pearson and Spearman's correlations and the estimation of the regression coefficients of EPD obtained with the ssRR-SHM on those obtained with their corresponding pedigree-based RRM.

Results of this study indicated that the implementation of ssRR-SHM for the genomic evaluation of singly-observed binary fertility traits like HPG and FSCR, as well as for the evaluation of a longitudinally recorded binary trait such as STAY was feasible. Nonetheless, an overestimation of genomic predictions occurred with these models when phenotypic records of pre-selected genotyped animals were included in the evaluation. Additionally, inaccurate imputation of genotypes for non-genotyped animals also impacted resulting genomic predictions, although this issue was restricted to this subgroup of animals only. In all cases, the removal of phenotypic records from preselected animals and the maintenance of closely related individuals in the pedigree ameliorated problems associated to overestimation of genomic predictions and improved correlations among genomically-enhanced and pedigree-based EPD for all traits. Regarding GWAS analyzes, the application of ssRR-SHM identified single nucleotide polymorphisms that resulted located either within or relatively close to genes that have been previously associated with important reproductive processes and fertility traits in cattle.

# ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Scott Speidel for his guidance, patience and continuous support during my PhD studies. It is because of everything that you taught to me that I have achieved one of the most important goals in my life, so I will always be grateful to you Sir. I also want to express my heartfelt gratitude to my co-advisor Dr. Milt Thomas. His mentorship, advice, recommendations and continuous support during my graduate career were crucial to keep me on track and accomplishing this doctoral degree. There are not enough words to explain how much positive influence you have had on my personal and professional development, but it is fair to mention that you represent a great source of inspiration to me.

A special thanks to Dr. Mark Enns for his continuum willingness to help me in every possible way with my work, as well as with many other difficulties that I faced during my time at CSU. I always felt supported and listened, so thank you for being there all those times that I needed it Sir. I would also like to thank to Drs. Stephen Coleman and Marshall Frasier for serving as my committee members and providing valuable feedback and different points of view that significantly enhanced the quality of this research effort.

My deepest gratitude to my fellow graduate students within the Animal Sciences Department at CSU for their support, friendship and encouragement to keep going despite of any adversity. Lastly, but not least, I would like to acknowledge to the National Council of Science and Technology of México (CONACYT-MÉXICO) for giving me the opportunity and the financial support required for my doctoral studies at Colorado State University (scholar: 297759/scholarship: 440065).

# DEDICATION

I dedicate this dissertation to my family, especially to my mother Rosa Elvia Castro Valenzuela for encouraging me to always pursue my goals no matter how difficult they seem to be. To my brother Jose Antonio and sister Ana Laura (as well as their respective families), for the unconditional support that you have provided to me during all these years that I have been away from home studying. To my little sister Maria de la Luz (Lucero), for being the light of the family and teaching us that there are no valid excuses to not succeed, you are the living proof that limitations are just part of the challenge and the beauty of life.

I also dedicate this work to Andrea Del Rio for believing in me, even when I didn't believe in myself. Your love, your patience and your endless understanding were crucial during all these years of continuous effort to reach this goal. I consider myself extremely fortunate for having the opportunity of sharing my life with you, thank you for everything, I love you.

This dissertation is also dedicated to every friend, colleague and professor that have helped me throughout my graduate career. Collectively, each one of your supporting actions enabled me to always move forward and work hard to overcome any challenge during this journey.

Thank to everyone for teaching me that: "my only job in this life is to reach my dreams".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

# CHAPTER 1 - INTRODUCTION

Genetic improvement of beef cattle fertility is challenging due to a special combination of factors associated to the biological complexity of reproductive performance and the difficulties related to its statistical modeling (Thaller, 1997; Weigel, 2004; González-Recio and Alenda, 2005). Although fertility encompasses a great diversity of traits that can potentially serve as selection criteria, perhaps only those measuring the success or failure of key biological events like conception and calving are able to summarize the economically relevant outcomes of fertility (Cammack et al., 2009, Walmsley et al., 2018). Consequently, traits such as heifer pregnancy (HPG), first-service conception rate (FSCR) and stayability (STAY) have become important elements of the breeding objectives of many beef cattle enterprises (Golden et al., 2000).

The binary nature of these key-fertility traits poses several challenges to apply best linear unbiased prediction (BLUP) procedures for their evaluation. For instance, categorical response variables are not normally distributed, and typically, heterogeneity of variances exist (Gianola, 1982; Gianola and Foulley, 1983). Therefore, animal breeders have usually attributed the phenotypic expression of categorical traits to an underlying continuous unobservable and normally distributed trait, referred to as liability (Falconer and Mackay, 1996). Under this assumption, observed categorical responses (e.g., 1 = pregnant; 0 = nonpregnant) are the result of animals exceeding or not a particular threshold level of the underlying trait; which is why models used for genetic evaluations of binary traits are commonly referred to as threshold models.

Within threshold models, it is important to recognize that since liability is not a direct observation, solutions for animal random effects could not be given by the usual linear mixed model equations. Conversely, solutions for animal random effects are provided by a non-linear

system of equations that requires to be solved in an iterative way (Gianola and Foulley, 1983). Despite their non-linear nature, the ability of threshold models to yield BLUP by assuming a Gaussian distribution of the liability, made of them the method of choice to perform genetic evaluations of lowly-heritable binary traits (Mrode, 2014; Gianola and Rosa, 2015).

Even with their theoretical advantages, threshold models also have limitations worthy of discussion. Initially, due to the iterative nature of the procedure required to obtain solutions, the computational cost of solving these models was between three to five times higher than that of a linear model (Misztal et al., 1989). However, according to a recent report by Campos et al. (2019), although linear models still have faster convergence than threshold models, current computational advancements have overcomed computational demands and threshold models can still be routinely used. Either way, even with enough computational power, a major problem associated with threshold models is the one related to the Extreme Case Problem (ECP). In this situation, all observations in a given class or level of a fixed effect (typically contemporary group) fall in the same category (e.g., all females are pregnant or the opposite). When this happens, a slow or lack of convergence occurs for these fixed effects as solutions approach $\pm \infty$ or 0 (Misztal et al., 1989).

In order to overcome ECP-related issues, Harville and Mee (1984) recommended to treat these fixed effects as random variables or to delete observations experiencing ECP. Inherent problems of such suggestions involve the usage of different data for different models, since records to be deleted when treating a factor as fixed would not be disregarded when treating the same factor as random. Generally, the option of deleting records experiencing ECP has been more widely adopted for the evaluation of fertility traits (Golden et al., 2018). However, this can lead to distorted inferences because edited data would not be appropriate to perform population-wide genetic predictions (Misztal et al., 1989).

In addition, threshold models employed to analyze binary traits in cattle are often restricted to specific points in the life of the animal, which ignores that in some instances, binary responses can also be longitudinal (e.g., pregnancy status at different ages). As such, genetic predictions using these models often yield lower accuracies of prediction in comparison to statistical methods capable to incorporate all available observations of a longitudinal trait (Sánchez-Castro et al., 2019). Furthermore, as reviewed by Speidel et al. (2018), threshold models are not readily adaptable to the incorporation of genomic information through single-step methods other than genomic relationship matrices (Legarra et al., 2009). The previous may result in difficulties for their implementation in populations with a large number of genotyped animals since genomic relationship matrices require inversion in order to obtain solutions. Lastly, these models have not yet been adapted into the framework of the recently developed single-step hybrid marker effects models that do not require the computation of a genomic relationship matrix or its inverse (Fernando et al., 2014, 2016).

Random regression models (RRM) represent an alternative method to evaluate binary traits and can incorporate data from contemporary groups with no variation (Golden et al., 2018). As such, information from records experiencing ECP are not disregarded and distortions in resulting predictions created by artificially-edited data sets can significantly decrease. Furthermore, RRM are especially suitable for the analysis of longitudinal traits due to their greater flexibility to account for the covariance structure between serial observations of the same response variable on the same individual (Laird and Ware, 1982; Schaeffer, 2004). Interestingly, even when they were originally conceptualized to analyze longitudinal traits, the efficacy of RRM to evaluate traits with phenotypes observed only once has shown acceptable degrees of success using sire models (Englishby et al., 2016) and animal models (Speidel et al., 2018). Additionally, given their

similarities with the traditional linear mixed models, RRM can be relatively easily extended to accommodate genomic information not only in the form of genomic relationship matrices, but also in the form of marker effects super-hybrid models (Kang et al., 2017; Golden et al., 2018).

Considering the special combination between documented weaknesses of traditional threshold models and the potential capabilities of RRM to overcome such weaknesses, we hypothesized that the application of RRM for the genetic predictions of binary fertility traits in beef cattle could yield more accurate results. As such, the general objective of this dissertation was to assess the impact of using random regression models in the genetic evaluation of binary fertility traits in beef cattle.

Specific objectives are outlined below:

1) Comparison between threshold models and random regression models in pedigree-based genetic predictions of dichotomous and singly-observed fertility traits of beef heifers such as heifer pregnancy and first-service conception rate.

2) Comparison between threshold models and random regression models in pedigree-based genetic predictions of longitudinal but binary fertility traits of multiparous beef cows such as stayability and first-service conception rate.

3) Application of single-step genomic evaluations of beef cattle binary fertility traits using random regression super-hybrid models.

Within each study and each particular trait, genetic predictions were compared by means of Pearson's correlations, Spearman Rank correlations and the regression of predictions obtained with the random regression models on those obtained with the base genetic predictions obtained either with threshold or pedigree-based random regression models. Mean accuracies obtained with each statistical model were also compared to each other.

*Literature cited*

Cammack, K. M., M. G. Thomas, and R. M. Enns. 2009. Reproductive traits and their heritabilities in beef cattle. Prof. Anim. Sci. 25:517-528.

Campos, G. S., F. A. Reimann, P. I. Schimdt, L. L. Cardoso, B. P. Sollero, J. Braccini, M. J. Yokoo, A. A. Boligon, and F. F. Cardoso. 2019. Threshold and linear models for genetic evaluation of visual scores in Hereford and Braford cattle. Anim. Prod. Sci. 59:619-627

Englishby, T. M., G. Banos, K. L. Moore, M. P. Coffey, R. D. Evans, and D. P. Berry. 2016. Genetic analysis of carcass traits in beef cattle using random regression models. J. Anim. Sci. 94:1354-1364.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th Ed. Longman, Harlow. Essex, UK.

Fernando, R. L., J. C. M. Dekkers and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50

Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96

Gianola, D. 1982. Theory and analysis of threshold characters. J. Anim. Sci. 54:1079-1096.

Gianola, D., and J. L. Foulley. 1983. Sires evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201-224.

Gianola, D., and G. J. Rosa. 2015. One hundred years of statistical developments in animal breeding. Annu. Rev. Anim. Biosci. 3:19-56.

Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits, a framework for the next generation of EPDs. In: Proc. Beef Improvement Federation, Wichita, KS. p. 2–13.

Golden, B. L., S. Weerasinghe, B. Crook, S. Sanders, and D. J. Garrick. 2018. A Single-Step Hybrid Marker Effects Model Using Random Regression for Stayability in Hereford Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand.

González-Recio, O. and R. Alenda. 2005. Genetic parameters for female fertility traits and a fertility index in Spanish dairy cattle. J. Dairy Sci. 88:3282-3289.

Harville, D. A., and R. W. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. Biometrics. 393-408.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J. F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity. 119:459-467.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics. 38:963-974.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Misztal, I., D. Gianola, and J. L. Foulley. 1989. Computing aspects of a nonlinear method of sire evaluation for categorical data. J. Dairy Sci. 72:1557-1568.

Mrode, R. A. 2014. Linear models for the prediction of animal breeding values. 3rd Ed. CAB Int. Wallingford, UK.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Sánchez-Castro, M. A., M. G. Thomas, R. M. Enns, and S. E. Speidel. 2019. Stability of genetic predictions for stayability using random regression models that include end points beyond 6 yr of age. Trans. Anim. Sci. 3:1678-1682.

Speidel, S. E., R. M. Enns, and B. L. Golden. 2018. Use of a Random Regression model for the Evaluation of Heifer Pregnancy in Red Angus Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand.

Thaller, G. 1997. Genetics and breeding for fertility. Proc. Int. Workshop on Genetic Improvement of Functional Traits in Cattle, Bull. No. 18:55–61.

Walmsley, B. J., S. J. Lee, P. F. Parnell, and W. S. Pitchford. 2018. A review of factors influencing key biological components of maternal productivity in temperate beef cattle. Anim. Prod. Sci. 58:1-19.

Weigel, K. A. 2004. Improving the reproductive efficiency of dairy cattle through genetic selection. J. Dairy Sci. 87:E86-E92.

# CHAPTER 2 - LITERATURE REVIEW

## *2.1 Economic relevance of fertility in cattle*

Economic sustainability of most beef cattle enterprises is largely dependent on their reproductive efficiency because the quantity of beef produced relies on the number of calves born and raised per breeding cycle (Grossi et al., 2008; Speidel et al., 2018a). Specifically, the most limiting factor for producing the greatest number of calves each year on a herd level is the reproductive ability of the cows (Boldt, 2017); therefore, it is imperative to focus effort in improving reproductive performance. Within a genetic improvement context and in relationship to the profitability of conventional cow-calf operations selling calves at weaning, improvements in fertility traits have been estimated to be up to 4-fold more important than improvements in end-product traits (Melton, 1995; Formigoni et al., 2002).

Heritability estimates of reproductive traits commonly used to describe fertility in beef cattle are typically low (Cammack et al., 2009). Nonetheless, it is widely recognized that measures of reproductive efficiency should be included in the breeding objective of beef cattle operations in order to assure profitability (Barth, 1993; Olesen et al., 2000; Ball and Peters, 2004). Within US beef production systems, rearing and maintenance costs of animals are high, so any delay beyond two years to first calving, as well as, any increase in calving interval beyond 365 days, can cause a significant reduction in herd profitability (Ball and Peters, 2004; Walmsley et al., 2018). Selecting for a reduction of unproductive periods in cows could help enhance the economic viability of beef enterprises (Burns et al., 2010). In this regard, traits like heifer pregnancy (HPG), stayability (STAY) and first-service conception rate (FSCR), have been identified as economically relevant traits (ERT) for beef cattle (Golden et al., 2000; Minick Bormann et al., 2006).

*2.2 Fertility traits*

*2.2.1 Heifer pregnancy (HPG)*

Heifer pregnancy (HPG) has been defined as the probability of a female conceiving at the end of her first breeding season (Crews and Enns, 2008; Boldt et al., 2018). Within *Bos taurus* breeds, the previous definition involves a heifer's ability to become pregnant in order to calve at two-years of age (Cammack et al., 2009). The large investments of time and resources associated with replacement heifer development represent some of the reasons why this trait is relevant to beef cattle producers (Doyle et al., 2000). MacNeil and Vukasinovic (2011), suggested that HPG also influences profitability by impacting the number calves for sale (e.g., the more pregnant heifers, the more saleable calves). Furthermore, females that become pregnant as yearlings will typically have more calves over their lifetime (Champman et al., 1978; Núñez-Domínguez et al., 1985; Patterson et al., 1992). Phenotypes of HPG are recorded as binary, with a value of 1 for pregnant heifers and a value of 0 for nonpregnant heifers (Eler et al., 2002).

Heritability ($h^2$) estimates of HPG are typically low (<0.1) or moderate (0.1 to 0.3) and vary depending upon factors such as breed, scale in which the trait is analyzed (underlying *vs* linear) and statistical method employed for its variance component estimation (Buddenberg et al., 1989; Kadarmideen et al., 2000; Cammack et al., 2009). Evans et al. (1999) reported a $h^2$ of 0.14 for this trait in Hereford heifers, Doyle et al. (2000) described a $h^2$ of 0.21 for HPG in Angus cattle, whereas, the $h^2$ estimate for HPG in Nellore was reported to be 0.57 (Eler et al., 2002). In all of these studies, observations of HPG were transformed to an underlying scale and method R procedures were used for variance component estimations. Differences in estimates were related then, to the varying levels of selection pressure applied to HPG between *Bos taurus* and *Bos indicus* breeds. Variations relative to the scale in which the trait was analyzed were reported by

Buddenberg et al. (1989), since authors indicated that HPG $h^2$ estimates on the observed scale were consistently lower than estimates obtained on the underlying scale. For instance, in the case of Angus heifers, $h^2$ estimates were 0.17 and 0.34 on the observed and underlying scales, respectively. Whereas the estimates for Hereford and Polled Hereford heifers were 0.04 and 0.05 on the observed scale and 0.08 and 0.10 on the underlying scale, respectively.

*2.2.2 First-service conception rate (FSCR)*

Bormann et al. (2006) defined first-service conception rate (FSCR) as the probability that a cow will conceive in response to her first artificial insemination (AI). This trait provides producers an opportunity to identify females that become pregnant on their first service, from those that require multiple inseminations or that conceive by natural service (Cammack et al., 2009). Economic implications of FSCR include its relationship with the cost of semen, as well as, the costs associated with synchronization protocols, estrus detection and AI services (Bormann et al., 2006). This trait is also related to differences in the quality and value between AI-produced calves (e.g., superior genetics) and natural service calves. Females conceiving on their first service, calve earlier within the calving season, have more chances to breed postpartum within a year, and have more time to nurse and wean heavier calves (Lesmeister et al., 1973; Marshall et al., 1990).

Identifying cows with an improved ability to conceive with just one service, could trigger an increase in the use of AI within the beef industry. The low adoption of AI programs in beef cattle operations is likely to remain until a precise human control of conception becomes more feasible and cost effective. In the 1980's, the adoption of the AI biotechnology by cattle industries was widely different. In dairy operations, about 60 to 70% of the cows were bred through AI, while in the beef industry, only 3 to 5% of the females were artificially inseminated and most of the AI-derived calves were kept within the seedstock sector of the beef industry (Barber, 1983; Koch et

al., 1986). A more recent report by Colazo and Mapletoft (2014) explained that adoption of AI by the dairy industry has increased to 80% of all cows, whereas for beef cattle, the overall percentage of use of this biotechnology remained almost static at 4%. Later, Lamb and Mercadante (2016) reported that 7.6% of beef operations in the US use AI as a reproductive management tool. Phenotypes for FSCR are binary, values of 1 are assigned to females becoming pregnant with only one AI and, values of 0 are allocated to females failing to conceive in their first AI (Bormann et al., 2006). Heritability estimates for this trait have been reported to range between 0.03 and 0.22 (Dearborn et al., 1973; Bormann et al., 2006; Cammack et al., 2009). Some factors associated with the variability of the estimates of FSCR are breed composition, age of the females, and the scale in which the trait is analyzed.

*2.2.3 Stayability (STAY)*

Stayability (STAY) was originally defined as the ability of a cow to remain in a herd until a specific age given the opportunity to reach that age (Hudson and Van Vleck, 1981). A refined definition states that STAY represents the probability that a cow will remain in the herd until 6 years of age, given she first calved as a 2-year-old (Brigham et al., 2007). The age of 6 is considered as a financial breakeven point within the US beef industry, since cows that have produced 5 consecutive calves by this age, already recouped their development and maintenance costs (Snelling et al., 1995; Brigham et al., 2006). Cows staying in production longer benefit profitability of herds by reducing the need of additional female replacements, decreasing the incidence of dystocia and increasing the average weaning weight of marketed calves (Garrick, 2006). From a genetic improvement perspective, it has been determined a 1 unit increase in overall herd STAY results in an increase in profit of $2500 for herds with 40% of cows remaining in the herd to the age of 6 (Enns et al., 2005).

Stayability represents a measure of sustained fertility through the lifetime of a beef cow; therefore, it is a key driver of beef production efficiency (MacNeil and Vukasinovic, 2011). Heritability estimates for STAY have been estimated to range from 0.02 to 0.36, depending on the age endpoint chosen, the statistical methodology implemented for its estimation, breed, and the scale in which the trait was analyzed (Snelling et al., 1995; Cammack et al., 2009; Jamrozik et al., 2013). Martinez et al. (2005), reported $h^2$ estimates for STAY to consecutive ages (1, 2, and up to 6 years of age) ranging between 0.09 and 0.30 for threshold models and between 0.05 and 0.19 for linear models in Hereford cattle. In African Angus cattle, $h^2$ estimates for STAY to consecutive ages (4, 5, and up to 8 years of age) were reported to range between 0.24 to 0.26 using a sire threshold model and between 0.18 to 0.20 when using an animal threshold model (Maiwashe et al., 2009). Breed differences in $h^2$ estimates for this trait were reported by Brigham et al. (2007) when analyzing information from the American Gelbvieh Association (AGA), the American Simmental Association (ASA) and the Red Angus Association of America (RAAA). Authors reported that $h^2$ to consecutive ages (from 3 to 6 years of age) ranged between 0.15 to 0.18, from 0.17 to 0.21 and from 0.15 to 0.18, for Gelbvieh, Simmental and Red Angus cattle, respectively. Furthermore, Jamrozik et al. (2013) reported $h^2$ estimates for STAY to consecutive ages (from 2 to 8 years of age) that ranged between 0.36 to 0.12 in Canadian Simmental cattle, when data were analyzed in a longitudinal scale and using Bayesian methods with Gibbs sampling.

*2.3 Factors affecting reproductive efficiency in cattle*

Many of the factors influencing fertility of cattle populations have been recognized for more than four decades (Venter et al., 1973; De Kruif, 1978). Perhaps the simplest strategy to classify sources of variation in cattle fertility is by dividing them into environmental and genetic causes (Venter et al., 1973). Within the environmental causes, variations in climate conditions, nutritional

status, housing, management practices and any stimuli that demand a response from the animal to adapt to new circumstances could be included (Lee, 1993). Genetic sources of variation include the natural variability in performance according to the genetic make-up of animals, as well as, genetic correlations between reproductive and production traits such as milk yield (Berry et al., 2014).

### *2.3.1 Environmental causes*

#### *2.3.1.1 Climate conditions*

Climate has historically been recognized as a major factor affecting fertility in cattle (Thatcher, 1974; Gwazdauskas et al., 1975). Specifically, Gwazdauskas (1985) suggested that an animal's environment is dependent upon ambient temperature, humidity, radiation and wind, nonetheless, the first factor is typically the most influential on reproductive efficiency. Each species, breed or animal category, has an ambient temperature comfort zone in which the energy expenditure of the animal is minimal, constant, and independent of the environment (Nardone et al., 2006). However, extremely cold or hot temperatures increase or decrease the maintenance requirements because homeostasis is disrupted beyond the range of thermoneutrality, and reproductive efficiency results are compromised (Gwazdauskas, 1985).

Much of the research describing the impact of weather on cattle fertility has been executed in hot rather than cold environments, therefore, little is known about the effects of cold stress over the physiology of reproductive processes (Gwazdauskas, 1985; Lee, 1993). Among the few reports about negative effects of cold weather on cattle fertility, a historic study performed in Eastern Canada revealed that lower conception rates were registered in the coldest months of the year (Mercier and Salisbury, 1947). Authors concluded that changes of fertility were possibly results of a disruption of the circadian cycle due to reduced day-length during the winter, as well as, the

very low environmental temperatures. Later, Westra and Christopherson (1976) suggested that serum triiodothyronine ($T_3$) and thyroxine ($T_4$) concentrations increased significantly when animals were below the temperature comfort zone. Along with the increments in the concentrations of the aforementioned hormones, there was an increase in dry matter intake associated with the need of cows to produce more energy for warmth (Aceves et al., 1987; Huszenicza et al., 2002). Due to the previous, pregnant beef cows managed under severe cold conditions on pastures without supplementation, were reported to lose substantial body weight and produced weaker calves (Jordan et al., 1968). Hemsworth et al. (1995) informed that calves are especially susceptible to cold at birth since they have a lack of metabolic heat production coming from rumen fermentation.

Negative effects of heat stress over reproductive efficiency of cattle have been extensively reviewed (Wolfeson et al., 2000; Jordan, 2003, Takahashi, 2012). The way in which heat stress affects fertility in cattle is multifactorial and dependent upon the type of stress (e.g., acute or chronic) to which animals are subjected (Wolfeson et al., 1988; Correa-Calderón et al., 2014). Heat stress impairs reproductive processes such as oocyte competence, embryonic growth, gonadotropin secretion, ovarian follicular growth, steroidogenesis, development of corpus luteum, and uterine endometrial responses (Wolfeson and Roth, 2018). These deleterious effects are the result of either the hyperthermia associated with heat stress or the physiological adjustments made by heat-stressed animals to regulate body temperature (Hansen, 2009). In beef cows, exposure to high ambient temperatures has been shown to decrease the length and intensity of estrus, since pedometer measurements showed a reduction in the number of steps of cows grazing under these conditions (Takahashi, 2012). Moreover, high summer temperatures have been shown to decrease semen quality in bulls for up to 8 weeks after animals were stressed, which compromises fertility after AI or natural mating (Meyerhoeffer et al., 1985). Consequently, dramatic drops in pregnancy

rates have been commonly observed during warm seasons in places where high humidity was combined with high temperatures (Loyacano et al., 1972; Sprott, 1999; Sprott et al., 2001).

*2.3.1.2 Nutrition*

According to Short and Adams (1988), sub-standard nutritional management is the most limiting factor for reproduction in beef cattle. Most reproductive failures in beef females can be attributed to improper nutrition and/or thin body condition scores (BCS). The percentage of body fat at specific stages of a beef cow's production cycle is an important determinant of its reproductive performance and overall productivity (Herd and Sprott, 1986). Energy intake has effects on a wide variety of endocrine, neural and metabolic physiological mechanisms. Effects include changes in gonadotropic hormone secretion, synthesis and secretion of progesterone during both the estrous cycle and pregnancy, differential sensitivity of the pituitary-hypothalamus to steroids and releasing hormones and changes in ovarian activity measured by hormone secretion, follicular development and ovulation (Short and Adams, 1988). Energy restrictions during late pregnancy results in thin BCS at calving and extends the interval to first postpartum estrus in beef cows (Richards et al., 1986). Short et al. (1990) explained that postpartum infertility is affected by several minor factors (season, breed, presence of a bull, among others); however, the two major factors affecting postpartum anestrus are calf suckling and cow nutrition level. These two factors have direct effects on the reproductive ability of beef cows after calving, but also interact with one or more of the other factors to control postpartum anestrus. Regarding the effects of nutrition in pregnancy rate, Selk et al. (1988) suggested that BCS precalving and at the start of the breeding season, along with body weight changes between 2 and 4 months before parturition, had profound effects in pregnancy rates of range beef cows.

One of the major determinants of lifetime reproductive efficiency of beef cows is age at puberty, and nutrition has an inverse relationship with it. Specifically, effects of nutrition on sexual maturation are related to the timing of the prepubertal increase in LH secretion and seems to involve the LH pulse generating system located in the hypothalamus (Schillo et al., 1992). It is economically important that heifers first calve at 2 years of age and 20 to 30 days ahead of the main cow herd. To accomplish these managements targets, heifers must reach puberty at 14 or 15 months of age, and energy intake is the main factor influencing body weight gains by these ages. Commonly, a bench-mark used within the US beef industry establishes that heifers should reach about 66% of their mature weight before their first breeding season (Dziuk and Bellows., 1983; Mass, 1987; Patterson et al., 1992). As explained by Williams et al. (2002), a targeted body weight of about 66% represent a minimum level of adiposity and a threshold circulating level of the adipose-derived hormone leptin, which has a central role in the regulation of reproduction in cattle. Of the environmental elements influencing reproduction, nutrition commands the greatest attention because livestock producers can control nutritional inputs (Dunn and Moss, 1992). Appropriate nutritional strategies may afford beef cattle managers the opportunity to produce beef cattle more efficiently and become more sustainable (Hess et al., 2005).

*2.3.1.3 Management practices*

Management is the sum of decisions and actions made by a manager who then become the focal point for success or failure of any program. Reproductive management is desirable because of convenience, economics and disease control (Dziuk and Bellows, 1983). One of the most important decisions that beef cattle producers need to make is to define the appropriate length of a breeding season (Frasier and Pfeiffer, 1994). Limited breeding seasons generally result in increased calf production and greater efficiency of beef enterprises (Deutscher et al., 1991).

Usually, breeding seasons are restricted to the time of the year that optimize subsequent calf survival and growth under constraints imposed by feed costs. Timing of the breeding season is also influenced by marketing alternatives for the calves (Azzam et al., 1990). Although it has been reported that extending the breeding season up to 120 days may be beneficial due to the enhanced increment in the proportion of females becoming pregnant (Frasier and Pfeiffer, 1994); such practice can hide poor conception rates and prolonged periods of anestrus (Caldow et al., 2005). Breeding seasons of nine to ten weeks in length have been proposed as the most appropriate for beef cattle operations trying to keep calving intervals no longer than 365 days (Deutscher et al., 1991; Caldow et al., 2005; Walmsley et al., 2018). Once a breeding season is established, it is possible to enhance the overall herd fertility by initiating reproductive management of replacement heifers 20 days earlier than the cow herd (Wiltbank, 1970). Young dams nursing their first calf have postpartum intervals to first estrus 15 to 25 days greater than older dams (Dziuk and Bellows, 1983). Therefore, early breeding of heifers would allow them additional time to return to estrus and be rebred for the production of their second calf along with the older cows.

Another important decision to make by beef cattle producers is the use of natural mating or AI. Recent surveys suggested that more than 90% of the US beef cattle operations utilize natural mating as their primary reproductive strategy (Lamb and Mercadante, 2016). Within this scenario, a key practice to ensure acceptable and profitable conception rates is performing a breeding soundness exam (BSE) of the natural service bulls (Menegassi et al., 2011). Breeding soundness refers to a bull's ability to get cows pregnant and its importance relies on the fact that bulls account for over 90% of the genetics of herds, even though they represent only 5% of them. Normally, a bull can produce from 20 to 30 calves, depending on the bull-to-cow ratio and on pregnancy rates during the breeding period (Amaral et al., 2003). However, it has been reported that approximately

20 to 40% of bulls of unselected populations have some degree of subfertility (Kastelic and Thundathil, 2008). Caldow et al. (2005) explained that the minimum standard to define a fertile bull is that he should be able to get at least 45 out of 50 normal cycling females pregnant within nine weeks of the breeding season, and 60% of these should be pregnant within the first three weeks of breeding season. A BSE is based on a physical evaluation and acceptable thresholds for testicular development and functionality (Kastelic and Thundathil, 2008). According to the Society of Theriogenology BSE guidelines (Chenoweth et al., 1993), a set of minimum thresholds to evaluate yearling bulls are: scrotal circumference greater than 34 cm, more than 30% of progressively motile sperm and less of 30% of morphologically abnormal sperm.

Advancements in reproductive biotechnologies and a better understanding of the dynamics of the bovine estrus cycle have made possible the development of estrus-synchronization and ovulation-synchronization protocols (Seidel, 1995; Lamb and Mercadante, 2016). Synchronization protocols have the potential to shorten breeding and calving seasons, increase calf uniformity and facilitate the use of AI (Larson et al., 2006). Implementation of synchronization protocols by beef producers, however, depends largely on two main factors: limiting the frequency of handling cattle and the elimination of detection of estrus (Lamb and Mercadante, 2016). Early estrus-synchronization protocols focused on regressing the corpus luteum with an injection of prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) followed by estrus detection (Lauderdale et al., 1974; Burfening et al., 1978). However, estrus detection is a time-consuming repetitive task that is problematic to apply in commercial beef operations, since it needs to be carried out up to 5-times a day for the purpose of using AI (Stevenson et al., 1996; Diskin and Sreenan, 2000; Taponen, 2009). Later developments of protocols combined the use of $PGF_{2\alpha}$ and exogenous progestins, improving synchronization and pregnancy rates (Lucy et al., 2001), but still had no complete control over the

ovulation process. The addition of gonadotropin-releasing hormone (GnRH) to synchronization protocols overcame the previous issue allowing external control of follicular waves and synchronize ovulations (Pursley et al., 1995).

For a synchronization protocol to be useful in order to apply fixed-time artificial insemination (TAI) within the beef industry, it should be effective for inducing cyclicity, easy and inexpensive to administer, applied in a short period of time, and able to synchronize follicular development (Geary et al., 2001). Probably, the most widely used protocol to synchronize ovulation in both dairy and beef cattle is the Ovsynch (Pursley et al. 1995; Geary et al., 1998). As summarized by Taponen (2009), Ovsynch protocol consists of three hormonal treatments: the first one, GnRH, is intended to synchronize follicular waves, the second one, $PGF_{2\alpha}$, given 7 days later, induces luteolysis, and the third one, GnRH, given 36 to 48 hours after the $PGF_{2\alpha}$ administration, induces ovulation at a predetermined time. Artificial insemination is performed 16 to 24 hours after the second GnRH administration (Figure 1).

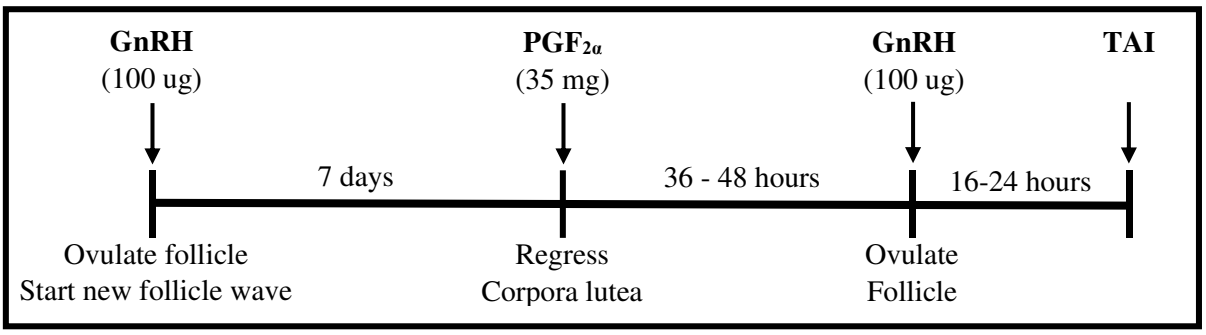


**Figure 2.1.** Description of the timing and physiological action of each hormonal injection applied in the Ovsynch protocol (Adapted from Pursley et al., 1995).

Modifications to the Ovsynch protocol led to the development of another protocol called Co-synch, in which $PGF_{2\alpha}$ is administered 7 days after GnRH followed by a second GnRH injection and TAI at 48 hours (Geary et al., 2001). This protocol has been proved to yield pregnancy rates of 52% compared with 54% obtained with the Ovsynch protocol. Furthermore,

the insertion of an intravaginal progesterone device (CIDR) during the 7 days interval between the initial GnRH and $PGF_{2\alpha}$ injections have been reported to enhance pregnancy rates by 9 to 10% (Lamb et al., 2010). Therefore, both protocols (Ovsynch and Co-synch) are reliable ovulation-synchronization protocols that eliminate the need of estrus detection and allow the performance of TAI in beef cattle (Larson et al., 2006).

*2.3.1.4 Herd health*

Herd health is another major factor that influences reproductive performance in beef cattle (Ball and Peters, 2004). Bovine reproductive diseases result in yearly economic losses that range between $441 to $502 million for US beef producers due to decreased production, delayed reproduction, and increased treatment and preventive measurement costs (Bellows et al., 2002). According to Sprott and Field (1998), the most common reproductive diseases in cattle are brucellosis, leptospirosis, vibriosis, trichomoniasis, infectious bovine rhinotracheitis (IBR) and bovine viral diarrhea (BVD). Givens (2006) provided a more complete list of infectious causes of infertility in cattle, noting that, most of the etiological agents of these diseases can be prevented and controlled with adequate surveillance, biosecurity and/or vaccination programs. Improvements in the reproductive performance of cow-calf systems that implemented vaccination protocols to prevent diseases such as bovine herpesvirus, BVD and leptospirosis have been documented (Aono et al., 2013). Furthermore, the instauration of eradication programs (especially in the case of zoonotic diseases such as brucellosis), have proved to be a successful avenue for eliminating those causes of infertility in cattle populations and preserve health in humans (Zhang et al., 2018).

### *2.3.2 Genetic variability and genetic correlations with other traits*

Regardless of the magnitude of the heritability of a trait, as long as the heritability does not equal zero, there will be genetic variation leading to the possibility of finding animals with high breeding values, average breeding values and low breeding values in a population (Bourdon, 2000). As such, despite the low heritability estimates for female fertility traits, enough genetic variability exists within cattle populations to make genetic improvement of fertility a feasible practice (Mackinnon et al., 1990a; Meyer et al., 1991; Thaller, 1997). Direct selection for cow fertility is challenging since it can only be practiced in females, with a limited selection intensity and a significant delay in phenotype collection. However, it has been reported that a favorable genetic correlation exists between cow and bull fertility; therefore, cow fertility could be genetically improved by indirect selection for improved bull fertility (Land, 1973; Mackinnon et al., 1990a). The opportunity to apply higher selection intensities in males, allows breeders to generate a faster genetic improvement in their female progeny. This has been proven in Droughtmaster cattle, since selection programs that applied high selection intensities over sires with high Estimated Breeding Values (EBV) for pregnancy rate, improved the fertility of heifers and 4-year-old lactating cows (Mackinnon et al., 1990b; Davis et al., 1993). Furthermore, it has been recently suggested that that improving bull's fertility is particularly critical to improve the overall reproductive efficiency in beef cattle, since with the advancements of reproductive biotechnologies, one bull can breed thousands of females through AI (Thundathil et al., 2016).

Another method of improving female fertility in beef cattle could be to select on a more heritable but genetically correlated trait (Morris et al., 2000). In contrast with female reproductive traits, testicular measurements are highly heritable and show favorable correlations with sperm production traits (Coulter et al., 1976; Neely et al., 1982). Land (1973) suggested that since gonads

in both males and females are regulated by the same endocrinologic factors (e.g., follicle stimulating hormone and luteinizing hormone), it is physiologically expected that correlations among male and female fertility traits exist. In the US, Brinks et al. (1978) were the first to study genetic correlations across sexes in beef cattle and they reported a favorable genetic correlation between scrotal circumference (SC) and age at puberty (AP) in heifers. Subsequent studies explored further the previously mentioned relationship in different breeds of cattle and in general, all agreed in that selection for increased testicular size would lead to improvements in female reproduction, particularly an increase in calving rate and a decrease in age at first breeding (Toelle and Robison, 1985; Morris et al., 1992; Vargas et al., 1998). Even when Martínez-Velázquez et al. (2003) found a small but favorable genetic correlation between SC and AP ($r_g$ = -0.15) in nine beef cattle breeds, authors suggested that genetic response in female reproductive traits through sire selection on yearling SC may not be as effective as previous reports stated. Discrepancies among results were attributed to the different variance components estimation methods employed across studies. Earlier reports could have been biased since variance components were estimated via regression or ANOVA based on sib covariances, without accounting for the selection of the parents. Whereas in Martínez-Velázquez et al. (2003), the REML method was implemented, mitigating the bias of previous parents' selection. Nonetheless, a recent report by Bonamy et al. (2018) in Angus cattle, supported the utility of SC as an indicator trait of female fertility, since favorable genetic correlations were found between SC and age at first calving (AFC) using REML procedures. Moreover, the same report suggested that early rather than late SC measurements (e.g., measurements taken at 300 days of age, as opposed to measurements taken at 400 or 630 days of age), better reflected female precocity in beef cattle.

Considering that growth rate remains the primary selection criterion for most beef cattle breeders, it is important to understand the consequences of selecting for growth traits in other economically relevant traits, including reproductive performance (Archer et al., 1998). Concerns about selecting for increased growth rate on the reduction of the reproductive efficiency stems from positive relationships with dystocia (Bellows et al., 1971; Smith et al., 1976). Specifically, growth traits such as weaning weight and yearling weight, hold positive and unfavorable genetic correlations with birth weight; therefore, selection for high mature weights was expected to increase birth weights (Bourdon and Brinks, 1982). In this regard, a strong and positive genetic correlation (e.g., $r_g = 0.9$) between birth weight and dystocia (Meijering, 1984), complicated even more the accommodation of reproduction and growth performance into the breeding objective of beef enterprises.

However, the application of the multiple-trait models (MTM) originally suggested by Henderson and Quaas (1976) in the genetic evaluations of beef cattle, allowed the finding of the so-called "curve benders". These are beef cattle that combine superior breeding values for birth weight (e.g., low or intermediate birth weights) and with acceptable or superior breeding values for weaning weight (Meyer et al., 1991; McNeil et al., 1998). Consequently, dystocia became a considerably less frequent problem and selection for high growth rate has not compromised reproductive performance of beef cattle (Archer, 1998; Bennet, 2008; Santana et al., 2012). With respect to the possible correlations among reproductive and carcass quality traits, a study performed in Wagyu cattle suggested that genetic relationships between these traits were generally low; therefore, selection for carcass traits would not compromise genetic progress of reproductive traits (Oyama et al., 2004).

*2.4 Genetic evaluations for fertility traits*

The binary nature of phenotypes for pregnancy status in cattle (e.g., HPG, FSCR and STAY) pose several challenges to apply the best linear unbiased prediction (BLUP) methodology of Henderson (1975). Among these challenges, the fact that categorical response variables are not normally distributed and do not possess homogeneous variance was noted (Gianola, 1982; Gianola and Foulley, 1983). As such, animal breeders have usually attributed the phenotypic expression of categorical traits to an underlying continuous unobservable and normally distributed trait, referred to as the liability (Falconer and Mackay, 1996). The observed categorical responses (e.g., 1 = pregnant; 0 = nonpregnant) are therefore due to animals exceeding particular threshold levels of the underlying trait, consequently, models used for genetic evaluations of these types of traits are commonly referred to as threshold models (Figure 2.2).

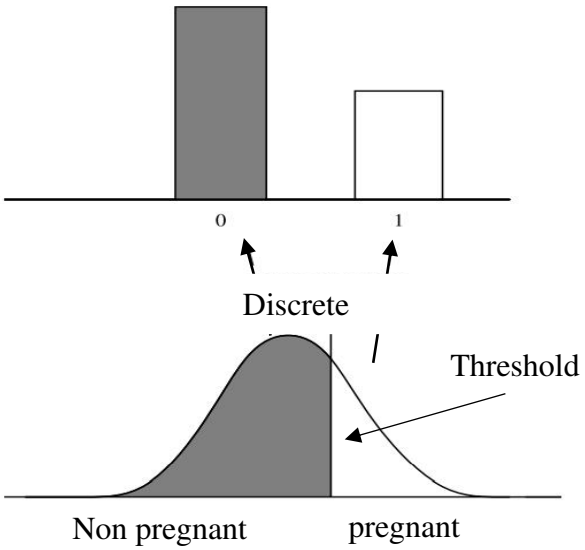


**Figure 2.2.** Schematic representation of the continuous distribution of the underlying liability, and the resulting discrete distribution of the observed phenotype (Adapted from Felsenstein, 2014).

*2.4.1 Threshold models (TM)*

Genetic evaluations using animal threshold models (TM) predict breeding values on the underlying scale. In practice, these predictions have been normally expressed as Expected Progeny

Differences (EPD) in the form of probabilities. For instance, EPD for probability of pregnancy can be used to select females with a higher probability of being fertile (Eler et al., 2002). A TM is often described in matrix form by the following equation:

$$\boldsymbol{y^* = Xb + Zu + e}$$

where $\boldsymbol{y^*}$ corresponds to a vector of transformed observations of the trait in question (e.g., HPG, FSCR or STAY) on the underlying scale, $\boldsymbol{b}$ is a vector of unknown solutions for fixed effects, $\boldsymbol{u}$ corresponds to a vector of unknown solutions of animal random effects. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known incidence matrices relating observations in $\boldsymbol{y^*}$ to both fixed and random effects, and $\boldsymbol{e}$ represents a vector of unknown residual errors. For this model, variances are assumed to be:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_a^2} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I\sigma_e^2}\end{bmatrix}$$

where **A** represents the Wright's numerator relationship matrix, **I** is an identity matrix and $\boldsymbol{\sigma_a^2}$ and $\boldsymbol{\sigma_e^2}$ are the additive and residual variances, respectively. The additive direct genetic variance ($\boldsymbol{\sigma_a^2}$) is trait specific and its units are expressed on the underlying scale, while, as explained by Gianola and Foulley (1983), the residual variance ($\boldsymbol{\sigma_e^2}$) is constrained to be equal to 1 in accordance to the specifications of the maximum *a posteriori* (MAP) probit threshold model. An important note within threshold models is that, since $\boldsymbol{y^*}$ is not observed, it is not possible to solve for $\boldsymbol{u}$ using the usual mixed model equations. Nonetheless, as reviewed by Mrode (2014), Gianola and Foulley (1983) provided the following non-linear system of equations that requires to be solved in an iterative way, in order to obtain solutions for the specific number of thresholds being considered, as well as, for fixed and random effects:

$$\begin{bmatrix} T^{[i-1]} & L'^{[i-1]}X & L'^{[i-1]}Z \\ X'L^{[i-1]} & X'W^{[i-1]}X & X'W^{[i-1]}Z \\ Z'L^{[i-1]} & Z'W^{[i-1]}X & Z'W^{[i-1]}Z + A^{-1}G^{-1} \end{bmatrix} \begin{bmatrix} \Delta t^{[i]} \\ \Delta b^{[i]} \\ \Delta u^{[i]} \end{bmatrix} = \begin{bmatrix} p^{[i-1]} \\ X'v^{[i-1]} \\ Z'v^{[i-1]} - A^{-1}G^{-1}u \end{bmatrix}$$

where **T**, **L**, **W**, **p** and **v** involve normal distribution functions, $\mathbf{G} = \mathbf{I}\sigma_a^2$ and **W** is a diagonal matrix. The above equations follow the Newton-Raphson or Fisher's scoring iterative algorithms, and **i** is the iterate number. These methods require solving a linear set of equations successively until solutions converge (when Δ are sufficiently small from one iteration to the next). This system resembles the mixed model equations; however, matrices and vectors due to threshold effects are created differentially, the right-hand sides are functions of **A**, **G** and **u**, and the equations are solved for differences between consecutive iterations (Misztal et al., 1989).

Historically, threshold models have been the method of choice to perform genetic evaluations for categorical traits with low heritabilities (Mrode, 2014); however, they have limitations worthy of discussion. Initially, due to the iterative nature of the procedure required to obtain solutions, the computational cost of solving these models was between three to five times higher than that of a linear model (Misztal et al., 1989). However, according to a recent report by Campos et al. (2019), even when linear models still have faster convergence than threshold models, recent computational advancements overcome computational demands and threshold models can still be used. Either way, even with enough computational power, a major problem associated with threshold models is the one related to the Extreme Case Problem (ECP). In this situation, all observations in a given class or level of a fixed effect (typically contemporary group) fall in the same category (e.g., all females are pregnant or the opposite). When this happens, a slow or lack of convergence occurs for these fixed effects as solutions approach ± ∞ or 0 (Misztal et al., 1989). In order to overcome this issue, Harville and Mee (1984) recommended to treat these fixed effects as random variables or to delete observations experiencing ECP. The second option has been more widely adopted for the evaluation of fertility traits (Golden et al., 2018); however, this can lead to

distorted inferences because edited data would not be appropriate to perform a population-wide genetic prediction (Misztal et al., 1989; Speidel et al., 2018b).

In addition, threshold models typically employed in beef cattle evaluations are often restricted to specific points in the life of the animal, ignoring that in some instances, binary responses can also be longitudinal (e.g., observations can be taken repeatedly during an individual's lifetime like pregnancy status at different ages). Consequently, genetic predictions using these models often yield lower accuracies in comparison to statistical methods capable of incorporating all available observations of a longitudinal trait (Sánchez-Castro et al., 2019). Furthermore, as reviewed by Speidel et al. (2018b), threshold models are not readily adaptable to the incorporation of genomic information through single-step methods other than genomic relationship matrices (Legarra et al., 2009). The previous may result in difficulties for their implementation in populations with a large number of genotyped animals since genomic relationship matrices require inversion in order to obtain solutions. Lastly, these models have not yet been adapted into the framework of the currently developed single-step hybrid marker effects models that do not require the computation of a genomic relationship matrix or its inverse and provide richer inferences (Fernando et al., 2014, Fernando et al., 2016).

*2.4.2 Multiple-trait models (MTM)*

As it was briefly alluded in previous sections, multiple-trait models (MTM) have the ability to simultaneously predict the genetic merit of the animals for two or more traits (Henderson and Quaas, 1976). The key feature of these models is that they incorporate the genetic and residual variances among the traits under study (Mrode, 2014). The way to set up a MTM consists basically in stacking as many single-trait models, as many different traits we want to analyze, simultaneously. For simplicity, an example of how to set up a two-trait model will be described.

Considering the subscripts $_1$ and $_2$, as references to "trait 1" and "trait 2", respectively; two single-trait models can be specified as follows:

$$\mathbf{y_1 = X_1 b_1 + Z_1 u_1 + e_1}$$

$$\mathbf{y_2 = X_2 b_2 + Z_2 u_2 + e_2}$$

Now, within the MTM a key aspect is that animals need to be ordered within traits, leading to the opportunity to represent the model in matrix form as follows:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{b_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{z_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{z_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_1} \\ \boldsymbol{u_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \end{bmatrix}$$

where $\mathbf{y_i}$ represents a vector of observations for i$^{th}$ trait, $\mathbf{b_i}$ corresponds to a vector for fixed effects for i$^{th}$ trait, $\mathbf{u_i}$ is a vector containing the animal random genetic effects for the i$^{th}$ trait, $\mathbf{e_i}$ is a vector of random residual effects for the i$^{th}$ trait. **Xi** and **Zi** are incidence matrices that relates observations in **y** to levels of fixed effects in **b** and random animal genetic effects in **u**, respectively. Regarding to the assumptions of this model (statistical moments), we have the following:

**E[y] = Xb**

**E[u] = E[e] = 0** → the mean of the random effects is assumed to be zero

Whereas the variances in general can be represented as follows:

$$\mathbf{Var} \begin{bmatrix} \boldsymbol{u_i} \\ \boldsymbol{e_i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G^*} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R^*} \end{bmatrix} \quad = \quad \mathbf{Var} \begin{bmatrix} \boldsymbol{u_1} \\ \boldsymbol{u_2} \\ \boldsymbol{e_1} \\ \boldsymbol{e_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A\sigma^2_{u_1}} & \boldsymbol{A\sigma_{u_1u_2}} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{A\sigma_{u_2u_1}} & \boldsymbol{A\sigma^2_{u_2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I\sigma^2_{e_1}} & \boldsymbol{I\sigma_{e_1e_2}} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I\sigma_{e_2e_1}} & \boldsymbol{I\sigma^2_{e_2}} \end{bmatrix}$$

Probably, a simpler way to represent the variances is separately, having the additive genetic variance and covariance matrix represented as:

$$\boldsymbol{G^*} = \mathbf{Var} \begin{bmatrix} \boldsymbol{u_1} \\ \boldsymbol{u_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma^2_{u_1}} & \boldsymbol{\sigma_{u_1u_2}} \\ \boldsymbol{\sigma_{u_2u_1}} & \boldsymbol{\sigma^2_{u_2}} \end{bmatrix} \otimes \mathbf{A}$$

where $\boldsymbol{\sigma^2_{u_1}}$ represents the additive genetic variance of trait 1, $\boldsymbol{\sigma_{u_1u_2}}$ and $\boldsymbol{\sigma_{u_2u_1}}$ corresponds to the additive covariances among the two traits and, $\boldsymbol{\sigma^2_{u_2}}$ is the additive genetic variance for trait 2. In the above, **A** again represents the Wright's numerator relationship matrix and $\otimes$ indicates the Kronecker product. The previous matrix, can be ultimately inverted as:

$$\mathbf{G^{-1}} = \boldsymbol{G^{*-1}} \otimes \mathbf{A^{-1}} = \begin{bmatrix} \boldsymbol{G^{11}} & \boldsymbol{G^{12}} \\ \boldsymbol{G^{21}} & \boldsymbol{G^{22}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g^{11}A^{-1}} & \boldsymbol{g^{12}A^{-1}} \\ \boldsymbol{g^{21}A^{-1}} & \boldsymbol{g^{22}A^{-1}} \end{bmatrix}$$

The variance and covariance matrix for the residual effects can be represented individually as follows:

$$\boldsymbol{R^*} = \mathbf{Var}\begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma^2_{e_1}} & \boldsymbol{\sigma_{e_1e_2}} \\ \boldsymbol{\sigma_{e_2e_1}} & \boldsymbol{\sigma^2_{e_2}} \end{bmatrix} \otimes \mathbf{I}$$

where $\boldsymbol{\sigma^2_{e_1}}$ represents the residual variance of trait 1, $\sigma_{e_1e_2}$ and $\sigma_{e_2e_1}$ are the residual covariances between the two traits and, $\boldsymbol{\sigma^2_{e_2}}$ is the residual variance for trait 2. Within the previous, **I** is an identity matrix whose order is equal to the number of animals within each respective trait and again, $\otimes$ indicates the Kronecker product. The previous matrix can be inverted as:

$$\mathbf{R^{-1}} = \boldsymbol{R^{*-1}} \otimes \mathbf{I^{-1}} = \begin{bmatrix} \boldsymbol{R^{11}} & \boldsymbol{R^{12}} \\ \boldsymbol{R^{21}} & \boldsymbol{R^{22}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R^{11}I} & \boldsymbol{R^{12}I} \\ \boldsymbol{R^{21}I} & \boldsymbol{R^{22}I} \end{bmatrix}$$

Finally, the mixed model equations (MME) for this example of a bivariate analysis can be presented as follows:

$$\begin{bmatrix} X'_1R^{11}X_1 & X'_1R^{12}X_2 & X'_1R^{11}Z_1 & X'_1R^{12}Z_2 \\ X'_2R^{21}X_1 & X'_2R^{22}X_2 & X'_2R^{21}Z_1 & X'_2R^{22}Z_2 \\ Z'_1R^{11}X_1 & Z'_1R^{12}X_2 & Z'_1R^{11}Z_1 + g^{11}A^{-1} & Z'_1R^{12}Z_2 + g^{12}A^{-1} \\ Z'_2R^{21}X_1 & Z'_2R^{22}X_2 & Z'_2R^{21}Z_1 + g^{21}A^{-1} & Z'_2R^{22}Z_2 + g^{22}A^{-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} X'_1(R^{11}y_1 + R^{12}y_2) \\ X'_2(R^{12}y_1 + R^{22}y_2) \\ Z'_1(R^{11}y_1 + R^{12}y_2) \\ Z'_2(R^{12}y_1 + R^{22}y_2) \end{bmatrix}$$

And solving for $\hat{b}_i$ and $\hat{u}_i$ we have:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} X'_1R^{11}X_1 & X'_1R^{12}X_2 & X'_1R^{11}Z_1 & X'_1R^{12}Z_2 \\ X'_2R^{21}X_1 & X'_2R^{22}X_2 & X'_2R^{21}Z_1 & X'_2R^{22}Z_2 \\ Z'_1R^{11}X_1 & Z'_1R^{12}X_2 & Z'_1R^{11}Z_1 + g^{11}A^{-1} & Z'_1R^{12}Z_2 + g^{12}A^{-1} \\ Z'_2R^{21}X_1 & Z'_2R^{22}X_2 & Z'_2R^{21}Z_1 + g^{21}A^{-1} & Z'_2R^{22}Z_2 + g^{22}A^{-1} \end{bmatrix} \begin{bmatrix} X'_1(R^{11}y_1 + R^{12}y_2) \\ X'_2(R^{12}y_1 + R^{22}y_2) \\ Z'_1(R^{11}y_1 + R^{12}y_2) \\ Z'_2(R^{12}y_1 + R^{22}y_2) \end{bmatrix}$$

The first description of a bivariate analysis combining categorical and quantitative traits was provided by Foulley et al. (1983). In that study, authors included birth weight as a continuous variable while calving difficulty was included as a binary trait (e.g., easy vs difficult calving). The limitation of this methodology was that it was only applicable with equal design matrices, was to say, when there were no missing observations for any trait. A year later, Foulley and Gianola (1984) described a method to perform bivariate analyzes including only categorical response variables by studying calf viability and calving ease data. However, their method still required that the two binary responses were recorded on every animal. Subsequently, the same research team developed models that overcame the issue of missing observations on some traits, but relying on the assumption that the same fixed effects were influencing all the response variables (Foulley and Gianola, 1986). Later, Foulley (1987) reported a method that supported the existence of different fixed effects affecting the traits in the multivariate analysis, but incapable to deal with missing data. It was not until 1993, when a methodology capable of dealing with unequal design matrices was presented, finally allowing the presence of trait-specific fixed effects and missing observations on the traits under study (Janss and Foulley, 1993). A further extension of the previous methodology was presented a couple of years later by Hoeschele et al. (1995). Authors generalized the multiple-trait genetic evaluation for binary and continuous traits, to an evaluation in which a categorical trait having more than two expressions (e.g., a polychotomous trait) and several continuous traits were included, allowing for missing data and unequal models. Hoeschele's methodology has been employed world-wide to perform multivariate genetic evaluations of several reproductive, productive and conformational traits in cattle (Lee et al., 2002; Matilainen et al., 2009; Jeyaruban et al., 2012).

Among the advantages of using MTM over single-trait models (STM) was that animals without observations for one of the traits considered in the multivariate analysis, could still have a prediction for that specific trait based on its genetic correlation with the rest of the traits considered in the evaluation (Schaeffer and Wilton, 1981). Furthermore, another good property of MTM was that they can remove the selection bias that may be present in single trait analysis (Pollak et al., 1984). The previous capability has been very relevant to the beef industry since often, one trait is used to decide whether animals should remain in the herd and be recorded for other traits (e.g., weaning weight performance may determine if an animal will still be considered for traits measured later in life). Additionally, perhaps the main advantage of MTM was that they increased the accuracy of genetic evaluations. As discussed by (Mrode, 2014), the gain in accuracy depends on the absolute difference between the genetic and residual correlations existent among the traits included in the analysis. The greater the absolute difference in correlations, the greater was the reduction in the prediction error variance for the traits under analysis and the larger the accuracy gains (Schaeffer, 1984). However, it is important to mention that the accuracy gains in the predictions of the traits evaluated using a MTM were inversely related to the similarity of their heritability. For instance, when the heritabilities of the traits included in a multivariate analysis were equal or close to each other, the predictions yielded by this approach were practically equivalent to the evaluations performed using a univariate methodology. Conversely, when differences in the heritabilities of the traits included in a MTM exist, predictions for the lowly-heritable trait result more benefited in terms of accuracy gains than those for the highly-heritable trait (Thompson and Meyer, 1986).

Several studies have been performed utilizing MTM in order to explore the possible genetic correlations among different reproductive traits, as well as, the possible genetic relationships

between reproductive and performance traits (Mwansa et al., 2002; Forni and Albuquerque, 2005; Rasali et al., 2005). Such research efforts have tried to elucidate what traits could form part of an ideal multivariate evaluation intended to promote faster genetic improvement in reproductive traits. In this context, a recent report suggested the possibility of enhance the accuracy of predictions for fertility traits by evaluating them in a multivariate approach that incorporate traits more densely recorded that possess favorable genetic correlations with reproductive traits (Boldt et al., 2018). Specifically, authors suggested the inclusion of preweaning gain records in genetic evaluations for HPG and, incorporating ultrasound back fat observations in genetic evaluations for STAY.

When molecular information became available in the form of single nucleotide polymorphisms (SNP), it was of special interest to identify genomic regions associated with particular quantitative traits of economic importance, the so-called "Quantitative Trait Loci" or simply "QTL" (Soller, 1990). In this regard, the superiority of MTM over STM in the mapping of QTL was demonstrated by Jiang and Zeng (1995), who explained that by taking into account the correlated structure of multiple traits, it was possible to increase the power of detection of QTL. In the same context, authors suggested that QTL mapping using MTM was an effective procedure to test a number of biologically interesting hypotheses concerning the nature of genetic correlations between different traits. More recently, several studies using field and simulated data, have demonstrated that the use of MTM for genomic selection based on single-step procedures, yielded higher prediction accuracies than their corresponding individual single-step STM procedures (Tsuruta et al., 2011, Calus and Verkaamp, 2011; Jia and Jannink, 2012; Guo et al., 2014).

Despite all the advantages of MTM, they also possess limitations that require consideration when performing genetic evaluations. The main limitation of this methodology comes from the

fact that, the dimensions of the set of equations to solve increases in a quadratic way (Gutiérrez, 2010). For instance, with a two-trait model, the blocks of equations to solve is equal to 4 ($2^2 = 4$), however, if a three-trait model is applied, the blocks of equations that will require solutions increases to 9 ($3^2 = 9$; assuming that only direct genetic effects are being estimated). Evidently, increments in the number of equations are followed by considerable increments in the computational cost required to solve them. Mrode (2014) explained that the cost of multiple analysis on *n* traits is always more than the cost of *n* single analysis. Furthermore, the susceptibility of MTM to become highly-dimensional rather quickly (over-parametrization), hampers their suitability to work in a feasible way with longitudinal traits (Speidel et al., 2010).

*2.4.3 Random regression models (RRM)*

As an alternative to overcome the problem of over-parametrization in multiple-trait analyses, random regression models (RRM) were introduced to the community of animal breeding by Henderson in September of 1982 (Henderson, 1982). About 3 months later, Laird and Ware (1982) reinforced the idea that the use of random effects models was the appropriate approach to study longitudinal data. Authors explained that statistical models intended to analyze longitudinal data must recognize the relationship (e.g., covariance structure) between serial observations of the same response variable on the same experimental unit. In this context, it was 1994 when RRM were employed to analyze field data in livestock. These models were applied to milk test-day records in dairy cattle, allowing the shape of the lactation curve to be different for individual cows by including random regression coefficients for each animal (Schaeffer and Dekkers, 1994).

A straightforward explanation about the theory behind the application of RRM to livestock data was provided by Jamrozik and Scheaffer (1997a). In general, when the trait of interest in a genetic evaluation was longitudinal in nature, it was possible to take measurements on an

individual of the same underlying trait at many different time points. Now, applying the same concept to a group of animals (for instance, a contemporary group), there was an opportunity to estimate the overall phenotypic trajectory for the trait using information from all the animals within the group (fixed regression). However, in a more interesting way, there was also the opportunity to model the deviations around that phenotypic trajectory for each one of the animals that was part of the contemporary group (random regressions). The estimation of those individual deviations from the group trajectory was the particular purpose of RRM, and represented the main reason of why RRM have become the method of choice for the analysis of longitudinal traits (Schaeffer, 2004). Even when the main application of these models was found in the analysis of milk test-day records in dairy cattle, other applications of RRM include growth traits in all species, genotype by environment interactions, as well as, the analysis of survival data and fertility data (Schaeffer, 2004). In matrix form, RRM can be specified as:

$$\mathbf{y = Xb + Z_1u + Z_2pe + e}$$

where **y** represents a vector of repeated observations of the trait of interest for each animal, **b** is a vector of fixed effects and fixed regressions, **u** corresponds to a vector of random regressions for animal additive genetic effects, **pe** is a vector of random permanent environmental regression coefficients for each animal. **X** is an incidence matrix relating observations in **y** to fixed effects and fixed regressions contained in **b**, $\mathbf{Z_1}$ represents an incidence matrix of covariates that relates observations in **y** to animal random additive genetic regression coefficients in **u**, $\mathbf{Z_2}$ corresponds to an incidence matrix of covariates relating observations in **y** to random permanent environmental regressions in **pe** and, **e** is a vector of random residual terms that include temporary environmental effects.

With respect to the assumptions of these models, as reviewed by Schaeffer (2004), when the response variable **y** is normally distributed and the variances and covariances are known, it is not necessary to make assumptions about the distributions of **y** and the other random variables in the model to derive best linear unbiased predictors (BLUP) or the MME (Goldberger, 1962; Henderson, 1984). However, the variance components required to solve the MME are not typically known in advance in practice, therefore, it is necessary to estimate them from the data set. Similar to other models, variance components of a RRM can be estimated using REML (Ghiasi and Carabaño, 2018) or Bayesian (Jamrozik and Schaeffer, 1997a) methods. Since historically, REML procedures have been adopted as the preferred method for estimating genetic parameters (Gianola and Rosa, 2015), the model assumptions to estimate the variance components for a RRM via the REML approach will be shown first. The model assumptions for a typical RRM can be described as:

$\mathbf{E[y] = Xb}$

$\mathbf{E[u] = E[pe] = E[e] = 0}$ → the mean of the random effects is assumed to be zero

and

$$\mathbf{V} = \mathbf{Var}\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{p_e} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \otimes \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R} \end{bmatrix}$$

where **A** represents the Wright's numerator relationship matrix, ⊗ indicates the Kronecker product, **G** corresponds a variance-covariance matrix of additive genetic random regression coefficients, **I** is an identity matrix with an order equal to the number of observations, **P** is a variance-covariance matrix of the permanent environmental random regression coefficients and finally, **R** can represent a diagonal matrix of temporary environmental variances that depending on the specifications of the model, they can vary or not depending on time (or the specific

continuous covariate implemented). When temporary environmental variances are allowed to vary, then a heterogeneous residual variance is assumed; therefore, the residual variance structure is $\mathbf{var[e] = diag\{\sigma^2_{e_i}\}}$, where $\boldsymbol{i}$ represents the total number of differing residual variances. Conversely, when a homogeneous random residual variance is assumed, then $\mathbf{R = I\sigma^2_e}$ (Schaeffer, 2004; Speidel et al., 2010; Oliveira et al. 2019a). Nonetheless, it's worth mentioning that when the assumption of a homogeneous residual variance does not hold across all the values of the specific continuous covariate implemented in the model, ignoring the necessity of modelling a heterogeneous residual variance could lead to over- or under-estimations of heritability values for the trait under study (Olori et al., 1999). Interestingly, the assumption of homogeneous residual variance has no effect on the estimation of the permanent environmental variance (López-Romero et al., 2003).

If a Bayesian method is implemented to estimate the variance components of a RRM (considering a heterogeneous residual variance), then normality of the random variables must be assumed as follows:

$$\mathbf{y \mid b, u, p_e, \sigma^2_{e_i} \sim N\ (Xb + Z_1u + Z_2pe, R),}$$

and $$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{p_e} \\ \boldsymbol{e} \end{bmatrix} \sim \mathbf{N\ [0, V]}$$

where $\mathbf{V = Var} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{p_e} \\ \boldsymbol{e} \end{bmatrix}$, has the same aforementioned structure:

$$\mathbf{V = Var} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{p_e} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \otimes \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R} \end{bmatrix}$$

Regardless of the method implemented to estimate the variance components, once they are obtained, the MME for a typical RRM are:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_1 & X'R^{-1}Z_2 \\ Z_1'R^{-1}X & Z'_1R^{-1}Z_1 + A^{-1} \otimes G^{-1} & Z'_1R^{-1}Z_2 \\ Z_2'R^{-1}X & Z_2'R^{-1}Z_1 & Z'_2R^{-1}Z_2 + I \otimes P^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \\ pe \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z_1'R^{-1}y \\ Z_2'R^{-1}y \end{bmatrix}$$

And solving for $\widehat{\boldsymbol{b}}$, $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{p_e}}$ we have:

$$\begin{bmatrix} \hat{b} \\ \hat{u} \\ \widehat{p_e} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_1 & X'R^{-1}Z_2 \\ Z_1'R^{-1}X & Z'_1R^{-1}Z_1 + A^{-1} \otimes G^{-1} & Z'_1R^{-1}Z_2 \\ Z_2'R^{-1}X & Z_2'R^{-1}Z_1 & Z'_2R^{-1}Z_2 + I \otimes P^{-1} \end{bmatrix} \begin{bmatrix} X'R^{-1}y \\ Z_1'R^{-1}y \\ Z_2'R^{-1}y \end{bmatrix}$$

In the above set of equations, some elements that deserve more emphasis in their description are the covariates included in the incidence matrices $\mathbf{Z_1}$ and $\mathbf{Z_2}$. In order to explain their inclusion to RRM, first is important to explain that the aforementioned covariates have a phenotypic correlation with the observations in **y**. The origin of this relationship is given by the fact that longitudinal traits are recorded multiple times during an individual's lifetime or physiological cycle; therefore, expressions of the phenotypes in these traits are linked to the specific time point (or age) in which they are recorded (Oliveira et al., 2019a). Alternatively, records in **y** can also be taken along some spatial scale, or any other continuous covariate capable of influence the phenotypic expression of our trait of interest (e.g., ambient temperature). In general, these covariates represent "control variables" (typically regarded as the **x** variables in regression analyzes) and our traits of interest (**y**) are complete curves or trajectories rather than individual data points (Meyer and Kirkpatrick, 2005). When applying RRM to analyze this type of information, we want to quantify genetic values and their dispersion structure among records of **y** for the complete range of values of the control variable (Shaeffer, 2004; Meyer and Kirkpatrick, 2005). As such, these covariates are used to perform both the fixed regressions (phenotypic trajectories within contemporary groups) and the random regressions (individual animal deviations around the phenotypic trajectory of their contemporary group).

Longitudinal traits could also be referred as function-valued traits, where the expression "function-valued" emphasizes that the corresponding biological curves could be described by a mathematical function (Meyer and Kirkpatrick, 2005). Mathematical functions have been employed to provide a smooth trajectory of observations in **y** over the covariate utilized in the analysis (Schaeffer, 2004). The idea of implementing mathematical functions to link observations in **y** to the covariate **x**, comes from the fact that observations in **y** are dependent upon the specific value of **x** associated with its measurement. As such, phenotypes of individuals evaluated with RRM can be better described by mathematical functions rather that a finite set of measurements (Kirkpatrick and Heckman, 1989). Mathematically, the goal of the implemented function is to describe the covariance among records measured at different values of the covariate being used, therefore, they are referred as covariance functions (Kirkpatrick and Heckman, 1989). Covariance functions using Legendre polynomials have been commonly recommended to link observations between **x** and **y** within RRM, because they provide smooth curves similar to those observed in biological curves of interest (e.g., growth curves; Kirkpatrick et al., 1990). Since Legendre polynomials have been widely used in many studies, the procedure to calculate them will be presented as was summarized by Speidel et al. (2010):

Given $P_0$ (x) and $P_1$(x) are defined to be:

$$P_0(x) = 1 \text{ and } P_1(x) = x$$

Subsequent Legendre polynomials $P_{n+1}$(x) are of the form:

$$P_{n+1}(x) = \frac{1}{n+1}\left[(2n+1)xP_n(x) - nP_{n-1}(x)\right]$$

Which then are normalized as:

$$\Phi_n(x) = \sqrt{\frac{2n+1}{2}}\,P_n(x)$$

Table 2.1 displays how a fourth-order polynomial could be calculated using the aforementioned formulas for a normalized Legendre polynomial.

**Table 2.1.** Normalized Legendre polynomials up to a fourth order polynomial

| Order | Legendre Polynomial | Normalized Legendre Polynomial |
|---|---|---|
| n = 0 | $P_1(x) = x$ | $\Phi_0(x) = 0.7071$ |
| n = 1 | $P_2(x) = \frac{3}{2} x^2 - \frac{1}{2}$ | $\Phi_1(x) = 1.2247\ x$ |
| n = 2 | $P_3(x) = \frac{5}{2} x^3 - \frac{9}{6} x$ | $\Phi_2(x) = 2.3717\ x^2 - 0.7906$ |
| n = 3 | $P_4(x) = \frac{35}{8} x^4 - \frac{45}{12} x^2 \frac{3}{8}$ | $\Phi_3(x) = 4.6771\ x^3 - 2.8062x$ |
| n = 4 | $P_5(x) = \frac{63}{8} x^5 - \frac{35}{4} x^3 \frac{15}{8} x$ | $\Phi_4(x) = 9.2808\ x^4 - 7.9550x^2 + 0.7955$ |

Then, the estimated normalized Legendre polynomials can be entered into a matrix called

$$\Lambda' = \begin{vmatrix} 0.7071 & 0 & 0 & 0 & 0 \\ 0 & 1.2247 & 0 & 0 & 0 \\ -0.7906 & 0 & 2.3717 & 0 & 0 \\ 0 & -2.8062 & 0 & 4.6771 & 0 \\ 0.7955 & 0 & -7.9550 & 0 & 9.2808 \end{vmatrix}$$

Once the Legendre polynomials are normalized, they are combined with standardized values of the covariate being used. Legendre polynomials are defined within the range of -1 to 1 (Kirkpatrick et al., 1990); therefore, the covariate needs to be standardized within the same range. Shaeffer (2004) presented the formula to standardize covariates as follows:

$$t_i^* = -1 + 2\left[\frac{t_i - t_{min}}{t_{max} - t_{min}}\right]$$

where $\boldsymbol{t_i^*}$ represents the standardized value of the covariate implemented (here the letter t was used in reference to "time"), $\mathbf{t_i}$ represents the value of the covariate subjected to the

standardization and, $\mathbf{t_{min}}$ and $\mathbf{t_{max}}$ corresponds to the lowest and the highest value of the covariate contained the data, respectively. The standardized covariate values are placed into a matrix named **M**, whose order depends on the maximum number of values of the covariate within the dataset (rows) and the order of Legendre polynomial implemented (columns). There are several orders of Legendre polynomials that can be used (e.g., linear, quadratic, cubic, etc.), however, often simpler orders are preferred. As an example of how to form an **M** matrix, consider a linear RRM with Legendre polynomials as the base function, then, let an age vector of ten consecutive parities be:

$$t_i = [3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12]^T$$

The **M** matrix of standardized covariate values would be:

$$\mathbf{M} = \begin{bmatrix} 1 & -1.0000 \\ 1 & -0.7778 \\ 1 & -0.5556 \\ 1 & -0.3333 \\ 1 & -0.1111 \\ 1 & 0.1111 \\ 1 & 0.3333 \\ 1 & 0.5556 \\ 1 & 0.7778 \\ 1 & 1.0000 \end{bmatrix}$$

The first column of the **M** matrix is a column of ones representing the intercept of the curve, whereas the second column corresponds to the standardized ages. The final combination between the standardized covariate values and a set of linear Legendre polynomials could be performed by forming the matrix **Φ**, since **Φ = MΛ**. For instance:

$$\mathbf{\Phi} = \begin{bmatrix} 1 & -1.0000 \\ 1 & -0.7778 \\ 1 & -0.5556 \\ 1 & -0.3333 \\ 1 & -0.1111 \\ 1 & 0.1111 \\ 1 & 0.3333 \\ 1 & 0.5556 \\ 1 & 0.7778 \\ 1 & 1.0000 \end{bmatrix} \begin{bmatrix} 0.7071 & 0 \\ 0 & 1.2247 \end{bmatrix} = \begin{bmatrix} 0.7071 & -1.2247 \\ 0.7071 & -0.9526 \\ 0.7071 & -0.6804 \\ 0.7071 & -0.4083 \\ 0.7071 & -0.1361 \\ 0.7071 & 0.1361 \\ 0.7071 & 0.4083 \\ 0.7071 & 0.6804 \\ 0.7071 & 0.9526 \\ 0.7071 & 1.2247 \end{bmatrix}$$

The values contained in the **Φ** matrix would be the covariate values to implement in the incidence matrices relating observations in **y** to both fixed and random effects, within the MME of a RRM that uses the number of parities as the unit of time.

An important characteristic of RRM is that regardless of whether or not phenotypes of each animal are recorded at all points within the range of values of the covariate implemented, these models can obtain EBV for all individuals included in the pedigree at any value inside of the range of the covariate (White, 1999; Speidel et al., 2010; Stinchcombe et al., 2012). For instance, as detailed by Oliveira et al. (2019a), the vector of estimated breeding values ($\mathbf{EBV_j}$) of animal j, including all possible values of the specific covariate used, can be obtained as follows:

$$\mathbf{EBV_j} = \mathbf{\Phi}\hat{\boldsymbol{a}}_j$$

where **Φ** is a matrix of independent covariates for all time points (e.g., all ages) associated with the function used, and $\hat{\boldsymbol{a}}_j$ is the vector of EBV for the covariance function coefficients of animal j.

Benefits of the previous feature of RRM relates to the possibility of performing selection decisions based on each animal's entire trajectory over a relevant biological performance curve (Kirkpatrick et al., 1990; Meyer and Kirkpatrick, 2005; Stinchcombe et al., 2012). Moreover, possibilities exist to refine breeding programs by selecting over specific points within the performance curves in where more genetic variability exist in our trait of interest (De Haas et al., 2007, Tsuruta et al., 2009; Yin et al., 2014). The previous imply that it could be feasible to modify the patterns of performance curves in order to create curves with a more desirable shape depending on the production system (Oliveira et al., 2019b).

With respect to the application of RRM to analyze fertility data, Schaeffer (2004) provided a description of how these models may be used to analyze reproductive traits in cattle. The author

suggested that genetic merit for reproductive performance could change over time, so he recommended the use of parity number as the unit of time being observed. Then, he suggested to limit the number of parities to a maximum of ten and to standardize the parity numbers from -1 (for first parity) to +1 (for tenth parity). Finally, after explaining the procedure to standardize the age covariate, he suggested a set of fixed and random effects that could be included in a RRM intended to analyze longitudinal fertility traits. A couple of years later, Averill et al. (2006) applied RRM for male and female fertility evaluations in dairy cattle using longitudinal binary data. Genetic correlations between BCS and fertility traits such as days to first service, days between first and last insemination, calving interval, number of services per conception and FSCR were analyzed in Holstein cattle using a series of bivariate RRM (De Haas et al., 2007). In 2009, again using dairy cattle records, Tsuruta et al. (2009) applied bivariate analysis of conception rates and test-day milk yields using a threshold-linear model with random regressions. Similarly, Brügemann et al. (2013) implemented a bivariate threshold-linear sire RRM to assess the impact of the temperature and humidity index over female fertility (e.g., conception rate) on the phenotypic as well as on the genetic scale. Conception rates of Thai dairy cows were analyzed with random regression threshold models as a function of the days in milk during the lactation period (Buaban et al., 2016).

In the case of beef cattle, RRM have also been employed to perform genetic evaluations of fertility-related traits. Jamrozik et al. (2013) estimated genetic parameters for STAY to consecutive calvings in Canadian Simmentals using these models. Later, a series of bivariate RRM were employed to estimate the genetic association of SC with female reproductive traits such as first calving interval (FCI), AFC, HPG, and STAY in Nelore cattle (Santana et al., 2015a). Santana et al. (2015b) applied RRM to describe the pattern of phenotypic expression different economically

relevant traits in Nelore cattle reared under varying tropical conditions. Among the traits included in the previous study, the fertility-related trait was the one referred as "PRODAM", which according to Santana et al. (2013), was defined as the weight (in kg) of weaned calves produced annually by a cow during the time she stayed in the herd (similar to STAY). In 2017, Sánchez-Castro et al. (2017) estimated expected progeny differences for STAY in Angus cattle using RRM with Legendre polynomials as the base function. Genetic correlations dependent on environmental conditions between growth traits (hip height, body weight at 18 months of age and post weaning gain) and HPG, were explored by Santana et al. (2018) by applying a multiple-trait RRM in Nelore cattle. Silva et al. (2018) applied linear RRM to fit STAY to consecutive calvings of Guzerá, Nelore and Tabapuã cows in order to estimate genetic parameters for this trait. Speidel et al. (2018b) developed a prototype RRM genetic prediction for HPG utilizing Red Angus cattle data. Whereas Golden et al. (2018) extended the use of random regressions to the framework of single-step hybrid marker effects models when analyzing STAY data of Hereford cattle. More recently, Sánchez-Castro et al. (2019) analyzed the stability of genetic predictions for STAY in Angus cattle using RRM that included endpoints beyond 6 years of age. Altogether, previous reports suggest that the application of RRM to the analysis of longitudinal traits in beef cattle is feasible and may help to improve the accuracy of genetic predictions for fertility-related traits (Speidel et al., 2010; Jamrozik et al., 2013; Sánchez-Castro et al., 2019).

Several other advantages exist when using RRM as opposed to other statistical methodologies when analyzing longitudinal traits. Perhaps the most relevant advantage of RRM rely on the fact that they enable fitting random genetic and environmental effects over time, while accounting for different time-dependent nongenetic effects affecting the trait of interest over the course of the phenotypic curve (Meyer, 1998; Swalve, 2000; Mark, 2004). These capabilities

ultimately lead to higher accuracies of estimated breeding values compared with other statistical approaches (Oliveira et al., 2019a). In this regard, utilizing simulated data, Meyer (2004) reported that accuracies obtained with RRM were consistently higher than those estimated through MTM. Later, Boligon et al. (2011) reported that breeding value accuracy estimates for growth traits using RRM were more reliable than those obtained with MTM in Nelore cattle. More recently, Sánchez-Castro et al. (2019) suggested that the mean accuracy for 6-year STAY EPD estimated with various RRM, was about 4.6 times higher than the accuracy obtained with a TM.

Meyer (1998) suggested that since RRM more adequately use the covariance structure of traits that change gradually along some continuous scale, they overcome problems associated with oversimplifications incurred when using repeatability models or the typical overparameterizations faced by MTM. Wilson et al. (2005) gave a straightforward explanation of how RRM can reduce the number of parameters to estimate when posing a hypothetical evaluation considering a series of age-specific traits linked by a covariance structure. Specifically, the authors used "size" as the underlying trait subjected changes related to age (continuous covariate). They explained that for five age-specific size assessments, the additive genetic variance-covariance matrix will contain 15 parameters to estimate (five variances and 10 covariances). However, if the additive genetic values can be adequately modeled as a first-order linear function of time using RRM, then this number of parameters can be reduced to three (corresponding to the variances in intercept and slope and the covariance between them). In this sense, Stinchcombe et al. (2012) explained that by using a function of a lower order than the number of observations per individual, fewer parameters were required to be estimated, resulting in enhanced power and accuracy. In addition, the RRM approach do not require records be measured at the same time in all individuals or a minimum number of observations per animal, which enables the use of all the phenotypic data available in

all individuals when performing genetic analyzes (Schaeffer, 2004; Boligon et al., 2011; Sánchez-Castro et al., 2019). Additionally, another advantage of RRM is their flexibility to include censored data, a feature that has been proven to be beneficial in terms of accuracy gains when performing genetic evaluations for traits measured late in an animal's life as survival, longevity and STAY (Verkaamp et al. 2001; Jamrozik et al., 2013; van Pelt et al., 2015).

Domínguez-Viveros et al. (2015) noted as a good property of RRM that they can analyze directly raw phenotypes of each animal, without transformations or arbitrary adjustments that reduce the natural time-dependent variability of longitudinal traits. In contrast, others consider attractive that RRM can also be employed using the threshold methodology. According to Averill et al. (2006) longitudinal threshold animal models offer the possibility of computing quantities of interest to animal breeders that could not be obtained using cross-sectional analyses, such as the probability of observing a success or failure within a specific period. Furthermore, in situations when a limited number of binary outcomes exist per animal, it is possible to adapt the additive genetic numerator relationship matrix typically included in RRM in such way that, sire relationship matrices are built in order to apply sire-threshold RRM to perform the genetic analyzes (Tsuruta et al., 2009; Yin et al., 2012; Brügemann et al., 2013). In addition, the easy extension of single-trait RRM to multiple-trait RRM has been also recognized as good property of these models. Perhaps the most noted advantage that multiple-trait RRM have over single-trait RRM, is the possibility to estimate genetic correlations between different traits over time. The first application of a multiple-trait RRM involved Canadian Holstein cattle traits such as milk yield, milk fat yield, milk protein yield and somatic cell score (Jamrozik et al., 1997b). According to Oliveira et al. (2016), such types of estimates could allow the identification of the most feasible time periods to perform indirect selection and boost genetic gains thru correlated responses.

Some controversy exists with respect to how to utilize covariance functions in multiple-trait RRM analyzes. Originally, Meyer and Hill (1997) suggested that on certain occasions, it could be desired to fit more than one covariance function when applying a multiple-trait RRM, since measurements taken were representative of different characters or physiological processes. However, as reviewed by Oliveira et al. (2019a), the majority of the genetic evaluations that have been made using multi-trait RRM have used the same functions (e.g., Legendre polynomials or splines) to model the random effects of all traits. The idea of using the same function to describe genetic and permanent environmental effects (PE) when employing RRM apparently was based in ensuring that both curves had equal flexibility (Pool and Meuwissen, 1999; Pool et al., 2000). Jamrozik et al. (2001) also suggested that the same mathematical function should be used to describe random effects of a trait and stated that its selection should be determined by its goodness of fit to the performance curve of interest at the phenotypic level. However, recent reports have suggested that the combination of different functions to describe different traits in multiple-trait RRM was feasible and might improve the breeding values and genetic parameter estimates (Oliveira et al., 2016, 2017). Furthermore, even when using the same function in the analysis, a more adequate fit to field data has been reported when using different orders of fit for the direct genetic and the PE effects, generally, a lower degree is needed for the direct genetic effects than for PE effects (Pool et al., 2000; López-Romero and Carabaño, 2003; Kheirabadi et al., 2014).

Finally, with a constantly increasing availability of SNP panels capable of span the entire genome of the major livestock species (Eggen, 2012), an important feature of RRM is that they can accommodate the inclusion of genomic data. Unfortunately, incorporation of genomic information within the framework of RRM for the prediction of longitudinal traits had not received extensive attention until recently (Koivula et al., 2015; Kang et al., 2017). The feasibility and

advantages of including genomic information within the RRM framework have been proven in both simulated and real data of plants, animals and even humans (Kang et al., 2017, Sun et al., 2017, Oliveira et al., 2019c). In general, the addition of genetic markers to the genetic evaluations performed using RRM have exhibited robust prediction ability in longitudinal trait analyses by achieving higher accuracies and unbiasedness (Koivula et al., 2015; Kang et al., 2018; Oliveira et al., 2019c). Flexibility to apply single-trait or multiple-trait genomic RRM either in a single-step or a two-steps approach (Jattawa et al., 2016, Baba et al., 2017; Oliveira et al., 2019b), represent a major advantage of these models and allows the selection of breeding animals based on the complete pattern of the performance curve using genomic information (Oliveira et al., 2019a). Moreover, it is also possible to identify QTL associated with time-dependent variations on economically important traits in livestock, which may help to better understand phenotypic variations in longitudinal traits over time and to have a better insight of the biological timeline of gene effects (Das et al., 2011; Strucken et al., 2011, Oliveira et al., 2019d). Additionally, as suggested by Speidel et al. (2018b) and demonstrated by Golden et al. (2018), given the similarity of RRM to the traditional linear mixed models, the incorporation of genomic information into an evaluation using the recently developed single-step hybrid marker effects models (Fernando et al., 2014, 2016), was not a difficult task.

Among the concerns and limitations related to the application of RRM in animal breeding, probably the most discussed issue was the selection of the appropriate mathematical function to fit the data of the performance curve of interest. For instance, Misztal et al. (2000) suggested that the choice of the mathematical function to describe the lactation curve of dairy cattle was a key element when fitting RRM. As it was previously mentioned, Legendre polynomials have been widely used in genetic evaluations because they provide smooth curves similar to those observed

in biological curves of interest (Kirkpatrick et al., 1990). However, Legendre polynomials have potential limitations worth of noting, for instance, that higher-order polynomials are “wiggly” and do not have asymptotes (Pletcher and Geyer, 1999). The problem associated with higher-order polynomials was mathematically known as "Runge's phenomenon" and suggested that the error of a polynomial approximation of a curve increases with the polynomial order of fit, with errors predominantly located at the extremes of the curve (de Boor, 2001; Meyer, 2005a). In the context of quantitative genetics, Shaeffer and Jamrozik (2008) reported that when using Legendre polynomials, the estimated covariance matrices used to calculate genetic variances over the range of data, tend to result in genetic variances that were much higher at the beginning and end of the data range than in the middle. As reviewed by Speidel et al. (2010), the previous could be due to the fact that polynomials place a large emphasis on observations at the extremes, aggravating the appearance of the Runge’s phenomenon. Besides the rapid changes of high-order terms at the extremes and poor modelling capability of asymmetrical functions, Misztal (2006) also mentioned that other problems of Legendre polynomials were their lack of information to estimate a very large number of parameters and their sensitivity to each of the many different (co)variance parameters.

In order to overcome the issues related to the use of Legendre polynomials as the base function of RRM, several research efforts have been conducted in order to find better covariance functions to fit performance curves typical of livestock species (White et al., 1999; Torres and Quaas, 2001; Robert-Granié et al., 2002). Meyer (2005a) explained that an alternative to the use of high degree polynomials are to use “piece-wise polynomials”, which basically are curves constructed from pieces of lower degree polynomials commonly referred as “splines”, joined smoothly at selected points known as “knots”. Splines have demonstrated to be advantageous since

knots have the ability to connect the different segments of biological curves in a better way (White et al., 1999; de Boor, 2001). Meyer (2005b) investigated the use of a particular type of spline called "B-splines" (the "B" stands for basis) to model growth in Australian Angus cattle and reported that B-splines outperformed the results obtained using Legendre polynomials. Misztal (2006) explored the properties of RRM using linear splines as the base function by comparing their performance to the one obtained with Legendre polynomials. In summary, the author reported that linear splines had better convergence than Legendre polynomials when solving the MME. However, Misztal also explained that when using splines, potential drawbacks are the depression of variances and predictions in the middle of intervals between the knots, and inflation of predictions close to knots. According to the authors, the previous issues could be greatly reduced by adjusting the number and positions of knots, so he provided a useful guide to do it:

1) The first two knots to choose must be those harboring all points on the trajectory occurring in the data
2) The remainder knots should be added in such way that correlations between adjacent knots are in the range of 0.6–0.8

Speidel et al. (2010) noted that the aforementioned suggestions will result in knots being placed close together around areas that have the largest data density, as well as, in areas where the data are changing more rapidly. Even when splines seem to have clear advantages over Legendre polynomials, only small differences in accuracies (about 2.5%) favored the use of splines as opposed to Legendre polynomials in a simulation study (Bohmanova et al., 2005). Furthermore, a study performed with real data of Brazilian Gyr cattle, suggested that, in practice, RRM using either splines or Legendre polynomials as base functions were able to rank animals almost exactly,

since Spearman's correlations between EBV obtained with both types of models were in the range of 0.946 to 0.998 (Pereira et al., 2013).

*2.5 Genomic selection*

In 2001, a revolutionary paper not only for animal but also for plant breeding was published. Even anticipating the technology constraints of that time, Meuwissen et al. (2001) proposed a novel approach in where the breeding values could be estimated from markers spanning the entire genome (Boichard et al., 2014; Van Eenennaam et al., 2014). As explained by Gianola and Rosa (2015), the mathematical rationale behind the proposal of Meuwissen et al. (2001) was relatively simple: given a battery of p SNP and a sample of n individuals genotyped for such markers, the fitting of a multiple linear regression on the number of copies of a reference allele at each one of the p loci could predict the total genetic value of an individual. Due to the fact that the breeding value predictions relied on the use of genome-wide dense marker platforms, this type of selection was subsequently termed "genomic selection" (Eggen, 2012; Van Eenennaam et al., 2014). Goddard and Hayes (2007) explained that genomic selection was a form of marker-assisted selection (MAS) in which genetic markers covering the whole genome were used under the assumption that all QTL were in linkage disequilibrium (LD) with at least one marker. As opposed to MAS in where a prior knowledge of gene or marker associations with the traits of interest is required, the genomic selection approach infers that there will always be a SNP in close proximity to a particular gene or DNA fragment of interest. Therefore, by means of indirect associations based on LD assumptions, a significant fraction of the variation in a trait of interest could be explained (Eggen, 2012). Genomic Selection has a tremendous potential to accelerate the rate of genetic improvement within any specie and, in order to better understand how it does it, it is important to recall some basic principles of animal breeding. Considering that the purpose of any

selection program is to accelerate the rate of genetic change or selection response per unit of time (ΔG) toward a given breeding objective, the classic equation for explaining ΔG, as described by Falconer and Mackay (1996) is:

$$\Delta G = \frac{i * r * \sigma_A}{L}$$

where $\boldsymbol{i}$ represents the selection intensity, $\boldsymbol{r}$ represents the accuracy of selection, $\boldsymbol{\sigma_A}$ represents the genetic variability of the trait of interest and $\boldsymbol{L}$ represents the generation interval. Within the previous formula, any technology capable of increasing accuracy, intensity, and/or genetic variation or decreasing the generation interval, has the potential to accelerate the rate of genetic gain (Van Eenennaam et al., 2014). Genomic Selection has been shown to improve the accuracy of traditional genetic evaluations based on pedigree and phenotypes alone in several livestock species (Wolc et al., 2011; Saatchi et al., 2015; Hidalgo et al., 2015). Within the case of dairy cattle, adoption of genomic selection has resulted in high intensities of selection and shorter generation intervals (Pryce and Daetwyler, 2012). In general, two main approaches have been developed in order to perform genomic selection in livestock: a multiple-step approach (VanRaden, 2008; Hayes et al., 2009) and the single-step approach (Legarra et al., 2009; Aguilar et al., 2010; Legarra et al., 2014). Brief descriptions of each procedure are provided below.

*2.5.1 Multiple-step genomic selection approach*

Within the multiple-step genomic selection procedure, the first step consists in the calculation of genomic breeding values (GEBV) as the sum of the effects of a multitude of genetic markers or QTL (quantitative trait loci = genes affecting a quantitative trait) across the entire genome (Hayes et al., 2009; Lourenco et al., 2017). Estimation of such GEBV require a large population for which phenotype and genotype (typically SNP) data must be available. This population if often referred as the "reference or training population", since prediction equations

are trained and calibrated within this subset of animals (Van Eenennaam et al., 2014). Once the marker effects have been estimated, the previously mentioned prediction equations can be used to predict the genetic value of another population of individuals with genotypes but without phenotypes, the so-called "validation population" (Goddard et al., 2016; Koivula et al., 2016). Genetic values produced for members of the validation population were then based solely on their molecular make up and, therefore, referred to as Molecular Breeding Values (MBV). These MBV required to subsequently be blended with traditional EPD or used as correlated traits in multivariate analyses (Kachman et al., 2013). A general overview of the multiple-step genomic selection approach applied in livestock is shown in Figure 2.3 (Van Eenennaam et al., 2014).

Possibilities of applying molecular-based predictions to estimate genetic values of animals virtually since the moment of their birth, represented an attractive feature of this method due to the opportunity of reducing the generation interval (Daetwyler, 2009). Furthermore, significant increments in accuracy of breeding value estimations of genotyped animals as well as a more optimal utilization of available genetic resources through genome-guided mate selection were also interesting benefits of genomic selection (Daetwyler et al., 2013).

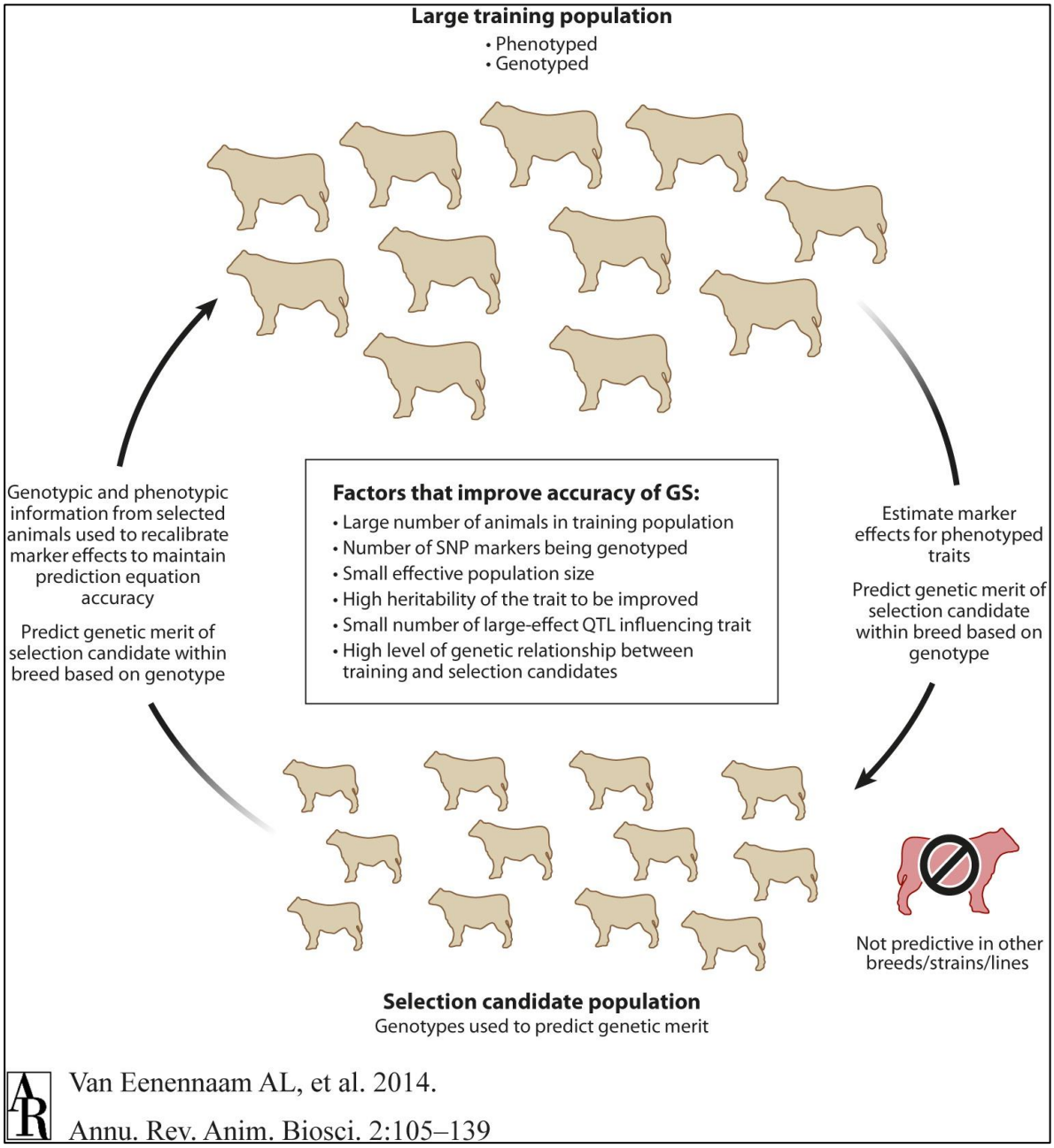


**Figure 2.3.** Overview of the multiple-step genomic selection approach applied in livestock (Abbreviations: QTL, quantitative trait loci; SNP, single-nucleotide polymorphism)

Moreover, the multiple-step genomic also contained drawbacks that were summarized by Koivula et al. (2015). Some of the problems of this method were that parent averages (PA) of progeny of genomically selected animals do not automatically include genomic information. Additionally, when animals were selected by their GEBV, the future estimation of unbiased EBV

was difficult because genomic information was not taken into account in the traditionally employed methods of genetic prediction (preselection bias; Misztal et al.,2009). Moreover, genomic selection using the multiple-step approach is prone to bias since includes several approximations (e.g., blending), all of which reduce the accuracy and can inflate the resultant GEBV. Another important drawback of the multiple-step procedure is that prediction equations require to be periodically updated or retrained as a consequence of using SNPs in LD with phenotypes, rather than causal mutations. The previous leads to a decay in the accuracy of genomic predictions as the number of generations separating the training population from the validation population increases (Taylor, 2014). Furthermore, MBV (or GEBV) could only be generated for simple models, for instance single-trait models with no maternal components (Lourenco et al., 2017). Given all these deficiencies, it has been stated that the elimination of multiple-step methods and the migration to single-step genomic prediction procedures represent one of the largest evolutions of the utilization of genomic information within current genetic evaluation systems (Spangler, 2018).

*2.5.2 Single-step genomic selection approach*

With the goal of simplifying the multiple-step genomic selection procedure, a methodology capable of incorporating molecular information (e.g., marker data) into the traditional mixed model equations using phenotypes and pedigree data was proposed and termed single-step procedure. A couple of papers by Legarra et al. (2009) and Misztal et al. (2009) detailed that such procedure was based on the modification of the typical numerator relationship matrix **A** to include genomic information. Essentially, the main idea was to adjust the relationships between animals based on the similarities among their genotypes. By doing so, genomic information was incorporated into

the classical BLUP methodology originating what is known as single-step-GBLUP (Figure 2.4; Legarra et al., 2014; Lourenco et al., 2017).

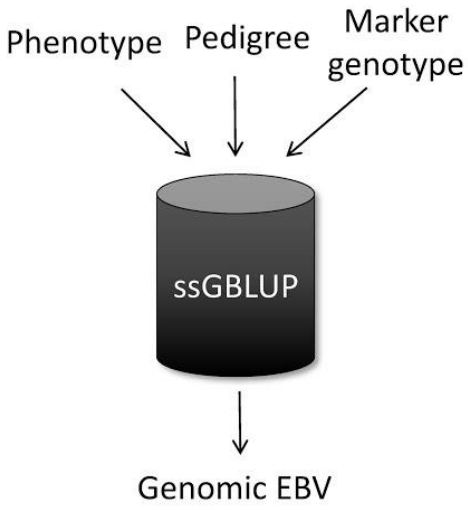


**Figure 2.4**. Simplistic overview of the single-step genomic selection approach (Abbreviations: ssGBLUP, single-step genomic best linear unbiased predictor; EBV, estimated breeding value. Lourenco et al., 2017).

Using selection index principles, Legarra et al. (2009) outlined the procedure to blend the complementary information of molecular markers to the historically recorded and available pedigree data. In parallel, Misztal et al. (2009) detailed the required computational methods to achieve such combination of information. For their part and practically at the same time, Christensen and Lund (2010) proposed the same idea (combination of pedigree and DNA markers), but departing from a different perspective based on the imputation of missing genotypes within non-genotyped individuals. Following Legarra's derivations (Legarra et al., 2009, 2014), a brief summary of the methodology to combine the numerator relationship matrix **A,** with a marker-derived genomic relationship matrix **G** in order to create a modified genetic relationships matrix **H** will be presented.

Considering that 1 and 2 refer to non-genotyped and genotyped individuals, respectively; authors started from the situation that before markers are observed in a subset of a population, the joint distribution of breeding values of the future genotyped animals and non-genotyped individuals is multivariate normal:

$$p\begin{pmatrix}u_1\\u_2\end{pmatrix} = N(0, \sigma_u^2 A)$$

with a covariance matrix of the form:

$$Var\begin{pmatrix}u_1\\u_2\end{pmatrix} = \sigma_u^2 A = \sigma_u^2 \begin{pmatrix}A_{11} & A_{12}\\A_{21} & A_{22}\end{pmatrix}$$

where $A_{11}$, $A_{12}/A_{21}$ and $A_{22}$ are partitions of the numerator relationship matrix **A** (based only in pedigree) and $\boldsymbol{\sigma_u^2}$ representing the additive genetic variance. In a Bayesian context, the joint distribution presented before can be split into the product of a marginal and a conditional density as follows:

$$p(u_1|u_2) = \underbrace{p(u_1|u_2)}_{\text{marginal}}\underbrace{p(u_2)}_{\text{conditional}}$$

As such, this joint distribution is distributed as:

$$p(u_1|u_2) = N\big(\underbrace{A_{12}A_{22}^{-1}u_2}_{\text{mean}}, \underbrace{\sigma_u^2(A_{11} - A_{12}A_{22}^{-1}A_{21})}_{\text{variance}}\big)$$

However, after observing the genotypes of the markers for the subset of genotyped individuals within the initial population, their relationships are no longer based on pedigree averages, but instead they are fully informative observed genomic relationships. Such relationships are then agglutinated into what is known as the genomic relationship matrix **G**. Therefore, after observing the genotypes we have that:

$$p(u_2|genotypes) = N(0, \sigma_u^2 G)$$

Interestingly, genotypes also influence the relationships among non-genotyped animals and between non-genotyped and genotyped individuals. In consequence, the joint distribution of both kinds of individuals conditional on the observed genotypes converts to:

$$p(u_1, u_2|genotypes) = \underbrace{p(u_1|u_2)}_{\text{marginal}} \underbrace{p(u_2|genotypes)}_{\text{conditional}}$$

From here, elements of the final matrix **H** can be derived and the original covariance matrix can be appropriately modified as follows:

$$Var\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \sigma_u^2 H = \sigma_u^2 \begin{pmatrix} A_{11-}A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21} & A_{12}A_{22}^{-1}G \\ GA_{22}^{-1}A_{21} & G \end{pmatrix}$$

Finally, although matrix **H** looks complicated and is completely dense, the form of its inverse ($H^{-1}$) is much simpler (Aguilar et al., 2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

At this point, the $\boldsymbol{H^{-1}}$ matrix could easily replace the regular $\boldsymbol{A^{-1}}$ matrix and all the framework within Henderson's best linear unbiased prediction methodology holds. Consequently, any model utilizing relationships matrices can be fitted using the combined relationship matrix **H**. This means that evaluations using the single-step procedure would not be restricted to simple models (e.g., single-trait models with no maternal components), but rather, the method allows the application of more complex evaluations such as those performed using multivariate models with maternal components, threshold models and/or even random regression models (Misztal et al., 2009; Legarra et al., 2014; Kang et al., 2017). Other advantages of the single-step over the multiple-step procedure include the ability of the single-step approach to automatically account for all relatives of genotyped individuals and their performance. Accordingly, increments in the accuracy of predictions are not limited only to genotyped individuals since their non-genotyped

relatives also result benefitted (Christensen et al., 2012). Additionally, elimination of the extra evaluation steps contributes to evade the loss of information (Legarra et al., 2009).

Despite of the great benefits of the single-step procedure, potential challenges and drawbacks have been also acknowledged in the comprehensive review of Legarra et al. (2014). For instance, it is recognized that the method increases the programming complexity to fit complicated marker effects models such as those based on Bayesian regressions. Additionally, given that the nature and assumptions of the method rely on Fisher's infinitesimal model, the single-step procedure does not have a way to handle appropriately QTL with major effects (although this is also a limitation within multiple-step procedures). Furthermore, and perhaps more importantly, since the method requires explicitly the inverse of a dense G matrix, the constantly increasing number of genotyped individuals could eventually reach a limit in where the computation of such inverse would become impossible (e.g., >100,000), threatening the feasibility of this methodology (Fernando et al., 2014, 2016).

In order to overcome potential limitations of single-step procedures due to the increasing number of genotyped individuals, procedures such as the Algorithm for Proven and Young (APY) have been developed (Misztal et al., 2014). Based on the recursive algorithm of Henderson (1976b) implemented to obtain ($\mathbf{A}^{-1}$) without explicitly creating $\mathbf{A}$, the APY methodology builds $\mathbf{G}^{-1}$ directly for a subset of the most influential genotyped animals denominated "core animals". To do so, APY assumes that the genomic recursions for young animals (non-core individuals) contain coefficients only for proven animals (core individuals); therefore, it is possible to ignore the relationships among non-core animals in the construction of $\mathbf{G}^{-1}$ at the cost of a negligible impact on the estimation of genomic breeding values (Fragomeni et al., 2015). From a different angle, an alternative strategy to avoid limitations imposed by the increasing number of genotyped animals

is the implementation of methods which do not require computing **G** or its inverse, deriving in the development of the single-step Bayesian regression marker effects models, also known as single-step hybrid models (Fernando et al., 2014, 2016).

*2.5.3 Single-step hybrid models*

As acknowledged by creators of the ssGBLUP methodology, an alternative method to combine phenotype, pedigree and genotype data for populations with genotyped and non-genotyped individuals, was to impute markers in ungenotyped animals via marker and pedigree information and estimate marker effects after imputation was completed (Legarra et al., 2009). In this regard, this avenue was explored by Fernando et al. (2013) who proposed a methodology in where after genotypes were imputed for the non-genotyped proportion of the population, all animals were subsequently treated as genotyped individuals. This methodology was termed marker effects hybrid model and as important notes, breeding values of animals were expressed as the sum of marker effects estimated within the analysis. However, an important consideration of an extra term in the model to account for imputation errors was made for the originally non-genotyped animals, for which their breeding values were expressed as the sum of the effects of their imputed marker genotypes plus their corresponding imputation residuals (Fernando et al., 2014).

Although the uncertainty associated with the genotype imputation process received some criticism, it was the requirement of the method to store large intermediate data files corresponding to the imputed genotypes as well as the software restrictions to accomplish that task the factors considered as the crucial limitations for hybrid models (Misztal et al., 2014). Nonetheless, appropriate considerations of the model to account for imputation residuals and, more importantly, the development of computational strategies to avoid storing large and dense blocks of the mixed model equations involving imputed genotypes, allowed the consolidation of the single-step super

hybrid model (ss-SHM; Fernando et al., 2016). The key feature of the refinement of hybrid models consisted in utilizing a breeding value type model for animals with missing genotypes, rather than expressing their breeding values as the sum of the effects of their imputed marker genotypes plus their separate imputation residuals $\boldsymbol{\epsilon}$ (reaffirming even more the hybrid nature of the method).

As stated by its creators, the ss-SHM is computationally attractive for pedigree files containing millions of animals with a large proportion of genotyped individuals; essentially, because the method does not require computing the G matrix or its inverse (Fernando et al., 2016). Importantly, accompanying the establishment of the statistical theory behind the ss-SHM, the required software developments for its application were attained in parallel (Golden et al., 2016); allowing its rapid employment within genetic evaluation procedures (Garrick et al., 2018). According to Fernando et al. (2016), for a single-trait evaluation a ss-SHM as the following form:

$$\begin{bmatrix} y_n \\ y_g \end{bmatrix} = \begin{bmatrix} X_n \\ X_g \end{bmatrix} \mathrm{b} + \begin{bmatrix} 0 & Z_n \\ Z_g M_g & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ u_n \end{bmatrix} + \mathrm{e},$$

where the subscripts **n** and **g** refer to non-genotyped and genotyped individuals, respectively; $\boldsymbol{X_n}$ and $\boldsymbol{X_g}$ are appropriate incidence matrices relating fixed effects in **b** to observations in **y** (specifically sorted with **n** individuals first and **g** individuals after). Similarly, $\boldsymbol{Z_n}$ and $\boldsymbol{Z_g}$ correspond to incidence matrices relating random marker effects in **α** (e.g., breeding values) and $\boldsymbol{u_n}$ (where $\boldsymbol{u_n = M_n \alpha + \epsilon}$) to observations in **y**. **M** denote a matrix of centered marker values (typically coded as -1, 0 or 1); and **e** representing a vector of random errors. Then, after computing the inverse of the covariance matrix for random effects the MME are given by:

$$\begin{bmatrix} X'X & X'Z_g M_g & X_n'Z_n \\ M_g'Z_g'X_g & Q & M_g'A^{gn}\frac{\sigma_e^2}{\sigma_g^2} \\ Z_n'X_n & A^{ng}M_g\frac{\sigma_e^2}{\sigma_g^2} & Z_n'Z_n + A^{nn}\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \\ \hat{u}_n \end{bmatrix} = \begin{bmatrix} X'y \\ M_g'Z_g'y_g \\ Z_n'y_n \end{bmatrix}$$

where Q = $M_g' Z_g' Z_g M_g + \mathrm{I}\frac{\sigma_e^2}{\sigma_\alpha^2} + M_n' A^{nn} M_n \frac{\sigma_e^2}{\sigma_g^2}$, a matrix with dimensions equal to the number of marker covariate (usually <50,000) that can be stored in memory during the iterative procedure implemented to solve the equation's system (Fernando et al., 2016; Mäntysaari et al., 2019).

Among the main advantages of the ss-SHM is their ability to allow for alternative prior distributions for marker effects (although only for genotyped animals). This means that within this particular approach, inabilities of ssGBLUP and multiple-step genomic selection procedures to appropriately account for the presence of QTL with major effects is partially overcomed. Interestingly, when considering the predecessor of the ss-SHM (e.g., regular hybrid model), since in that procedure all animals are treated as genotyped individuals after the imputing process is done; then it is possible to utilize different prior distributions for marker effects for all animals (not just genotyped), at the cost of dealing with the large and dense blocks of the MME pertaining to the imputed genotypes. Given this important feature, it has been indicated that if in the future it becomes useful to give different weights for different SNP effects, or to fit different SNP for different traits in multi-trait models, the marker effects hybrid models have a clear advantage over methods relying on the genomic relationship matrix G (Mäntysaari et al., 2019).

*Literature cited*

Aceves, C., C. Romero, L. Sahagún, and R. Valverde-R. 1987. Thyroid hormone profile in dairy cattle acclimated to cold or hot environmental temperatures. Eur. J. Endocrinol. 114:201-207.

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. J. Dairy Sci. 93:743-752.

Amaral, T. B., E. S. Correa, and F. P. Costa. 2003. Aspectos económicos do uso de touros melhoradores em um sistema de produção de gado de corte. In: 40ª Reunião Anual da Sociedade Brasileira de Zootecnia, Santa Maria, São Paulo.

Aono, F. H., R. F. Cooke, A. A. Alfieri, and J. L. M. Vasconcelos. 2013. Effects of vaccination against reproductive diseases on reproductive performance of beef cows submitted to fixed-timed AI in Brazilian cow-calf operations. Theriogenology. 79:242-248.

Archer, J. A., P. F. Arthur, P. F. Parnell, and R. J. Van de Ven. 1998. Effect of divergent selection for yearling growth rate on female reproductive performance in Angus cattle. Livest. Prod. Sci. 57:33-40.

Averill, T., R. Rekaya, and K. Weigel. 2006. Random regression models for male and female fertility evaluation using longitudinal binary data. J. Dairy Sci. 89:3681-3689.

Azzam, S. M., J. E. Kinder, and M. K. Nielsen. 1990. Modelling reproductive management systems for beef cattle. Agric. Syst. 34:103-122.

Ball, P. J. H., and A. R. Peters. 2004. Reproduction in Cattle. 3rd ed. Blackwell Publishing, Oxford, UK.

Barber, K. A. 1983. Maximizing the impact of dairy and beef bulls through breeding technology. J. Dairy Sci. 66:2661-2671.

Barth, A. D. 1993. Factors affecting fertility with artificial insemination. Vet. Clin. N. Am-Food A. 9:275-289.

Bellows, R. A., R. E. Short, D. C. Anderson, B. W. Knapp, and O. F. Pahnish. 1971. Cause and effect relationships associated with calving difficulty and calf birth weight. J. Anim. Sci. 33:407-415.

Bellows, D. S., S. L. Ott, and R. A. Bellows. 2002. Cost of reproductive diseases and conditions in cattle. Prof. Anim. Sci. 18:26-32.

Bennett, G. L. 2008. Experimental selection for calving ease and postnatal growth in seven cattle populations. I. Changes in estimated breeding values. J. Anim. Sci. 86:2093-2102.

Berry, D. P., E. Wall, and J. E. Pryce. 2014. Genetics and genomics of reproductive performance in dairy and beef cattle. Animal 8:105-121.

Bohmanova, J., I. Misztal, and J. K. Bertrand. 2005. Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. J. Anim. Sci. 83:62-67.

Boichard, D., V. Ducrocq, P. Croiseau, and S. Fritz. 2016. Genomic selection in domestic animals: principles, applications and perspectives. CR. Biol. 339:274-277.

Boldt, R. J. 2017. Genetic parameters for fertility and production traits in Red Angus cattle. Master's Thesis, Colorado State University, Fort Collins.

Boldt, R. J., S. E. Speidel, M. G. Thomas and R. M. Enns .2018. Genetic parameters for fertility and production traits in Red Angus cattle. J. Anim. Sci. 96:4100-4111.

Boligon, A. A., F. Baldi, M. E. Z. Mercadante, R. B. Lobo, R. J. Pereira, and L. G. D. Albuquerque. 2011. Breeding value accuracy estimates for growth traits using random regression and multi-trait models in Nelore cattle. Genet. Mol. Res.10:1227-1236.

Bonamy, M., J. J. de Iraola, A. Baldo, A. Prando, G. Giovambattista, and S. Munilla. 2018. Early rather than late scrotal circumference measurements better reflect female precocity in beef cattle. Livest. Sci. 218:79-84.

Bormann, J. M., L. R. Totir, S. D. Kachman, R. L. Fernando, and D. E. Wilson. 2006. Pregnancy rate and first-service conception rate in Angus heifers. J. Animal Sci. 84:2022-2025.

Bourdon, R. M., and J. S. Brinks. 1982. Genetic, environmental and phenotypic relationships among gestation length, birth weight, growth traits and age at first calving in beef cattle. J. Anim. Sci. 55:543-553.

Bourdon, R. M. 2000. Understanding animal breeding (second edition). Upper Saddle River, NJ: Prentice Hall.

Brigham, B. W., S. E. Speidel and R. M. Enns. 2006. Alternative Definitions of Stayability. In Proc. Beef Imp. Fed. 38th Annu. Res. Symp. Annu. Meet., Chotcaw, MS. pp. 75-78.

Brigham, B. W., S. E. Speidel, R. M. Enns, and D. J. Garrick. 2007. Stayability to alternate ages. Proc. West. Sec. Amer. Soc. Anim. Sci. 58:27-30.

Brinks, J. S., J. M. Mclnerney and P. J. Chenoweth. 1978. Relationships of age at puberty in heifers to reproductive traits in young bulls. Proc. West. Sec. Amer. Soc. of Anim. Sci. 29:28-30.

Brügemann, K., E. Gernand, U. U. von Borstel, and S. König. 2013. Application of random regression models to infer the genetic background and phenotypic trajectory of binary conception rate by alterations of temperature× humidity indices. Livest. Sci. 157:389-396.

Buaban, S., K. Kuchida, M. Suzuki, Y. Masuda, W. Boonkum, and M. Duangjinda. 2016. Genetic analysis of the rates of conception using a longitudinal threshold model with random regression in dairy crossbreeding within a tropical environment. Anim. Sci. J. 87:961-971.

Buddenberg, B. J., C. J. Brown, Z. B. Johnson, J. E. Dunn, and H. P. Peterson. 1989. Heritability estimates of pregnancy rate in beef cows under natural mating. J. Anim. Sci. 67:2589-2594.

Burfening, P. J., D. C. Anderson, R. A. Kinkie, J. Williams, and R. L. Friedrich. 1978. Synchronization of estrus with PGF2α in beef cattle. J. Anim. Sci. 47:999-1003.

Burns, B.M., G. Fordyce, and R. G. Holroyd. 2010. A review of factors that impact on the capacity of beef cattle females to conceive, maintain a pregnancy and wean a calf—Implications for reproductive efficiency in northern Australia. Anim. Reprod. Sci. 122:1-22.

Caldow, G., B. Lowman, and I. Riddell. 2005. Veterinary intervention in the reproductive management of beef cow herds. In Practice. 27:406-411.

Calus, M. P., and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43:26.

Cammack, K. M., M. G. Thomas, and R. M. Enns. 2009. Reproductive traits and their heritabilities in beef cattle. Prof. Anim. Sci. 25:517-528.

Campos, G. S., F. A. Reimann, P. I. Schimdt, L. L. Cardoso, B. P. Sollero, J. Braccini, M. J. Yokoo, A. A. Boligon, and F. F. Cardoso. 2019. Threshold and linear models for genetic evaluation of visual scores in Hereford and Braford cattle. Anim. Prod. Sci. 59:619-627.

Chapman, H. D., J. M. Young, E. G. Morrison, and N. C. Edwards, Jr. 1978. Differences in lifetime productivity of Herefords calving first at 2 and 3 years of age. J. Anim. Sci. 46:1159.

Chenoweth, P. J., F. M. Hopkins, J. C. Spitzer, and J. C. Larsen. 1993. Guidelines for using bull breeding soundness evaluation form. Theriogenology Elsevier Science Inc. Handbook B-10.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Gen. Sel. Evol. 42:2.

Colazo, M. G., and R. J. Mapletoft. 2014. A review of current timed-AI (TAI) programs for beef and dairy cattle. Can. Vet. J. 55:772-780.

Correa-Calderón, A., C. I. Gómez-Álvarez, L. Avendaño-Reyes, R. Diaz-Molina, S. Medina-Cervantes, F. Rivera-Acuña, P. Luna-Nevárez y U. Macias-Cruz. 2014. Efecto de la suplementación de progesterona y enfriamiento artificial post-inseminación sobre la eficiencia reproductiva de vaquillas Holstein durante el verano. Arch. Med. Vet. 46:189-196.

Coulter, G. H., T. R. Rounsaville, and R. H. Foote. 1976. Heritability of testicular size and consistency in Holstein bulls. J. Anim. Sci. 43:9-12.

Crews, D. H., and R. M. Enns. 2008. Genetic improvement of Heifer pregnancy and performance in beef cattle. In Proc. Appl. Rep. Strategies in Beef Cattle. http://beefrepro.unl.edu/proceedings/2008ftcollins/12_csu_genetics_crews.pdf (Accessed 7 August 2019).

Das, K., Li, J., Wang, Z., Tong, C., Fu, G., Li, Y., Xu, M., Ahn, K., Mauger, D., Li, R. and Wu, R. 2011. A dynamic model for genome-wide association studies. Hum. Genet. 129:629-639.

Daetwyler, H. D., 2009 Genome-wide evaluation of populations. Ph.D. Thesis, Wageningen University, Wageningen, The Netherlands.

Daetwyler, H. D., M. P. Calus, R. Pong-Wong, G. de los Campos, and J. M. Hickey. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics. 193:347-365.

Davis, G. P., N. J. Corbet, M. J. Mackinnon, D. J. S. Hetzel, K. W. Entwistle, and R. Dixon. 1993. Response in female fertility and calf growth to selection for pregnancy rate in tropical beef cattle. Aust. J. Agric. Res. 44:1509-1521.

Dearborn, D. D., R. M. Koch, L. V. Cundiff, K. E. Gregory, and G. E. Dickerson. 1973. An analysis of reproductive traits in beef cattle. J. Anim. Sci. 36:1032-1040.

de Boor, C. 2001. A practical guide to splines 2nd Ed. Springer. Verlag, NY. USA.

De Haas, Y., L. L. G. Janss, and H. N. Kadarmideen. 2007. Genetic correlations between body condition scores and fertility in dairy cattle using bivariate random regression models. J. Anim. Breed. Genet. 124:277-285.

De Kruif, A. 1978. Factors influencing the fertility of a cattle population. J. Reprod. Fert. 54:507-518.

Deutscher, Gene H., J. A. Stotts, and Merlyn K. Nielsen. 1991. Effects of breeding season length and calving season on range beef cow productivity. J. Anim. Sci. 69:3453-3460.

Diskin, M. G. and J. M. Sreenan. 2000. Expression and detection of oestrus in cattle. Reprod. Nutr. Dev. 40:481-491.

Domínguez-Viveros, J., F. A. Rodríguez-Almeida, R. Núñez-Domínguez, R. Ramírez-Valverde, J. Á. Ortega-Gutiérrez, and A. Ruíz-Flores. 2015. Estimation of genetic parameters in tropicarne cattle with random regression models using B-splines. Trop. Subtrop. Agroecosyst. 18:27-32.

Doyle, S. P., B. L. Golden, R. D. Green and J. S. Brinks. 2000. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. J. Anim. Sci. 78:2091-2098.

Dunn, T. G., and G. E. Moss. 1992. Effects of nutrient deficiencies and excesses on reproductive efficiency of livestock. J. Anim. Sci. 70:1580–1593.

Dziuk, P. J., and R. A. Bellows. 1983. Management of reproduction of beef cattle, sheep and pigs. J. Anim. Sci. 57(Suppl. 2):355-379.

Eggen, A. 2012. The development and application of genomic selection as a new breeding paradigm. Anim. Front. 2:10-15.

Eler, J. P., J. I. V. Silva, J. B. S. Ferraz, F. Dias, H. N. Oliveira, J. L. Evans and B. L. Golden. 2002. Genetic evaluation of the probability of pregnancy at 14 months for Nellore heifers. J. Anim. Sci. 80:951-954.

Enns, R. M., D. J. Garrick, and B. W. Brigham. 2005. Variability in economic value is dependent upon herd average stayability. J. Anim. Sci. 83 (Suppl. 2):112-115.

Evans, J. L., B. L. Golden, R. M. Bourdon, and K. L. Long. 1999. Additive genetic relationships between heifer pregnancy and scrotal circumference in Hereford cattle. J. Anim. Sci. 77:2621-2628.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th Ed. Longman, Harlow. Essex, UK.

Fragomeni, B. O., D. A. L. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Taylor and I. Misztal. 2015. Hot topic: use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. J. Dairy Sci. 98: 4090-4094.

Fernando, R. L., D. Garrick, and J. C. M. Dekkers. 2013. Bayesian regression method for genomic analyses with incomplete genotype data. Pages 26–30 in Proc. 64th Annual Meeting of the European Federation of Animal Science, August 2013, Nantes, France, Vol. 225.

Fernando, R. L., J. C. M. Dekkers and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50

Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96

Formigoni, I. B., J. A. I. D. V. Silva, R. C. Brumatti, J. B. S. Ferraz, and J. P. Eler. 2002. Economic aspects of stayability as selection criterion in beef cattle industry in Brazil. In: 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France.

Forni, S., and L. G. Albuquerque. 2005. Estimates of genetic correlations between days to calving and reproductive and weight traits in Nelore cattle. J. Anim. Sci. 83:1511-1515.

Foulley, J. L., D. Gianola, and R. Thompson. 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. Genet. Sel. Evol. 15:401-424.

Foulley, J. L. and D. Gianola. 1984. Estimation of genetic merit from bivariate «all or none» responses. Genet. Sel. Evol. 16:285-306.

Foulley, J. L., and D. Gianola. 1986. Sire evaluation for multiple binary responses when information is missing on some traits. J. Dairy Sci. 69:2681-2695.

Foulley J. L. 1987. Methods of genetic evaluation for discrete polygenic traits in animal breeding. ScD thesis, Université de Paris-Sud, centre d'Orsay.

Frasier, W. M., and G. H. Pfeiffer. 1994. Optimal replacement and management policies for beef cows. Am. J. Agr. Econ. 76:847-858.

Garrick, D. 2006. Evaluation, interpretation and relevance of stayability in beef cattle genetic improvement. J. Anim. Sci. 84(Suppl. 2):50.

Garrick, D. J., D. P. Garrick, and B. L. Golden. 2018. An introduction to BOLT software for genetic and genomic evaluations. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand.

Geary, T. W., J. C. Whittier, E. R. Downing, D. G. LeFever, R. W. Silcox, M. D. Holland, T. M. Nett, and G. D. Niswender. 1998. Pregnancy rates of postpartum beef cows that were synchronized using Syncro-Mate-B® or the Ovsynch protocol. J. Anim. Sci. 76:1523-1527.

Geary, T. W., J. C. Whittier, D. M. Hallford, and M. D. MacNeil. 2001. Calf removal improves conception rates to the Ovsynch and CO-Synch protocols. J. Anim. Sci. 79:1-4.

Ghiasi, H., and M. J. Carabaño. 2018. Eigenvector index for two female fertility traits based in random regression coefficients matrix in Holstein cows. Span. J. Agric. Res. 16:e401.

Gianola, D. 1982. Theory and analysis of threshold characters. J. Anim. Sci. 54:1079-1096.

Gianola, D., and J. L. Foulley. 1983. Sires evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201-224.

Gianola, D., and G. J. Rosa. 2015. One hundred years of statistical developments in animal breeding. Annu. Rev. Anim. Biosci. 3:19-56.

Givens, M. D. 2006. A clinical, evidence-based approach to infectious causes of infertility in beef cattle. Theriogenology. 66:648-654.

Grossi, D. A., O. G. Frizzas, C. C. P. Paz, L. A. F. Bezerra, R. B. Lôbo, J. A. Oliveira, and D. P. Munari. 2008. Genetic associations between accumulated productivity, and reproductive and growth traits in Nelore cattle. Livest. Sci. 117:139-146.

Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet. 124:323-330.

Goddard, M. E., K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, and B. J. Hayes. 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proc. R. Soc. B: Biol. Sci. 283:20160569.

Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits, a framework for the next generation of EPDs. In: Proc. Beef Improvement Federation, Wichita, KS. p. 2–13.

Golden, B. L. and D. J. Garrick. 2016. An introduction to BOLT. Theta solutions LLC. Pp 1-21.

Golden, B. L., S. Weerasinghe, B. Crook, S. Sanders, and D. J. Garrick. 2018. A Single-Step Hybrid Marker Effects Model Using Random Regression for Stayability in Hereford Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/single-step-hybrid-marker-effects-model-using-random-regression-stayability-hereford-cattle.pdf

Gutiérrez, J. P. 2010. Iniciación a la valoración Genética animal, Metodología adaptada al EEES. UCM Editorial Complutense, Madrid, SP.

Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. 15:30.

Gwazdauskas, F.C, C. J. Wilcox, and W. W. Thatcher. 1975. Environmental and managemental factors affecting conception rate in a subtropical climate. J. Dairy Sci. 58:88-92.

Gwazdauskas, F. C. 1985. Effects of climate on reproduction in cattle. J. Dairy Sci. 68:1568-1578.

Hansen, P. J. 2009. Effects of heat stress on mammalian reproduction. Phil. Trans. R. Soc. B: Biological Sciences. 364:3341–3350.

Harville, D. A., and R. W. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. Biometrics. 393-408.

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. 92:433–443.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics. 423-447.

Henderson, C. R., and R. L. Quaas. 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. 43:1188-1197.

Henderson, C. R. 1976b. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–93.

Henderson, C. R. 1982. Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. Biometrics. 3:623-640.

Hemsworth, P. H., J. L. Barnett, L. Beveridge, and L. R. Matthews. 1995. The welfare of extensively managed dairy cattle: A review. Appl. Anim. Behav. Sci. 42:161-182.

Herd, D. B., and L. R. Sprott. 1986. Body condition, nutrition and reproduction of beef cows. Texas Agric. Ext. Serv. Bull. 1526, College Station.

Hess, B. W., S. L. Lake, E. J. Scholljegerdes, T. R. Weston, V. Nayigihugu, J. D. C. Molle, and G. E. Moss. 2005. Nutritional controls of beef cow reproduction. J. Anim. Sci. 83(E. Suppl.):E90-E106.

Hidalgo, A. M., J. W. Bastiaansen, M. S. Lopes, B. Harlizius, M. A. Groenen, and D. J. de Koning. 2015. Accuracy of predicted genomic breeding values in purebred and crossbred pigs. G3: Genes, Genomes, Genetics, 5(8), pp.1575-1583.

Hoeschele, I., B. Tier, and H. U. Graser. 1995. Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. J. Anim. Sci. 73:1609-1627.

Hudson, G. F. S., and L. D. Van Vleck. 1981. Relationship between production and stayability in Holstein cattle. J. Dairy Sci. 64:2246-2250.

Huszenicza, G., M. Kulcsar, and P. Rudas. 2002. Clinical endocrinology of thyroid gland function in ruminants. Vet. Med. Czech. 47:199–210.

Janss, L. L. G., and J. L. Foulley. 1993. Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty. Livest. Prod. Sci. 33:183-198.

Jamrozik, J., and L. R. Schaeffer. 1997a. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J. Dairy Sci. 80:762-770.

Jamrozik, J., L. R. Schaeffer, Z. Liu, and G. Jansen. 1997b. Multiple trait random regression test day model for production traits. Interbull Bulletin. 16:43-47.

Jamrozik, J., D. Gianola, and L. R. Schaeffer. 2001. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. Livest. Prod. Sci. 71:223-240.

Jamrozik, J., S. McGrath, R. A. Kemp, and S. P. Miller. 2013. Estimates of genetic parameters for stayability to consecutive calvings of Canadian Simmentals by random regression models. J. Anim. Sci. 91:3634-3643.

Jeyaruban, G., B. Tier, D. Johnston, and H. Graser. 2012. Genetic analysis of feet and leg traits of Australian Angus cattle using linear and threshold models. Anim. Prod. Sci. 52:1-10.

Jia, Y. and Jannink, J.L. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics. 192:1513-1522.

Jiang, C., and Z. B. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics. 140:1111-1127.

Jordan, W. A., E. E. Lister, and G. J. Rowlands. 1968. Effect of planes of nutrition on wintering pregnant beef cows. Can. J. Anim. Sci. 48:145-154.

Jordan, E. R. 2003. Effects of heat stress on reproduction. J. Dairy Sci. 86:E104-E114.

Kachman, S. D., M. L. Spangler, G. L. Bennett, K. J. Hanford, L. A. Kuehn, W. M. Snelling, R. M. Thallman, M. Saatchi, D. J. Garrick, R. D. Schnabel, J. F. Taylor and J. Pollack. 2013. Comparison of molecular breeding values based on within-and across-breed training in beef cattle. Gen. Sel. Evol. 45:30.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J. F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity. 119:459-467.

Kastelic, J. P., and J. C. Thundathil. 2008. Breeding soundness evaluation and semen analysis for predicting bull fertility. Reprod. Domest. Anim. 43:368-373.

Kheirabadi, K., A. Rashidi, S. Alijani, and I. Imumorin. 2014. Modeling lactation curves and estimation of genetic parameters in Holstein cows using multiple-trait random regression models. Anim. Sci. J. 85:925-934.

Koivula, M., I. Strandén, J. Pösö, G. P. Aamand, and E. A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775-2784.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 10:1061-1066.

Koch, R. M., J. B. Gibb, and J. A. Gosey. 1986. Evaluation of beef cattle industry breeding programs: breeders and breed associations. In: 3th World Congr. Genet. Appl. Livest. Prod. Nebraska, United States of America.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics. 38:963-974.

Lamb, G. C., C. R. Dahlen, J. E. Larson, G. Marquezini, and J. S. Stevenson. 2010. Control of the estrous cycle to improve fertility for fixed-time artificial insemination in beef cattle: a review. J. Anim. Sci.88(Suppl. 13):E181-E192.

Lamb, G. C., and V. R. Mercadante. 2016. Synchronization and artificial insemination strategies in beef cattle. Vet. Clin. Food Anim. 32:335-347.

Land, R. B. 1973. The expression of female sex-limited characters in the male. Nature. 241:208-209.

Larson, J. E., G. C. Lamb, J. S. Stevenson, S. K. Johnson, M. L. Day, T. W. Geary, D. J. Kesler, J. M. DeJarnette, F. N. Schrick, A. DiCostanzo, and J. D. Arseneau. 2006. Synchronization of estrus in suckled beef cows for detected estrus and artificial insemination and timed artificial insemination using gonadotropin-releasing hormone, prostaglandin F2α, and progesterone. J. Anim. Sci. 84:332-342.

Lauderdale, J. W., B. E. Seguin, J. N. Stellflug, J. R. Chenault, W. W. Thatcher, C. K. Vincent, and A. F. Loyancano. 1974. Fertility of cattle following PGF2α injection. J. Anim. Sci. 38:964-967.

Lee, C. N. 1993. Environmental stress effects on bovine reproduction. Vet. Clin. N. Am-Food A. 9:263-273.

Lee, D., I. Misztal, J. K. Bertrand, and R. Rekaya. 2002. National evaluation for calving ease, gestation length and birth weight by linear and threshold model methodologies. J. Appl. Genet. 43:209-216.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54-65.

Lesmeister, J. L., P. J. Burfening, and R. L. Blackwell. 1973. Date of first calving in beef cows and subsequent calf production. J. Anim. Sci. 36:1-6.

López-Romero, P., R. Rekaya, and M. J. Carabaño. 2003. Assessment of homogeneity vs. heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. J. Dairy Sci. 86:3374-3385.

López-Romero, P., and M. J. Carabaño. 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein–Friesian cattle. Livest. Prod. Sci. 82:81-96.

Lourenco, D., K. Bertrand, H. Bradford, S. Miller, and I. Misztal. 2017. The Promise of Genomics for Beef Improvement. In BIF Conference. In: Proc. Beef Imp. Fed. 49th Annu. Res. Symp. Annu. Meet., Athens, GA. pp. 30-42.

Loyacano. A. L., C. K. Vincent, W. A. Nipper, and J. Pontif. 1972. Effect of supplemental energy during breeding and season of breeding on reproductive performance in beef cattle. Louisiana Livestock Producer's Day Rep.12:127.

Lucy, M. C., H. J. Billings, W. R. Butler, L. R. Ehnis, M. J. Fields, D. J. Kesler, J. E. Kinder, R. C. Mattos, R. E. Short, W. W. Thatcher, R. P. Wettemann, J. V. Yelich, and H. D. Hafs. 2001. Efficacy of an intravaginal progesterone insert and an injection of PGF2α for synchronizing estrus and shortening the interval to pregnancy in postpartum beef cows, peripubertal beef heifers, and dairy heifers. J. Anim. Sci. 79:982-995.

Maas, J. 1987. Relationship between nutrition and reproduction in beef cattle. Vet. Clin. N. AM-FOOD A. 3:633-646.

Mackinnon, M. J., J. F. Taylor, and D. J. S. Hetzel. 1990a. Genetic variation and covariation in beef cow and bull fertility. J. Anim. Sci. 68:1208-1214.

Mackinnon, M. J., D. J. S. Hetzel, N. J. Corbet, R. P. Bryan, and R. Dixon. 1990b. Correlated responses to selection for cow fertility in a tropical beef herd. Anim. Prod. 50:417-424.

MacNeil, M. D., J. J. Urick, and W. M. Snelling. 1998. Comparison of selection by independent culling levels for below-average birth weight and high yearling weight with mass selection for high yearling weight in Line 1 Hereford cattle. J. Anim. Sci. 76:458-467.

MacNeil, M. D. and N. Vukasinovic. 2011. A prototype national cattle evaluation for sustained reproductive success in Hereford cattle. J. Anim. Sci. 89:1712-1718.

Maiwashe, A., K. A. Nephawe, and H. E. Theron. 2009. Analysis of stayability in South African Angus cattle using a threshold model. S. Afr. J. Anim. Sci. 39:55-60.

Mark, T. 2004. Applied genetic evaluations for production and functional traits in dairy cattle. J. Dairy Sci. 87:2641-2652.

Marshall, D. M., W. Minqiang, and B. A. Freking. 1990. Relative calving date of first-calf heifers as related to production efficiency and subsequent reproductive performance. J. Anim. Sci. 68:1812-1817.

Martínez, G. E., R. M. Koch, L. V. Cundiff, K. Gregory, S. D. Kachman and L. D. Van Vleck. 2005. Genetic parameters for stayability, stayability at calving, and stayability at weaning to specified ages for Hereford cows. J. Anim. Sci. 83:2033-42.

Martínez-Velázquez, G., K. E. Gregory, G. L. Bennett, and L. D. Van Vleck. 2003. Genetic relationships between scrotal circumference and female reproductive traits. J. Anim. Sci. 81:395-401.

Matilainen, K., R. Mrode, I. Strandén, R. Thompson, and E. A. Mäntysaari. 2009. Linear–threshold animal model for birth weight, gestation length and calving ease in United Kingdom Limousin beef cattle data. Livest. Sci. 122:143-148.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. J. Dairy Sci. 103:5314-5326.

Meijering, A. 1984. Dystocia and stillbirth in cattle—A review of causes, relations and implications. Livest. Prod. Sci. 11:143-177.

Menegassi, S. R. O., J. O. J. Barcellos, V. D. N. Lampert, J. B. S. Borges, and V. Peripolli. 2011. Bioeconomic impact of bull breeding soundness examination in cow-calf systems. R. Bras. Zootec. 40:441-447.

Mercier, E., and G. W. Salisbury. 1947. Seasonal variations in hours of daylight associated with fertility level of cattle under natural breeding conditions. J. Dairy Sci. 30:747.

Melton, B. E. 1995. Conception to consumption: The economics of genetic improvement. In: Proc. Beef Improvement Federation 27th Research Symposium and Annual Meeting, Sheridan, Wyoming. pp. 40−47.

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

Meyer, K., K. Hammond, M. J. Mackinnon, and P. F. Parnell. 1991. Estimates of covariances between reproduction and growth in Australian beef cattle. J. Anim. Sci. 69:3533-3543.

Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. Genet. Sel. Evol. 30:221-240.

Meyer, K. 2004. Scope for a random regression model in genetic evaluation of beef cattle for growth. Livest. Prod. Sci. 86:69-83.

Meyer, K. 2005a. Advances in methodology for random regression analyses. Aust. J. Exp. Agr. 45:847-858.

Meyer, K. 2005b. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37:473-500.

Meyer, K., and M. Kirkpatrick. 2005. Up hill, down dale: quantitative genetics of curvaceous traits. Phil. Trans. R. Soc. B. 360:1443-1455.

Meyerhoeffer, D. C., R. P. Wettemann, S. W. Coleman, and M. E. Wells. 1985. Reproductive criteria of beef bulls during and after exposure to increased ambient temperature. J. Anim. Sci. 60:352–357.

Misztal, I., D. Gianola, and J. L. Foulley. 1989. Computing aspects of a nonlinear method of sire evaluation for categorical data. J. Dairy Sci. 72:1557-1568.

Misztal, I. 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. 123:74-80.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648-4655.

Morris, C. A., J. A. Wilson, G. L. Bennett, N. G. Cullen, S. M. Hickey, and J. C. Hunter. 2000. Genetic parameters for growth, puberty, and beef cow reproductive traits in a puberty selection experiment. New Zeal. J. Agr. Res. 43:83-91.

Mrode, R. A. 2014. Linear models for the prediction of animal breeding values. 3rd Ed. CAB Int. Wallingford, UK.

Morris, C. A., R. L. Baker, and N. G. Cullen. 1992. Genetic correlations between pubertal traits in bulls and heifers. Livest. Prod. Sci. 31:221–233.

Mwansa, P. B., D. H. Crews Jr, J. W. Wilton, and R. A. Kemp. 2002. Multiple trait selection for maternal productivity in beef cattle. J. Anim. Breed. Genet. 119:391-399.

Nardone, A., B. Ronchi, N. Lacetera, and U. Bernabucci. 2006. Climatic effects on productive traits in livestock. Vet. Res. Commun. 30:75-81.

Neely, J. D., B. H. Johnson, E. U. Dillard, and O. W. Robison. 1982. Genetic parameters for testes size and sperm number in Hereford bulls. J. Anim. Sci. 55:1033-1040.

Núñez-Dominquez, R., L. V. Cundiff, G. E. Dickerson, K. E. Gregory, and R. Koch. 1985. Effects of managing heifers to calve first at two vs three years of age on longevity and lifetime production of beef cows. Beef Res. Resp. 2. US. MARC.

Olesen, I., A. F. Groen, and B. Gjerde. 2000. Definition of animal breeding goals for sustainable production systems. J. Anim. Sci. 78:570-582.

Oliveira, H. D., F. F. Silva, O. H. G. B. D. Siqueira, N. O. Souza, V. S. Junqueira, M. D. V. Resende, R. R. A. Borquis, and M. T. Rodrigues. 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models. J. Anim. Sci. 94:1865-1874.

Oliveira, H. R., F. F. Silva, M. V. G. B. Silva, O. H. G. B. D. Siqueira, M. A. Machado, J. C. C. Panetto, L. S. Glória, and L. F. Brito. 2017. Bayesian models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. Livest. Sci. 201:78–84.

Oliveira, H. R., L. F. Brito, D. A. L. Lourenco, F. F. Silva, J. Jamrozik, L. R. Schaeffer, and F. S. Schenkel. 2019a. Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. J. Dairy Sci. 102:7664-7683.

Oliveira, H. R., L. F. Brito, F. F. Silva, D. A. L. Lourenco, J. Jamrozik, and F. S. Schenkel. 2019b. Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Holstein cattle. J. Dairy Sci. 102:452-463.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, and F. S. Schenkel. 2019c. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. J. Dairy Sci. 102:2365-2377.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019d. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:9995-10011.

Olori, V. E., W. G. Hill, and S. Brotherstone. 1999. The Structure of the Residual Error Variance of Test Day Milk Yield in Random Regression Models. In: Proceeding of Computational Cattle Breeding Workshop, March 18-20, 1999, Interbull Bulletin, Tuusula, 103-108.

Oyama, K., T. Katsuta, K. Anada, and F. Mukai. 2004. Genetic parameters for reproductive performance of breeding cows and carcass traits of fattening animals in Japanese Black (Wagyu) cattle. Anim. Sci. 78:195-201.

Patterson, D. J., R. C. Perry, G. H. Kiracofe, R. A. Bellows, R. B. Staigmiller, and L. R. Corah. 1992. Management considerations in heifer development and puberty. J. Anim. Sci. 70:4018-4035.

Pereira, R. J., A. B. Bignardi, L. El Faro, R. S. Verneque, A. E. Vercesi Filho, and L. G. Albuquerque. 2013. Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (Bos indicus) cattle. J. Dairy Sci. 96:565-574.

Pletcher, S. D., and C. J. Geyer. 1999. The genetic analysis of age-dependent traits: modeling the character process. Genetics. 151:825-835.

Pollak, E.J., J. Van der Werf, and R. L. Quaas. 1984. Selection bias and multiple trait evaluation. J. Dairy Sci. 67:1590-1595.

Pool, M. H., and T. H. E. Meuwissen. 1999. Prediction of daily milk yields from a limited number of test days using test day models. J. Dairy Sci. 82:1555-1564.

Pool, M. H., L. L. G. Janss, and T. H. E. Meuwissen. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. J. Dairy Sci. 83:2640-2649.

Pryce, J.E. and H. D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim. Prod. Sci. 52:107-114.

Pursley, J. R., M. O. Mee, and M. C. Wiltbank. 1995. Synchronization of ovulation in dairy cows using PGF2α and GnRH. Theriogenology. 44:915-923.

Rasali, D. P., G. H. Crow, J. N. B. Shrestha, A. D. Kennedy, and A. Brûlé-Babel. 2005. Multiple trait estimates of genetic parameters for juvenile growth and calving traits in Canadian Angus cattle. Can. J. Anim. Sci. 85:309-316.

Richards, M. W., J. C. Spitzer, and M. B. Warner. 1986. Effect of varying levels of postpartum nutrition and body condition at calving on subsequent reproductive performance in beef cattle. J. Anim. Sci. 62:300-306.

Robert-Granié, C., E. Maza, R. Rupp, and J. L. Foulley. 2002. Use of fractional polynomial for modelling somatic cell scores in dairy cattle. In: 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France.

Sánchez-Castro, M. A., R. J. Boldt, M. G. Thomas, R.M. Enns, and S. E. Speidel. 2017. Expected progeny differences for stayability in Angus cattle using a random regression model. J. Anim. Sci. 95(Suppl. 4):88.

Sánchez-Castro, M. A., M. G. Thomas, R. M. Enns, and S. E. Speidel. 2019. Stability of genetic predictions for stayability using random regression models that include end points beyond 6 yr of age. Trans. Anim. Sci. 3:1678-1682.

Santana, M. L., J. P. Eler, J. B. S. Ferraz, and E. C. Mattos. 2012. Genetic relationship between growth and reproductive traits in Nellore cattle. Animal. 6:565-570.

Santana, M. L., J. P. Eler, A. B. Bignardi, and J. B. S. Ferraz. 2013. Genetic associations among average annual productivity, growth traits, and stayability: A parallel between Nelore and composite beef cattle. J. Anim. Sci. 91:2566-2574.

Santana, M. L., J. P. Eler, A. B. Bignardi, and J. B. S. Ferraz. 2015a. Two-trait random regression model to estimate the genetic association of scrotal circumference with female reproductive performance in Nelore cattle. Theriogenology. 83:1534-1540.

Santana, M. L., J. P. Eler, A. B. Bignardi, A. Menéndez-Buxadera, F. F. Cardoso, and J. B. S. Ferraz. 2015b. Multi-trait linear reaction norm model to describe the pattern of phenotypic expression of some economic traits in beef cattle across a range of environments. J. Appl. Genet. 56:219-229.

Santana, M. L., J. P. Eler, A. B. Bignardi, A. A. Boligon, and J. B. Ferraz. 2018. Genetic correlation between growth and reproductive performance of beef females depends on environment. Anim. Prod. Sci. 58:1201-1209.

Saatchi, M., McClure, M.C., McKay, S. D., Rolf, M. M., Kim, J., Decker, J. E., Taxis, T. M., Chapple, R. H., Ramey, H. R., Northcutt, S. L. and Bauck, S. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet. Sel. Evol. 43:40.

Schaeffer, L. R. and J. W. Wilton. 1981. Comparison of single and multiple trait beef sire evaluations. Can. J. Anim. Sci. 61:565-573.

Schaeffer, L. R. 1984. Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67:1567-1580.

Schaeffer, L. R., and Dekkers, J. C. M. 1994. Random regressions in animal models for test-day production in dairy cattle. In: Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Ontario, Canada.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Schaeffer L. R. and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145-146.

Schillo, K. K., J. B. Hall, and S. M. Hileman. 1992. Effects of nutrition and season on the onset of puberty in the beef heifer. J. Anim. Sci. 70:3994-4005.

Seidel, G. E., Jr. 1995. Reproductive biotechnologies for profitable beef production. In: Proc. Beef Improvement Federation 27th Research Symposium and Annual Meeting, Sheridan, WY. Pp. 28-39.

Selk, G. E., R. P. Wettemann, K. S. Lusby, J. W. Oltjen, S. L. Mobley, R. J. Rasby, and J. C. Garmendia. 1988. Relationships among weight change, body condition and reproductive performance of range beef cows. J. Anim. Sci. 66:3153-3159.

Short, Robert E., and D. C. Adams. 1988. Nutritional and hormonal interrelationships in beef cattle reproduction. Can. J. Anim. Sci. 68:29-39.

Silva, D. O., M. L. Santana, D. R. Ayres, G. R. O. Menezes, L. O. C. Silva, P. R. C. Nobre, and R. J. Pereira. 2018. Genetic parameters for stayability to consecutive calvings in Zebu cattle. Animal. 12:1807-1814.

Smith, G. M., D. B. Laster, and K. E. Gregory. 1976. Characterization of biological types of cattle I. Dystocia and preweaning growth. J. Anim. Sci. 43:27-36.

Snelling, W. M., B. L. Golden, and R. M. Bourdon. 1995. Within-herd genetic analyses of stayability of beef females. J. Anim. Sci. 73:993-1001.

Soller, M. 1990. Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. J. Dairy Sci. 73:2628-2646.

Spangler, M. 2018. 311 Genomic Selection: Historical perspective and development of methods in use today. J. Anim. Sci. 96(suppl_3):118-119.

Speidel, S. E., R. M. Enns, and D. H. Crews Jr. 2010. Genetic analysis of longitudinal data in beef cattle: a review. Genet. Mol. Res. 9:19-33.

Speidel, S. E., Buckley, B. A., Boldt, R. J., Enns, R. M., Lee, J., Spangler, M. L. and Thomas, M. G. 2018a. Genome-wide association study of Stayability and Heifer Pregnancy in Red Angus cattle. J. Anim. Sci. 96:846-853.

Speidel, S. E., R. M. Enns, and B. L. Golden. 2018b. Use of a Random Regression model for the Evaluation of Heifer Pregnancy in Red Angus Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/use-random-regression-model-evaluation-heifer-pregnancy-red-angus-cattle.pdf

Sprott, L. R., and R. W. Field. 1998. Reproductive diseases in cattle. Reproductive diseases in cattle. Texas Agricultural extension Service, produced by Agricultural Communications, The Texas A & M University System, College Station, Texas.

Sprott, L. R. 1999. Choosing the time of year to breed and calve cows in Texas. Proc. TAMU Beef Cattle Shortcourse. pp III–16.

Sprott, L. R., G. E. Selk, and D. C. Adams. 2001. Factors affecting decisions on when to calve beef females. Prof. Anim. Sci. 17:238-246.

Stinchcombe, J. R., M. Kirkpatrick M and Function-valued Traits Working Group. 2012 Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. Trends Ecol. Evol. 27:637-647.

Stevenson, J. S., M. W. Smith, J. R. Jaeger, L. R. Corah, and D. G. LeFever. 1996. Detection of estrus by visual observation and radiotelemetry in peripubertal, estrus-synchronized beef heifers. J. Anim. Sci. 74:729-735.

Strucken, E. M., D. J. De Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442-449.

Sun, J., J. E. Rutkoski, J. A. Poland, J. Crossa, J. L. Jannink, and M. E. Sorrells. 2017. Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. Plant Genome. 10:1-12.

Swalve, H. H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. J. Dairy Sci. 83:1115–1124.

Takahashi, M. 2012. Heat stress on reproductive function and fertility in mammals. Reprod. Med. Biol. 11:37-47.

Taponen, J. 2009. Fixed-time artificial insemination in beef cattle. Acta Vet. Scand. 51:1-6.

Taylor, J. F. 2014. Implementation and accuracy of genomic selection. Aquaculture, 420:S8-S14.

Thaller, G. 1997. Genetics and breeding for fertility. Interbull Bull.18:55–61.

Thatcher, W.W. 1974 Effects of season, climate and temperature on reproduction and lactation. J. Dairy Sci. 57:360-368.

Thompson, R., and K. Meyer. 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. Livest. Prod. Sci. 15:299-313.

Thundathil, J. C., A. L. Dance, and J. P. Kastelic. 2016. Fertility management of bulls to improve beef cattle productivity. Theriogenology. 86:397-405.

Toelle, V. D., and O. W. Robison. 1985. Estimates of genetic correlations between testicular measurements and female reproductive traits in cattle. J. Anim. Sci. 60:89-100.

Torres, R. A. A., & Quaas, R. L. 2001. Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines. J. Anim. Sci. 79(Suppl. 1):112.

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. J. Dairy Sci. 92:2922-2930.

Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198-4204.

Van Pelt, M. L., T. H. E. Meuwissen, G. De Jong, and R. F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117-4130.

Vargas, C. A., M. A. Elzo, C. C. Chase Jr, P. J. Chenoweth, and T. A. Olson. 1998. Estimation of genetic parameters for scrotal circumference, age at puberty in heifers, and hip height in Brahman cattle. J. Anim. Sci. 76:2536-2541.

Van Eenennaam, A. L., K. A. Weigel, A. E. Young, M. A. Cleveland, and J. C. Dekkers. 2014. Applied animal genomics: results from the field. Annu. Rev. Anim. Biosci. 2:105-139.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423

Venter, H. A. W., J. C. Bonsma, and J. D. Skinner. 1973. The influence of climate on the reproduction of cattle. Int. J. Biometeorol. 17:147-151.

Veerkamp, R. F., S. Brotherstone, and B. Engel. 2001. Analysis of censored survival data using random regression models. Anim. Sci. 72:1-10.

Walmsley, B. J., S. J. Lee, P. F. Parnell, and W. S. Pitchford. 2018. A review of factors influencing key biological components of maternal productivity in temperate beef cattle. Anim. Prod. Sci. 58:1-19.

Westra, R., and R. J. Christopherson. 1976. Effects of cold on digestibility, retention time of digesta, reticulum motility and thyroid hormones in sheep. Can. J. Anim. Sci. 56:699-708.

White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82:632-638.

Williams, G. L., M. Amstalden, M. R. Garcia, R. L. Stanko, S. E. Nizielski, C. D. Morrison, and D. H. Keisler. 2002. Leptin and its role in the central regulation of reproduction in cattle. Domest. Anim. Endocrinol. 23:339-349.

Wilson, A. J., L. E. Kruuk, and D. W. Coltman. 2005. Ontogenetic patterns in heritable variation for body size: using random regression models in a wild ungulate population. Am. Nat. 166:E177-E192.

Wiltbank, J. N. 1970. Research needs in beef cattle reproduction. J. Anim. Sci. 31:755-762.

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, and J. C. Dekkers. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. Genet. Sel. Evol. 43:5.

Wolfenson, D., I. Flamenbaum, and A. Berman. 1988. Hyperthermia and body energy store effects on estrous behavior, conception rate, and corpus luteum function in dairy cows. J. Dairy Sci. 71:3497-3504.

Wolfenson, D., Z. Roth, and R. Meidan. 2000. Impaired reproduction in heat-stressed cattle: basic and applied aspects. Anim. Reprod. Sci. 60:535-547.

Wolfenson, D., and Roth, Z. 2018. Impact of heat stress on cow reproduction and fertility. Anim. Front. 9:32-38.

Yin, T., B. Bapst, U. U. Von Borstel, H. Simianer, and S. König. 2012. Genetic parameters for Gaussian and categorical traits in organic and low input dairy cattle herds based on random regression methodology. Livest. Sci. 147:159-169.

Yin, T., B. Bapst, U. U. Von Borstel, H. Simianer, and S. König. 2014. Genetic analyses of binary longitudinal health data in small low input dairy cattle herds using generalized linear mixed models. Livest. Sci. 162:31-41.

Zhang, N., D. Huang, W. Wu, J. Liu, F. Liang, B. Zhou, and P. Guan. 2018. Animal brucellosis control or eradication programs worldwide: A systematic review of experiences and lessons learned. Prev. Vet. Med. 160:105-115.

# CHAPTER 3 – COMPARISON OF THRESHOLD MODELS AND RANDOM REGRESSION MODELS IN THE GENETIC EVALUATION OF DICHOTOMOUS FERTILITY TRAITS IN ANGUS HEIFERS

*Summary*

Traditional evaluations of binary traits in cattle involve the use of threshold models (TM) that convert categorical phenotypes to an underlying normally distributed range of genotypic values known as liabilities. Despite the successful influence that TM have had on genetic trends of categorically evaluated traits within livestock species, their susceptibility to the extreme category problem (ECP) limits the ability to use all available information for genetic evaluation and Expected Progeny Differences (EPD). Random regression models (RRM) represent an alternative method to evaluate binary traits—a method not affected by ECP. Nevertheless, RRM were originally developed to analyze longitudinal traits, so their usefulness to evaluate traits with singly observed phenotypes requires further exploration. Objectives of this study were then to evaluate the feasibility of RRM genetic predictions for heifer pregnancy (HPG) and first-service conception rate (FSCR) by comparing its resulting EPD and genetic parameters to those obtained with traditional TM. Breeding and reproductive ultrasound records of 4,334 Angus heifers (progeny of 354 sires and 1,626 dams) collected between 1992 to 2019 at the John E. Rouse Colorado State University Beef Improvement Center were utilized. Observations for HPG and FSCR (1, successful; 0, unsuccessful) were defined by fetal age at pregnancy diagnosis performed approximately 130 d post-AI. Traditional evaluations for both traits were performed using univariate BLUP, threshold animal models, whereas alternative evaluations were performed by regressing HPG and FSCR on age at first exposure (AFE) using linear RRM with Legendre

Polynomials as the base function. Heritability ($h^2$) estimates on the underlying scale were 0.04 and 0.03 for HPG and FSCR using TM; whereas RRM derived $h^2$ estimates on the observed scale were 0.02 and 0.006 for the average AFE for HPG and FSCR, respectively. Pearson and rank correlations between EPD obtained with each methodology were 0.97 and 0.96 for HPG, while for FSCR were 0.75 and 0.72, respectively. Regression coefficients from RRM predictions on those obtained with TM were 0.27 and 0.15 for HPG and FSCR, respectively. Differences in mean accuracies of prediction calculated at the average AFE were minimal between methodologies; however, RRM produced consistently higher accuracies than TM. In conclusion, these results suggested that RRM genetic predictions for singly-observed fertility traits in beef heifers were feasible. More importantly, moderate to strong degrees of concordance were found between predictions obtained with both methodologies for both traits, implying that RRM could substitute for TM in genetic evaluations of binary fertility traits. Potential advantages of utilizing RRM in evaluations of categorical traits include the utilization of all available information to generate EPD and the ability to produce age-specific genetic predictions.

*3.1 Introduction*

Among the traits relevant to beef cattle breeding practices, those related to key fertility events such as conception are binary in nature. Genetic evaluations for these binary traits differ from those used for continuous traits since categorical variables violate many assumptions (e.g., normality and homogeneity of variances) of the linear mixed-models used to obtain best linear unbiased predictions (Henderson, 1975; Gianola, 1982; Abdel-Azim and Berger, 1999). Consequently, genetic evaluations of binary traits have been traditionally performed assuming their phenotypic expression is attributable to an underlying continuous unobservable and normally distributed range of genotypic values, referred to as liability (Gianola and Foulley, 1983; Falconer

and Mackay, 1996). Such assumption allows the application of non-linear systems of equations capable of predicting breeding values, and in turn, have made of threshold models (TM), the method of choice to perform routine genetic evaluations of categorical traits (Foulley, 1992; Gianola and Rosa, 2015).

Threshold models are not completely free of limitations. Originally, the higher computational demands of TM associated with their iterative procedures to yield solutions impaired their widespread application to large data sets until appropriate software developments were accomplished (Misztal et al., 1989). Furthermore, even with increased technological and computing capabilities, one of the major problems associated with TM is their susceptibility to the Extreme Category Problem (ECP). With ECP all observations in a given class or particular level of a fixed effect (typically contemporary group), fall within the same extreme category (e.g., all females are pregnant or vice versa in this study). When this happens, convergence of the algorithms are slowed and often there is a lack of convergence for the fixed effects (Misztal et al., 1989). To overcome ECP-related issues, usually observation groups with this condition are omitted (Harvile and Mee, 1984; Golden et al., 2018); nonetheless, this can lead to distorted inferences because edited data would not be representative of the entire population (Misztal et al., 1989).

Interestingly, only small differences have been reported between threshold and linear models when analyzing both field and simulated categorical data (Meijering and Gianola, 1985; Weller et al., 1988; Hagger and Hofer, 1989). Random regression models (RRM) represent an alternative method to evaluate binary traits and can incorporate data from systematic effects with no variation (Jamrozik et al., 2013; Golden et al., 2018); thereby overcoming the ECP. As such, information from class levels of fixed effects experiencing ECP are not required to be disregarded and distortions created by data editing processes previous to genetic evaluations can be avoided.

Consequently, potential increases in accuracy of prediction may occur by incorporating more information into genetic evaluation procedures. Even though RRM were originally conceptualized to analyze longitudinal traits, their efficacy to evaluate traits with phenotypes observed only once has shown acceptable degrees of success using sire models (Englishby et al., 2016) and animal models (Speidel et al., 2018a). Considering the special combination between documented weaknesses of traditional TM and the potential capabilities of RRM to overcome these weaknesses; we hypothesized that the application of RRM for the genetic predictions of binary fertility traits observed only once in beef heifers are feasible and could achieve higher accuracies of prediction than the TM using edited data. Therefore, the objective of this chapter was to perform a comparison of the two approaches using pedigree-based genetic predictions of heifer pregnancy and first-service conception rate.

*3.2 Materials and Methods*

Although data used in the present study were obtained from an existing database; animals within the experimental location were managed according to the Institutional Animal Care and Use Committee (IACUC) guidelines, covered in most recent years by IACUC number 18-8367A.

*3.2.1 Data collection and description*

Breeding and ultrasound records of 4,334 Angus heifers (progeny of 354 sires and 1,626 dams) collected from 1992 to 2019 at the Colorado State University Beef Improvement Center (CSU-BIC) were used for the study. Within each breeding year, heifers were estrus synchronized and subjected to AI only once before they were exposed to natural service sires approximately 2-wk after insemination. In the present study, heifer pregnancy (HPG) was defined as the ability of a heifer to produce a calf by 24 mo of age, given she conceived within a 60-d breeding season length. Considering every year's specific AI date as the beginning of its respective breeding season,

observations for HPG (1, successful; 0, unsuccessful) were defined by fetal age obtained from ultrasound pregnancy exams performed approximately 60 and 130 d post-AI. Although HPG is a once-in-life recorded phenotype, its expression is likely to be dependent on age of onset of puberty among other factors. Even though no direct measurements of age at puberty were available, the age-related pubertal status of heifers during their first breeding exposure was considered by including age at AI as explanatory variable. Age at first exposure (age at AI) was calculated as the difference between an individual's birthdate and the date when they were subjected to AI.

The same breeding and ultrasound records of all the CSU-BIC heifers previously described were used for heifer first-service conception rate (FSCR) analyses. Within this study, FSCR was defined as the probability of a heifer conceiving in response to her first artificial insemination (AI) and maintaining such pregnancy after the end of the breeding season. Observations for FSCR (1, successful; 0, unsuccessful) were defined by fetal age obtained from ultrasound or manual pregnancy exams performed 130 d post-AI. Although FSCR in 12- to 15-mo-old heifers is a singly observed phenotype, similar to HPG, its expression is likely also dependent on age of onset of puberty. Following the same rationale than for HPG analyzes, age at first exposure (age at AI) was also considered as explanatory variable.

*3.2.2 Testing fixed and random effects*

Systematic effects influencing HPG genetic evaluations have been outlined within the Beef Improvement Federation's guidelines (BIF, 2020); however, no official recommendations were found for genetic evaluations of heifer FSCR. Consequently, in order to identify the important factors influencing the traits of interest, incremental Wald F-tests were performed in the statistical software package ASREML 3.0 (Gilmour et al., 2009) according to equation 3.1 shown below:

$$\boldsymbol{W} = \left[\hat{\theta} - \theta\right]^{-T} \left[Var\left(\hat{\theta}\right)\right]^{-1} \left[\hat{\theta} - \theta\right] \qquad \text{Eq. 3.1}$$

Where $\hat{\theta}$ represented the maximizing argument of an unconstrained likelihood function (i.e., maximum likelihood estimate obtained when all the potential explanatory variables were included in the model), "Var" denoted the variance and $\theta$ corresponded to a hypothesized maximum likelihood estimate produced when assuming the null hypothesis true ($H_0 = \theta_k = 0$), meaning that the $k_{th}$ parameter does not help to explain variation in the response variable. Asymptotically, this test has a $X^2$ distribution with (n – k) degrees of freedom that within ASREML are adjusted using the Kenward-Roger method (Kenward and Roger, 1997). Possible fixed effects examined for HPG were age of dam (AOD), breeding year, breeding pasture combined with service sire and age at first exposure (AFE). In the case of FSCR, potential fixed effects included AOD, breeding year, semen type (e.g., sexed *vs* conventional), AI technician (with at least 5 insemination events recorded) and AFE. Results from the Wald F-tests performed for all the potential fixed effects of both traits are shown in Table 3.1.

**Table 3.1.** Results of Wald F tests for fixed effects for heifer pregnancy (HPG) and First-Service Conception Rate in Angus heifers.

| Trait | Effect | NumDF† | DenDF† | F-inc† | P-inc† |
|---|---|---|---|---|---|
| HPG | Age at first exposure | 1 | 17.1 | 6.3 | 0.022 |
| | Age of dam | 8 | 179.3 | 916.2 | <0.001 |
| | Breeding pasture and service sire | 52 | 2860.6 | 2.0 | <0.001 |
| | Breeding year | 28 | 1149.1 | 34.9 | <0.001 |
| FSCR | Age at first exposure | 1 | 24.0 | 5.2 | 0.032 |
| | Age of dam | 8 | 2189.0 | 17.2 | <0.001 |
| | AI technician | 69 | 3867.9 | 2.1 | <0.001 |
| | Breeding year | 28 | 154.6 | 26.7 | <0.001 |
| | Semen type | 15 | 1473.6 | 6.0 | <0.001 |

†NumDF = Numerator degrees of freedom (number of non-singular equations involved in the term); DenDF = denominator degrees of freedom (estimated according to the adjustments recommended by Kenward and Roger, 1997); F-inc = additional variation explained by the term being tested when added lastly to the model. P-inc = probability value.

All systematic effects tested for each trait were identified as significant sources of variation; however, before their direct incorporation into the genetic evaluations, a grouping strategy was imposed to form contemporary groups (CG). According to Bourdon (2000), when CG are correctly formed, they can help to increase the heritability (and repeatability when repeated measures are available) of the traits under evaluation; whereas, when they are not formed appropriately, the opposite occurs. A CG is a group of animals that have been managed alike and, in this sense, breeding pasture and service sire designations were specific to each breeding year; therefore, all these effects were combined in order to create a more precise definition of contemporary group for HPG. Similarly, in the case of FSCR the effects of breeding year and semen type were also combined to better represent the management decisions particular to each year of data. Forming contemporary groups in this way resulted in a total of 75 and 43 unique contemporary groups for HPG and FSCR, respectively. Summary statistics of contemporary groups for each trait within this study are shown in Table 3.2.

**Table 3.2.** Summary statistics outlining the number of animals represented per contemporary group definition in both heifer fertility traits

| | **HPG[1]** | **FSCR[2]** |
|---|---|---|
| N | 75 | 43 |
| Average | 57.7 | 100.1 |
| SD | 40.6 | 60.5 |
| Minimum | 2 | 4 |
| Maximum | 159 | 196 |
| Average pregnancy rate | 0.84 | 0.45 |

[1]HPG = Heifer pregnancy
[2]FSCR = First-service conception rate

There were two possible extra random effects to be considered in the genetic evaluation of FSCR (besides the animal random additive effects). These variables were mating group (e.g., heifers being inseminated in heat or during a mass mate) and AI sire (sire that produced the semen

straw used during the specific AI event). As previously described by Beckman et al. (2007), the utility of the inclusion of these variables to the model was tested using a likelihood ratio test (LRT, equation 3.2).

$$\mathbf{LRT} = 2|\mathrm{LogL_F} - \mathrm{LogL_R}| \qquad \text{Eq. 3.2}$$

where LRT represented the absolute difference between a full model REML log-likelihood ($\mathrm{LogL_F}$) and the REML log-likelihood of a reduced model ($\mathrm{LogL_R}$). For this test, the null hypothesis established that full models (i.e., containing all possible random effects such as animal, AI sire and mating group) did not fit significantly better than simpler models (models that excluded a particular random effect being tested). The LRT test statistic was distributed approximately as a $X^2$ with degrees of freedom equivalent to the difference in the number of parameters fit for the full and reduced models. These analyses were performed utilizing the package "ordinal" (Christensen, 2015) within the statistical software R (R Development Core Team, 2013). Results of the LRT are presented in Table 3.3.

**Table 3.3.** Results of log-likelihood ratio tests for random effects for first-service conception rate (FSCR) in Angus heifers of the CSU-BIC

| **Effect*** | **$\mathbf{LogL_F}$**† | **$\mathbf{LogL_R}$**† | **LRT**† | **df** | **P-value** |
|---|---|---|---|---|---|
| AI sire | -2667.8 | -2695.3 | 55.0 | 49 | 0.2005 |
| Mating group | -2667.8 | -2676.9 | 18.2 | 2 | <0.0001 |

*AI sire = sire that produced the semen straw used during the specific AI event; Mating group = heifers inseminated during heat or during a mass mate.
†$\mathrm{LogL_f}$ = log-likelihood value of the full model that included all effects; $\mathrm{LogL_r}$ = log-likelihood value of the reduced model that included all effects except for the one being tested (indicated in the effect column); LRT = likelihood ratio test.

Results from the LRT suggested that mating group accounted for a significant portion of the variation of heifer FSCR; notwithstanding, the AI sire effects did not show a significant influence in the conception rate of the heifers. Although, from a statistical point of view, the previous imply that AI sire could be excluded from the model, Averill et al. (2004) suggested that

it is preferred to perform a joint evaluation for males and females, since biologically, the outcome of an insemination depends on both male and female fertility. Consequently, both variables were kept for subsequent genetic evaluations and, in an attempt to visually depict the influences of AI sire and mating group, the conception rate associated with each particular level of each effect are shown in Figures 3.1 and 3.2, respectively.

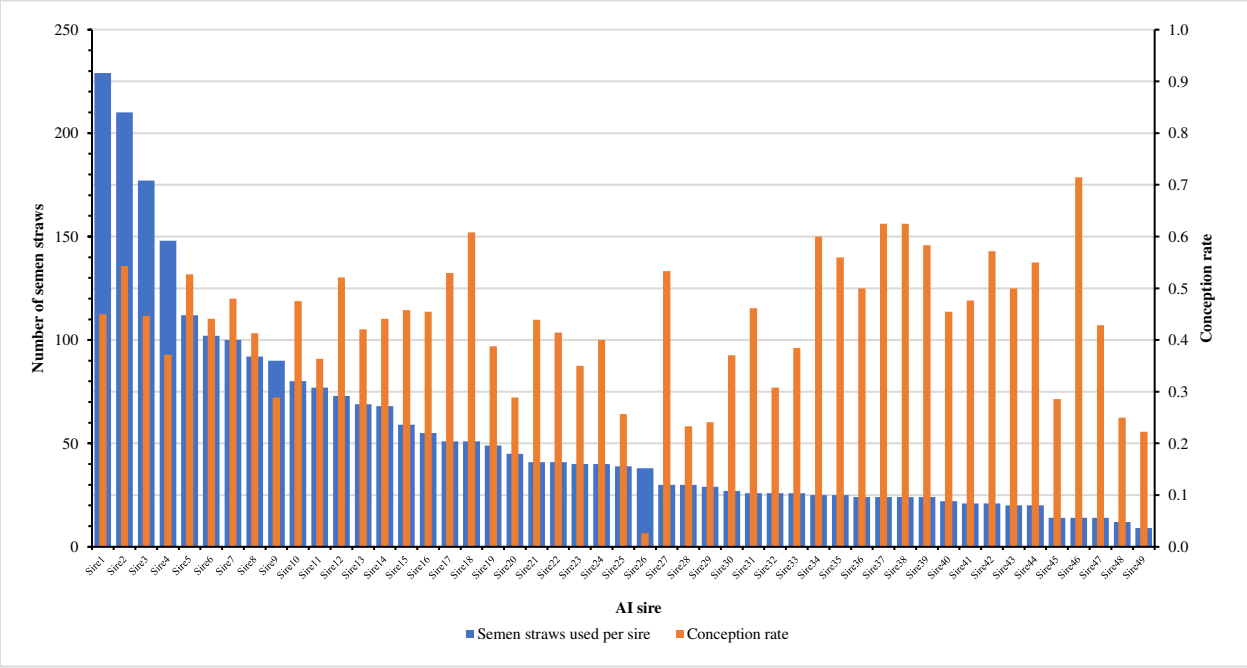


**Figure 3.1.** Average conception rate of AI sires utilized in heifers from the Colorado State University Beef Improvement Center (blue bars represent the number of semen straws used per sire and orange bars represent the conception rate).

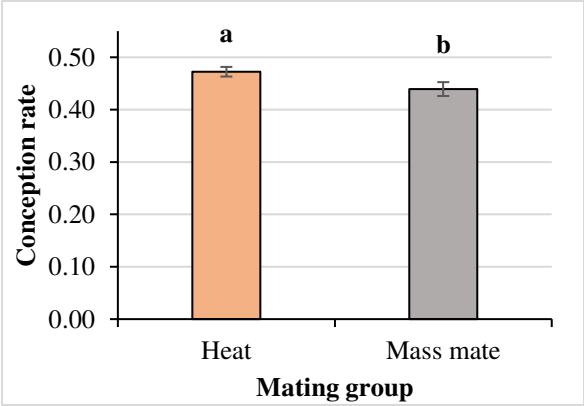


**Figure 3.2.** Average conception rate per mating group in heifers from the Colorado State University Beef Improvement Center (2,934 were inseminated in heat while 1,400 were subjected to AI during a mass mate).

*3.2.3 Genetic evaluations for Heifer Pregnancy*

Traditional EPD calculation for HPG was performed using a univariate BLUP animal TM using a probit link function to convert binary observations to an underlying normal distribution. The model Equation (3.3) was:

$$\boldsymbol{y^* = Xb + Zu + e} \qquad \text{Eq. 3.3}$$

where $\boldsymbol{y^*}$ corresponded to a vector of transformed HPG observations on the underlying scale, $\boldsymbol{b}$ was a vector of unknown solutions for fixed effects that included breeding contemporary group (defined as a combination between breeding year and breeding pasture), age of dam (expressed in the categories recommended by the Beef Improvement Federation); and the individual's AFE included as a linear covariate, $\boldsymbol{u}$ corresponded to a vector of unknown random additive genetic solutions of animal random effects. $\boldsymbol{X}$ and $\boldsymbol{Z}$ were known incidence matrices relating observations in $\boldsymbol{y^*}$ to both fixed ($\boldsymbol{b}$) and random effects ($\boldsymbol{u}$), whereas $\boldsymbol{e}$ was a vector of unknown residual errors. Random effects were assumed to have a mean of 0 and variances equal to:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_a^2} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I\sigma_e^2}\end{bmatrix}$$

where **A** corresponded to the Wright's numerator relationship matrix and $\boldsymbol{I}$ was an identity matrix with an order equal to the number of observations, respectively. The $\boldsymbol{\sigma_a^2}$ and $\boldsymbol{\sigma_e^2}$ were the additive and residual variances, respectively. In agreement with the specifications of a *maximum a posteriori* (MAP) probit threshold model, the residual variance ($\boldsymbol{\sigma_e^2}$) was constrained to be equal to 1.

Additionally, HPG was regressed on AFE using a linear RRM with Legendre polynomials as the base function. The model equation in matrix form is presented below (Equation 3.4):

$$\boldsymbol{y = Xb + Zu + e} \qquad \text{Eq. 3.4}$$

where $\boldsymbol{y}$ corresponded to a vector of binary observations of HPG, $\boldsymbol{X}$ was an incidence matrix relating HPG observations in $\boldsymbol{y}$ to unknown solutions for categorical fixed effects (breeding contemporary group and age of dam using BIF classes) and a linear fixed regression of HPG on AFE in $\boldsymbol{b}$, $\boldsymbol{Z}$ was an incidence matrix consisting of intercept and linear age covariates relating HPG observations in $\boldsymbol{y}$ to the animal random additive genetic regression coefficients (intercept and linear) in $\boldsymbol{u}$, and $\boldsymbol{e}$ was the vector of unknown residual errors. The mean of random effects was assumed to be 0 and variances were assumed to be:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A} \otimes \mathbf{G} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}\boldsymbol{\sigma}_e^2\end{bmatrix}$$

where $\mathbf{A}$ represented the Wright's numerator relationship matrix, $\otimes$ was the Kronecker product, and $\mathbf{G}$ corresponds to a modified variance-covariance matrix of additive genetic random regression coefficients where the covariance between the intercept and the linear term was assumed to be zero, given no heifer had more than one observation for HPG (Speidel et al., 2018a). $\boldsymbol{I}$ and $\boldsymbol{\sigma}_e^2$ remained as described for the TM. A pedigree file from the CSU-BIC consisting of 14,140 individual animals, with 971 and 3725 unique sires and dams, respectively, was used for the estimation of genetic parameters. The average inbreeding coefficient of the pedigree was 0.009.

Within the TM methodology, heritability estimates on the underlying scale for both traits were obtained by calculating the ratio of the additive to the phenotypic variance. Conversely, variance estimates obtained using RRM are not directly comparable to those obtained with conventional models used in animal breeding evaluations (Speidel, 2011). Particularly, the estimation of (co)variances for genetic evaluations using RRM yield genetic and phenotypic variances for the shape of the polynomial implemented. For instance, in this particular study given a linear order was used for the RRM used to evaluate HPG, the resulting variance estimates corresponded to estimates for the intercept and linear term (slope) of the random polynomial.

Nonetheless, through a relatively straightforward conversion process, it is possible to use the RRM variance estimates to calculate observed variance estimates for each value of the covariate used. Therefore, it was possible to calculate heritabilities of HPG for every AFE. The formula utilized to perform such transformations is shown in equation 3.5:

$$\widehat{\boldsymbol{G}} = \boldsymbol{\Phi}\ \mathbf{K}\ \boldsymbol{\Phi}' \qquad \text{Eq. 3.5}$$

where $\widehat{\boldsymbol{G}}$ represents a (co)variance matrix of HPG observations at *t* given AFE. **K** is a matrix of order *k* containing the variance components for the RRM coefficients contemplated in the model (e.g., intercept and linear) and, **Φ** is a matrix of order *t* x *k* containing orthogonal polynomial coefficients evaluated at *t* standardized AFE with elements $\Phi_{ij} = \Phi_j(x_i)$, being the $j^{th}$ polynomial coefficient for the $i^{th}$ AFE (Fischer et al., 2004; Speidel, 2011).

The predictions obtained with each method varied, where the TM predicted a single breeding value per animal on the underlying scale, and predictions obtained with RRM result in a vector equal to the order of the Legendre polynomials (e.g., each animal had a prediction for the intercept and the linear term of the random regression). Consequently, in order to compare predictions between the two methods, first it was necessary to condense the RRM predictions into single values per animal expressed on an observed scale for each AFE. Equation 3.6 shows the procedure to perform the conversions of the random regression coefficients obtained per each animal (e.g., intercept and linear term) back to specific AFE EPD:

$$\widehat{\boldsymbol{EPD}}_{\boldsymbol{observed}} = \frac{\boldsymbol{a_m} * \Phi_i}{\mathbf{2}} \qquad \text{Eq. 3.6}$$

where $\boldsymbol{\Phi_i}$ corresponded to the coefficients of Legendre polynomials standardized to the $\mathbf{i^{th}}$ AFE and $\mathbf{a_m}$ represented the random regression solutions (intercept and the linear terms) for the $\mathbf{m^{th}}$ animal. Once the EPD for each particular age was obtained, the prediction corresponding to the

average AFE (422 d) was chosen as the reference age point to compare RRM predictions to those produced by the TM.

Since TM predicted genetic merits on an underlying scale and RRM did it on an observed scale, both types of predictions were converted to a pseudo-probability scale as deviations from 50% (random chance of conception) following the procedure outlined by Speidel et al. (2018a). Briefly, EPD obtained with each methodology were converted to a Z-score by dividing by the HPG phenotypic standard deviation; subsequently, each set of predictions were transformed utilizing a normal cumulative distribution function and then they were multiplied by 100 to express them as probabilities. The resulting predictions were compared through the calculation of Pearson ($r_p$) and Spearman's ($r_s$) correlations and the estimation of the regression coefficient of EPD obtained with the RRM on those obtained with the TM. Analyses were performed using ASREML 3.0 (Gilmour et al., 2009), the Animal Breeder's Tool Kit (Golden et al., 1992) and BOLT (Garrick et al., 2018).

Accuracy (ACC) calculations were performed according to the guidelines of the Beef Improvement Federation (2020) using Equation 3.7:

$$\boldsymbol{ACC = 1 - \sqrt{\frac{PEVi}{(1 + F_i) * \sigma_a^2}}} \qquad \text{Eq. 3.7}$$

where $\boldsymbol{\sigma_a^2}$ denoted the additive genetic variance for HPG, $\boldsymbol{PEV_i}$ corresponded to the prediction error variance for the i$^{th}$ individual and $\boldsymbol{F_i}$ represented the inbreeding coefficient of the i$^{th}$ animal. In the case of the TM, $\boldsymbol{PEVi}$ was obtained by squaring the standard error of prediction reported next to the BLUP for the i$^{th}$ animal on the ASREML output solutions file (Gilmour et al., 2009). Conversely, given the MME for the RRM were assembled manually using the BOLT software (Garrick et al., 2018) and the size of the equation systems were not prohibitive (e.g., 28365 for HPG and 28454 for FSCR, respectively), a direct inversion of the coefficient matrix was

performed. Then, the inverse elements of the diagonal block for each set of RR coefficients provided an estimate of the prediction error variance (***PEVi***) of breeding values at age i (e.g., average AFE of 422 d). Finally, once the PEV of each methodology was obtained, mean accuracies were calculated and were then compared to each other.

*3.2.4 Genetic evaluations for first-service conception rate in heifers*

Traditional EPD calculation for FSCR was performed using a univariate BLUP animal TM with a probit link function to convert binary observations to an underlying normal distribution. The model Equation (3.8) was:

$$\boldsymbol{y^* = Xb + Zu + Qm + Ws + e} \qquad \text{Eq. 3.8}$$

where $\boldsymbol{y^*}$ corresponded to a vector of transformed observations of FSCR on the underlying scale, $\boldsymbol{b}$ was a vector of unknown solutions for fixed effects that included breeding contemporary group defined as a combination between breeding year and semen type (e.g., conventional *vs* sexed), AI technician, age of dam (BIF classes), and the individual's AFE as a linear covariate, $\boldsymbol{u}$ corresponded to a vector of unknown solutions of animal random effects, $\boldsymbol{m}$ was a vector of unknown solutions of mating group (e.g., inseminated in heat or during a mass mate) random effects and, $\boldsymbol{s}$ was a vector of unkwnown solutions for AI sire random effects. The matrices ***X, Z, Q*** and ***W*** were known incidence matrices relating observations in $\boldsymbol{y^*}$ to fixed effects in $\boldsymbol{b}$, as well as animal, mating group and service sire random effects in $\boldsymbol{u}$, $\boldsymbol{m}$ and $\boldsymbol{s}$, respectively. Finally, $\boldsymbol{e}$ was a vector of unknown residual errors. The mean of random effects was assumed to be 0 whereas variances were assumed to be equal to:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{m}\\ \boldsymbol{s}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_a^2} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_m\sigma_m^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_w\sigma_s^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

where **A** corresponded to the Wright's numerator relationship matrix, $\boldsymbol{I_m}$, $\boldsymbol{I_w}$ and $\boldsymbol{I_n}$ were identity matrices whose orders were equal to the number of mating groups, AI sires and observations, respectively. $\boldsymbol{\sigma_a^2}$, $\boldsymbol{\sigma_m^2}$, $\boldsymbol{\sigma_s^2}$ and $\boldsymbol{\sigma_e^2}$ were the additive, mating group, AI sire and residual variances, respectively. Within this model, the residual variance ($\boldsymbol{\sigma_e^2}$) was constrained to be equal to 1.

Additionally, FSCR was regressed on AFE using a linear RRM with Legendre polynomials as the base function. The model in matrix form is presented in Equation 3.9 below:

$$\boldsymbol{y = Xb + Zu + Qm + Ws + e} \qquad \text{Eq. 3.9}$$

where $\boldsymbol{y}$ corresponded to a vector of binary observations of FSCR, $\boldsymbol{b}$ was a vector of unknown solutions for categorical fixed effects (breeding contemporary group, AI technician and BIF age of dam classes) and a linear fixed regression of FSCR on AFE, $\boldsymbol{u}$ corresponded to a vector of unknown solutions of animal random regression coefficients (intercept and linear) for additive genetic effects, $\boldsymbol{m}$ was a vector of unknown solutions for mating group random effects and, $\boldsymbol{s}$ was a vector of unknown solutions for AI sire random effects. $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{Q}$ and $\boldsymbol{W}$ were known incidence matrices relating observations in $\boldsymbol{y}$ to fixed ($\boldsymbol{b}$), animal ($\boldsymbol{u}$) mating group ($\boldsymbol{m}$) and AI sire ($\boldsymbol{s}$) random effects, respectively. Lastly, $\boldsymbol{e}$ was the vector of unknown residual errors. Random effects were assumed to have a mean of 0 and variances equal to:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{m}\\ \boldsymbol{s}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\otimes\mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_m\sigma_m^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_w\sigma_s^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

where **A** represented the Wright's numerator relationship matrix, $\otimes$ was the Kronecker product, and **G** corresponded to a modified variance-covariance matrix of additive genetic random regression coefficients where the covariance between the intercept and the linear term was fitted to zero since no heifer had more than one observation for FSCR (Speidel et al., 2018a). The $\boldsymbol{I_m}$,

$\boldsymbol{I_w}$, $\boldsymbol{I_n}$, $\boldsymbol{\sigma_m^2}$, $\boldsymbol{\sigma_s^2}$ and $\boldsymbol{\sigma_e^2}$ terms remained as described for the previous TM. For the estimation of the genetic parameters, the same pedigree file implemented in HPG analysis was utilized. Similarly, the same procedures described for HPG evaluations were followed for FSCR analyses in order to estimate heritabilities and obtain EPD within each statistical method (e.g., see equations 3.5 and 3.6). The same comparative strategies previously mentioned for HPG predictions were followed to compare the outputs of each statistical model implemented to evaluate FSCR. Accuracy estimations were performed using Equation 3.7 and all analyses were performed using the same statistical packages previously described.

*3.3 Results and discussion*

*3.3.1 Pregnancy percentages*

Percentages of pregnant heifers per breeding year and pregnancy type are shown in Figure 3.3**.** Considering all years, the average heifer pregnancy was 89.2%, with the highest percentage of pregnant heifers occurring in 1997 with a 97.6% pregnancy rate and the lowest in 2002 with 64% pregnancy rate. A similar average heifer pregnancy rate was reported by Doyle et al. (2000) when analyzing HPG in a slightly overlapping, although previous time window (1985 through 1993), at the CSU-BIC. Authors of the previous study emphasized that fertility plays an important role within the breeding objective of the CSU-BIC, explaining that replacement heifers have been historically selected on the basis of their own performance as well as their dam's record, retaining only fertile animals. Knowledge about the selection pressure placed on fertility within the herd was helpful to understand the high heifer pregnancy rates observed within the 28 yr span of data analyzed in the present study. Furthermore, considering that the Angus herd at the CSU-BIC corresponds to an experimental beef cattle population with short controlled breeding season lengths, results of this study agree with a previous report suggesting that high pregnancy rates are

fully achievable in well-managed, purebred cattle populations with breeding seasons of short duration (Brinks et al., 1990).

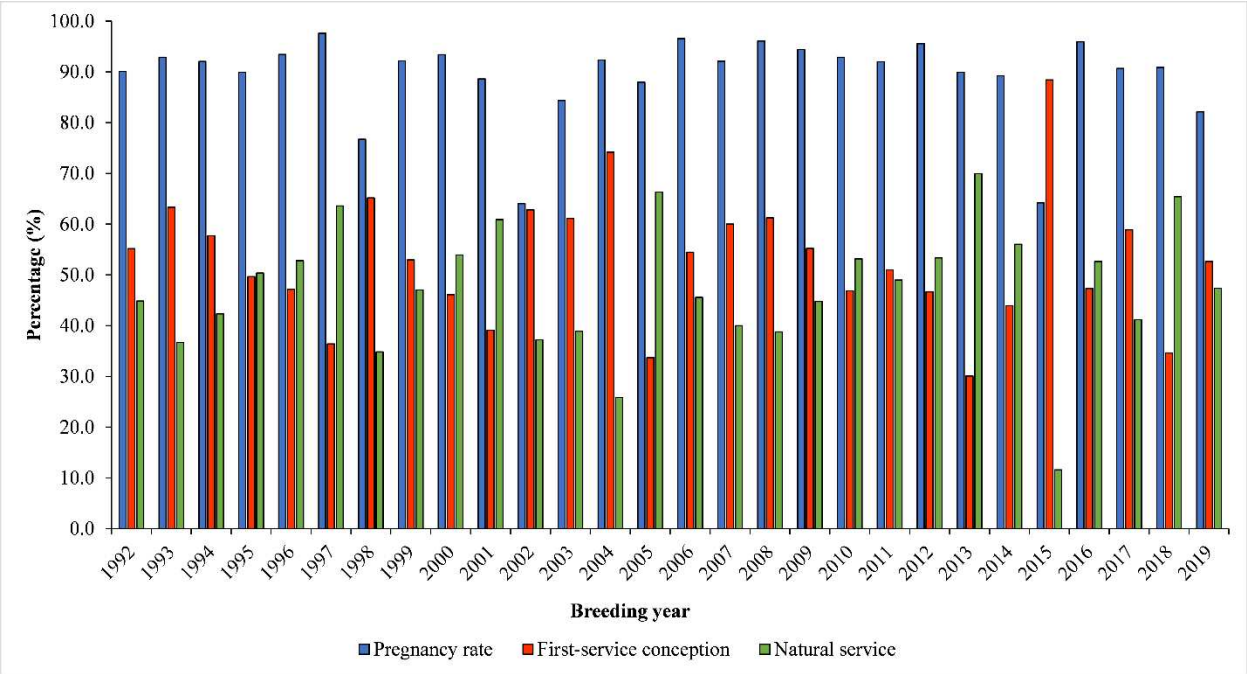


**Figure 3.3.** Percentages of pregnant heifers per breeding year and pregnancy type.

Among all heifer's pregnancies recorded between 1992 and 2019, 46.2% were pregnancies obtained in response to a unique AI (e.g., first-service conception rate), while the remaining 53.8% were pregnancies obtained by natural service. The average FSCR in our study was slightly smaller than the 54.9% reported by Foxworthy (2019) when using information from the CSU-BIC research herd; nonetheless, important differences in age grouping strategies existed among the two studies. In Foxworthy (2019), females (heifers, primiparous and multiparous cows) up to 4 yr of age were lumped into a general group of "immature cows" following the BIF age of dam classification recommendations; whereas in the present study, only 12- to 15-mo-old heifers entering to their first breeding season were considered. In an additional study, Bormann et al. (2006) reported a FSCR of 60% when analyzing data of 3,144 Angus replacement heifers coming from 6 different herds spread across 5 states within the US; while the average FSCR reported for a Brangus cattle

population (n = 830) was of 53% (Peters et al., 2013). Differences among the aforementioned reports with the average FSCR obtained in the present study (46.2%) could be attributed to the heterogeneity of reproductive managements across herds (e.g., different estrus synchronization protocols), effectiveness of estrus detection programs and AI-technician's expertise, since all these factors have a strong influence on this trait (Bormann et al., 2006).

As expected within the present study, some classes within the categorical fixed effects included in the genetic models exhibited no variation (ECP problems); therefore, they were removed from the data for TM analyzes. Consequently, slightly smaller datasets were used with TM evaluations in comparison to those used with RRM. Summary statistics of the final number of observations available for each trait and each evaluation methodology are shown in Table 3.4. Specifically, a total of 21 observations (19 successful/2 unsuccessful) were removed for HPG threshold evaluations, while 144 observations (66 successful/78 unsuccessful) were deleted to implement the TM evaluation for FSCR. In this particular study, the possible bias introduced by pre-analytical editing processes within the TM methodology was minimal for both traits; however, the problem has been reported to have a more marked severity when working with larger datasets like those used in national genetic evaluations (Misztal et al., 1989; Phocas and Laloë, 2003). In this regard, the capability of RRM to incorporate all the available information has been reported as an attractive feature of that statistical methodology (Golden et al., 2018, Speidel et al., 2018a).

**Table 3.4.** Heifer pregnancy, first-service conception rate, age at first exposure and age of dam summary statistics per statistical methodology

| Methodology | Trait | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| TM[1] | Heifer pregnancy | 4313 | 0.85 | 0.362 | 0 | 1 |
| | Age at first exposure (d) | 4313 | 422.1 | 21.1 | 347 | 479 |
| | Age of dam (yr) | 4115 | 4.9 | 2.8 | 2 | 13 |
| | First-service conception rate | 4190 | 0.47 | 0.499 | 0 | 1 |
| | Age at first exposure (d) | 4190 | 422.0 | 21.2 | 347 | 479 |
| | Age of dam (yr) | 3996 | 4.9 | 2.8 | 2 | 13 |
| RRM[2] | Heifer pregnancy | 4334 | 0.85 | 0.362 | 0 | 1 |
| | Age at first exposure (d) | 4334 | 422.1 | 21.1 | 347 | 479 |
| | Age of dam (yr) | 4136 | 4.9 | 2.8 | 2 | 13 |
| | First-service conception rate | 4334 | 0.46 | 0.499 | 0 | 1 |
| | Age at first exposure (d) | 4334 | 422.1 | 21.1 | 347 | 479 |
| | Age of dam (yr) | 4136 | 4.9 | 2.8 | 2 | 13 |

[1]TM = threshold model, [2]RRM = random regression model.

*3.3.2 Heritabilities*

Heritability estimates for HPG and FSCR obtained with each statistical methodology are presented in Table 3.5. The heritability ($h^2$) estimate for HPG of 0.04 obtained with the TM was considerably smaller than the moderate heritabilities for Red Angus cattle reported by McAllister et al. (2011) and Boldt et al. (2018) of 0.17 and 0.12, respectively. Similarly, it was smaller than the one communicated by Doyle et al. (2000) for the CSU-BIC black Angus heifer population (0.21) and the $h^2$ estimate of 0.14 reported for Hereford cattle (Evans et al., 1999). Nonetheless, the estimate obtained in the current study was in a smaller range of estimates (0.03 to 0.06) reported for Angus and Hereford cattle by other authors (Toelle and Robinson, 1985; Mathiews et al., 1995).

**Table 3.5.** Heritability estimates ($h^2$ ± SE) for heifer pregnancy (HPG) and first-service conception rate (FSCR) according to the statistical method employed (these are not transformed heritabilities).

| Methodology | Trait | | $h^2$ ± SE |
|---|---|---|---|
| TM[1] | HPG | | 0.04 ± 0.03 |
| RRM[2] | | Intercept | 0.02 ± 0.02 |
| | | Linear | 0.11 ± 0.09 |
| TM[1] | FSCR | | 0.03 ± 0.02 |
| RRM[2] | | Intercept | 0.003 ± 0.012 |
| | | Linear | 0.133 ± 0.079 |

[1]TM = threshold model, [2]RRM = random regression model.

A possible explanation of the low $h^2$ estimate for HPG obtained in this study could be the extreme phenotype incidences recorded for this trait across the time period analyzed (1992 to 2019). Roughly, frequencies of 90% success and 10% failure in heifer pregnancy rates were observed at the CSU-BIC within the 28 yr period of data included, with 15 years having pregnancy rates even higher than 90%. Although Dempster and Lerner (1950) suggested that when heritability estimates are calculated on an underlying scale, they become independent of the frequency of the trait; Meijering and Gianola (1985) explained that $h^2$ estimates obtained using the threshold theory are unstable when the frequencies of a binary response variable are extreme (e.g., when frequencies surpass a 80:20 ratio). In this regard, Lopes et al. (2000) graphically showed the influence of phenotypic incidences on the heritability estimates obtained using Dempster and Lerner's method (1950) using simulation techniques. Briefly, when phenotype incidence was intermediate (between 20 to 80%), heritability estimates were closer to the simulated true heritability of a binary trait; however, when the incidences were extreme (below 20 or higher than 80%), an underestimation of the heritability occurred.

Regarding the heritability estimates obtained for the random regression coefficients for HPG, similar results were reported for Red Angus cattle. Specifically, the estimate for the intercept term in the present study was lower (0.02 *vs* 0.10); but the estimate obtained for the linear term was higher (0.11 *vs* 0.10) than those reported by Speidel et al. (2018a). Transforming the RRM variance estimates obtained for HPG, it was possible to calculate $h^2$ estimates for HPG across all the range of AFE included in the dataset (Figure 3.4). In general, all $h^2$ estimates fell within previous reports indicating that up to 14% of the variation within this trait were attributable to differences in additive genetics (Mathiews et al., 1995; Evans et al., 1999; Boldt et al., 2018). A particular estimate of interest was the one obtained at the average AFE (422 d) of the Angus heifers population contemplated in this study, since it served as a reference point to compare it to the estimate produced by the TM.

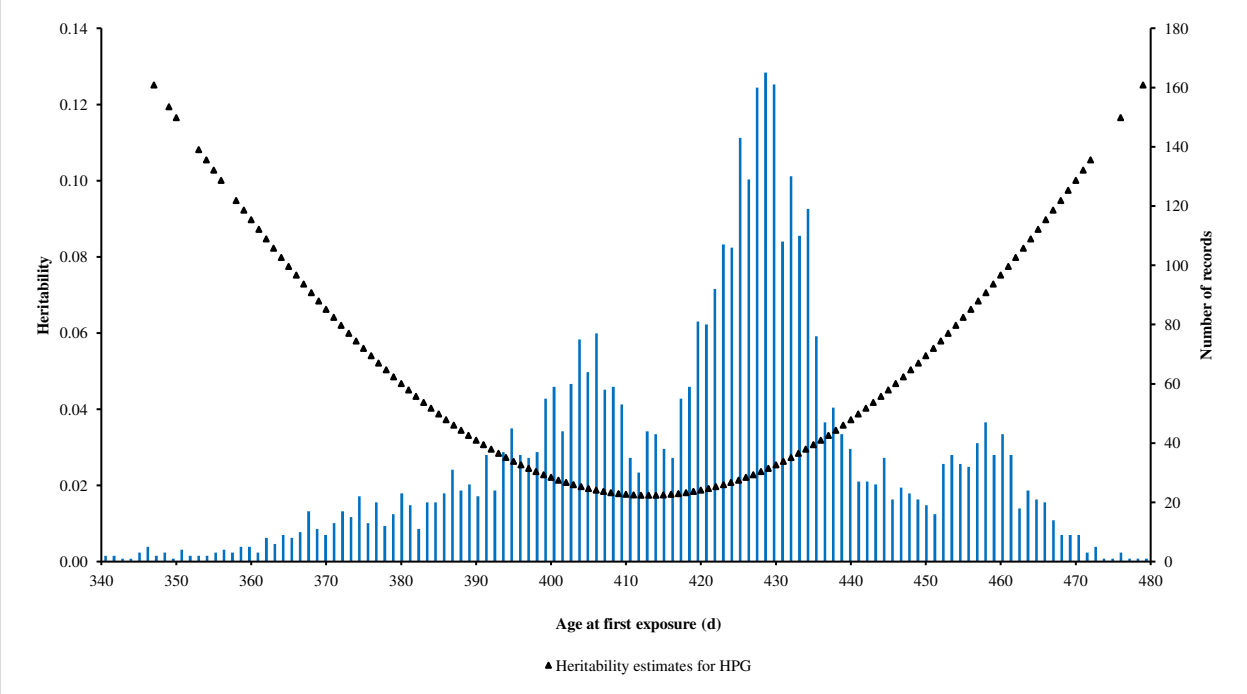


**Figure 3.4.** Changes in heritability estimates for heifer pregnancy and their relationship with the number of records of ages at first exposure in Angus heifers.

The observed $h^2$ estimate at 422 d was 0.02, which was lower than the 0.04 obtained on the underlying scale with the TM. Variation relative to the scale in which the trait was analyzed has been reported by Buddenberg et al. (1989), for instance, HPG $h^2$ estimates on the observed scale were consistently lower than estimates obtained on the underlying scale. Particularly, authors reported that in the case of Angus heifers, $h^2$ estimates were 0.17 and 0.34 on the observed and underlying scales, respectively. Whereas the estimates for Hereford and Polled Hereford heifers were 0.04 and 0.05 on the observed scale and 0.08 and 0.10 on the underlying scale, respectively. Focusing exclusively on results obtained using random regression procedures, Speidel et al. (2018a) reported a similar outcome to the present study from an analysis of Red Angus data, since the $h^2$ estimate obtained at 460 d of age was equal to 0.09, a value considerably lower than the 0.24 used by the Red Angus American Association for HPG genetic evaluations using TM. Interestingly, the $h^2$ estimate obtained in the present study at the age of 460 d of age was 0.08, similar to the previously discussed estimate obtained using RRM for Red Angus.

In the case of FSCR, the $h^2$ estimate obtained on the underlying scale using the TM was 0.03, which agrees with a previous report in Angus heifers indicating the same value (0.03; Bormann et al., 2006). Other reports have indicated FSCR $h^2$ estimates ranging between 0.06 to 0.22, however, such estimates have been obtained mainly in crossbred populations in which a greater phenotypic variation may exist within the *Bos indicus*-influenced heifers relative to purebred animals (Dearborn et al., 1973; Fortes et al., 2012; Peters et al., 2013). Conversely, heritability estimates for the resulting intercept and linear term of the RRM were 0.003 and 0.133, respectively. After transforming these RRM variance estimates, a $h^2$ of 0.006 for FSCR at the average AFE (422 d) was observed (just as an extra, the $h^2$ at 460 was 0.075). Using a bivariate sire random regression model, De Haas et al. (2007) reported a $h^2$ estimate for conception rate at

first insemination in Holstein heifers of 0.01 (using body condition score as the secondary trait), such result is considerably close to the one obtained in the present study. Similar to HPG results, the $h^2$ estimate obtained with the RRM was lower than that obtained using the TM. This result agrees with a plethora of reports suggesting that heritabilities calculated on an observed scale are consistently lower than those obtained on the underlying scale (Meijering, 1984; Johnston et al., 2014; Silvestre et al., 2019). Nonetheless, the pattern of changes in $h^2$ estimates for FSCR across the range of AFE contemplated in this study agreed with previous reports for this trait (Figure 3.5). Furthermore, regardless of the age-associated variations, it was clear that environmental conditions greatly influence the ability of a heifer to become pregnant in response to her first service.

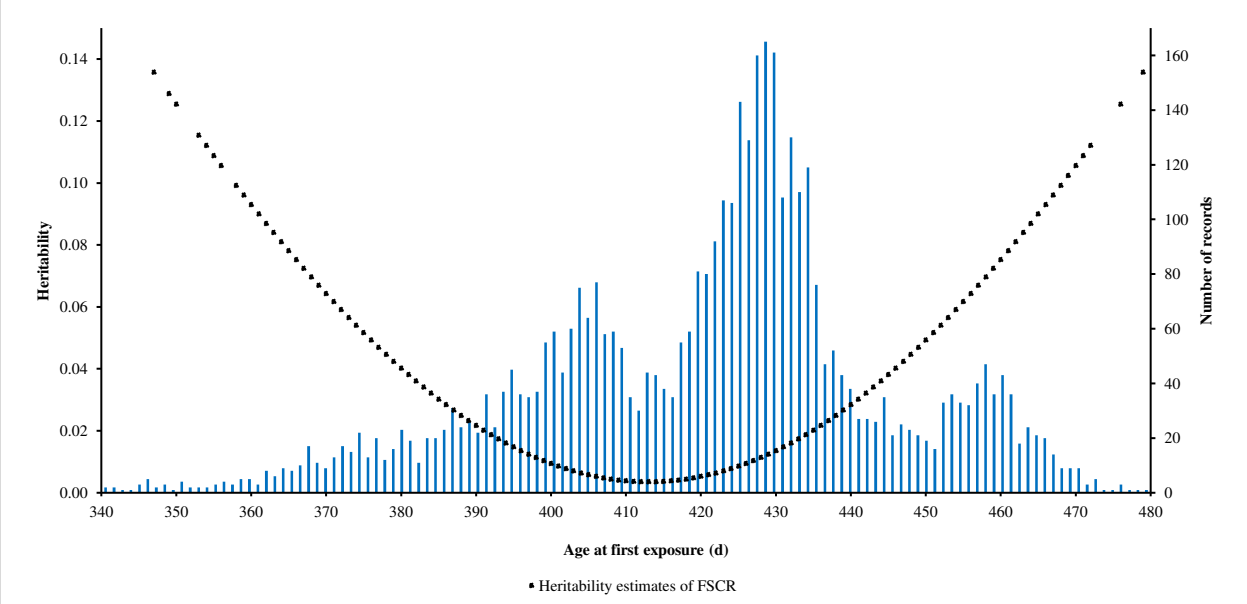


**Figure 3.5.** Changes in heritability estimates for heifer first-service conception rate and their relationship with the number of records ages at first exposure in Angus heifers.

Even though the $h^2$ estimates obtained for both HPG and FSCR using random regression techniques appeared to be reasonable and within ranges reported in literature, it is important to acknowledge that estimates were considerably higher at the extremes of the age prediction range than in the middle. As a possible explanation for such results, it has been previously reported that a common artefact of RRM using Legendre Polynomials as the base function is their tendency to

inflate the genetic variances at the beginning and the end of the data range (Shaeffer and Jamrozik, 2008). This occurs because RRM are sensitive to changes in data distribution, particularly, to reductions in the number of records associated with the covariate implemented (Brügemann et al., 2013). Figures 3.4 and 3.5 show the distribution of HPG and FSCR records associated with the ages at first exposure of the Angus heifers from the CSU-BIC. The significant reductions in the number of observations registered at the extremes of the data range could explain the substantial increases in $h^2$ estimates for these ages. Similar data structures have led to comparable variations in $h^2$ estimates for traits like days open and conception rate of dairy cattle when implementing random regression models (Yin et al., 2012; Brügemann et al., 2013).

*3.3.3 Comparison of genetic predictions*

With respect to the genetic predictions performed for both traits and with each methodology, EPD summary statistics are presented in Table 3.6. For both traits, results for the mean EPD were similar between models; however, a wider range in prediction values was observed with the TM. The lower spread in EPD observed within the RRM prediction could be explained by the smaller $h^2$ estimates obtained with this methodology for the average AFE. Speidel et al. (2018a) reported a similar outcome when applying RRM in the genetic prediction of HPG in Red Angus cattle.

**Table 3.6.** Heifer pregnancy and first-service conception rate expected progeny differences (at the average age at first exposure) summary statistics according to the statistical method implemented.

| Methodology | Trait | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| TM[1] | HPG | 14,140 | 0.411 | 2.03 | -8.88 | 9.61 |
| | FSCR | 14,140 | -0.119 | 1.33 | -7.88 | 5.85 |
| RRM[2] | HPG | 14,140 | 0.197 | 0.57 | -2.59 | 2.95 |
| | FSCR | 14,140 | 0.019 | 0.27 | -1.12 | 1.45 |

[1]TM = threshold model, [2]RRM = random regression model.

Pearson ($r_p$) and Spearman's ($r_s$) correlations among EPD obtained with the TM and RRM for both traits are shown in Table 3.7. In general, results suggested that predictions were moderately to highly correlated and similar animal rankings were obtained with both methods. Despite the differences between TM and RRM, almost the exact same ranking of animals was obtained for HPG. The previous suggested that RRM could potentially substitute for TM in the genetic evaluations for HPG in beef cattle. Similar results (e.g., $r_p = 0.87$ and $r_s = 0.89$) were reported by Speidel et al. (2018a) in a study that compared genetic predictions for HPG obtained with TM and RRM.

**Table 3.7.** Pearson correlation, rank correlation and regression coefficients of predictions obtained with each statistical method

| Trait | Pearson correlation | Rank correlation | Regression coefficient |
|---|---|---|---|
| HPG[1] | 0.97 | 0.96 | 0.27 |
| FSCR[2] | 0.75 | 0.72 | 0.15 |

[1]HPG = Heifer pregnancy
[2]FSCR = First-service conception rate

Regarding FSCR, correlation results ($r_p$ and $r_s$) were lower than for HPG, but were strong enough to consider a possible preference for RRM over TM in future genetic evaluations given the ability of RRM to utilize all available data. The previous could be especially useful when considering the larger list of categorical fixed effects influencing FSCR and the increased chances of occurrence of ECP potentially leading to greater information loses within the TM approach. Similar degrees of agreement between predictions obtained with TM and RRM have been reported for stayability (STAY) in Angus cattle. Specifically, Sánchez-Castro et al. (2017) reported Pearson correlations ranging from 0.59 to 0.83 for STAY genetic predictions at different ages (from 3 to 6 yr of age), suggesting that predictions obtained with both methodologies were similar. The same authors reported Spearman's correlations between 0.64 and 0.65 when analyzing STAY at the age of 6 using TM and RRM that included endpoints beyond 6 yr of age (Sánchez-Castro et al., 2019).

Such results indicated a significant reranking of animals between methodologies; however, considerable differences in the amount of data incorporated with each statistical method could explain the lower rank correlations reported in that study, when compared to those found in the present investigation. For their part, Lewis and Brotherstone (2002) reported Pearson correlations ranging from 0.81 to 0.91, and rank correlations from 0.77 to 0.78 when predicting breeding values for growth traits utilizing RRM and a univariate animal model. Authors of that study concluded that, depending on the age of interest, for genetic predictions, it was possible that the same individuals resulted in the best selection candidates with both methodologies. The regression of predictions obtained with the RRM on those obtained with TM for both traits of interest in the present study showed slight underestimations of the genetic merit using the RRM in comparison to TM ($\beta_1$ for HPG = 0.27 and $\beta_1$ for FSCR = 0.15). Recalling that both types of predictions (TM and RRM) were converted to a pseudo-probability scale as deviations from 50% (Speidel et al., 2018a); these underestimations were not likely a result of different prediction scales, but were probably reflections of the lower $h^2$ estimates obtained for both traits using RRM.

*3.3.4 Comparison of accuracies of prediction*

Lastly, mean accuracies for HPG and FSCR calculated at the average AFE within each methodology are shown in Figures 3.6 (A-D) and 3.7 (A-D), respectively. In general, mean accuracies for both traits and with both methodologies were low (<0.035). When considering all animals (Figure 3.6 A), the mean accuracy for HPG predictions obtained with the TM was 0.020 with a minimum of 0.000 and a maximum of 0.151. Alternatively, the mean accuracy for the same trait when analyzed using a RRM was 0.024 with values that ranged between 0.001 and 0.179. As expected, mean accuracies estimated for sires were higher than those obtained for the general population in both methodologies; however, mean values were slightly higher within the RRM

methodology for sires that have produced progeny in the last 5 (Figure 3.6 B) and 3 years (Figure 3.6 C), respectively. Similarly, when considering the youngest selection candidates (1-yr-old males), accuracies obtained with the RRM yielded higher values in comparison to those obtained with the TM (Figure 3.6 D). In the case of the accuracy values associated with the genetic predictions of FSCR, when taking into account the entire pedigree (Figure 3.7 A), the mean value obtained with the TM was 0.021 with a minimum of 0.000 and a maximum of 0.182. For the same trait but within the RRM methodology, the mean accuracy value was 0.029 with a range between 0.013 and 0.151.

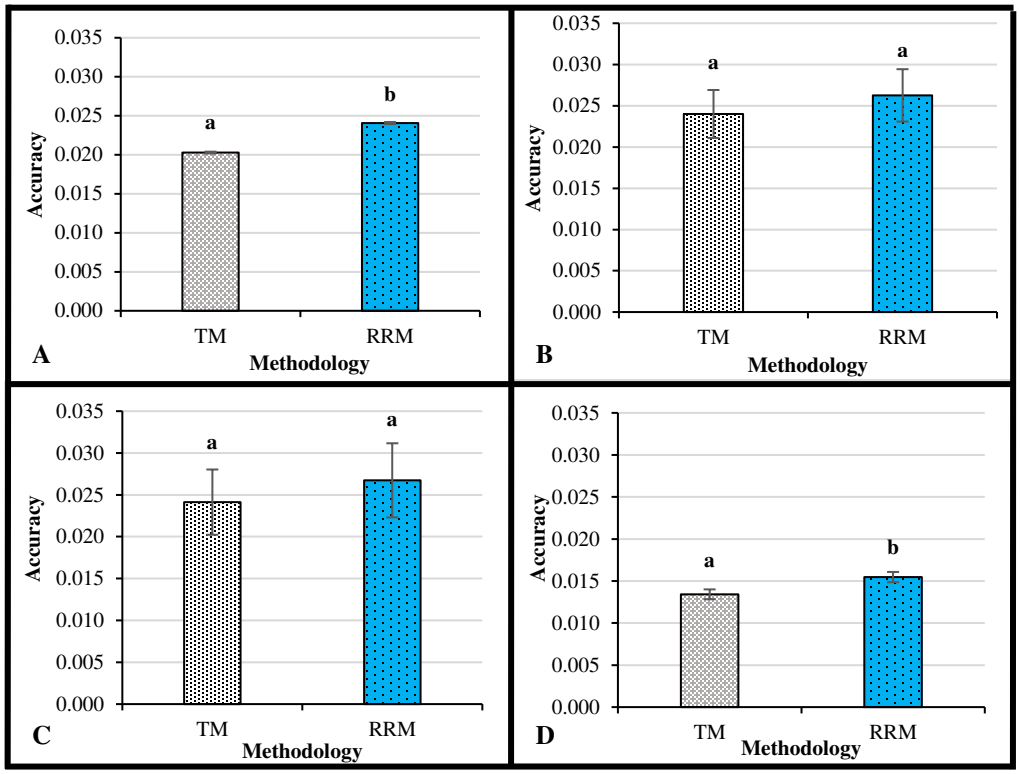


**Figure 3.6.** Mean accuracies for heifer pregnancy genetic predictions at the average age of first exposure (422 d) obtained with each statistical methodology. **A)** Mean accuracy for all animals in the pedigree (n = 14,140), **B)** Mean accuracies for sires that have produced progeny in the last five yr (n = 85), **C)** Mean accuracies for sires that have produced progeny in the last three yr (n = 51), **D)** Mean accuracies of the 1-yr-old males potential selection candidates (n = 180). Different letters indicate a statistical difference at the $P < 0.05$ level among methodologies according to the Fisher's least significant difference test.

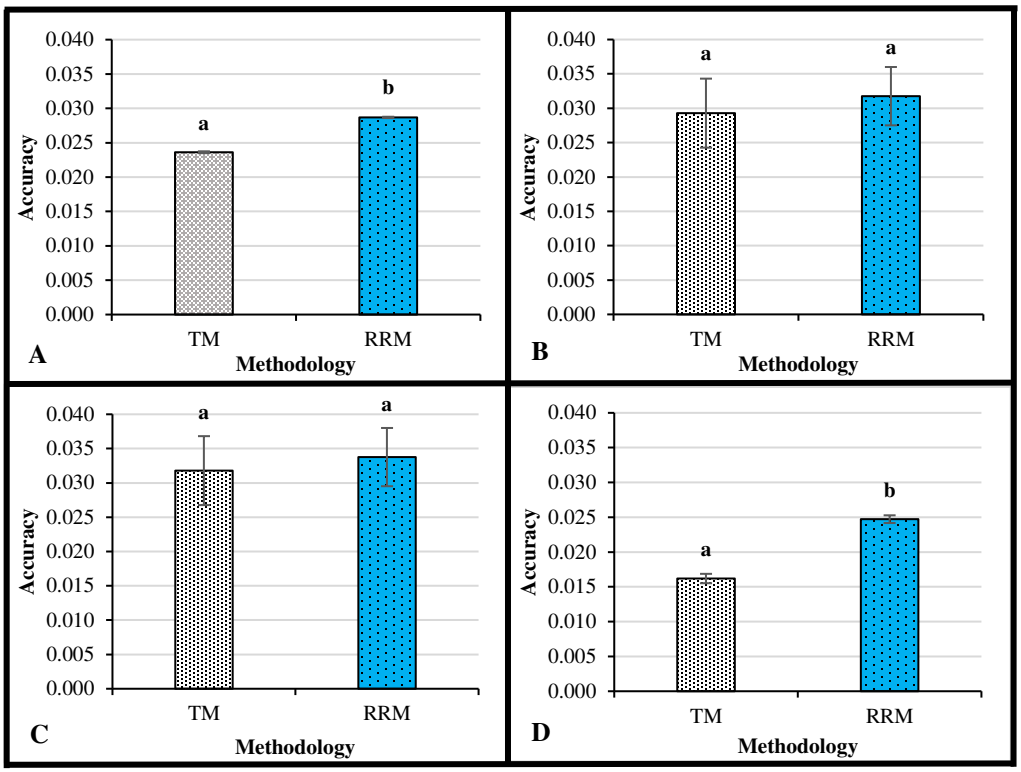


**Figure 3.7.** Mean accuracies for heifer first-service conception rate predictions at the average age of first exposure (422 d) obtained with each statistical methodology. **A)** Mean accuracy for all animals in the pedigree (n = 14,140), **B)** Mean accuracies for sires that have produced progeny in the last five yr (n = 85), **C)** Mean accuracies for sires that have produced progeny in the last three yr (n = 51), **D)** Mean accuracies of the 1-yr-old males potential selection candidates (n = 180). Different letters indicate a statistical difference at the P <0.05 level among methodologies according to the Fisher's least significant difference test.

Essentially, the same superiorities described for the mean accuracy values obtained with RRM in HPG analyzes were found for FSCR evaluations. Sires that have produced progeny in the last 5 years (Figure 3.7 B), was well as sires producing progeny in the 3 last years (Figure 3.7 C) and the youngest selection candidates (1-yr-old males; Figure 3.7 D), ended up with higher accuracies of predictions within the RRM method than with the TM method. The low accuracy values attained with both methodologies represent a result somewhat expected when taking into account the main factors affecting accuracy of genetic predictions: number of records, heritability of the trait, as well as, pedigree relationships (Bourdon, 2000). Among the previous factors, the

low heritability estimates obtained for the traits under study, as well as the relatively limited number of records available for the predictions may represent the main reasons for such low accuracies. In this regard, it has been mentioned that heritability plays a central role in formulas applied for the calculations of accuracy of predictions (Korsgaard et al., 2002). Basically, if $h^2$ is high, then accuracies of prediction are also high, whereas if $h^2$ is low, accuracies of prediction are low until a considerable number of observations are recorded. The reason for the previous is simple and was explained by Bourdon (2000), $h^2$ measures the strength of the relationship between breeding values and phenotypic values, as such, the stronger the relationship, each animal's performance record is a better indicator of that animal's breeding value.

Speidel et al. (2018b) reported an average accuracy of 0.604 for a HPG genetic prediction in Red Angus cattle that used a total of 104,100 phenotypic observations and a heritability on the underlying scale equal to 0.10. As an important note, within that study, the authors included a 3-generation pedigree of animals with valid phenotypes; therefore, no male half-sibs of heifers producing observations were considered within the evaluation. Conversely, in the current study, the vast majority of males within the pedigree were evaluated based solely in single observations of their female collateral relatives and not with progeny derived observations. Additionally, accuracy values reported in Speidel et al. (2018b) were presented as true accuracies ($r_{TI}$), whereas in the present study results were BIF accuracies ($r_{BIF}$), whose rate of increase towards 1 is much less pronounced than the rates of true accuracies or reliabilities (Van Vleck, 2016). Transforming the true accuracy for HPG reported by Speidel and coworkers to a BIF accuracy $\left(r_{BIF} = 1 - \sqrt{1 - r_{TI}^2}\right)$, results in a $r_{BIF}$ of 0.203, which is still being greater than the mean ACC obtained in this study for the TM. However, evident differences in number of records, data

structures and genetic parameters utilized among studies, are helpful to understand the big differences between contrasting results.

Regarding the mean accuracies for FSCR, no specific reports for accuracies of genetic predictions for this trait were found in literature. Perhaps the most comparable study was performed by Veerkamp et al. (2001), in which accuracies for the interval from first to second calving in Holsteins cows was reported including FSCR as a correlated trait using RRM. In that study, authors emphasized the benefits of using correlated traits to add accuracy in genetic predictions for dairy sires with relatively small number of daughters, nonetheless, no specific accuracies for FSCR were reported. Despite the previous, an interesting description of the variation in breeding value accuracy as a function of the number of daughters per sire was provided for calving interval. According to the authors, when more than 100 daughters of a particular sire had phenotypic observations of calving interval, minimal increments in accuracy of prediction were achieved when adding a secondary trait.

Applying such conclusions to the present study, it is highly probable that analyzing FSCR (as well as HPG) in a multiple-trait approach represented a good strategy to increase accuracy of prediction, since that is one of the main benefits of multivariate analyses; however, that particular analytic scenario was beyond of the scope of the current investigation. Further investigation of such a strategy might be more difficult considering that within beef cattle, the achievement of high numbers of progeny records per sire represents a much more daunting task than for dairy cattle. For instance, in the data available in the present study, only 9 sires had more than 50 daughters with phenotypic records for HPG and FSCR (Table 3.8). A similar number of daughter records by sire was reported by Bormann et al. (2006) when evaluating HPG and FSCR in Angus heifers from 6 different herds.

**Table 3.8.** Number of daughter records of heifer fertility traits per sire

| Number of daughters | Sires |
|---|---|
| <5 | 141 |
| 5 to 9 | 93 |
| 10 to 14 | 39 |
| 15 to 19 | 27 |
| 20 to 29 | 22 |
| 30 to 39 | 14 |
| 40 to 49 | 9 |
| >50 | 9 |

Finally, is important to acknowledge that increments in accuracy of predictions obtained with RRM were minimal in comparison to the accuracies obtained using TM in both traits of interest. A plausible cause of the similarities found between the mean accuracies obtained with each methodology could be that no extreme differences within datasets used with each method were present. Furthermore, another possibility relies in the fact that the non-repeated nature of the traits might constrained one of the main capabilities of RRM, which is precisely to more appropriately model the covariance structure of longitudinal traits (Meyer, 2004; Schaeffer, 2004; Schaeffer and Jamrozik et al., 2008).

Increments in mean accuracies of prediction ranged from 0.005 and 0.01 for HPG and FSCR, respectively. For both traits, the slight increases in accuracy of predictions were more evident in sires producing progeny within the last 5 or 3 yr and within the youngest selection candidates (1-yr-old males) of the CSU-BIC. Similar, but higher increments in accuracy of predictions of young selection candidates, were reported in a simulation study that compared the outcomes of RRM and multiple-trait models (MTM) in the genetic evaluation of weight traits (Meyer, 2004). Using field data, Boligon et al. (2011) reported that the highest gains in accuracy when using RRM as opposed to MTM were obtained at ages with a low number of weight records

(e.g., young animals). Although differences in accuracies found in the present study were minimal, and maybe from a practical perspective even insignificant, it was highly likely that distortions originated from the removal of observations suffering from ECP problems within the TM method were not actually severe. Nonetheless, even with such small differences among datasets, results of this study support the hypothesis that RRM could yield more accurate results than TM. Evidently, additional research including larger datasets and preferably utilizing information coming from different herds would be necessary to confirm or refute the current results.

*3.4 Conclusion*

This study compiled evidence regarding the feasibility of the application of random regression techniques in genetic evaluations of singly-observed binary traits like HPG and FSCR. Furthermore, the moderate to strong Pearson and Spearman's correlations found between predictions obtained with RRM and TM, suggested that RRM represent a viable option to substitute traditional genetic evaluation procedures of heifer binary traits. Lastly, even when differences in accuracies of predictions for both traits were minimal, RRM demonstrated their ability to overcome ECP problems and utilize all available information to produce more accurate results in comparison to TM.

*Literature cited*

Abdel-Azim, G. A. and P. J. Berger. 1999. Properties of threshold model predictions. J. Animal Sci. 77:582-590.

Averill, T. A., R. Rekaya, and K. Weigel. 2004. Genetic analysis of male and female fertility using longitudinal binary data. J. Dairy Sci. 87:3947-3952.

Beef Improvement Federation (BIF). 2010. Guidelines for uniform beef improvement programs. 9th ed. BIF, North Carolina State University, Raleigh, NC.

Beckman, D. W., R. M. Enns, S. E. Speidel, B. W. Brigham, and D. J. Garrick. 2007. Maternal effects on docility in Limousin cattle. J. Anim. Sci. 85:650-657.

Boldt, R. J., S. E. Speidel, M. G. Thomas, and R. M. Enns. 2018. Genetic parameters for fertility and production traits in Red Angus cattle. J. Anim. Sci. 96:4100-4111.

Boligon, A. A., F. Baldi, M. E. Z. Mercadante, R. B. Lobo, R. J. Pereira, and L. G. D. Albuquerque. 2011. Breeding value accuracy estimates for growth traits using random regression and multi-trait models in Nelore cattle. Genet. Mol. Res. 10:1227-1236.

Bormann, J. M., L. R. Totir, S. D. Kachman, R. L. Fernando, and D. E. Wilson. 2006. Pregnancy rate and first-service conception rate in Angus heifers. J. Animal Sci. 84:2022-2025.

Bourdon, R. M. 2000. Understanding animal breeding (second edition). Upper Saddle River, NJ: Prentice Hall.

Brinks, J. S., N. C. Speer, D. W. Schafer, D. G. LeFever and J. L. Moon. 1990. Factors affecting pregnancy rates to synchronized breeding: repeatability and heritability estimates. Beef Program Report, Colorado State University, Department of Animal Sciences, Ft. Collins, CO.

Brügemann, K., E. Gernand, U. U. von Borstel, and S. König. 2013. Application of random regression models to infer the genetic background and phenotypic trajectory of binary conception rate by alterations of temperature×humidity indices. Livest. Sci. 157:389-396.

Buddenberg, B. J., C. J. Brown, Z. B. Johnson, J. E. Dunn, and H. P. Peterson. 1989. Heritability estimates of pregnancy rate in beef cows under natural mating. J. Anim. Sci. 67:2589-2594.

Christensen, R. H. B. 2015. Package 'ordinal'.
Retrieved from cran.rproject.org/web/packages/ordinal/ordinal.pdf

Dearborn, D. D., R. M. Koch, L. V. Cundiff, K. E. Gregory, and G. E. Dickerson. 1973. An analysis of reproductive traits in beef cattle. J. Anim. Sci. 36:1032-1040.

Dempster, E. R. and I. M. Lerner. 1950. Heritability of threshold characters. Genetics. 35:212-236.

De Haas, Y., L. L. G. Janss, and H. N. Kadarmideen. 2007. Genetic correlations between body condition scores and fertility in dairy cattle using bivariate random regression models. J. Anim. Breed. Genet. 124:277-285.

Doyle, S. P., B. L. Golden, R. D. Green, and J. S. Brinks. 2000. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. J. Anim. Sci. 78:2091-2098.

Englishby, T. M., G. Banos, K. L. Moore, M. P. Coffey, R. D. Evans, and D. P. Berry. 2016. Genetic analysis of carcass traits in beef cattle using random regression models. J. Anim. Sci. 94:1354-1364.

Evans, J. L., B. L. Golden, R. M. Bourdon, and K. L. Long. 1999. Additive genetic relationships between heifer pregnancy and scrotal circumference in Hereford cattle. J. Anim. Sci. 77:2621-2628.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th Ed. Longman, Harlow. Essex, UK.

Fischer, T. M., A. R. Gilmour, and J. H. van der Werf. 2004. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. Gen. Sel. Evol. 36:1-8.

Fortes, M.R.S., W. M. Snelling, A. Reverter, S. H. Nagaraj, S. A. Lehnert, R. J. Hawken, K. L. DeAtley, S. O. Peters, G. A. Silver, G. Rincon, and J. F. Medrano. 2012. Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J. Anim. Sci. 90:2894-2906.

Foulley, J. L. 1992. Prediction of selection response for threshold dichotomous traits. Genetics. 132:1187-1194.

Foxworthy, H. 2019. Heritability and Repeatability Estimates of First Service Conception and First Cycle Calving in Angus Cattle. Master's Thesis, Colorado State University, Fort Collins.

Garrick, D.J., Garrick, D.P., Golden, B.L. 2018. An introduction to bolt software for genetic and genomic evaluations. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod. pp. 973.

Gianola, D. 1982. Theory and analysis of threshold characters. J. Anim. Sci. 54:1079-1096.

Gianola, D., and J. L. Foulley. 1983. Sires evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201-224.

Gianola, D., and G. J. Rosa. 2015. One hundred years of statistical developments in animal breeding. Annu. Rev. Anim. Biosci. 3:19-56.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, and D. Buttler. 2009. ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK*.

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's toolkit user' s guide and reference manual. Agric. Exp. Stn. Tech. Bull. LTB92-2. Colorado State University, Fort Collins, CO.

Golden, B. L., S. Weerasinghe, B. Crook, S. Sanders, and D. J. Garrick. 2018. A Single-Step Hybrid Marker Effects Model Using Random Regression for Stayability in Hereford Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand.

Hagger, C., and A. Hofer. 1989. Correlations between breeding values of dairy sires for frequency of dystocia evaluated by a linear and a nonlinear method. J. Anim. Sci. 67(Suppl. 1):88.

Harville, D. A., and R. W. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. Biometrics. 393-408.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics. 423-447.

Horan, B., J. F. Mee, P. O'connor, M. Rath, and P. Dillon. 2005. The effect of strain of Holstein-Friesian cow and feeding system on postpartum ovarian function, animal production and conception rate to first service. Theriogenology. 63:950-971.

Jamrozik, J., S. McGrath, R. A. Kemp, and S. P. Miller. 2013. Estimates of genetic parameters for stayability to consecutive calvings of Canadian Simmentals by random regression models. J. Anim. Sci. 91:3634-3643.

Johnston, D. J., S. A. Barwick, G. Fordyce, R. G. Holroyd, P. J. Williams, N. J. Corbet, and T. Grant. 2014. Genetics of early and lifetime annual reproductive performance in cows of two tropical beef genotypes in northern Australia. Anim. Prod. Sci. 54:1-15.

Kenward, M. G. and J. H. Roger. 1997. The precision of fixed effects estimates from restricted maximum likelihood. Biometrics. 53:983–997.

Korsgaard, I. R., A. H. Andersen, and J. Jensen. 2002. Prediction error variance and expected response to selection, when selection is based on the best predictor–for Gaussian and

threshold characters, traits following a Poisson mixed model and survival traits. Genet. Sel. Evol. 34:1-27.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics. 38:963-974.

Lewis, R. M., and S. Brotherstone. 2002. A genetic evaluation of growth in sheep using random regression techniques. Animal Sci. 74:63–70.

Lopes, U. V., D. A. Huber, and T. L. White. 1990. Comparison of methods for prediction of genetic gain from mass selection on binary threshold traits. Silvae Genet. 49:50-56.

Marshall, D. M., W. Minqiang, and B. A. Freking. 1990. Relative calving date of first-calf heifers as related to production efficiency and subsequent reproductive performance. J. Anim. Sci. 68:1812-1817.

Mathiews, G. L., R. D. Green, J. S. Brinks, B. L. Golden, R. M. Enns, and D. G. LeFever. 1995. Genetic parameters for conception rate to synchronized breeding in Angus cattle. Proc. West. Sec. Am. Soc. Anim. Sci. 46:217.

McAllister, C. M., S. E. Speidel, D. H. Crews Jr., and R. M. Enns. 2011. Genetic parameters for intramuscular fat percentage, marbling score, scrotal circumference, and heifer pregnancy in Red Angus cattle. J. Anim. Sci. 89:2068-2072.

Meijering, A. 1984. Dystocia and stillbirth in cattle—A review of causes, relations and implications. Livest. Prod. Sci. 11:143-177.

Meijering, A., and D. Gianola. 1985. Linear versus nonlinear methods of sire evaluation for categorical traits: a simulation study. Genet. Sel. Evol. 17:115-132.

Meyer, K. 2004. Scope for a random regression model in genetic evaluation of beef cattle for growth. Livest. Prod. Sci. 86:69-83.

Misztal, I., D. Gianola, and J. L. Foulley. 1989. Computing aspects of a nonlinear method of sire evaluation for categorical data. J. Dairy Sci. 72:1557-1568.

Peters, S. O., K. Kizilkaya, D. J. Garrick, R. L. Fernando, J. M. Reecy, R. L. Weaber, G. A. Silver, and M. G. Thomas. 2013. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. J. Anim. Sci. 91:605-612.

Phocas, F., and D. Laloë. 2003. Evaluation models and genetic parameters for calving difficulty in beef cattle. J. Anim. Sci. 81:933-938.

R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Sánchez-Castro, M. A., R. J. Boldt, M. G. Thomas, R.M. Enns, and S. E. Speidel. 2017. Expected progeny differences for stayability in Angus cattle using a random regression model. J. Anim. Sci. 95(Suppl. 4):88.

Sánchez-Castro, M. A., M. G. Thomas, R. M. Enns, and S. E. Speidel. 2019. Stability of genetic predictions for stayability using random regression models that include end points beyond 6 yr of age. Trans. Anim. Sci. 3:1678–1682.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Schaeffer L. R. and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145-146.

Silvestre, A., Â. Martins, V. Santos, and J. Colaço. 2019. Genetic parameters of calving ease in dairy cattle using threshold and linear models. Ital. J. Anim. Sci. 18:80-87.

Snelling, W. M., B. L. Golden, and R. M. Bourdon. 1995. Within-herd genetic analyses of stayability of beef females. J. Anim. Sci. 73:993-1001.

Speidel, S.E. 2011. Random regression models for the prediction of days to finish in beef cattle. Ph.D. diss. Colorado State Univ., Ft. Collins.

Speidel, S. E., R. M. Enns, and B. L. Golden. 2018a. Use of a Random Regression model for the Evaluation of Heifer Pregnancy in Red Angus Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand.

Speidel, S. E., B. A. Buckley, R. J. Boldt, R. M. Enns, J. Lee, M. L. Spangler, and M. G. Thomas. 2018b. Genome-wide association study of Stayability and Heifer Pregnancy in Red Angus cattle. J. Anim. Sci. 96:846-853.

Thaller, G. 1997. Genetics and breeding for fertility. Interbull Bull.18:55–61.

Toelle, V. D., and O. W. Robison. 1985. Estimates of genetic correlations between testicular measurements and female reproductive traits in cattle. J. Anim. Sci. 60:89-100.

Van Vleck, L. D. 2016. EPDs and Risk. In: Proc. Beef Imp. Fed. 48th Annu. Res. Symp. Annu. Meet., Manhattan, KS. pp. 107-114.

Veerkamp, R. F., E. P. C. Koenen, and G. De Jong. 2001. Genetic correlations among body condition score, yield, and fertility in first-parity cows estimated by random regression models. J. Dairy Sci. 84:2327-2335.

Weller, J. I., I. Misztal, and D. Gianola. 1988. Genetic analysis of dystocia and calf mortality in Israeli-Holsteins by threshold and linear models. J. Dairy Sci. 71:2491-2501.

Yin, T., B. Bapst, U. U. V. Borstel, H. Simianer, and S. König. 2012. Genetic parameters for Gaussian and categorical traits in organic and low input dairy cattle herds based on random regression methodology. Livest. Sci. 147:159-169.

# CHAPTER 4 – THRESHOLD AND RANDOM REGRESSION MODELS FOR THE GENETIC EVALUATION OF LONGITUDINAL BINARY FERTILITY TRAITS IN MULTIPAROUS BEEF COWS

*Summary*

Stayability (STAY) and first-service conception rate (FSCR) are two economically relevant traits in beef cows associated with longevity and superior fertility. Given the binary outcomes of their phenotypes, genetic evaluations for both traits rely on the use of threshold models (TM). Nonetheless, since their binary observations can be assigned to various discrete points in time during a cow's lifetime, the implementation of random regression models (RRM) might be attractive because of their ability to include any range of age endpoints for which phenotypic data is available. Few formal comparisons have been reported between RRM and TM genetic evaluations for binary fertility longitudinal traits in beef cattle. Therefore, the objectives of this chapter were to compare genetic evaluations for STAY and FSCR using RRM by contrasting resulting EPD and genetic parameters to those obtained with TM. Additionally, differences in accuracies of prediction between methodologies were also evaluated. Calving data, as well as, breeding and reproductive ultrasound records of multiparous Angus cows from the John E. Rouse Colorado State University Beef Improvement Center collected between 1990 to 2019 were used for the study. Ten STAY endpoints defined as whether a cow calved at age 3, 4, and up to 12 yr given she calved as a 2-yr-old were assigned observations (1, successful; 0, unsuccessful). Similarly, ten FSCR age specific observations were assigned depending on the age of exposure of the females (ages ranged from 2 to 11 yr) and were defined by fetal age at pregnancy inspections performed approximately 130 d post artificial insemination. Traditional evaluation of STAY was

performed using a TM that only considered the success/failure of females reaching the age of 6 (STAY06), since this age is considered as the financial break-even for cows within the beef industry. Conversely, given there is no specific age of interest for a multiparous cow to conceive in response to her first AI, the traditional evaluation for FSCR was performed using a repeatability TM. Alternative evaluations for both traits were performed by regressing each trait on its corresponding age specific endpoints using univariate linear RRM with Legendre Polynomials as the base function. Heritability ($h^2$) estimates obtained for STAY06 were 0.10 and 0.04 for the TM and the RRM, respectively. In the case of FSCR, $h^2$ estimates were 0.03 for the TM and ranged between 0.02 to 0.05 for all the ages at exposure considered in the RRM. Pearson ($r_p$) and Spearman's ($r_s$) correlations between EPD obtained with each method for STAY06 were 0.84 and 0.86. For FSCR, correlations were calculated between the EPD obtained with the repeatability TM and each one of the age-specific EPD obtained with the RRM; therefore, results for the $r_p$ ranged between 0.70 to 0.99; whereas results for $r_s$ ranged between 0.69 to 0.99, depending on the age of exposure considered in the RRM. Although mean accuracies of prediction were higher using RRM than using TM for both traits, increments were much more relevant for STAY than for FSCR. These results suggested that a RRM genetic prediction for STAY06 is more efficient than the traditional TM evaluation for this trait, since it yielded higher accuracy of prediction. More importantly, the strong degrees of concordance found between predictions obtained with both methodologies for STAY06, suggested that RRM could effectively substitute TM in genetic evaluations of this trait. For FSCR, no tangible improvements were achieved by evaluating the trait using random regression techniques, mainly due to the lack of influence that age had on the success or failure of cows to conceive in response to their first AI at consecutive ages.

*4.1 Introduction*

Female reproductive efficiency represents a major profitability driver of beef cattle operations since the quantity of beef produced relies on the number of calves born and raised per breeding cycle (Grossi et al., 2008). Producing the greatest number of calves each year depends on the ability of the cows to achieve calving intervals of 365 d (Walmsley et al., 2018). Such successful female reproductive ability is a key element for an important beef cattle fertility trait like stayability (STAY), since the trait measures the probability of a cow to produce one calf per year up to a specific age endpoint, normally established at 6 yr of age (Snelling et al., 1995). Although a very low proportion (~7.6%) of beef operations in the US use artificial insemination (AI) as a reproductive management tool (Lamb and Mercadante, 2016); superior first-service conception rates (FSCR) may also contribute to accomplish calving intervals of less than a year. Females conceiving in response to their first AI will calve earlier in the following calving season, and consequently, will have a greater chance to re-breed within a year (Deutscher et al., 1991).

The biology of traits like STAY and FSCR establishes that their phenotypes be recorded on a binary scale where values of 1 represent successful observations (i.e., pregnant) and values of 0 represent the opposite (i.e., non-pregnant). Given the categorical nature of their phenotypes, genetic predictions for these fertility traits have been performed using threshold models (TM; Gianola and Foulley, 1983; Harville and Mee, 1984). Although no formal genetic evaluations have been implemented for FSCR within any US breed association; in the case of STAY, significant genetic improvements have been made through the use of TM in several beef breeds (Snelling et al., 1995; Van Melis et al., 2007; Crews and Enns, 2008). Nonetheless, the inability of TM to include information from subclasses of categorical fixed effects with no variation represents an important limitation of the utilization of all phenotypic records available (Golden et al., 2018).

Furthermore, restrictions related to the age-specific definition of traits like STAY have also been a motive of concern due to the considerable amount of time (e.g., 6 yr) needed to collect observations useful to accurately evaluate sires (Brigham et al., 2007; Speidel et al., 2018a).

Even when the reproductive success or failure of a female is a binary observation within a year, depending on the number of years that a cow has an opportunity to express such performance within the herd, those events are also longitudinal. Consequently, the application of more robust analytical techniques like random regression models (RRM) represent an alternative for the evaluation of binary fertility traits to consecutive ages (Schaeffer, 2004; Jamrozik et al., 2013). Interesting features of such approach are the capabilities of RRM to include any range of age endpoints for which phenotypic data is available and their flexibility to incorporate information from class levels of categorical fixed effects with no variation (Golden et al., 2018). Furthermore, possibilities of generating age-specific genetic predictions and the easy modeling of time-dependent environmental effects represents also attractive characteristics of this method (Jamrozik et al., 2013). Given RRM have a superior ability to utilize all available phenotypic information when compared to TM; we hypothesized that RRM genetic predictions for STAY and FSCR could yield more accurate predictions in comparison to those obtained using a TM. Hence, the objective of this chapter was to perform a comparison between pedigree-based genetic predictions of STAY and FSCR, obtained using both TM and RRM.

*4.2 Materials and Methods*

Data used in this study were obtained from an existing database; however, animals within the experimental location were managed according to the Institutional Animal Care and Use Committee (IACUC) guidelines, covered in most recent years by the IACUC number 18-8367A.

*4.2.1 Data collection and description*

Calving performance data from 1,713 Angus females (progeny of 302 sires and 1,068 dams) collected between 1993 and 2019 at the John E. Rouse Colorado State University Beef Improvement Center (CSU-BIC) were used for the STAY study. Stayability observations were assigned to dams according to their age in days at each calving. Given every female calved as a 2-yr-old, starting from their third calving, the value of 1 (successful) or 0 (unsuccessful) was attributed to cows that either produced a calf or did not produce a calf within each particular age endpoint (ages ranging from 3 to 12 yr). A total of ten STAY endpoints were defined for the study, ranging from STAY03 through STAY12, forming a final data set of 8,907 observations.

A slightly larger dataset was available for FSCR analyses, since observations were based on breeding and ultrasound records collected between 1990 to 2019 from a total of 2,179 dams (progeny of 353 sires and 1,342 dams). Within each breeding year, cows were estrous synchronized and subjected to a single AI event prior to their placement into single-sire breeding pastures approximately 2-wk after the insemination (Crawford et al., 2016). Considering each year's specific AI date as the beginning of its respective breeding season (usually around June 15), FSCR was defined as the probability of a cow conceiving in response to her first AI service and maintaining such pregnancy after the end of a 60-d breeding season. Observations for FSCR (1, successful; 0, unsuccessful) were defined by fetal age obtained from 2 ultrasound pregnancy exams performed approximately 65 ± 5 d and 105 ± 5 d post-AI. Age at exposure of every female were calculated as the difference between an individual's birthdate and all its respective recorded breeding dates (ages ranged from 2 to 11 yr). Then, a total of ten FSCR age specific observations were assigned depending on the age of exposure of the females, ranging from FSCR02 through FSCR11. The final data set contained a total of 9,584 observations.

*4.2.2 Testing fixed and random effects*

Recommendations regarding potential fixed effects influencing STAY in genetic evaluations have been outlined within the Beef Improvement Federation's guidelines (BIF, 2020); however, no official recommendations were found for genetic evaluations of FSCR. Therefore, with the objective of the identification of important factors influencing these traits, incremental Wald F-tests were performed using the statistical software package ASREML 3.0 (Gilmour et al., 2009) according to equation 3.1. Possible fixed effects investigated for STAY were age at first calving (AFC), calving ease score of the immediate previous calving (CE), post-partum interval, breeding weight, breeding year, breeding pasture (confounded with service sire) and age at calving. In the case of FSCR, potential fixed effects included AFC, CE, post-partum interval, breeding weight, breeding year, synchronization protocol, semen type (e.g., sexed *vs* conventional), AI technician and breeding age. Results from the Wald F-tests performed for all the potential fixed effects of both traits are shown in Table 4.1.

**Table 4.1.** Results of Wald F tests for fixed effects for stayability (STAY) and first-service conception rate (FSCR) in Angus cows.

| Trait | Effect | NumDF† | DenDF† | F-inc† | P-inc† |
|---|---|---|---|---|---|
| STAY | Age at first calving (mo) | 7 | 20.5 | 5061.75 | <0.001 |
| | Age at calving (yr) | 1 | 58.1 | 15.38 | <0.001 |
| | Breeding pasture (and/or service sire) | 72 | 6853.9 | 7.07 | <0.001 |
| | Breeding weight (lbs) | 1 | 522.1 | 9.29 | 0.003 |
| | Breeding year | 26 | 3440.0 | 217.81 | <0.001 |
| | Post-partum interval (d) | 1 | 3004.4 | 1.06 | 0.305 |
| | Previous calving ease score | 4 | 3518.4 | 1170.52 | <0.001 |
| FSCR | Age at first calving (mo) | 1 | 1631.0 | 14.04 | <0.001 |
| | Age at exposure (yr) | 1 | 42.0 | 0.01 | 0.928 |
| | AI technician | 96 | 5838.0 | 1.90 | <0.001 |
| | Breeding weight (lbs) | 1 | 6157.9 | 25.25 | <0.001 |
| | Breeding year | 27 | 439.2 | 31.42 | <0.001 |
| | Post-partum interval (d) | 1 | 8177.1 | 4.38 | 0.038 |
| | Previous calving ease score | 6 | 3604.4 | 2.85 | 0.009 |
| | Semen type | 14 | 2126.7 | 10.07 | <0.001 |
| | Synchronization protocol | 30 | 4112.6 | 2.00 | 0.001 |

†NumDF = Numerator degrees of freedom (number of non-singular equations involved in the term); DenDF = denominator degrees of freedom (estimated according to the adjustments recommended by Kenward and Roger, 1997); F-inc = additional variation explained by the term being tested when added lastly to the model. P-inc = probability value.

Among the systematic effects tested for STAY, all of them resulted as significant sources of variation with the exception of post-partum interval; therefore, this variable was not included as predictor in the genetic evaluations for STAY. Possibly the reason of why post-partum interval did not explained variations in STAY is that, phenotypes for this trait include females conceiving as a result of a natural service. As such, even if cows fail to conceive during their first exposure within a particular breeding season, they still have opportunities to conceive naturally. Regarding the effects examined for FSCR, the only one that did not account for a significant amount of

variability in the phenotype was age at exposure. Such result suggests that the application of a linear RRM for the evaluation of cow FSCR may not be appropriate, since no evidence of variation associated with age were identified in this particular data set. This finding is opposite to a report by Azzam et al. (1989) in where age did significantly influence FSCR in beef cows; however, important differences in the age range for which observations were considered could explain contrasting results between studies. For instance, in the present study observations spanned ages from 2 to 11 yr, whereas in Azzam's and colleagues' study, the age range considered was between 1 and 3.5 yr. Acknowledging that according to the Wald F test, age could be excluded from the model to analyze FSCR, with the purpose of preserving the longitudinal nature of the trait it was decided to include this effect for further analysis; however, results of the RRM must be interpreted with caution.

The next step taken before performing the genetic evaluations was the determination of contemporary groups (CG) for each trait. Since a CG is a group of animals that have experienced a similar environment with respect to the expression of a given trait (Bourdon, 2000); the effects of breeding year and breeding pasture (specific to each breeding year and confounded with natural service sire) were combined to form CG for STAY. In the case of FSCR, given synchronization protocols and semen types were specific to each breeding year, all these effects were combined to create CG for this trait. Forming contemporary groups in this manner resulted in a total of 139 and 77 unique CG for STAY and FSCR, respectively. Summary statistics of the contemporary groups formed for each trait in this study are shown in Table 4.2.

**Table 4.2.** Summary statistics outlining the number of cows represented per contemporary group definition in both fertility traits.

| | STAY[1] | FSCR[2] |
|---|---|---|
| N | 139 | 77 |
| Average | 52.7 | 124.4 |
| SD | 30.4 | 122.1 |
| Minimum | 6 | 4 |
| Maximum | 217 | 438 |
| Average pregnancy rate | 0.92 | 0.46 |

[1]STAY = Stayability.
[2]FSCR = First-service conception rate.

Before performing the genetic evaluations for FSCR, two extra random effects were also tested. These variables were mating group (e.g., cows being inseminated 12 h after they were seen in heat or during a mass mate) and AI sire (sire that produced the semen straw used during the specific AI event). Following the procedure described by Beckman et al. (2007), the utility of the inclusion of these variables to the model was tested using a likelihood ratio test (LRT), implementing equation 3.2. Analyses were performed utilizing the package "ordinal" (Christensen, 2015) within the statistical software R (R Development Core Team, 2013). Results of the LRT are shown in Table 4.3.

**Table 4.3.** Results of log-likelihood ratio tests for random effects for first-service conception rate (FSCR) in Angus heifers.

| Effect* | $LogL_f$† | $LogL_r$† | $-2(LogL_r - LogL_f)$ | df | P-value |
|---|---|---|---|---|---|
| AI sire | 1847.52 | 1845.64 | 3.76 | 306 | <0.0001 |
| Mating group | 1847.52 | 1763.57 | 167.90 | 2 | <0.0001 |

*AI sire = sire that produced the semen straw used during the specific AI event; Mating group = cows inseminated during heat or during a mass mate.
†$LogL_f$ = log-likelihood value of the full model that included all effects; $LogL_r$ = log-likelihood value of the reduced model that included all effects except for the one being tested (indicated in the effect column).

Results from the LRT suggested both mating group and AI sire accounted for a significant portion of the variation of cow FSCR; therefore, the two variables were kept for subsequent genetic evaluations. A visual appreciation of the variability in conception rate associated with each individual AI sire and mating group are presented in Figures 4.1 and 4.2, respectively.

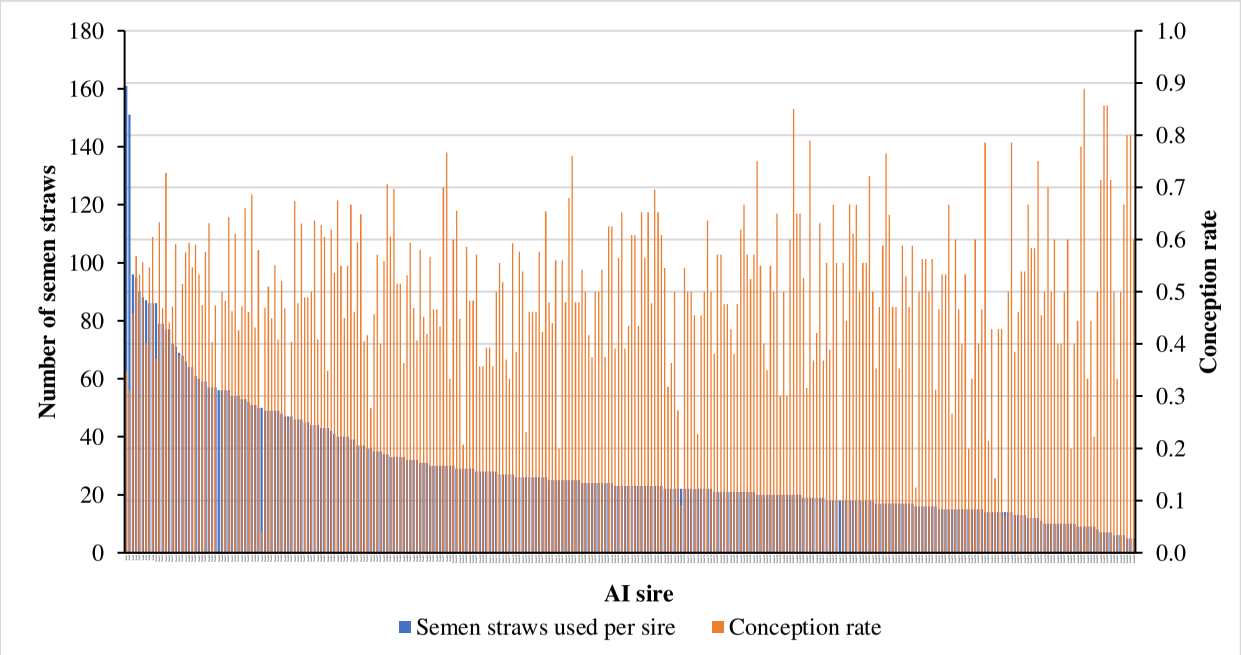


**Figure 4.1.** Average conception rate of AI sires utilized in cows from the Colorado State University Beef Improvement Center (blue bars represent the number of semen straws used per sire and orange bars represent the conception rate of each sire).

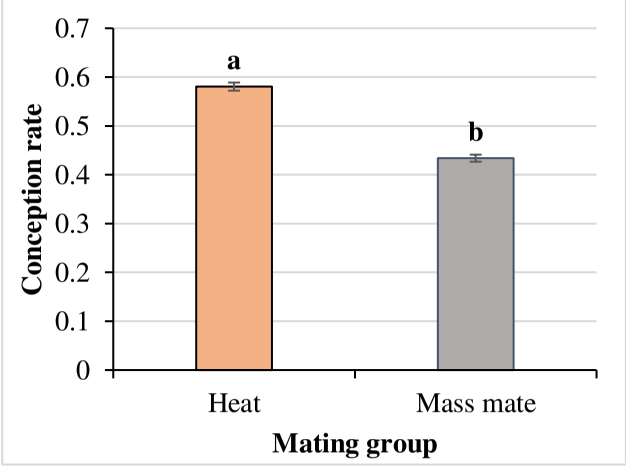


**Figure 4.2.** Average conception rate per mating group in cows from the Colorado State University Beef Improvement Center (3,543 observations belonged to cows inseminated in heat while 4,719 observations were from cows subjected to AI during a mass mate).

*4.2.3 Genetic evaluations for Stayability*

Traditional EPD calculation for STAY06 was performed using a univariate BLUP animal TM along with a probit link function to convert binary observations to an underlying normal distribution. The model Equation (4.1) was:

$$\boldsymbol{y^* = Xb + Zu + Qcg + e} \qquad \text{Eq. 4.1}$$

where $\boldsymbol{y^*}$ corresponded to a vector of transformed observations of STAY06 on the underlying scale; $\boldsymbol{b}$ was a vector of unknown solutions for fixed effects, which included AFC, CE and the individual's breeding weight as a linear covariate; $\boldsymbol{u}$ corresponded to a vector of unknown solutions of animal random effects; $\boldsymbol{cg}$ represented a vector of unknown solutions of contemporary group random effects; $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{Q}$ were known incidence matrices relating observations in $\boldsymbol{y^*}$ to fixed ($\boldsymbol{b}$), animal random ($\boldsymbol{u}$) and contemporary group random ($\boldsymbol{cg}$) effects; and $\boldsymbol{e}$ was the vector of unknown residual errors. The mean for random effects was assumed to be 0 while variances were assumed to be distributed as:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{cg}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_a^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_{cg}\sigma_{cg}^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

where **A** represented the additive numerator relationship matrix amongst animals included in the pedigree; $\boldsymbol{I_{cg}}$ and $\boldsymbol{I_n}$ were identity matrices with orders equal to the number of contemporary groups and observations, respectively. The $\boldsymbol{\sigma_a^2}$, $\boldsymbol{\sigma_{cg}^2}$ and $\boldsymbol{\sigma_e^2}$ denoted the additive, contemporary group and residual variances, respectively. Importantly, the additive variance ($\boldsymbol{\sigma_a^2}$) was specific for the evaluated age endpoint (STAY06) and the residual variance ($\boldsymbol{\sigma_e^2}$) was constrained to be equal to 1.

Alternatively, all STAY endpoints (STAY03 through STAY12) were evaluated together using a linear RRM with Legendre polynomials as its base function. The model in matrix form is presented in Equation 4.2 below:

$$\boldsymbol{y = Xb + Z_1u + Z_2p + Qcg + e} \qquad \text{Eq. 4.2}$$

where $\boldsymbol{y}$ corresponded to a vector of binary STAY observations, $\boldsymbol{X}$ was an incidence matrix relating STAY observations in $\boldsymbol{y}$ to AFC, CE, breeding weight and fixed regression coefficients of STAY on age at calving to their solutions in $\boldsymbol{b}$; $\boldsymbol{Z_1}$ represented an incidence matrix of age covariates relating the STAY observations in $\boldsymbol{y}$ to the random additive genetic regression coefficients (intercept and linear) in $\boldsymbol{u}$; $\boldsymbol{Z_2}$ was an incidence matrix of age covariates relating STAY observations in $\boldsymbol{y}$ to the permanent environmental linear random regression coefficients for each animal in $\boldsymbol{p}$; $\boldsymbol{Q}$ was a known incidence matrix relating STAY observations in $\boldsymbol{y}$ to their corresponding random contemporary group effects in $\boldsymbol{cg}$; and $\boldsymbol{e}$ was the vector of unknown residual errors. Random effect means were assumed to be 0 whereas variances were assumed to be:

$$\mathbf{Var}\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{p} \\ \boldsymbol{cg} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \otimes \mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I_p} \otimes \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_{cg}\sigma^2_{cg}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{R} \end{bmatrix}$$

where **A** represented the Wright's numerator relationship matrix, $\otimes$ was the Kronecker product; **G** corresponded to a (co)variance matrix of additive genetic random regression coefficients; **P** was a (co)variance matrix of permanent environmental random regression coefficients and, $\mathbf{R} = \mathbf{diag}\{\boldsymbol{\sigma^2_{e_k}}\}$ was a diagonal matrix of temporary environmental variances that themselves vary depending on the k$^{th}$ age endpoint. $\boldsymbol{I_p}$ and $\boldsymbol{I_{cg}}$ represented identity matrices whose order were equal to the number of observations and contemporary groups, respectively. Lastly, the $\boldsymbol{\sigma^2_{cg}}$ remained as described for the TM.

This RRM predicted the genetic merit of the presence of a calf at each particular age endpoint; therefore, EPD were summed to obtain the individual's genetic merit for the presence of a calf at 3, 4, and up to 12 years of age. A pedigree file from the CSU-BIC consisting of 14,140 individual animals, with 971 and 3,725 unique sires and dams, respectively, was used for the estimation of genetic parameters within both statistical methodologies (TM and RRM). The average inbreeding coefficient of the pedigree was 0.009, with a minimum of 0 and a maximum of 0.263. Within the TM methodology, the heritability estimate on the underlying scale was obtained by calculating the ratio of the additive to the phenotypic variance $\left(\sigma_A^2 \Big/ \sigma_P^2\right)$. Conversely, variance estimates obtained using RRM were utilized to calculate observed variance estimates for each age endpoint utilizing equation 3.5. The previous transformation allowed the possibility to calculate heritabilities of STAY at every age endpoint considered within the study (from 3 to 12 yr of age).

Regarding predictions obtained with each method, the TM predicted a single breeding value per animal on the underlying scale whereas predictions obtained with RRM resulted in a vector equal to the order of the Legendre polynomials (e.g., each animal had a prediction for the intercept and the linear term). Therefore, for comparison purposes, the RRM predictions were condensed into single values per animal expressed on an observed scale for each age endpoint using equation 3.6. After such conversion, particularly the prediction on the observed scale at the age of 6 was chosen to be compared to the prediction obtained with the TM. Once that both sets of predictions were available, a homogenization of the prediction scales was performed for both methods following the procedure outlined by Speidel et al. (2018b). Briefly, predictions were converted to a pseudo-probability scale and expressed as deviations from 50% (random chance of conception). Resulting predictions were compared through the calculation of Pearson ($r_p$) and

Spearman's ($r_s$) correlations, as well as the estimation of the regression coefficient of EPD obtained with the RRM on those obtained with the TM. Analyses were performed using ASREML 3.0 (Gilmour et al., 2009), the Animal Breeder's Tool Kit (Golden et al., 1992) and BOLT (http://www.thetasolutionsllc.com/bolt-software.html).

In order to calculate the accuracy of the predictions obtained with both methodologies, it was necessary to obtain the prediction error variance (PEV). In the case of the TM, the prediction error variance of the $i^{th}$ animal (*PEVi*) was obtained by squaring the standard error reported next to the BLUP of each individual evaluated on the ASREML output solutions file (Gilmour et al., 2009). These values represented approximations of the diagonal elements of the inverse of the coefficient matrix assembled in the final iteration round performed by the statistical software package. Conversely, given the mixed-model equations for the RRM evaluations were assembled and solved using the BOLT software (Garrick et al., 2018), the PEV of each animal was estimated via Markov Chain Monte Carlo procedures (MCMC) using Gibbs sampling. In summary, a total of 100,000 samples were obtained after disregarding the first 5,000 samples during the burn-in period in order to obtain the estimates of PEV. Finally, once the PEV of each methodology was obtained, mean accuracies (ACC) were calculated according to the guidelines of the Beef Improvement Federation (2020) using Equation 3.7 and then compared to each other.

*4.2.4 Genetic evaluations for cows first-service conception rate*

The EPD calculation for FSCR was performed using a univariate repeatability TM that included a probit link function to convert binary observations to an underlying normal distribution. The model Equation (4.3) was:

$$\boldsymbol{y^* = Xb + Z_1u + Z_2p + Qm + Ws + e} \qquad \text{Eq. 4.3}$$

where $\boldsymbol{y^*}$ corresponded to a vector of transformed observations of FSCR on the underlying scale; $\boldsymbol{b}$ was a vector of unknown solutions for fixed effects that included AFC, CE, breeding contemporary group (defined as a combination between breeding year, synchronization protocol and semen type), AI technician, and the individual's post-partum interval, breeding weight and age at exposure as a linear covariates; $\boldsymbol{u}$ corresponded to a vector of unknown solutions of animal random effects; $\boldsymbol{p}$ denoted a vector of unknown random permanent environmental effects; $\boldsymbol{m}$ was a vector of unknown solutions of mating group (e.g., inseminated in heat or during a mass mate) random effects and, $\boldsymbol{s}$ was a vector of unkwnown solutions for AI sire random effects. The matrices $\boldsymbol{X}$, $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$, $\boldsymbol{Q}$ and $\boldsymbol{W}$ were known incidence matrices relating observations in $\boldsymbol{y^*}$ to fixed effects in $\boldsymbol{b}$, as well as animal, permanent environment, mating group and AI sire random effects in $\boldsymbol{u}$, $\boldsymbol{p}$, $\boldsymbol{m}$ and $\boldsymbol{s}$, respectively. Finally, $\boldsymbol{e}$ was a vector of unknown residual errors. It was assumed that the mean of random effects was equivalent to 0 whereas variances were assumed to be equal to:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{p}\\ \boldsymbol{m}\\ \boldsymbol{s}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_a^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_p\sigma_p^2} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_m\sigma_m^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_w\sigma_s^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

where $\mathbf{A}$ denoted the Wright's numerator relationship matrix, $\boldsymbol{I_p}$, $\boldsymbol{I_m}$, $\boldsymbol{I_w}$ and $\boldsymbol{I_n}$ were identity matrices whose orders were equal to the number of animals, mating groups, AI sires and observations, respectively. The $\boldsymbol{\sigma_a^2}$, $\boldsymbol{\sigma_p^2}$, $\boldsymbol{\sigma_m^2}$, $\boldsymbol{\sigma_s^2}$ and $\boldsymbol{\sigma_e^2}$ were the additive, permanent environment, mating group, AI sire and residual variances, respectively. Within this model, the residual variance ($\boldsymbol{\sigma_e^2}$) was constrained to be equal to 1. For this TM, heritability was calculated as the ratio of the additive to the phenotypic variance $\left(\sigma_A^2 \big/ \sigma_P^2\right)$, whereas repeatability was estimated as the ratio of

the variance in producing ability ($\sigma^2_{PA}$ = sum of the genetic and permanent environmental variances) to the phenotypic variance $\left(\sigma^2_{PA}/\sigma^2_P\right)$, as described by Foxworthy (2019b).

In addition, FSCR was regressed on age at exposure using a linear RRM with Legendre polynomials as the base function. The model in matrix form is presented in Equation 4.4 below:

$$\boldsymbol{y = Xb + Z_1u + Z_2p + Qm + Ws + e} \qquad \text{Eq. 4.4}$$

where $\boldsymbol{y}$ represented a vector of binary observations of FSCR, $\boldsymbol{X}$ was an incidence matrix relating FSCR observations in $\boldsymbol{y}$ to the vector of unknown solutions of AFC, CE, breeding contemporary group (combination between breeding year, synchronization protocol and semen type), AI technician, post-partum interval, breeding weight and a set of fixed regression coefficients of age at exposure contained in $\boldsymbol{b}$; $\boldsymbol{Z_1}$ was an incidence matrix of age covariates relating FSCR observations in $\boldsymbol{y}$ to the unknown animal random additive genetic regression coefficients (intercept and linear) in $\boldsymbol{u}$; $\boldsymbol{Z_2}$ was an incidence matrix of age covariates relating FSCR observations in $\boldsymbol{y}$ to the permanent environmental linear random regression coefficients for each animal in $\boldsymbol{p}$; $\boldsymbol{Q}$ was an incidence matrix relating FSCR observations in $\boldsymbol{y}$ to random mating group effects in $\boldsymbol{m}$; $\boldsymbol{W}$ was a known incidence matrix relating FSCR observations in $\boldsymbol{y}$ to random AI sire effects in $\boldsymbol{s}$; and lastly, $\boldsymbol{e}$ was the vector of unknown residual errors. The mean of random effects was assumed to be 0 while the variances were assumed to be:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{p}\\ \boldsymbol{m}\\ \boldsymbol{s}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\otimes\mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_p}\otimes\mathbf{P} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_m\sigma^2_m} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_w\sigma^2_s} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma^2_e}\end{bmatrix}$$

where **A** represented the Wright's numerator relationship matrix, $\otimes$ was the Kronecker product, **G** corresponded to a variance-covariance matrix of additive genetic random regression coefficients

and **P** was a (co)variance matrix of permanent environmental random regression coefficients. $\boldsymbol{I_p}$, $\boldsymbol{I_m}$, $\boldsymbol{I_w}$, $\boldsymbol{I_n}$, $\boldsymbol{\sigma_m^2}$, $\boldsymbol{\sigma_s^2}$ and $\boldsymbol{\sigma_e^2}$ remained as described for the previous TM. For the estimation of the genetic parameters, the same pedigree file implemented in the STAY analyzes was utilized. Similarly, the same procedures described for STAY evaluations were followed for the FSCR analyses in order to estimate heritabilities (as well as repeatabilities) and to obtain EPD within each statistical method (e.g., refer to equations 3.5 and 3.6). The same comparative strategies previously mentioned for STAY predictions were followed to compare the outputs of each statistical model implemented to evaluate FSCR. Prediction error variances and accuracy estimations were performed as described for STAY and all analyses were performed using the same statistical packages previously described.

*4.3 Results and discussion*

*4.3.1 Phenotypic trends and data availability*

The percentages of dams receiving a successful observation at each of the STAY endpoints included in the study, as well as, the percentages of conception rate at each age of exposure are shown in Figures 4.3 A and 4.3 B, respectively. The tendency of the average percentage of STAY to consecutive calvings is clearly negative, which was consistent with previous reports and the biology of cow production in general (Van der Westhuizen et al., 2001; Jamrozik et al., 2013; Silva et al., 2018). In the case of FSCR, a relatively consistent average conception rate close to 50% was observed across all ages at exposure contemplated in this study. The trend observed for FSCR provides further evidence that age had no influence in the probability of multiparous cows to conceive in their first AI attempt at consecutive ages. As such, is important to reiterate that at least for the particular data set of this study, a RRM evaluation using age as the longitudinal descriptor does not seem to be appropriate; nonetheless, the evaluation was performed just for exploratory

purposes. Perhaps a better suitability of a RRM genetic prediction for cows' FSCR could be achieved by substituting age for another biologically relevant continuous descriptor associated with AI outcomes, such as breeding weight (Snelling et al., 2019). However, no attempt to do that was done in this study given only 51% of the females had breeding weight records; furthermore, such records belonged mainly to earlier years of data (e.g., prior to 2000).

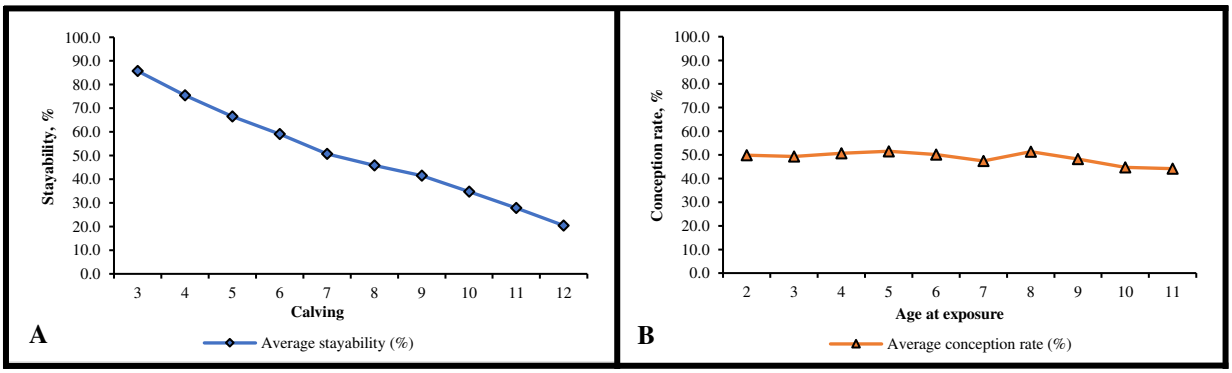


**Figure 4.3. A)** Average stayability (%) to consecutive calvings in the Colorado State University Beef Improvement Center. **B)** Average conception rate (%) to consecutive ages at exposure.

In general, the average FSCR rate of this study agreed with the 52% AI conception rate reported by Lamb et al. (2001) in a study that included multiparous suckled beef cows of different breeds (e.g., Angus, Angus crosses, Simmental and Hereford). Similarly, conception rates ranging from 50.6 to 59.7% in multiparous Angus and crossbred Angus cows (n = 901) were reported by Peel et al. (2012) in a study that tested the effects of intervals of 2, 4, or 6 h, between 2 prostaglandin F2α injections administered in a 5-d CO-Synch + controlled internal drug-release device (CIDR) estrus synchronization protocol. Within such study, 414 multiparous cows from the CSU-BIC were part of the experimental population and their pregnancy percentages in response to the fixed-time AI (FTAI) were 38.1, 34.1 and 46.3% for animals allocated within the 2-h, 4-h and 6-h intervals, respectively. For their part, Whittier et al. (2013) reported AI pregnancy rates of 58.1 and 55.1% for cows synchronized with the 5-day COSynch + CIDR protocol and cows

synchronized with 7-day CO-Synch + CIDR protocol, respectively. Together, these results demonstrated the evident connection between the final outcomes of a FTAI program (e.g., pregnancy rates) and the specific synchronization protocol implemented for its accomplishment.

In relationship to the number of observations available to perform genetic evaluations for STAY06, after the removal of 2 (successful) observations that showed no variation within a specific class level of the fixed factor AFC (all other classes of categorical fixed effects showed variation), a total of 1542 records remained useful for the TM evaluation. Particularly for the STAY06 analysis, it is important to clarify that CG effects were decided to be treated as random instead of fixed, precisely due to the severity of the reduction in data size when attempting to remove subclasses with no variation. Specifically, 32% of reduction in the available data set (490 observations) required to be disregarded if trying to model CG as a categorical fixed effect. A similar decision was made by González-Recio and Alenda (2005) based on the recommendations of Moreno et al. (1997) when performing threshold evaluations for binary fertility traits in Holstein cattle. Such decisions become even more relevant when recalling that the traditional TM evaluation only considered the success/failure of females reaching the age of 6, explicitly ignoring the information from females younger than 6 yr and cows that were still producing beyond that age endpoint. Conversely, the totality of the 8,907 observations spanning all age endpoints included in the study were included for the RRM since there was no need for record removal.

After the data editing process for STAY observations, it is possible that the removal of just a couple of records for the TM evaluation had not represented a significant pre-analytical distortion for STAY06 traditional evaluation. However, even when almost all information available for the STAY06 endpoint was kept, a considerable increase of information (~5.8 times more data) was achieved by the RRM approach when considering STAY endpoints from 3 to 12 yr old. The

previous is a reflection of the significant loss of data related to the restrictive age-specific definition of STAY to the particular endpoint of 6 yr. Problems associated with the age-related trait definition have been recognized for a long time (Hudson and Van Vleck, 1981); especially when considering that waiting 6 yr until a female receives an observation represents a considerable delay to gather information useful to evaluate the sire of such female (Brigham et al., 2007). This delay in collection of phenotypes reduces the accuracy of sire's genetic predictions at early ages and therefore slows genetic progress for STAY. In this context, the inclusion of both earlier and later ages by RRM represents a feasible avenue to improve the accuracy of sire evaluations (Jamrozik et al., 2013; Silva et al., 2018).

Regarding FSCR, a total of 9448 observations were kept for the TM analyzes after removing 136 records (28 successful/108 unsuccessful) coming from specific classes of categorical fixed effects (e.g., AFC, CE, CG and AI technician) with no variation. For RRM analyzes, the 9,584 FSCR observations originally available were included in the evaluation. The small difference in the total number of observations available within each methodology suggested that the possible bias introduced by pre-analytical editing processes within the TM methodology was small for this particular analysis. Nonetheless, González-Recio and Alenda (2005) illustrated the severity of extreme category problems (subclasses of fixed effects without variation) when analyzing the same trait in Holstein cattle. Particularly, the authors explained that they opted to consider their CG definition as random variable within their analyzes in order to minimize the disadvantages of losing a significant proportion of data. Interestingly, even when considering CG as a random effect, authors acknowledged the necessity of removing observations of other categorical fixed effects within their data set, something experienced in the present study as well.

In this regard, the flexibility of RRM to incorporate all available information has been reported as an attractive feature of this statistical methodology (Golden et al., 2018).

*4.3.2 Heritabilities and correlations*

Heritability ($h^2$) estimates for STAY06 obtained with each statistical method are presented in Table 4.4. The $h^2$ estimate obtained on the underlying scale with the TM was 0.10, which is smaller than the $h^2$ estimates previously reported for this trait within the CSU-BIC Angus population. For instance, in 1995 Snelling and coworkers reported an $h^2$ of 0.14 for STAY06; while 5 years later, Doyle reported an $h^2$ estimate of 0.15 for this trait (Snelling et al., 1995; Doyle et al., 2000). Nonetheless, the STAY06 $h^2$ estimate found in this study agrees with an estimate of 0.10 recently reported for Red Angus cattle (Boldt et al., 2018). In such study, authors acknowledged that their estimate was lower than expected, but also emphasized how $h^2$ normally varies between breeds, as well as, with the model used for its estimation and the trait definition. For instance, when defining STAY06 as the ability of a female to produce 5 consecutive calves (equivalent to the trait definition in the present study) and using a marginal maximum likelihood animal model, Snelling et al. (1995) reported heritabilities of 0.11 and 0.14 for a Red Angus and a Black Angus population, respectively. However, using the same trait definition but implementing Method R, $h^2$ estimates reported in the same study were 0.12 and 0.23 for Red Angus and Black Angus, respectively. Martinez et al. (2005) defined the trait equivalently and reported a $h^2$ estimate on the underlying scale of 0.30 in Hereford cattle when estimating variance components using an animal TM; conversely, in the same study, an $h^2$ estimate of 0.19 was reported for STAY06 when using a linear model. For its part, Maiwashe et al. (2009) used a TM to estimate $h^2$ for STAY06 but defined the trait as the probability of a cow to remain in the herd until the age of 6, given she

had an opportunity to reach that age and was a dam producing at least one calf; with such definition, authors reported an $h^2$ estimate of 0.20 for STAY06 in Angus cattle.

**Table 4.4.** Heritability estimates ($h^2 \pm$ SE) for stayability according to the statistical method employed.

| Methodology | | $h^2 \pm$ SE |
|---|---|---|
| TM[1] | | 0.10 ± 0.03 |
| RRM[2] | Intercept | 0.048 ± 0.009 |
| | Linear | 0.013 ± 0.008 |

[1]TM = threshold model, [2]RRM = random regression model.

The $h^2$ estimates obtained for the intercept and linear random regression coefficients included in the analysis (specific for the age endpoint of 6 yr) were 0.048 and 0.013, respectively. These estimates were smaller than the 0.24 and 0.16 reported for the intercept and linear terms of a third order degree Legendre polynomial RRM that estimated genetic parameters for STAY in Simmental cattle (Jamrozik et al., 2013). Discrepancies among reports may be related to the differences of the order of Legendre polynomials implemented and the considerable disparities in data availability among studies. The interconnection among these factors was explained by Meyer et al. (2005), whose work suggested that RRM using high order polynomials (e.g., ≥ cubic orders) can produce erratic estimates of variance components and genetic parameters when data sets contain many more records at earlier than later ages (as typically occurs with STAY). In this sense, the data in our study included 8,907 observations generated within a single herd, while the final STAY data file of Jamrozik et al. (2013) contained 1,164,319 records of cows from different herds. The greater availability of records in Jamrozik's and colleagues research justified their implementation of higher Legendre polynomial orders with a minimum risk of producing implausible $h^2$ estimates; nonetheless, the risk of producing erratic values was higher in our study and therefore a lower order was utilized (e.g., linear order).

Additionally, the distinct CG definitions between studies may also be influencing the way the variance is being partitioned for the trait. In the present study, all females belonged to the CSU-BIC and within the herd there are pre-established breeding and calving seasons. Consequently, all females within a year were born approximately within the same time window (between February and April), as such, not an appropriate adjustment of variability was being made by defining CG as a combination between birth year and birth season as indicated by Jamrozik et al. (2013). Conversely, it was opted to define CG as a combination of breeding year and breeding pasture (an effect that was confounded with natural service sire) in order to better capture the variations in fertility associated with each respective year of data. Another possible reason for the differences in results between studies could be that genetic variation for STAY differs among breeds (e.g., Simmental *vs* Angus), since differences in $h^2$ estimates have been reported in literature when evaluating STAY data coming from different breeds using random regression techniques (Silva et al., 2018). Once the $h^2$ estimates of the intercept and linear term of the RRM were transformed to an observed scale associated to each age endpoint, changes in $h^2$ estimates for STAY across all the age endpoints were plotted and are shown in Figure 4.4.

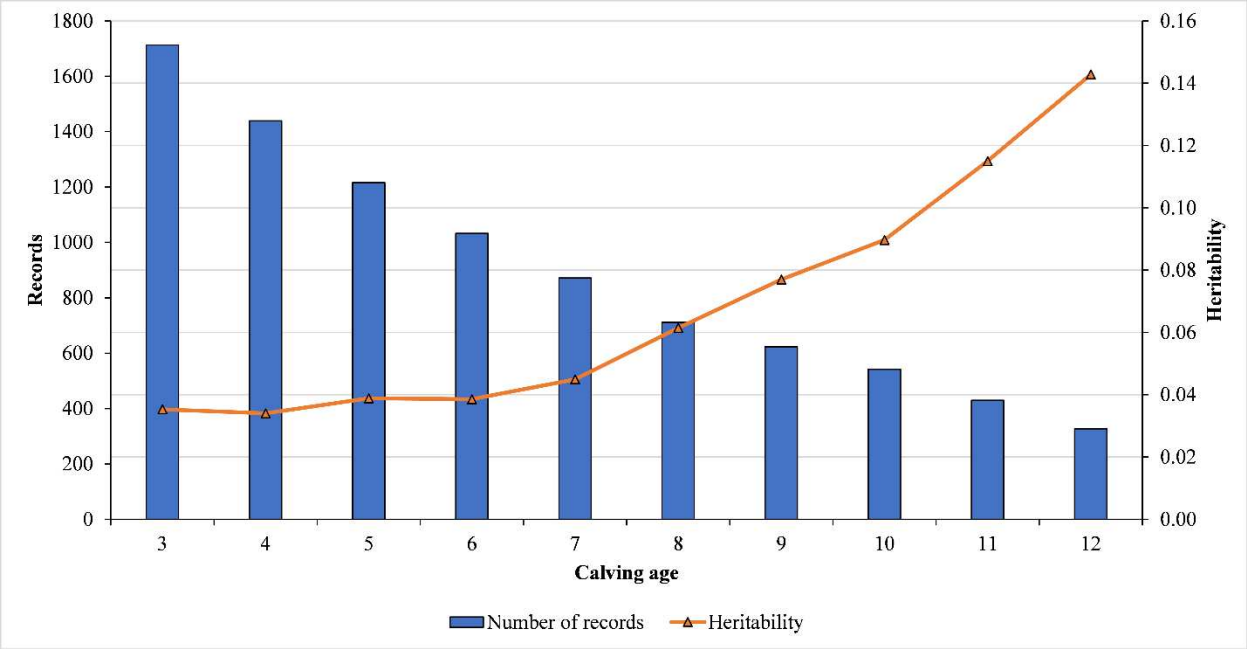


**Figure 4.4.** Changes in heritability estimates for Stayability and their relationship with the number of records at each endpoint in Angus cows.

The observed $h^2$ estimate at the age of 6 yr was 0.04, which was lower than the 0.10 obtained with the TM. In this regard, several studies have noted that heritabilities calculated on the observed scale are typically lower than heritabilities on the underlying scale (Kadarmideen et al., 2000; Johnston et al., 2014; Silvestre et al., 2019). Nonetheless, restricting our comparisons only to $h^2$ calculated in the observed scale and by RRM; our results were smaller in comparison to those reported by Jamrozik et al. (2013) in Simmental cattle. In such study, it was reported that $h^2$ for STAY decreased from the age of 2 (0.36) to the age of 8 (0.12). Discrepancies in results between studies may be related to previously discussed differences among both investigations. Conversely, results obtained in this study were similar to the $h^2$ estimates to consecutive ages (from 2 to 8 yr of age) reported by Silva et al. (2018) for three different *Bos indicus* cattle breeds (Tabapuã, Nellore and Guzerá). In that study, $h^2$ estimates for STAY in Tabapuã and Nellore breeds tended to increase with calving number and ranged from 0.03 to 0.07 and from 0.03 to 0.08, respectively. In the case of the Guzerá breed, $h^2$ estimates ranged from 0.05 to 0.08 showing a

quadratic trend with a peak between the fourth and sixth calving. In the present study, $h^2$ estimates for STAY ranged from 0.03 to 0.14 and also tended to increase with the number of calvings. Focusing on $h^2$ from age of 3 up to the age of 8 (the final endpoint reported by Silva et al., 2018), $h^2$ estimates found in the present study ranged from 0.03 to 0.06, resembling to those reported for Tabapuã and Nellore cows.

An interesting aspect worthy of discussion is that increments in $h^2$ estimates of STAY in the present investigation corresponded with reductions in the availability of records of the later age endpoints considered. This is a commonly reported mathematical artifact of RRM fitting polynomial regressions originated for reductions in the number of observations associated to the covariate (Meyer et al., 2005). Speidel et al. (2010) explained that Legendre polynomials place a large amount of emphasis on observations at the extremes of the covariate; therefore, when severe reductions in data availability occurs at these points, it's common to observe inflations of $h^2$ estimates. Therefore, it is possible that the $h^2$ estimates at later ages in this study were inflated due to the decreased availability of STAY records (especially at 11 and 12 yr of age).

Estimates of phenotypic and genetic correlations of STAY to consecutive calvings are shown in Table 4.5. In general, all correlations were positive with values ranging between 0.16 to 0.68 in the case of phenotypes and between 0.77 to 0.99 for genetic effects. Correlations increased when the calving events were close to each other, whereas the opposite occurred when the calvings were more distant on the longitudinal scale. These results compile more evidence that STAY defined at consecutive ages does not represent phenotypically nor genetically the same trait. However, STAY endpoints are correlated and the magnitude of such correlations depends on the proximity of the calving events within the time scale. The genetic correlation of 0.77 between STAY03 and STAY12; suggested that STAY03 could be a good indicator of STAY to later ages.

**Table 4.5.** Genetic (above diagonal) and phenotypic (below diagonal) correlations for stayabilities to consecutive calvings.

| Calving No | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | | 0.99 | 0.96 | 0.93 | 0.90 | 0.87 | 0.84 | 0.81 | 0.79 | 0.77 |
| 4 | 0.66 | | 0.99 | 0.98 | 0.95 | 0.93 | 0.91 | 0.89 | 0.87 | 0.85 |
| 5 | 0.51 | 0.65 | | 0.99 | 0.98 | 0.97 | 0.95 | 0.94 | 0.93 | 0.91 |
| 6 | 0.42 | 0.50 | 0.65 | | 0.99 | 0.99 | 0.98 | 0.97 | 0.96 | 0.95 |
| 7 | 0.35 | 0.40 | 0.48 | 0.59 | | 0.99 | 0.99 | 0.99 | 0.98 | 0.97 |
| 8 | 0.31 | 0.34 | 0.40 | 0.47 | 0.68 | | 0.99 | 0.99 | 0.99 | 0.98 |
| 9 | 0.28 | 0.31 | 0.35 | 0.40 | 0.55 | 0.65 | | 0.99 | 0.99 | 0.99 |
| 10 | 0.24 | 0.26 | 0.29 | 0.32 | 0.42 | 0.44 | 0.52 | | 0.99 | 0.99 |
| 11 | 0.20 | 0.21 | 0.23 | 0.26 | 0.32 | 0.32 | 0.36 | 0.59 | | 0.99 |
| 12 | 0.16 | 0.17 | 0.19 | 0.20 | 0.25 | 0.24 | 0.25 | 0.40 | 0.57 | |

Similar phenotypic and genetic correlations for STAY at consecutive ages (from 2 to 8 yr of age) have been reported by Jamrozik et al. (2013) in Simmental cattle and by Silva et al. (2018) in three Zebu cattle breeds. For the most part, Martinez et al. (2005) did not report genetic correlations but informed correlations among sire predictions for STAY at different endpoints (from the age of 3 to the age of 8) in Hereford cattle. In such study, correlations among EPD also declined as the calving events were more distant apart. In the case of the permanent environmental correlations obtained in this study, they followed the same pattern than genetic and phenotypic correlations, decreasing as calvings became more distant (Table 4.6).

**Table 4.6.** Permanent environmental correlations for stayabilities to consecutive calvings.

| Calving No | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 0.99 | | | | | | | | |
| 5 | 0.98 | 0.99 | | | | | | | |
| 6 | 0.93 | 0.96 | 0.99 | | | | | | |
| 7 | 0.83 | 0.88 | 0.93 | 0.98 | | | | | |
| 8 | 0.65 | 0.72 | 0.80 | 0.89 | 0.97 | | | | |
| 9 | 0.41 | 0.49 | 0.59 | 0.72 | 0.85 | 0.96 | | | |
| 10 | 0.15 | 0.24 | 0.36 | 0.51 | 0.68 | 0.85 | 0.96 | | |
| 11 | -0.07 | 0.02 | 0.15 | 0.31 | 0.50 | 0.71 | 0.88 | 0.98 | |
| 12 | -0.23 | -0.14 | -0.02 | 0.14 | 0.35 | 0.59 | 0.79 | 0.93 | 0.99 |

Perhaps the most intriguing result related to the correlations of permanent environmental effects are the negative correlations detected between the most distant calving events (e.g., STAY03 and STAY11 or STAY03 and STAY12; among others). Interpretation of these results is challenging since a negative permanent environmental correlation implies that a better than average environmental effect at earlier ages would be associated with a lower than average environmental effect at later ages. From a practical point of view, the previous means that animals receiving a "preferential treatment" at younger ages, will likely experience adverse environments at later ages. This does not seem to be realistic so the explanation of such results should probably be approached from a statistical point of view. It is highly probable that the results of these negative correlations are the product of the considerable reduction of observations available at later ages and the mathematical issues of Legendre Polynomials at the extreme of the data range. In order to test this explanation, an alternative RRM model was executed lumping observations of 11 and 12-yr-old cows into one single age category named "≥11". Results of such study are shown in appendix A in Figures A-1 and Tables A-1 and A-2. Briefly, we were able to confirm that increasing the number of observations at later ages, the apparent unrealistic permanent environmental correlation estimates between the more distant calving effects were mitigated. Also, it was possible to notice that a less inflated $h^2$ estimate (0.12) was obtained for the ≥11 age category.

In the case of FSCR, $h^2$ and repeatability (r) estimates obtained with each statistical methodology are shown in Table 4.7. The $h^2$ estimate obtained on the underlying scale using the TM (0.03) agrees with the report of Bormann et al. (2006) whose informed a $h^2$ of 0.03 in Angus heifers. Our result is also in line with the $h^2$ estimate of 0.029 reported by Ghiasi et al. (2011) when analyzing FSCR of multiparous Holstein cows. A slightly smaller $h^2$ estimate for this trait (0.015)

was reported by Rahbar et al. (2016) in Holstein dairy cows exposed to heat stress; however, the influence of warm environmental conditions in such study may explain the smaller $h^2$ estimate. Averill et al. (2004) reported a $h^2$ estimate of 0.028 for a trait defined as "the outcome of an artificial insemination"; however, within such study, the trait definition was not analogous to FSCR since a maximum of 3 AI events were allowed per cow if she failed to conceive within the first two services. With respect to the repeatability estimate obtained with the TM, the same estimate obtained for the $h^2$ of FSCR was obtained for this parameter (r = 0.03). This result agrees with a previous report suggesting that the permanent environmental effects that a female has experienced seem to not have an effect on her ability to conceive on the first insemination during the consecutive breeding seasons (Foxworthy et al., 2019a).

**Table 4.7.** Heritability ($h^2$ ± SE) and repeatability (r) estimates for cow first-service conception rate according to the statistical method employed.

| Methodology | Age | $h^2$ ± SE | r ± SE |
|---|---|---|---|
| TM[1] | All ages | 0.03 ± 0.01 | 0.03 ± 0.01 |
| RRM[2] | 2 | 0.037 ± 0.017 | - |
| | 3 | 0.031 ± 0.013 | - |
| | 4 | 0.025 ± 0.009 | - |
| | 5 | 0.022 ± 0.008 | - |
| | 6 | 0.020 ± 0.008 | - |
| | 7 | 0.024 ± 0.010 | - |
| | 8 | 0.026 ± 0.013 | - |
| | 9 | 0.031 ± 0.017 | - |
| | 10 | 0.038 ± 0.023 | - |
| | 11 | 0.049 ± 0.032 | - |

[1]TM = threshold model, [2]RRM = random regression model.

In the case of the $h^2$ estimates obtained for the random regression coefficients for the additive genetic effects, results were 0.0202 ± 0.0081 and 0.0213 ± 0.0151 for the intercept and

the linear terms, respectively. Random regression coefficients for permanent environmental effects were 0.0213 ± 0.0072 and 0.0260 ± 0.0164 for the intercept and the linear terms, respectively. Given the absence of a similar study where random regression techniques have been applied to evaluate FSCR to consecutive ages in beef cattle, direct comparisons of these random regression coefficients to other results were not possible. Perhaps the lack of reports of RRM applied to consecutive age FSCR observations in beef cattle derives from the combination of a limited use of AI in the industry and the apparent null influence of age in insemination outcomes. The fact that a RRM genetic evaluation for FSCR utilizing age as the longitudinal descriptor resulted in an inappropriate modeling of this trait, was verified by testing the amount of variation accounted by the linear order of the RRM of FSCR in age by using a LRT as reported by Speidel et al. (2016). Briefly, the REML log-likelihood estimates for a zero order (just the intercept) and a linear RRM were 1844.94 and 1847.52, respectively, resulting in a test statistic of 5.16 ($P > 0.05$); suggesting that the linear term of the RRM did not account for any additional variation in FSCR.

Table 4.7 shows the heritabilities and repeatabilities of FSCR obtained with each statistical methodology contemplated in this study. In general, $h^2$ estimates for FSCR obtained for all ages were low (< 0.05), which is consistent with the reports made for this trait in beef and dairy cattle populations (Azzam et al., 1989; Cammack et al., 2009; Hossein-Zadeh and Ardalan, 2011). However, is important to acknowledge that the overparameterization that occurred in the RRM evaluation of FSCR apparently artificially added noise to the estimates of $h^2$ in the observed scale for this trait. Although the $h^2$ estimates obtained at the extremes of the data range (ages at exposure) do not seem to be unrealistic for a trait like FSCR (e.g., 0.037 for FSCR02 and 0.049 for FSCR11), the small fluctuations of the estimates across the age range are likely the result of forcing a regression with a slope not statistically different from zero (Figure 4.3 – B).

A graphical representation of the changes in $h^2$ estimates of FSCR and their associations with the number of records per age at exposure is shown in Figure 4.5. In general, $h^2$ estimates for FSCR obtained with the RRM decreased from 0.037 at the age of 2 to 0.020 at the age of 6 and then increased gradually with the age at exposure reaching a maximum value of 0.049 at the age of 11. Using a similar time descriptor (e.g., parity number) but a different trait (services per conception, SPC), Nishida et al. (2005) reported a similar trend in changes of $h^2$ estimates when applying random regression techniques. Authors of that study reported that $h^2$ estimates of SPC declined from 0.15 in parity 1 to 0.04 in parity 6 and then the estimates increased continuously with the parity number until ending at 0.22 in parity 10. Even when such reported trend had resemblance to the one obtained in the present study, is important to clarify that within Nishida's investigation, parity number was an important model component that significantly accounted for variation of SPC.

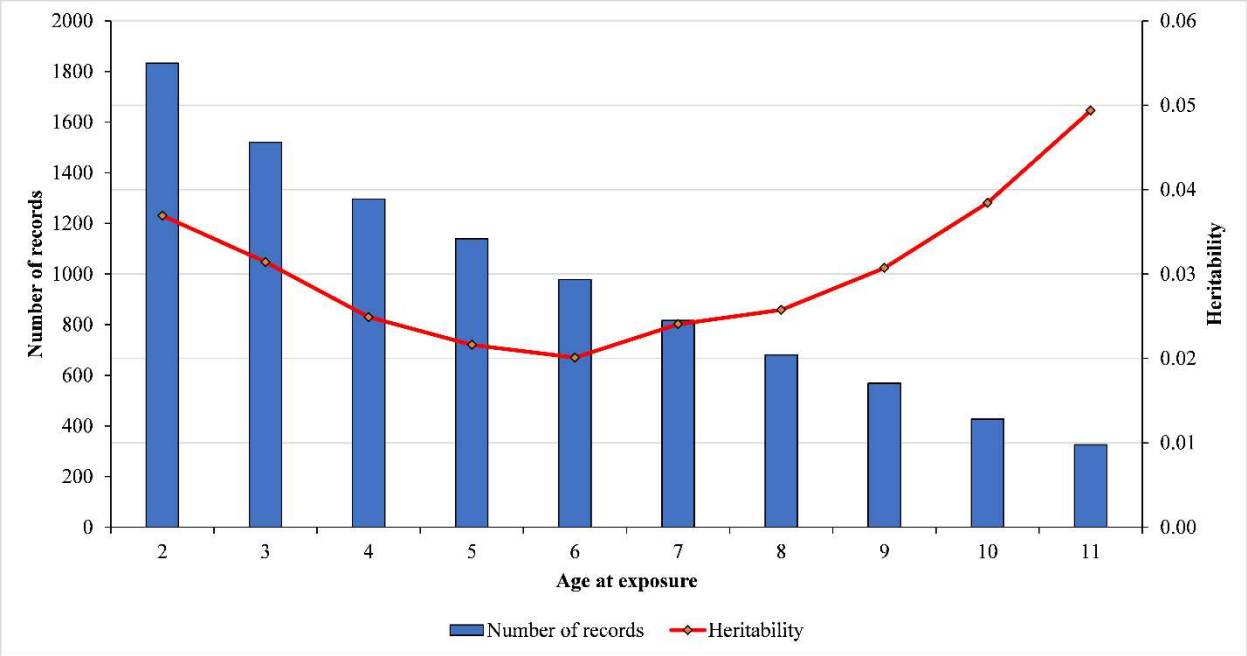


**Figure 4.5.** Changes in heritability estimates for first-service conception rate and their relationship with the number of records at each age of exposure in Angus cows.

Although RRM have been applied to analyses of insemination outcomes and conception rate in dairy cattle (Averill et al., 2006; Tsuruta et al., 2009; Buaban et al., 2016), direct comparisons of those reports with our results are difficult because generally the time descriptor used for dairy cows is days in milk (DIM), which are normally limited to only one specific age endpoint (e.g., 2-yr-old primiparous cows). Furthermore, differences in the reproductive management between beef and dairy cattle creates a challenge to have the same trait definition in both cattle types, since within dairy cattle it is normal that a cow be allowed ≥2 AI services. Within beef cattle, RRM have been successfully applied to examine genetic relationships among cow weight and productivity (Snelling et al., 2019). Therefore, depending upon on data availability, it is possible that a RRM using cow weight at each breeding event as a longitudinal descriptor may work better for a genetic evaluation for FSCR than the current model that uses age.

Given the trait definition, it is biologically impossible that a female could generate repeated records for the outcome of an insemination event within the same age; therefore, no repeatabilities were estimated with the random regression methodology for FSCR. Furthermore, previous research efforts involving the usage of a threshold repeatability model to analyze FSCR in mature cows without regressing such observations on an age covariate, have suggested that temporary environmental effects associated to each specific AI event have a greater degree of influence on FSCR than variations attributable to genetics or permanent environmental effects (Foxworthy et al., 2019a). In dairy cattle-based studies, the estimation of repeatabilities for the conception rate in response to insemination events using RRM is viable mainly due to two fundamental differences with beef cattle: dairy cows are typically allowed to have more than 1 AI and, usually, the age covariate implemented in the analysis is the number of DIM (normally restricted to one parity). For instance, Buaban et al. (2016) reported repeatabilities for conception rate in first-lactation

crossbred dairy cows when regressing it on DIM using RRM; specifically, repeatabilities ranged from 0.060 to 0.259, from 0.073 to 0.407 and from 0.078 to 0.579 when using RRM with 2-, 3- and 4-order Legendre polynomials as the base function. Perhaps further investigations in beef cattle could determine if a RRM can be applied using beef cattle data based on postpartum intervals or some similar measure.

Phenotypic and genetic correlations for all the age-specific FSCR observations considered in the RRM are shown in Table 4.8. At the phenotypic level, correlations were close to zero with values ranging between -0.04 to 0.08. In the case of the genetic correlations, they decreased considerably as the AI events were more distant within the range of ages at exposure. Specifically, genetic correlations were as high as 0.99 for immediate consecutive ages and as low as -0.03 for the more distant ages (e.g., 2 and 11 yr). The lack of concordance between phenotypic and genetic correlations is perhaps a product of the overfitting of the RRM evaluating FSCR. Nonetheless, at least at the genetic level, it has been reported that fertility traits are not necessarily the same traits within younger and older cows, since energy requirements are considerably higher in older females due to their higher milking ability (Roxström et al., 2001; Jamrozik et al., 2005).

**Table 4.8.** Genetic (above diagonal) and phenotypic (below diagonal) correlations for first service conception at consecutive ages at exposure.

| Calving No | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | | 0.99 | 0.95 | 0.87 | 0.73 | 0.56 | 0.38 | 0.22 | 0.08 | -0.03 |
| 3 | 0.04 | | 0.99 | 0.93 | 0.82 | 0.67 | 0.51 | 0.34 | 0.22 | 0.12 |
| 4 | 0.04 | 0.01 | | 0.98 | 0.91 | 0.79 | 0.65 | 0.51 | 0.39 | 0.29 |
| 5 | 0.01 | 0.07 | 0.03 | | 0.97 | 0.90 | 0.79 | 0.68 | 0.57 | 0.48 |
| 6 | -0.04 | -0.03 | -0.03 | 0.04 | | 0.98 | 0.91 | 0.83 | 0.74 | 0.66 |
| 7 | 0.00 | -0.03 | -0.03 | 0.02 | 0.00 | | 0.98 | 0.93 | 0.87 | 0.81 |
| 8 | 0.02 | 0.03 | 0.04 | -0.02 | 0.04 | 0.06 | | 0.99 | 0.95 | 0.91 |
| 9 | 0.02 | 0.02 | -0.03 | 0.00 | 0.01 | 0.01 | 0.06 | | 0.99 | 0.97 |
| 10 | 0.04 | -0.01 | 0.02 | 0.04 | 0.02 | -0.02 | 0.08 | -0.04 | | 0.99 |
| 11 | -0.03 | -0.02 | 0.00 | -0.02 | 0.02 | 0.02 | 0.04 | 0.06 | 0.07 | |

In the case of the permanent environmental correlations obtained for FSCR in this study, results are shown in Table 4.9. Generally, they changed from positive to negative mainly when comparing immature cows (≤5-yr-old) to mature cows (>5-yr-old); however, they remained positive when comparing only mature cows. Results of these correlations could be also biased due to the overfitting of the RRM as well as by fact that only females with an acceptable sustained fertility are maintained within the herd. For instance, if an environmental event that negatively impacts the fertility of a female (e.g., dystocia) occurs early in her life, it is highly probable that such female will fail to conceive in her subsequent exposure and therefore be culled from the herd. Furthermore, variations attributed to major deleterious fertility events, like dystocia were already accounted by including CE in the RRM (e.g., more evidence of overfitting).

**Table 4.9.** Permanent environmental correlations for first-service conception to consecutive ages at exposure.

| Calving No | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 0.98 | | | | | | | | |
| 4 | 0.84 | 0.93 | | | | | | | |
| 5 | 0.31 | 0.49 | 0.78 | | | | | | |
| 6 | -0.26 | -0.07 | 0.32 | 0.84 | | | | | |
| 7 | -0.53 | -0.36 | 0.02 | 0.64 | 0.96 | | | | |
| 8 | -0.65 | -0.50 | -0.13 | 0.52 | 0.90 | 0.99 | | | |
| 9 | -0.72 | -0.57 | -0.22 | 0.44 | 0.86 | 0.98 | 0.99 | | |
| 10 | -0.76 | -0.62 | -0.28 | 0.39 | 0.83 | 0.96 | 0.99 | 0.99 | |
| 11 | -0.78 | -0.65 | -0.31 | 0.35 | 0.80 | 0.94 | 0.98 | 0.99 | 0.99 |

*4.3.3 Comparison of genetic predictions*

Genetic predictions for STAY at the age of 6 obtained with each methodology are summarized in Table 4.10. The average EPD was close to zero and similar between methodologies; however, a wider range in prediction values was clearly observed with the TM. The lower range of EPD observed within the RRM prediction could be related to the smaller $h^2$ estimate obtained with this methodology for the age endpoint of 6 yr of age. Speidel et al. (2018b) also reported a

smaller range in random regression genetic predictions for HPG in comparison to those obtained with a TM in Red Angus cattle, discussing that the $h^2$ estimate of the RRM was about 3 times smaller than the one in the TM.

**Table 4.10.** Stayability at the age of 6 expected progeny differences (EPD) summary statistics according to the statistical method implemented in Angus cows.

| Methodology | N | Average | SD | Min | Max |
|---|---|---|---|---|---|
| TM[1] | 14,140 | -0.865 | 6.44 | -27.40 | 34.07 |
| RRM[2] | 14,140 | -0.266 | 5.03 | -17.88 | 13.86 |

[1]TM = threshold model; [2]RRM = random regression model.

Summary statistics for FSCR genetic predictions resulting from each statistical approach are shown in Table 4.11. For this trait in particular, since there is no specific age of interest within the beef industry for a multiparous cow to conceive in response to her first AI, it was opted to show all the predictions obtained using the RRM (i.e., predictions for each age at exposure) in order to compare them with the prediction obtained with the TM. In general, EPD averages were close between methodologies; but again, the range of prediction values was greater for the TM methodology. Explanation for these results had resemblance to the already discussed results for STAY genetic predictions (e.g., differences in $h^2$ estimates among methodologies).

**Table 4.11.** First-service conception rate expected progeny differences (EPD) summary statistics according to the statistical method implemented in Angus cows.

| Methodology | N | Age | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| TM[1] | 14,140 | All ages | 2.24 | 2.53 | -8.04 | 15.84 |
| RRM[2] | 14,140 | 2 | 1.29 | 1.28 | -4.06 | 6.91 |
| | | 3 | 1.20 | 1.18 | -3.66 | 6.40 |
| | | 4 | 1.12 | 1.10 | -3.29 | 6.35 |
| | | 5 | 1.03 | 1.03 | -3.02 | 6.29 |
| | | 6 | 0.95 | 0.99 | -3.09 | 6.23 |
| | | 7 | 0.86 | 0.96 | -3.44 | 6.17 |
| | | 8 | 0.78 | 0.96 | -3.79 | 6.11 |
| | | 9 | 0.69 | 0.99 | -4.14 | 6.20 |
| | | 10 | 0.61 | 1.04 | -4.50 | 6.42 |
| | | 11 | 0.52 | 1.11 | -4.85 | 6.64 |

[1]TM = threshold model; [2]RRM = random regression model.

Although no EPD summary statistics for FSCR in beef cattle were found in literature in order to perform direct comparisons to those obtained in the present study, a similar average and range in estimated breeding values (EBV) were presented for a trait defined as the percent of calves born to AI (CAI) in a New Zealand national dairy cattle evaluation (Harris et al., 2005). Within such report, authors informed that the average CAI EBV across different cattle breeds (e.g., Holstein, Jersey, Ayrshire, Shorthorn, Guernsey and Brown Swiss) was -1.04, with a $\sigma = 3.61$ and minimum and maximum values of -24.9 and 14.6, respectively. Considering that such genetic predictions were presented as EBV, if those values are halved, they become closer to the predictions obtained in the present study for FSCR. Evidently, it must be acknowledged that predictions are not performed exactly for the same trait and that important differences in fertility may exist between beef and dairy cattle breeds. Additionally, Olori et al. (2002) performed genetic predictions for calving interval using a linear model within a Holstein population, reporting an

average predicted transmitting ability of 1.93, with a $\sigma = 1.61$ and minimum and maximum values of -3.36 and 8.10, respectively. Such predictions are close to those obtained in the present study for FSCR and, even when they belong to a different trait, it is interesting that Olori and colleagues reported a $h^2$ for calving interval of 0.04, which is very similar estimate to those obtained for FSCR in this investigation.

Similarities among predictions (EPD) and rankings of animals obtained with the TM and RRM for STAY06 were high. Specifically, the Pearson correlation ($r_p$) among predictions was 0.84, suggesting that both statistical methodologies predicted similar genetic merits for STAY06. The Spearman's rank correlation ($r_s$) was 0.86, suggesting a high degree of concordance between the ranking of animals for this trait. Similar results (e.g., $r_p = 0.77$ and $r_s = 0.79$) were reported by Sánchez-Castro et al. (2017) when comparing TM and RRM genetic predictions for STAY06 using a reduced data set from the CSU-BIC that contained observations collected between 1993 to 2012. Heringstad et al. (2003) reported a rank correlation larger than 0.99 for genetic predictions of clinical mastitis (binary trait) in Holstein dairy cattle obtained using sire TM and sire linear models that ignored the binary nature of the trait. Within such study, authors explained that the great similitude in the ranking of animals between methodologies could be attributed to the large progeny groups in their data set (664 daughters per sire), since larger progeny groups made averages of binary records more normally distributed. Essentially, the previous results imply that differences in inferences between TM and linear models can be more marked for estimating genetic parameters (e.g., $h^2$) than for ranking sires, at least for those with large progeny groups (Heringstad et al., 2003). This is relevant to the findings of the present study because even when the correlations ($r_p$ and $r_s$) were not as high as for dairy cattle-based studies, they were strong and were produced from a data set representative of the beef industry, where at least within a single herd, typically

few observations per sire exist. Nonetheless, perhaps with larger data sets as those used in national cattle evaluations where prominent AI sires with daughters in different herds may be included, larger progeny groups may exist and correlations among predictions could improve.

Regarding the regression of predictions obtained with the RRM on those obtained with TM for STAY06, an underestimation of the genetic merit occurred in the RRM in comparison to TM ($\beta_1 = 0.65$). An opposite result was reported by Sánchez-Castro et al. (2019) when comparing seven different linear RRM to a TM genetic prediction for STAY06 for the CSU-BIC Angus population. In that study, authors mentioned that an underestimation of the genetic merit for STAY06 occurred with the traditional TM when compared to the RRM evaluations. The key difference between such report and the results presented in this study is that no homogenization of prediction scales was performed in Sánchez-Castro et al. (2019) as was done in the current study. Recalling that both types of predictions (TM and RRM) were converted to a pseudo-probability scale as deviations from 50% (Speidel et al., 2018b); the underestimation of EPD that occurred with the RRM was not the result of different prediction scales, but more likely was caused by the fact that TM captures more genetic variance than linear models (Heringstad et al., 2003). However, it has been noted that although the $h^2$ is greater when defined on the underlying scale, selection based on predictions of TM may not yield higher genetic progress on the observed scale than selection on predictions derived from linear models (Boettcher et al., 1999).

Although predictions from the TM showed a high degree of concordance with RRM predictions specific to the age of 6 (STAY06), there is an alternative comparison between predictions that is also worth of discussing. Specifically, it has been suggested that predictions obtained specifically for the intercept ($\beta_0$) of a RRM could serve as the sole criteria of selection capable of modifying the entire STAY curve (Jamrozik et al., 2013). The mathematical rationale

behind such argument is that the intercept of a RRM has the same weight for STAY in all calvings (regardless of how many endpoints are included in the model) and normally exhibit the highest $h^2$ among all the regression coefficients (Silva et al., 2018). Interestingly, to our knowledge, comparisons of predictions specific to the intercept of a RRM for STAY at consecutive ages has not been directly compared to the predictions obtained with a traditional TM. Therefore, a complementary analysis was performed in an attempt to explore the possibility of using predictions for the intercept as the single selection criteria for STAY. Specifically, predictions obtained on the linear scale were transformed to a pseudo-probability scale as outlined by Speidel et al. (2018b) and then, comparisons were made among predicted progeny differences for $\beta_0$ and EPD obtained with the TM. Additionally, comparisons between predictions for $\beta_0$ and EPD for all the age-specific STAY endpoints evaluated in the RRM were also performed (Table 4.12).

**Table 4.12.** Pearson ($r_p$) and rank ($r_s$) correlations of predictions for the intercept of the random regression model with all the age specific predictions for STAY produced by both methodologies.

| | | | Age specific stayability endpoint within the RRM[2] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TM[1] | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $\beta_0$[†] | $r_p$ | 0.83 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 |
| | $r_s$ | 0.86 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 |

[1]TM = threshold model; [2]RRM = random regression model; [†]$\beta_0$ = intercept of the RRM.

Correlations among predictions for the intercept of the RRM and predictions from the TM (e.g., $r_p = 0.83$ and $r_s = 0.86$) were almost identical to those obtained when comparing the TM and the RRM prediction specific to the age of 6 (e.g., $r_p = 0.84$ and $r_s = 0.86$). In the case of the comparisons made for the predictions for $\beta_0$ and the rest of the age-specific predictions obtained with the RRM (e.g., from STAY03 to STAY12), correlation coefficients indicated that predictions and ranking of animals were essentially the same (≥0.98). In this regard, Jamrozik et al. (2013) reported that intercept predictions correlated quite well with all the age specific-predictions of STAY at consecutive ages in Simmental cattle (correlations ranged between 0.86 to 0.99). Silva

et al. (2018) also reported high correlations (ranging from 0.75 to 0.95) for intercept predictions and age-specific STAY predictions produced using random regression techniques. These results support the idea of considering the intercept of a RRM evaluating STAY at consecutive ages as the single most important selection criteria capable of improving all STAY endpoints. Furthermore, the correlations obtained between the intercept predictions and the EPD obtained with the TM, provide more evidence that the RRM methodology produces similar predictions and rankings of animals to those of the traditional method.

Pearson and Spearman's correlations, as well as the regression coefficient of EPD obtained with the TM on those yielded by the RRM for FSCR are shown in Table 4.13. Again, since there is no specific age of interest for a multiparous cow to conceive in response to her first AI, it was opted to compare all the predictions obtained with the RRM (i.e., predictions for each age at exposure) with the prediction obtained with the TM. Results suggested that predictions were highly correlated and almost the same animal rankings were obtained with both methodologies. Nonetheless, even when such results were obtained, they should not be considered totally reliable due to the noise introduced by using age as the longitudinal descriptor of FSCR.

**Table 4.13.** Pearson correlation ($r_p$), rank correlation ($r_s$) and regression coefficient ($\beta_1$) of predictions for first-service conception rate obtained with each statistical method.

| | Ages at exposure | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| $r_p$ | 0.89 | 0.93 | 0.97 | 0.99 | 0.99 | 0.98 | 0.93 | 0.87 | 0.79 | 0.70 |
| $r_s$ | 0.89 | 0.93 | 0.96 | 0.99 | 0.99 | 0.97 | 0.93 | 0.86 | 0.78 | 0.69 |
| $\beta_1$ | 0.45 | 0.44 | 0.42 | 0.40 | 0.39 | 0.37 | 0.36 | 0.34 | 0.32 | 0.31 |

The utility of using exclusively the predictions for the intercept term of the RRM as the unique selection criterion was also explored for this trait (Table 4.14). Correlations among predictions for the intercept of the RRM and predictions from the TM (e.g., $r_p = 0.99$ and $r_s = 0.99$) indicated a high degree of concordance between the predictions from both methods and that

animals were ranked in the same way; consequently, the result for the regression coefficient seems irrelevant. Although predictions for $\beta_0$ also correlated quite well (values ranging from 0.78 to 0.99) with the rest of the age-specific predictions obtained with the RRM, the null influence that age had on FSCR at consecutive ages negates the utility of applying a RRM using age as longitudinal descriptor on the first place. Consequently, results from this analysis should be considered just as a documentation of an unsuccessful attempt to model FSCR using age as a covariate.

**Table 4.14.** Pearson ($r_p$) and rank ($r_s$) correlations of predictions for the intercept of the random regression model with all the age specific predictions for first-service conception rate produced by both methodologies.

| | | | Range of ages at exposure considered in the RRM[2] prediction | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TM[1] | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| $\beta_0$[†] | $r_p$ | 0.99 | 0.84 | 0.89 | 0.94 | 0.97 | 0.99 | 0.99 | 0.97 | 0.92 | 0.85 | 0.78 |
| | $r_s$ | 0.99 | 0.84 | 0.89 | 0.93 | 0.97 | 0.99 | 0.99 | 0.97 | 0.92 | 0.85 | 0.77 |

[1]TM = threshold model; [2]RRM = random regression model; [†]$\beta_0$ = intercept of the RRM.

*4.3.4 Comparison of accuracies of prediction*

The last set of results to discuss are those related to the comparisons between mean accuracies of prediction between TM and RRM for both traits. In the case of the accuracy of predictions for STAY06 obtained with each method, results are shown in Figure 4.6 (A-D). Considering all animals in the pedigree (Figure 4.6 A), the mean accuracy for STAY06 predictions obtained with the TM was 0.045 with a minimum of 0.002 and a maximum of 0.340. Alternatively, the mean accuracy for the same trait when analyzed using the RRM was 0.108 with values that ranged between 0.000 and 0.520. Similar increments were evident for accuracies of predictions for all the sires in pedigree (Figure 4.6 B), where the mean, minimum and maximum accuracy values were 0.030, 0.002 and 0.340 for the TM, and 0.073, 0.020 and 0.520 for the RRM.

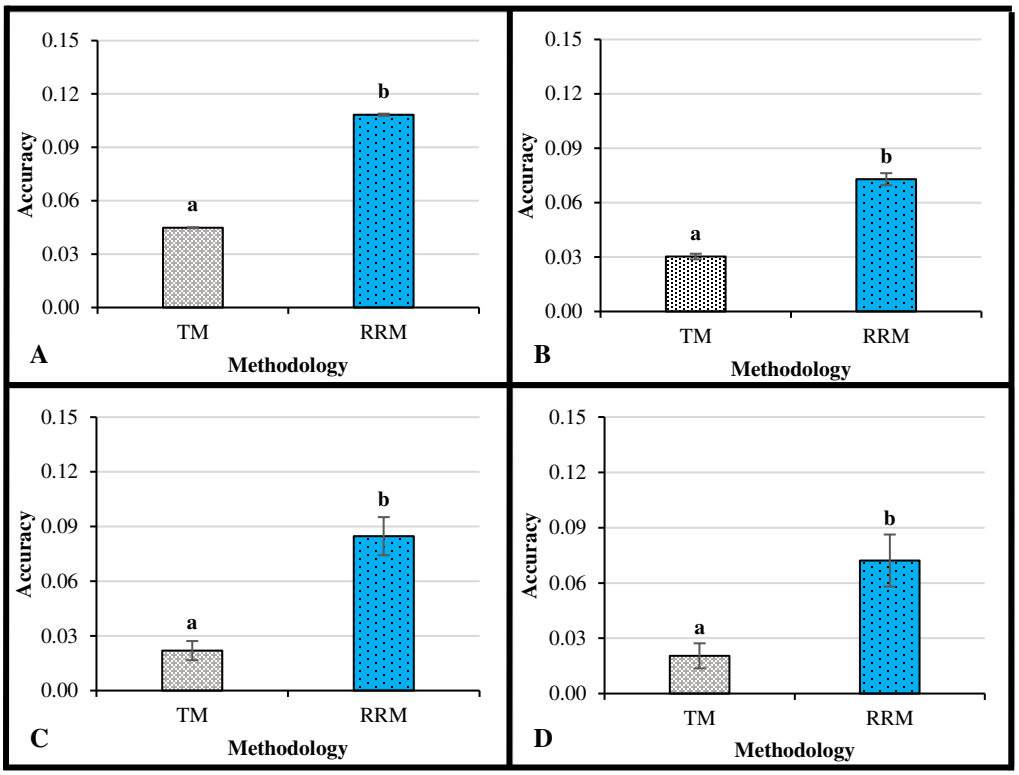


**Figure 4.6.** Mean accuracies for stayability predictions at the age of 6 yr obtained with each statistical methodology. **A)** Mean accuracy for all animals in the pedigree (n = 14,140), **B)** Mean accuracies for all sires in pedigree (n = 971), **C)** Mean accuracies for sires that have produced progeny in the last five yr (n = 85), **D)** Mean accuracies for sires that have produced progeny in the last three yr (n = 51). Different letters indicate a statistical difference at the P <0.05 level among methodologies according to the least significant difference test.

In the case of sires that have produced progeny in the last 5 yr within the CSU-BIC (Figure 4.6 C), the average accuracy was 0.022 for the TM and 0.085 for the RRM, with values that ranged between 0.002 to 0.340 and 0.001 to 0.520 for the TM and RRM, respectively. The last group animals whose mean accuracy values obtained by each method were compared was the sires that have produced progeny within the last 3 yr within the CSU-BIC (Figure 4.6 D). For this group of animals, the mean, minimum and maximum accuracy values were 0.020, 0.002 and 0.340 for the TM, and 0.072, 0.001 and 0.520 for the RRM. Such differences among average accuracies of prediction could be explained by two potentially important factors: the difference in the amount of

information included with each methodology and the better capability of RRM to model the time-dependent variations in fertility of cows (Meyer, 2004; Schaeffer, 2004).

The considerable increase in the amount of information incorporated to the evaluation by the RRM (~ 5.8 times more records in comparison to the TM) was one of the main motivations to explore this statistical methodology. The gains in prediction accuracy were originally glimpsed by Jamrozik et al. (2013), since authors mentioned that analyzing STAY through RRM had the advantage that a larger number of phenotypic records per cow could be included, which could lead to increments in the accuracy of predicted breeding values. Furthermore, the improved ability of RRM to more appropriately model the covariance structure of longitudinal traits, allowed this methodology to better account for the time-dependent fertility variations of cows, and therefore, yielded more accurate predictions. In order to prove that gains in accuracy were not only originated by the inclusion of more data, a complementary comparison between models was performed executing a repeatability TM for STAY (REP; Appendix A, equation A.1). Briefly, such model considered each STAY endpoint as a repeated measure of fertility over time and included a total of 8,890 STAY observations (~5.77 times more records than the traditional TM for STAY06). Comparisons between mean accuracies of prediction between all models (TM, REP and RRM) revealed that the RRM yielded a mean accuracy of prediction 2.4 times higher than the TM and 1.9 times higher than the REP when considering all animals (Appendix A, Figure A-2).

Comparable increments in accuracy of predictions for STAY06 were reported by Sánchez-Castro et al. (2019) when comparing TM and various RRM genetic predictions using a reduced data set from the CSU-BIC that contained observations collected between 1993 to 2016. Within that study, authors reported mean accuracies of prediction higher than those obtained in the present study (e.g., TM mean accuracy of 0.088 and RRM mean accuracies ≥0.386). Nonetheless, in that

particular study, the population of females with STAY phenotypes plus a 3-generations pedigree for those animals was used; whereas in the current study the entire pedigree of the CSU-BIC was utilized. This is relevant because a larger number of animals with more distant genetic relationships with the animals producing phenotypes received a genetic prediction in the present study and consequently, the overall mean accuracy for both methods was found to be smaller. Nonetheless, accuracy gains achieved by the RRM in the genetic evaluation for STAY06 undoubtedly remain and totally support our hypothesis that RRM could yield more accurate predictions than TM.

In the case of the accuracy values associated to the genetic predictions of FSCR, results of the comparisons between statistical methods are shown in Figure 4.7 (A-D). For this trait, it is important to clarify that the comparison of the mean accuracy of the TM was performed relative to the accuracy obtained for predictions at average age at exposure of the cows, which was 5 yr of age. This decision was made with the objective of avoiding the need to performing 10 different accuracy comparisons (e.g., one per each age at exposure). Nonetheless, theoretically, accuracy of prediction could be obtained for any particular age of interest within the data range. When taking into account the entire pedigree (Figure 4.7 A), the mean value obtained with the TM was 0.056 with a minimum of 0.002 and a maximum of 0.392. For the same trait but within the RRM methodology, the mean accuracy value was 0.079 with a range between 0.033 and 0.359. In the case of accuracy comparisons for all the sires in the pedigree (Figure 4.7 B), the mean, minimum and maximum accuracy values were 0.035, 0.0002 and 0.392 for the TM, and 0.062, 0.033 and 0.359 for the RRM. For sires that produced progeny in the last 5 yr within the CSU-BIC (Figure 4.7 C), a mean accuracy of 0.024 was achieved by the TM (with values ranging between 0.0002 and 0.241), whereas the RRM yielded a mean accuracy of 0.054 (with a range between 0.033 and 0.213). Lastly, for the group of sires that have produced progeny within the last 3 yr at the herd

(Figure 4.7 D), the TM had an average accuracy of 0.025 (range between 0.002 and 0.241) and the RRM had a mean accuracy of 0.052 (range between 0.033 and 0.213).

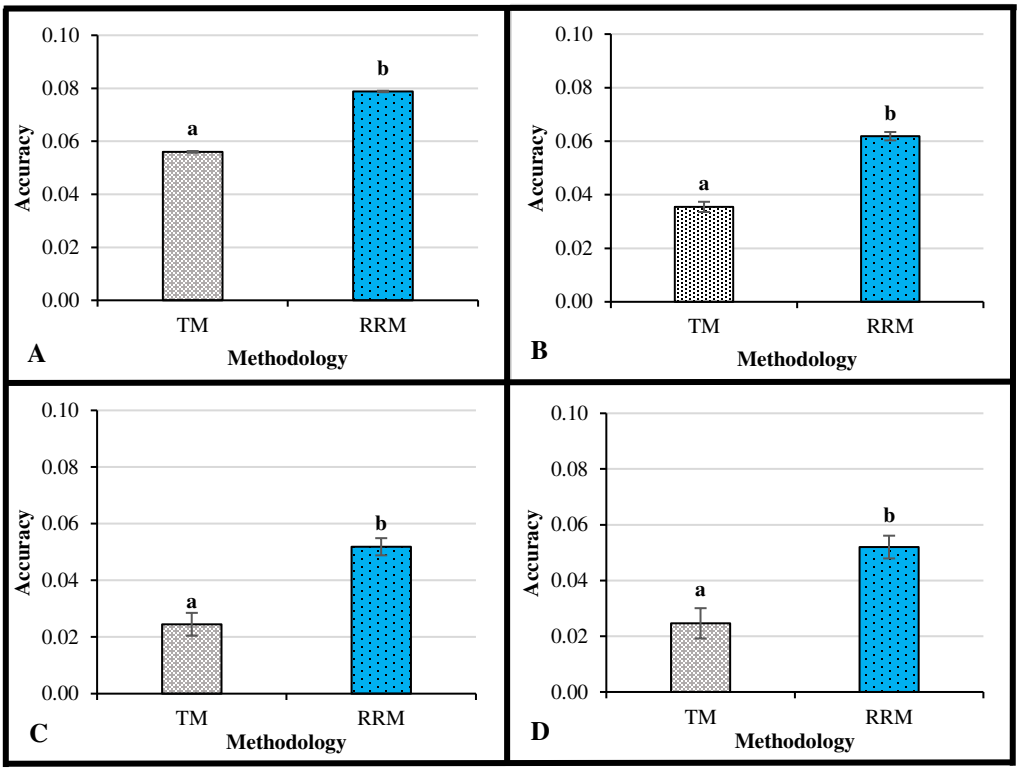


**Figure 4.7.** Mean accuracies for cow first-service conception rate predictions obtained at the average age of exposure (5 yr of age) with each statistical methodology. **A)** Mean accuracy for all animals in the pedigree (n = 14,140), **B)** Mean accuracies for all sires in pedigree (n = 971), **C)** Mean accuracies for sires that have produced progeny in the last five yr (n = 85), **D)** Mean accuracies for sires that have produced progeny in the last three yr (n = 51). Different letters indicate a statistical difference at the P <0.05 level among methodologies according to the least significant difference test.

Although the average increment of 2.6% in accuracy of prediction in favor of the RRM for all the comparisons performed could be explained by the slightly higher number of records utilized by this method (e.g., 9,448 records for the TM and 9,584 records for RRM); the inappropriateness of using age as the time descriptor in the RRM for FSCR prevented us from concluding that this increment was valid. The low accuracy values obtained with both methodologies was considered normal considering the low $h^2$ and repeatability estimates obtained for FSCR. First, there is a positive correlation between accuracy of predictions and the $h^2$ of the trait evaluated; therefore, if

$h^2$ is high, accuracies of prediction are high, whereas if $h^2$ is low, accuracies of prediction are low. Furthermore, even when considering all the AI events to which a cow was subjected during her life as repeated records, the practically null repeatability estimate of FSCR in this study implied that the expected accuracy gains due to the incorporation of more records to the analysis were still low (Bourdon, 2000). Additionally, the large number of bulls with small groups of daughters could also be an influencing factor in both methodologies to achieve low accuracies of prediction. For instance, about 62% of sires with female progeny had less than 5 daughters with FSCR observations (Table 4.15). Similar numbers of daughter records by sire were reported by Bormann et al. (2006) when evaluating HPG and FSCR in Angus heifers from 6 different herds; however, no accuracies for breeding value predictions were reported within such study.

**Table 4.15.** Number of daughters producing records of first-service conception rate by sire

| Number of daughters | Sires |
|---|---|
| <5 | 220 |
| 5 to 9 | 73 |
| 10 to 14 | 31 |
| 15 to 19 | 10 |
| 20 to 29 | 12 |
| 30 to 39 | 6 |
| >40 | 1 |

*4.4 Conclusion*

This study confirmed the capabilities of RRM to include any range of age endpoints for which phenotypic data was available and their flexibility to incorporate information from class levels of fixed effects with no variation in genetic evaluations of longitudinal binary fertility traits of multiparous beef cows. Particularly in the case of STAY, the strong Pearson and Spearman's correlations found between predictions obtained with RRM and TM, suggested that RRM could

effectively substitute to traditional TM genetic evaluations for this in beef cattle. In this case, even when some degree of re-ranking could be expected if RRM substitute traditional TM evaluation procedures, the important gains accuracies of prediction yielded by RRM may offset the possible inconvenience of such re-ranking. Furthermore, this study also compiled evidence about the possibilities of using predictions of the intercept of a RRM as the sole criterion in selection for STAY to any age of interest.

Conversely, in the case of FSCR evaluations, results of this study suggested that the age of the cows was not an appropriate longitudinal descriptor to model this trait using an RRM. As such, the predictions obtained via the execution of a repeatability TM remain as a base point from which improvements should be sought. The flexibility of RRM to accommodate any biologically relevant covariate associated to the trait of interest as the longitudinal descriptor, allows for future research where perhaps breeding weight or body condition scores could be used as a continuous descriptor of variations in FSCR. Evidently, such type of investigations would be dependent upon data availability and it may take some time before substantial information could be collected. Therefore, in the case of FSCR, it is imperative to keep altering the environment of cows thru reproductive managements focused to maximize reproductive success of females subjected to AI events.

*Literature cited*

Averill, T.A., R. Rekaya, and K. Weigel. 2004. Genetic analysis of male and female fertility using longitudinal binary data. J. Dairy Sci. 87:3947-3952.

Averill, T., R. Rekaya, and K. Weigel. 2006. Random regression models for male and female fertility evaluation using longitudinal binary data. J. Dairy Sci. 89:3681-3689.

Azzam, S. M., J. W. Keele, and M. K. Nielsen. 1988. Expectations of heritability estimates for non-return rate of bulls and conception rate of cows. J. Anim. Sci. 66:2767-2783.

Azzam, S. M., J. E. Kinder, and M. K. Nielsen. 1989. Conception rate at first insemination in beef cattle: effects of season, age and previous reproductive performance. J. Anim. Sci. 67:1405-1410.

Beckman, D. W., R. M. Enns, S. E. Speidel, B. W. Brigham, and D. J. Garrick. 2007. Maternal effects on docility in Limousin cattle. J. Anim. Sci. 85:650-657.

Beef Improvement Federation (BIF). 2020. Guidelines for Uniform Beef Improvement Programs: New, Improved and Wiki. http://guidelines.beefimprovement.org/index.php/Special:AllPages Accessed 28 July 2020).

Boettcher, P. J., L. K. Jairath, and J. C. M. Dekkers. 1999. Comparison of methods for genetic evaluation of sires for survival of their daughters in the first three lactations. J. Dairy Sci. 82:1034-1044.

Boldt, R.J., S. E. Speidel, M. G. Thomas, and R. M. Enns. 2018. Genetic parameters for fertility and production traits in Red Angus cattle. J. Anim. Sci. 96:4100-4111.

Bormann, J. M., L. R. Totir, S. D. Kachman, R. L. Fernando, and D. E. Wilson. 2006. Pregnancy rate and first-service conception rate in Angus heifers. J. Anim. Sci. 84:2022-2025.

Bourdon, R. M. 2000. Understanding animal breeding (second edition). Upper Saddle River, NJ: Prentice Hall.

Brigham, B. W., S. E. Speidel, and R. M. Enns. 2006. Alternative Definitions of Stayability. In: Proc. Beef Imp. Fed. 38th Annu. Res. Symp. Annu. Meet., Athens, GA. pp. 75-78.

Brigham, B. W., S. E. Speidel, R. M. Enns, and D. J. Garrick. 2007. Stayability to alternate ages. Proc. West. Sec. Amer. Soc. Anim. Sci. 58:27-30.

Buaban, S., K. Kuchida, M. Suzuki, Y. Masuda, W. Boonkum, and M. Duangjinda. 2016. Genetic analysis of the rates of conception using a longitudinal threshold model with random regression in dairy crossbreeding within a tropical environment. Anim. Sci. J. 87:961-971.

Cammack, K. M., M. G. Thomas, and R. M. Enns. 2009. Reproductive traits and their heritabilities in beef cattle. Prof. Anim. Sci. 25:517-528.

Christensen, R. H. B. 2015. Package 'ordinal'.
Retrieved from cran.rproject.org/web/packages/ordinal/ordinal.pdf

Crawford, N. F., M. G. Thomas, T. N. Holt, S. E. Speidel, and R. M. Enns. 2016. Heritabilities and genetic correlations of pulmonary arterial pressure and performance traits in Angus cattle at high altitude. J. Anim. Sci. 94:4483-4490.

Crews, D. H., and R. M. Enns. 2008. Genetic improvement of Heifer pregnancy and performance in beef cattle. In Proc. Appl. Rep. Strategies in Beef Cattle.
http://beefrepro.unl.edu/proceedings/2008ftcollins/12_csu_genetics_crews.pdf
(Accessed 4 July 2020).

Deutscher, G. H., J. A. Stotts, and M. K. Nielsen. 1991. Effects of breeding season length and calving season on range beef cow productivity. J. Anim. Sci. 69:3453-3460.

Doyle, S. P., B. L. Golden, R. D. Green, and J. S. Brinks. 2000. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. J. Anim. Sci. 78:2091-2098.

Foxworthy, H. M., R. M. Enns, M. G. Thomas, and S. E. Speidel. 2019a. The estimation of heritability and repeatability of first service conception and first cycle calving in Angus cattle. Trans. Anim. Sci. 3(Suppl_1):1646-1649.

Foxworthy, H. 2019b. Heritability and Repeatability Estimates of First Service Conception and First Cycle Calving in Angus Cattle. Master's Thesis, Colorado State University, Fort Collins.

Garrick, D.J., Garrick, D.P., Golden, B.L. 2018. An introduction to bolt software for genetic and genomic evaluations. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod. pp. 973.

Ghiasi, H., A. Pakdel, A. Nejati-Javaremi, H. Mehrabani-Yeganeh, M. Honarvar, O. González-Recio, M. J. Carabaño, and R. Alenda, R. 2011. Genetic variance components for female fertility in Iranian Holstein cows. Livest. Sci. 139:277-280.

Gianola, D., and J. L. Foulley. 1983. Sires evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201-224.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, and D. Buttler. 2009. ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK.*

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's toolkit user' s guide and reference manual. Agric. Exp. Stn. Tech. Bull. LTB92-2. Colorado State University, Fort Collins, CO.

Golden, B. L., S. Weerasinghe, B. Crook, S. Sanders, and D. J. Garrick. 2018. A Single-Step Hybrid Marker Effects Model Using Random Regression for Stayability in Hereford Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/single-step-hybrid-marker-effects-model-using-random-regression-stayability-hereford-cattle.pdf

González-Recio, O. and R. Alenda. 2005. Genetic parameters for female fertility traits and a fertility index in Spanish dairy cattle. J. Dairy Sci. 88:3282-3289.

Grossi, D. A., O. G. Frizzas, C. C. P. Paz, L. A. F. Bezerra, R. B. Lôbo, J. A. Oliveira, and D. P. Munari. 2008. Genetic associations between accumulated productivity, and reproductive and growth traits in Nelore cattle. Livest. Sci. 117:139-146.

Harris, B. L., J. E. Pryce, Z. Z. Xu, and W. A. Montgomerie. 2005.Fertility breeding values in New Zealand, the next generation. Interbull Bull. 33:47–50.

Harville, D. A., and R. W. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. Biometrics. 393-408.

Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2003. Genetic change for clinical mastitis in Norwegian cattle: a threshold model analysis. J. Dairy Sci. 86:369-375.

Hossein-Zadeh, N.G. and M. Ardalan. 2011. Genetic relationship between milk urea nitrogen and reproductive performance in Holstein dairy cows. Animal. 5: 26-32.

Hudson, G.F.S. and L. D. Van Vleck. 1981. Relationship between production and stayability in Holstein cattle. J. Dairy Sci. 64:2246-2250.

Jamrozik, J., J. Fatehi, G. J. Kistemaker, and L. R. Schaeffer. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. J. Dairy Sci. 88:2199-2208.

Jamrozik, J., S. McGrath, R. A. Kemp, and S. P. Miller. 2013. Estimates of genetic parameters for stayability to consecutive calvings of Canadian Simmentals by random regression models. J. Anim. Sci. 91:3634-3643.

Johnston, D. J., S. A. Barwick, G. Fordyce, R. G. Holroyd, P. J. Williams, N. J. Corbet, and T. Grant. 2014. Genetics of early and lifetime annual reproductive performance in cows of two tropical beef genotypes in northern Australia. Anim. Prod. Sci. 54:1-15.

Kadarmideen, H. N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 71:411-419.

Kenward, M. G. and J. H. Roger. 1997. The precision of fixed effects estimates from restricted maximum likelihood. Biometrics. 53:983–997.

Lamb, G. C., and V. R. Mercadante. 2016. Synchronization and artificial insemination strategies in beef cattle. Vet. Clin. N AM-Food A. 32:335-347.

Maiwashe, A., K. A. Nephawe, and H. E. Theron. 2009. Analysis of stayability in South African Angus cattle using a threshold model. S. Afr. J. Anim. Sci. 39:55-60.

Martinez, G. E., R. M. Koch, L. V. Cundiff, K. E. Gregory, Stephen D. Kachman, and L. Dale Van Vleck. 2005. Genetic parameters for stayability, stayability at calving, and stayability at weaning to specified ages for Hereford cows. J. Anim. Sci. 83:2033-2042.

Meyer, K. 2004. Scope for a random regression model in genetic evaluation of beef cattle for growth. Livest. Prod. Sci. 86:69-83.

Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Gen. Sel. Evol. 37:1-28.

Moreno, C., D. Sorensen, L. A. García-Cortés, L. Varona, and J. Altarriba. 1997. On biased inferences about variance components in the binary threshold model. Genet. Sel. Evol. 29:145-160.

Nishida, A., M. A. Aziz, S. Nishida, and K. Suzuki. 2006. Modelling number of services per conception of Japanese Black cattle by random regression. J. Anim. Breed. Genet. 123:56-63.

Olori, V.E., T.H.E. Meuwissen, and R. F. Veerkamp. 2002. Calving interval and survival breeding values as measure of cow fertility in a pasture-based production system with seasonal calving. J. Dairy Sci. 85:689-696.

Peel, R. K., J. L. Seabrook, G. E. Seidel Jr, J. C. Whittier, A. V. Grove, and J. K. Ahola. "Effect of 2, 4, and 5-hour intervals between 2 prostaglandin F2α injections administered with 5-day CO-Synch+ CIDR protocol on pregnancy rate in beef cows. Prof. Anim. Sci. 28:623-627.

Rahbar, R., M. Aminafshar, R. Abdullahpour, and M. Chamani. 2016. Genetic analysis of fertility traits of Holstein dairy cattle in warm and temperate climate. Acta Sci. Anim. Sci. 38:333-340.

R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Roxström, A., E. Strandberg, B. Berglund, U. Emanuelson, and J. Philipsson. 2001. Genetic and environmental correlations among female fertility traits and milk production in different parities of Swedish Red and White dairy cattle. Acta Agr. Scand. Section A-An. 51:7-14.

Sánchez-Castro, M.A., R. J. Boldt, M. G. Thomas, R. M. Enns, and S. E. Speidel. 2017. Expected progeny differences for stayability in Angus cattle using a random regression model. J. Anim. Sci. 95:88-89.

Sánchez-Castro, M. A., M. G. Thomas, R. M. Enns, and S. E. Speidel. 2019. Stability of genetic predictions for stayability using random regression models that include end points beyond 6 yr of age. Trans. Anim. Sci. 3:1678–1682.

Santana, M. L., J. P. Eler, A. B. Bignardi, and J. B. S. Ferraz. 2015. Two-trait random regression model to estimate the genetic association of scrotal circumference with female reproductive performance in Nelore cattle. Theriogenology. 83:1534-1540.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Schaeffer L. R. and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145-146.

Silva, D. O., M. L. Santana, D. R. Ayres, G. R. O. Menezes, L. O. C. Silva, P. R. C. Nobre, and R. J. Pereira. 2018. Genetic parameters for stayability to consecutive calvings in Zebu cattle. Animal. 12:1807-1814.

Silvestre, A., Â. Martins, V. Santos, and J. Colaço. 2019. Genetic parameters of calving ease in dairy cattle using threshold and linear models. Ital. J. Anim. Sci. 18:80-87.

Snelling, W. M., B. L. Golden, and R. M. Bourdon. 1995. Within-herd genetic analyses of stayability of beef females. J. Anim. Sci. 73:993-1001.

Snelling, W. M., L. A. Kuehn, R. M. Thallman, G. L. Bennett, and B. L. Golden. 2019. Genetic correlations among weight and cumulative productivity of crossbred beef cows. J. Anim. Sci. 97:63-77.

Speidel, S. E., R. M. Enns, and D. H. Crews. 2010. Genetic analysis of longitudinal data in beef cattle: a review. Genet. Mol. Res. 9:19-33.

Speidel, S. E., R. K. Peel, D. H. Crews, and R. M. Enns. 2016. Random regression models for the prediction of days to weight, ultrasound rib eye area, and ultrasound back fat depth in beef cattle. J. Anim. Sci. 94:471-482.

Speidel, S. E., Buckley, B. A., Boldt, R. J., Enns, R. M., Lee, J., Spangler, M. L. and Thomas, M. G. 2018a. Genome-wide association study of Stayability and Heifer Pregnancy in Red Angus cattle. J. Anim. Sci. 96:846-853.

Speidel, S. E., R. M. Enns, and B. L. Golden. 2018b. Use of a Random Regression model for the Evaluation of Heifer Pregnancy in Red Angus Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/use-random-regression-model-evaluation-heifer-pregnancy-red-angus-cattle.pdf

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. J. Dairy Sci. 92:2922-2930.

Van der Westhuizen, R. R., S. J. Schoeman, G. F. Jordaan, and J. B. Van Wyk. 2001. Heritability estimates derived from threshold analyses for reproduction and stayability traits in a beef cattle herd. S. Afr. J. Anim. Sci. 31:25-32.

Van Melis, M. H., J. P. Eler, H. N. Oliveira, G. J. M. Rosa, J. A. V. Silva, J. B. S. Ferraz, and E. Pereira. 2007. Study of stayability in Nellore cows using a threshold model. J. Anim. Sci: 85:1780-1786.

Walmsley, B. J., S. J. Lee, P. F. Parnell, and W. S. Pitchford. 2018. A review of factors influencing key biological components of maternal productivity in temperate beef cattle. Anim. Prod. Sci. 58:1-19.

Whittier, W. D., J. F. Currin, H. Schramm, S. Holland, and R. K. Kasimanickam. 2013. Fertility in Angus cross beef cows following 5-day CO-Synch+ CIDR or 7-day CO-Synch+ CIDR estrus synchronization and timed artificial insemination. Theriogenology. 80:963-969.

# CHAPTER 5 – SINGLE-STEP GENOMIC EVALUATIONS OF BEEF CATTLE BINARY FERTILITY TRAITS USING RANDOM REGRESSION SUPER HYBRID MODELS

*Summary*

Female fertility is a key economic driver of cow-calf operations; however, the achievement of genetic improvement of female fertility traits is challenging due to the biological complexity of reproduction and the difficulties related to statistical modeling of categorical and binary traits. Traits like heifer pregnancy (HPG), first-service conception rate (FSCR) and stayability (STAY) are important elements of the breeding objective for beef cattle producers. Nonetheless, their categorical nature has dictated that these traits be evaluated under a threshold theory that currently has not been adapted to incorporate genomic information in the form of single-step super-hybrid marker effects models. In contrast, random regression models (RRM) have emerged as an feasible alternative to evaluate binary fertility traits and they are flexible enough to accommodate any of the single-step genomic evaluation procedures that have been developed. There is a paucity of studies applying single-step random regression super-hybrid models (ssRR-SHM) for the genomic evaluation of beef cattle fertility traits, and given their recent development, a scarcity of reports prevail relative to the behavior of these models under different data structure scenarios. Moreover, the benefits related to the parallel detection of influential chromosomal regions while obtaining genomic predictions offered by marker effects models are limited. Therefore, objectives of this chapter were 1) to explore the feasibility of implementing ssRR-SHM for the genomic evaluation of HPG, FSCR and STAY; 2) to assess the impact of differing data structures in the resulting genomic predictions of ssRR-SHM for the traits; 3) identify quantitative trait loci (QTL) associated with the binary fertility traits contemplated in this study. Two types of genetic evaluations were

implemented for each trait. The first type of evaluation that was implemented was a pedigree-based RRM that utilized Legendre polynomials as the base function in where the phenotype of interest was regressed on an appropriate age covariate. The second evaluation type was a ssRR-SHM that also used Legendre polynomials as the base function and regressed observations of the trait of interest on its appropriated age covariate, but that included random effects of marker and extra polygenic effects. Within each trait, four different data structure scenarios were created depending on the phenotypic performance of the genotyped and non-genotyped subsets of animals. The behavior of the genomic predictions was assessed through the calculation of Pearson and Spearman's correlations and the estimation of the regression coefficients of EPD obtained with the ssRR-SHM on those obtained with their corresponding pedigree-based RRM. Results of this study suggested that the implementation of ssRR-SHM for the genomic evaluation of singly-observed binary fertility traits like HPG and FSCR, as well as for the evaluation of a longitudinally recorded binary trait such as STAY was feasible. Nonetheless, an overestimation of genomic predictions occurred with these models when phenotypic records of pre-selected genotyped animals were included in the evaluation. Additionally, inaccurate classify of genotypes for non-genotyped animals also impacted resulting genomic predictions, although this issue was restricted to this subgroup of animals only. In all cases, the removal of phenotypic records from preselected animals and the maintenance of closely related individuals in the pedigree ameliorated problems associated with the overestimation of genomic predictions and improved correlations among genomically-enhanced and pedigree-based EPD for all traits. Regarding GWAS analyses, the application of ssRR-SHM identified single nucleotide polymorphisms that resulted located either within or relatively close to genes that have been previously associated with important reproductive processes and fertility traits in cattle.

*5.1 Introduction*

Female reproductive performance represents one of the most relevant factors associated with the economic viability of beef cattle operations (Toghiani et al., 2017; Speidel et al., 2018a; Chudleigh et al., 2019). In fact, the enhancement of fertility-related traits has been estimated to be up to 4 times more important than improvements in end-product characteristics (Melton, 1995). Although fertility encompasses a variety of traits, only those measuring the success or failure of key biological events like conception and calving summarize the economically relevant outcomes of reproduction (Cammack et al., 2009, Walmsley et al., 2018). Consequently, traits such as heifer pregnancy (HPG), first-service conception rate (FSCR) and stayability (STAY) represent important elements of the breeding objectives of cow-calf enterprises (Golden et al., 2000). Nonetheless, factors hindering a rapid genetic progress of livestock populations for these characteristics include their sex-limited and discrete phenotypic expression, as well as their low heritability and the considerable amount of time required to collect phenotypes for their evaluation (Dekkers, 2010; Kluska et al., 2018; Hayes et al., 2019).

Several US beef cattle breed associations calculate expected progeny differences (EPD) for traits like HPG and STAY based on genetic evaluations performed using threshold models (Boldt, 2017). However, even when threshold models are theoretically superior for the evaluation of discrete response variables, these models do not consistently yield better results than linear models and some authors have suggested that it is not strictly necessary to use them in univariate genetic evaluations involving categorical traits (Ramírez-Valverde et al., 2001; Vostrý et al., 2014). Furthermore, among other limitations, threshold models have received criticism due to their lack of flexibility to incorporate genomic information differently than using genomic relationship matrices whose inverse is difficult to obtain when the number of genotyped animals is high

(Speidel et al., 2018b). Random regression models (RRM) represent an alternative method to evaluate binary fertility traits and their extension to genomic evaluation procedures that avoid the need of inverting dense relationship matrices has been already demonstrated (Jamrozik et al., 2013; Golden et al., 2018a, 2018b). Such procedures have been referred to as super-hybrid models and basically, they are Bayesian regression models capable of combining all available data from genotyped and non-genotyped animals in a single-step evaluation (Fernando et al., 2014, 2016). Super-hybrid models permit any *a priori* assumption for marker effects and allow a parallel Quantitative Trait Loci (QTL) detection while resolving for genomic enhanced breeding values using a single-step methodology (Golden and Garrick, 2016; Misztal and Lourenco, 2018).

The feasibility of a single-step random regression super-hybrid model (ssRR-SHM) genetic evaluation for a longitudinal binary reproductive trait like STAY has been documented in a large population of Hereford cattle (Golden et al., 2018a). However, no reports of the application of such statistical approach for the same trait exist in Angus cattle and neither for phenotypes observed once in the life of an animal (e.g., HPG and/or FSCR). Furthermore, although pedigree-based RRM have been successfully applied to evaluate binary reproductive traits in populations of any size (Averill et al., 2006; Speidel et al., 2018b); challenges associated with single-herd data structures have not been documented when extending this methodology for the inclusion of genomic information in the form of marker effects models. Therefore, objectives of this chapter were 1) to explore the feasibility of implementing ssRR-SHM for the genomic evaluation of HPG, FSCR and STAY in a purebred seedstock population of Angus cattle; 2) to assess the impact of differing data structures in the resulting genomic predictions of ssRR-SHM for all traits; 3) to identify QTL associated with the binary fertility traits contemplated in this study.

*5.2 Materials and Methods*

Data used in this study were obtained from an existing database; however, animals within the experimental location were managed according to the Colorado State University Institutional Animal Care and Use Committee (IACUC) guidelines, covered in most recent years by the IACUC number 18-8367A.

*5.2.1 Heifer phenotypic data collection and description*

A full description of the original dataset containing the heifer fertility phenotypes of HPG and FSCR was provided in the Materials and Methods section of Chapter 3. Briefly, phenotypic information for both traits was extracted from breeding and ultrasound records of 4,334 Angus heifers (progeny of 354 sires and 1,626 dams) collected between 1992 to 2019 at the Colorado State University Beef Improvement Center (CSU-BIC). Heifer pregnancy was defined as the ability of a heifer to produce a calf by 24 mo of age, given she conceived within a 60-d breeding season length. First-service conception rate was defined as the probability of a heifer conceiving in response to her first artificial insemination (AI) and maintaining such pregnancy until the end of the breeding season. For both traits, successful observations were coded as 1 and unsuccessful observations were coded as 0. Within each particular trait, different filtering processes of phenotypes were performed in order to identify the most suitable data for the application of a ssRR-SHM genomic evaluation. After such filtering processes, a pedigree containing 3 generations of ancestors was built for the animals with phenotypes that remained in each one of the data subsets by extracting information from the historical pedigree of the CSU-BIC (the full pedigree contained 14,140 individuals in total).

*5.2.2 Cow phenotypic data collection and description*

Similar to the original heifer phenotypic files, a complete description of the data containing the calving performance of multiparous cows of the John E. Rouse CSU-BIC between 1993 and 2019 was provided in the Materials and Methods section of Chapter 4. In summary, the original phenotypic information included 8,907 calving records of 1,713 Angus cows (progeny of 302 sires and 1,068 dams). Stayability observations were assigned to dams according to their age in days at each calving. Given every female calved as a 2-yr-old, starting from their third calving, the value of 1 (successful) or 0 (unsuccessful) was assigned to cows that either produced a calf or did not produced a calf within each particular age endpoint (ages ranging from 3 to 12 yr). Again, distinct filtering processes of STAY phenotypes were conducted with the objective of identifying the most appropriate data structure for the application of a ssRR-SHM genomic evaluation for this trait. Afterwards, a 3-generation pedigree was constructed for the females with STAY phenotypes that remained in each one of subsets of data retrieving such information from the historical pedigree records of the CSU-BIC.

*5.2.3 Genotypic data collection and description*

In 2011, a whole-herd genotyping process was begun at the CSU-BIC. Initially, all animals were genotyped using a 50k SNP panel (Illumina Bovine SNP50 v2.0, Illumina Inc., San Diego, CA); whereas 65 steers were genotyped with the Illumina High-Density Bovine SNP chip for a total of 777,962 markers (Illumina Bovine high density (HD); Illumina Inc., San Diego, CA). Since that time, a continuous effort has been made to genotype every year's calf crops (~400 calves per year) using SNP arrays of varying densities and from different genotyping laboratories (e.g., 50k or i50k from Zoetis; GGP from GeneSeek). Until 2017, the total number of genotyped animals at the CSU-BIC was 3,621 (Appendix B, Table B-1).

Given the heterogeneity of the SNP panels used across different years at the CSU-BIC, an imputation process using the findhap software (VanRaden, 2011) was implemented to homogenize the number of available markers to those contained in the Illumina's Bovine SNP50 v2.0 array (~54,000 SNP per animal). As part of the marker quality control analyses, genetic markers with an average call rate lower than 0.85, minor allele frequency less than 0.01, deviated from Hardy–Weinberg equilibrium (P<0.0001) and in extreme linkage disequilibrium ($r^2 > 0.99$) were removed from the data using the PLINK 1.9 software (Chang et al., 2015). After this filtration process, a total of 33,862 SNP genotypes remained available for subsequent analyzes.

*5.2.4 Data structure exploration of HPG phenotypes*

Considering that phenotypic information for HPG dates back to 1992 and that the overall genotyping process of the Rouse Angus herd did not begin until 2011, heifers born in the early years within the data did not possessed genotypes. Figure 5.1 shows the number of heifers with fertility phenotypic records, as well as their genotyping status according to their year of birth.

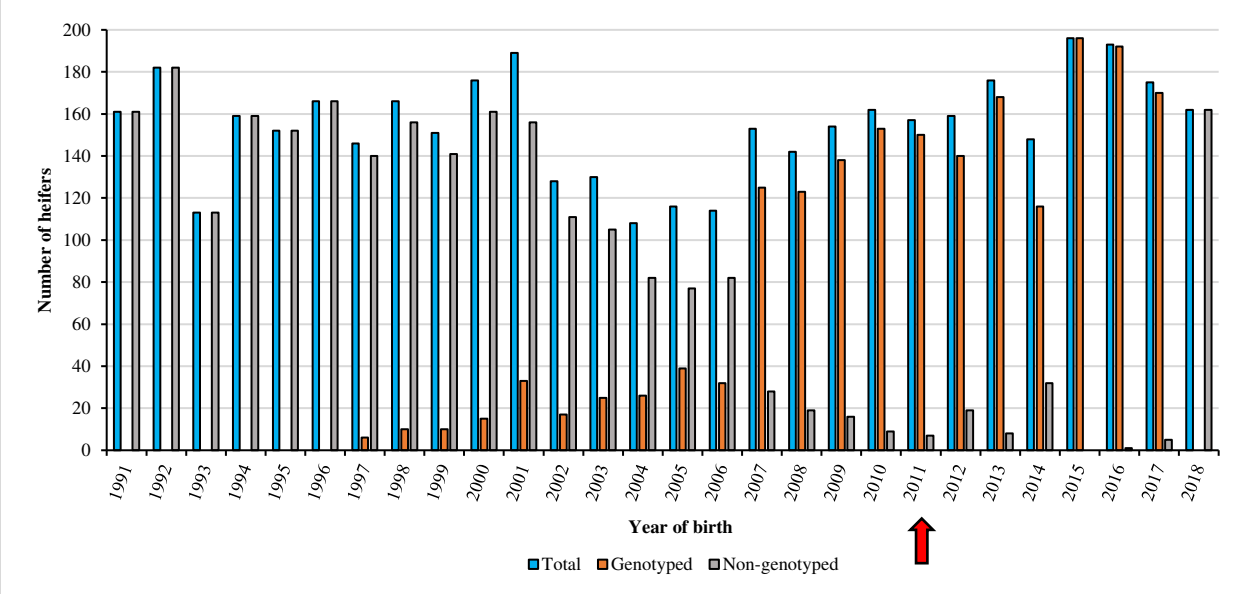


**Figure 5.1.** Summary of the genotyping status per year of birth of all heifers with fertility phenotypic information at the Colorado State University Beef Improvement Center (red arrow indicates the first genotyping year).

The first dataset implemented to test the viability of a ssRR-SHM genomic evaluation for HPG (DS-1_HPG) contained phenotypic records of all 4,334 heifers. The final pedigree in DS-1_HPG consisted in a total of 6,773 individuals, with 832 unique sires and 2,904 dams. The average inbreeding coefficient of the final pedigree was 0.008. Among the animals included in the final pedigree file, a total of 2,037 individuals had genotypes available, while the remaining 4,736 animals were non-genotyped (a graphical description of DS-1_HPG is shown in Figure 5.2).

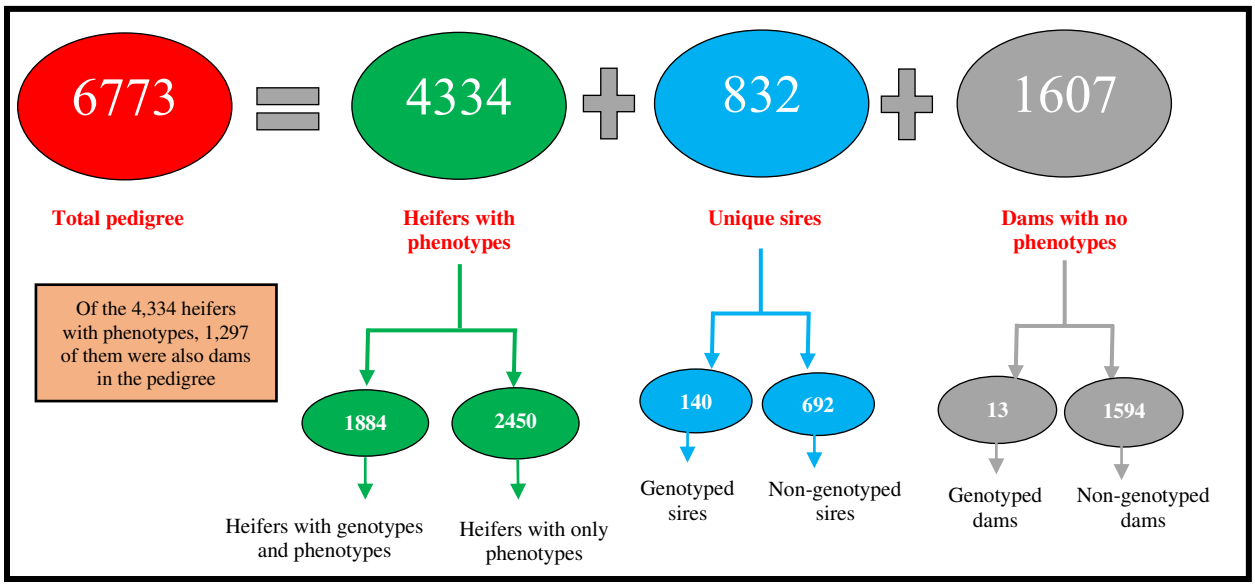


**Figure 5.2.** Data structure within the first dataset (DS-1_HPG) used in the random regression super-hybrid model genomic evaluation for heifer pregnancy.

According to each year's specific culling rate within the CSU-BIC, a variable number of heifers were kept as replacements annually, and consequently, they appeared in the pedigree as dams of some other heifers that produced phenotypes for HPG in later years. A detailed graphical description of the number of heifers according to their genotyping status and their fate within the herd is provided in Figure 5.3. In general, 1,297 out of the 4,334 heifers with HPG phenotypes were kept as replacements at the CSU-BIC; from which, 620 had genotypes and phenotypes and 677 only possessed phenotypes.

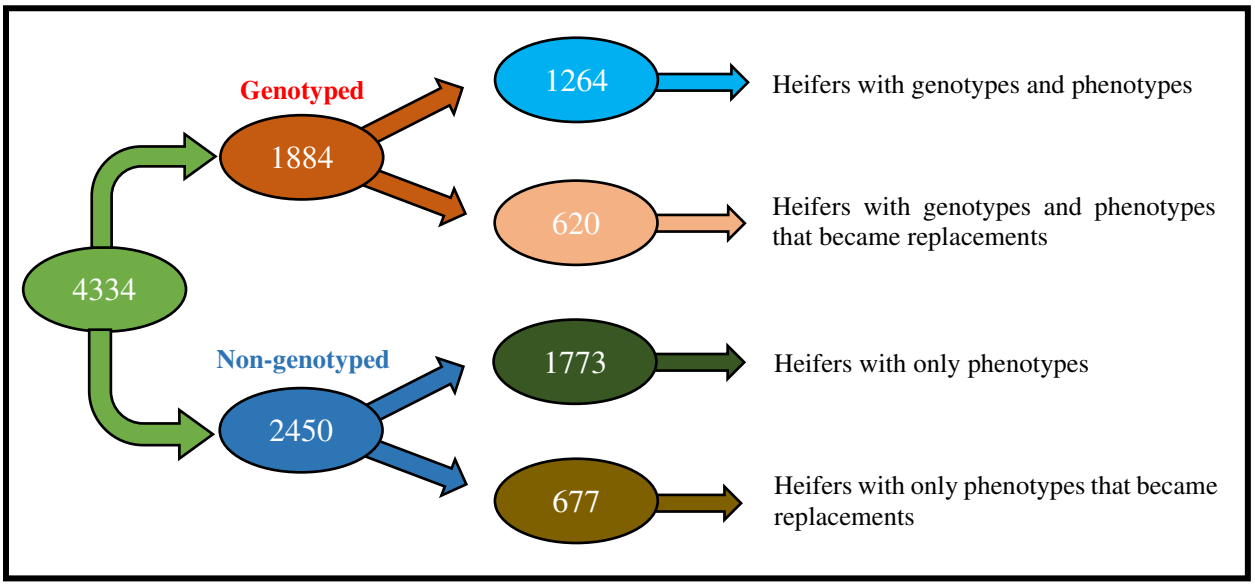


**Figure 5.3.** Number of heifers according to their genotyping status and their fate within the Colorado State University Beef Improvement Center.

Given the strong selection pressure that the CSU-BIC places on fertility, replacement heifers have been historically selected on the basis of their genetic merit and proved performance, retaining only fertile animals (e.g., pregnant heifers). Figure 5.4 illustrates such selection policy by showing the pregnancy rates of each year's selected group of replacements (n = 1,297).

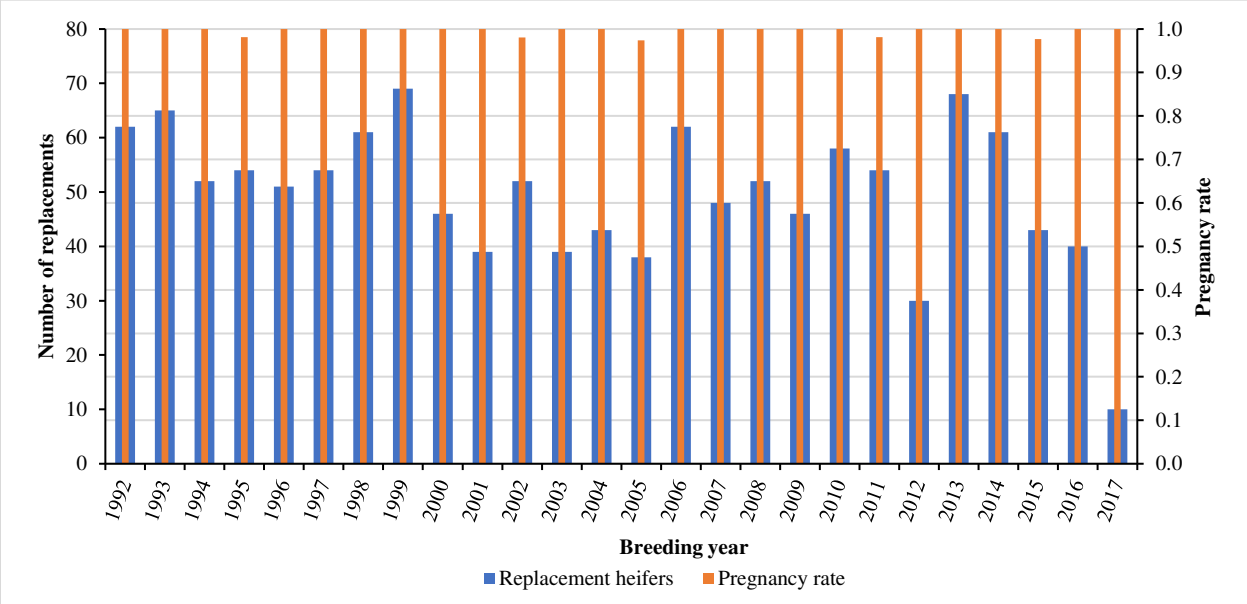


**Figure 5.4.** Pregnancy rate of selected replacement heifers at the Colorado State University Beef Improvement Center.

The fact that only heifers with successful HPG observations became replacements and only this subset of females appeared as dams of other heifers in the pedigree could affect resulting genomic predictions due to preselection bias. Therefore, a second dataset was created to evaluate the feasibility of a ssRR-SHM genomic evaluation for HPG (DS-2_HPG). This data set consisted of the removal of phenotypic records from the 1,297 heifers that eventually became replacements in the herd (620 genotyped and 677 non-genotyped). Such filtering processes resulted in a file containing phenotypic records of 3,037 heifers. The final pedigree in DS-2_HPG consisted in a total of 6,738 individuals, with 827 unique sires and 2,874 dams. The average inbreeding coefficient of this pedigree was 0.008. Among the animals included in the final pedigree file, 2,037 individuals had genotypes available, while the remaining 4,701 animals were non-genotyped (a graphical description of the DS-2_HPG is shown in Figure 5.5).

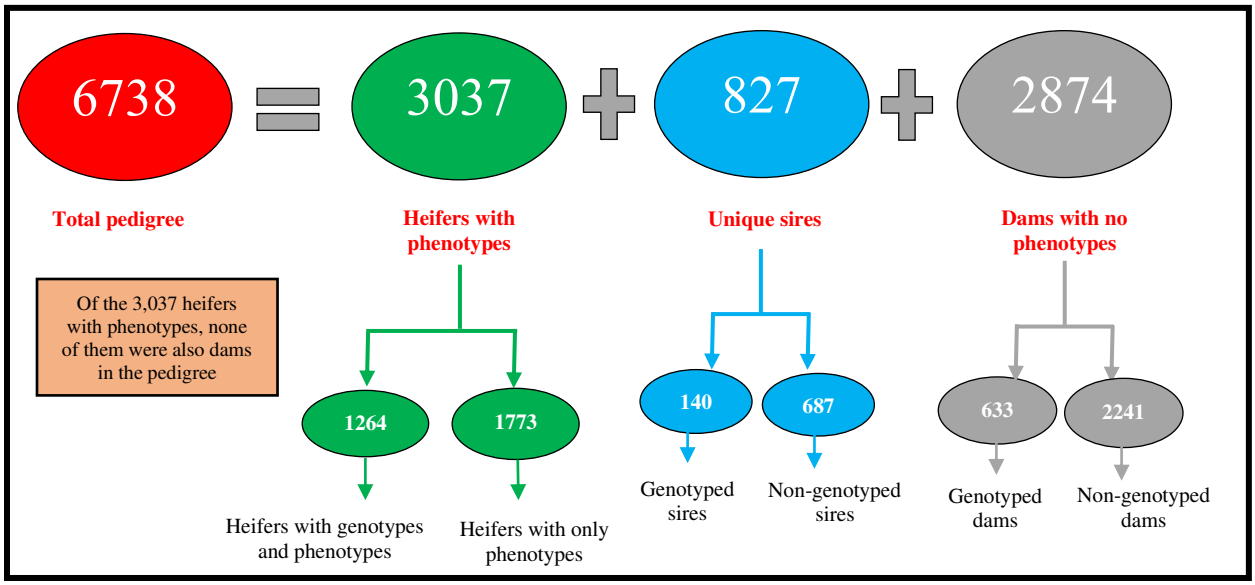


**Figure 5.5.** Data structure within the second dataset (DS-2_HPG) used in the random regression super-hybrid model genomic evaluation for heifer pregnancy.

The next exploration procedure consisted of investigating the pregnancy rates of all the genotyped heifers (n = 1,884) of the CSU-BIC. Figure 5.6 shows the average pregnancy rate of genotyped heifers grouped according to their specific year of birth. Overall, pregnancy rate of

genotyped heifers ranged between 0.64 to 1; however, the majority of the average pregnancy rates that were ≥0.9 belonged to heifers that were born before 2007 and such values were highly influenced by the reduced numbers of animals of born in those years of data (e.g., <40).

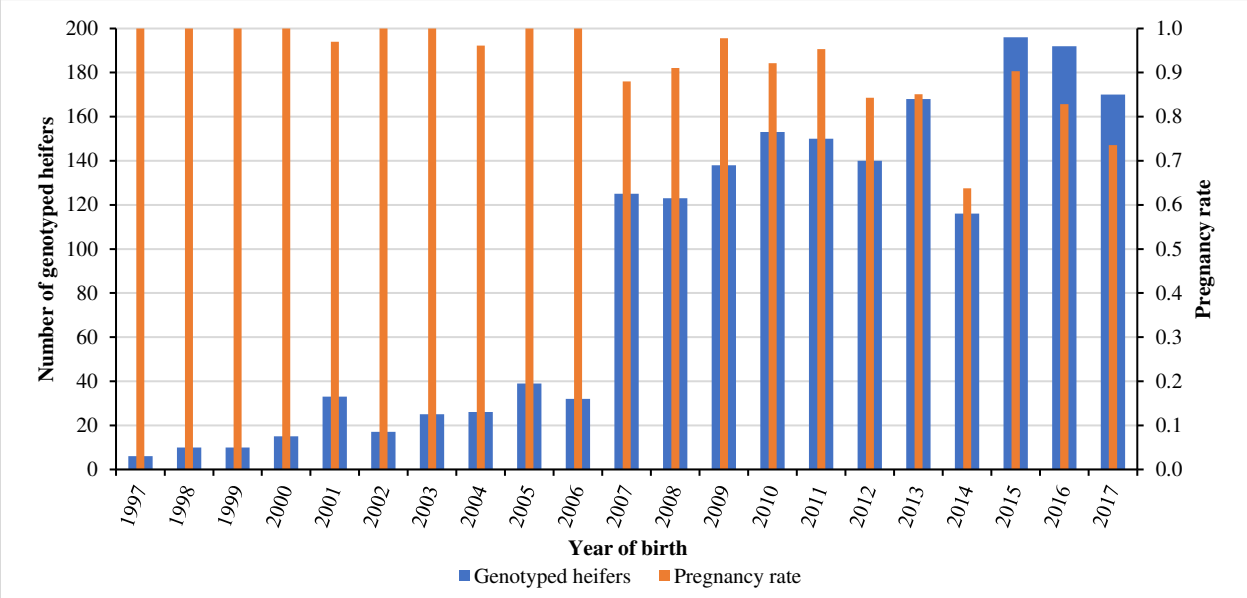


**Figure 5.6.** Pregnancy rate of genotyped heifers grouped by year of birth at the Colorado State University Beef Improvement Center.

The third dataset generated to evaluate the feasibility of a ssRR-SHM genomic evaluation for HPG (DS-3_HPG) was created by removing phenotypic records of the genotyped heifers born before 2007 (n = 213). The previous resulted in a file containing phenotypic records of 4,121 heifers and a pedigree file formed for a total of 6,764 individuals, with 832 unique sires and 2,902 dams. The average inbreeding coefficient of this pedigree was 0.008. Among the animals included in this particular pedigree, a total of 2,029 individuals had genotypes, while the remaining 4,735 animals were non-genotyped (a graphical description of the DS-3_HPG is shown in Figure 5.7).

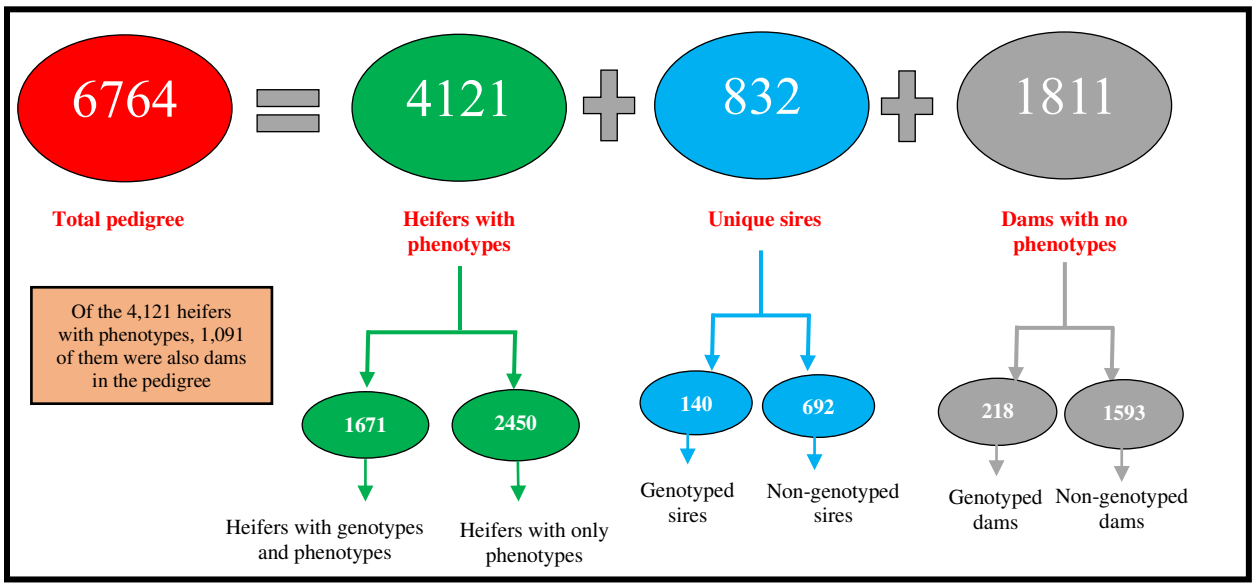


**Figure 5.7.** Data structure within the third dataset (DS-3_HPG) used in the random regression super-hybrid model genomic evaluation for heifer pregnancy.

The last dataset created to evaluate the feasibility of a ssRR-SHM evaluation for HPG (DS-4_HPG) included only the phenotypes belonging to the 1,884 genotyped heifers with phenotypic records. For DS-4_HPG the final pedigree had an average inbreeding coefficient of 0.007 and included 3,934 individuals, as well as 695 unique sires and 1,944 dams. Within this pedigree, 2,032 animals had genotypes available and 1,902 were ungenotyped animals (Figure 5.8).

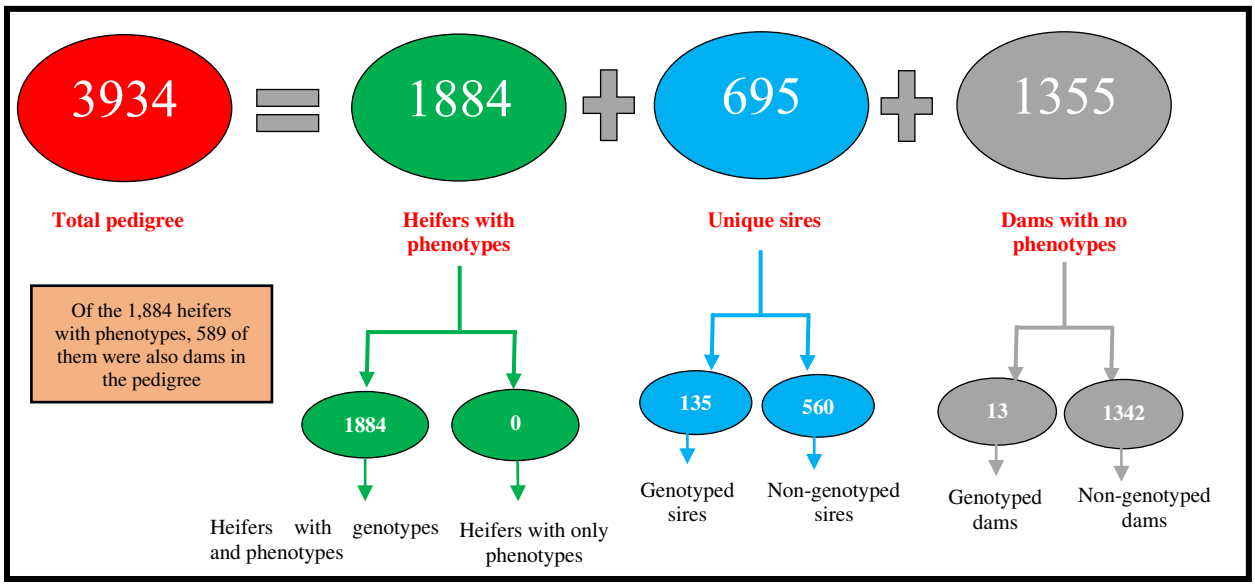


**Figure 5.8.** Data structure within the third dataset (DS-4_HPG) used in the random regression super-hybrid model genomic evaluation for heifer pregnancy.

*5.2.5 Data structure exploration of heifer FSCR phenotypes*

The same cohort of heifers with phenotypic information for HPG also possessed FSCR records, therefore, the same information and data structure relative to the genotyping status of these heifers existed for FSCR analyses (e.g., Figure 5.1). In this context, all the previously described datasets created to test the feasibility of a ssRR-SHM genomic evaluations for HPG were also used to test the viability of this statistical approach for FSCR. Such datasets were just renamed and handled appropriately to include the systematic and random effects biologically associated to FSCR.

Specifically, the first dataset was DS-1_FSCR (contained phenotypic information of the original 4,334 heifers with FSCR phenotypes along with its respective 3-generations pedigree file, e.g., Figure 5.2). The second dataset was DS-2_FCSR and contained phenotypic records of 3,037 heifers along with its corresponding 3-generation pedigree file (e.g., Figure 5.5). The third dataset (DS-3_FSCR) included phenotypic records of the 4,121 heifers born after 2007 and its 3-generation pedigree file (Figure 5.5). The last dataset (DS-4_FSCR) was formed with the information of only genotyped heifers (n = 1,884) along with their respective 3-generation pedigree file (e.g., Figure 5.8). As an important clarification, the same rationale followed to remove phenotypic records of heifers born before 2007 in HPG was followed for the FSCR evaluations given the same group animals showed considerably inflated averages for FSCR due to the reduced number of females represented in the data (e.g., <40). Figure 5.9 shows the average FSCR of genotyped heifers grouped according to their corresponding year of birth.

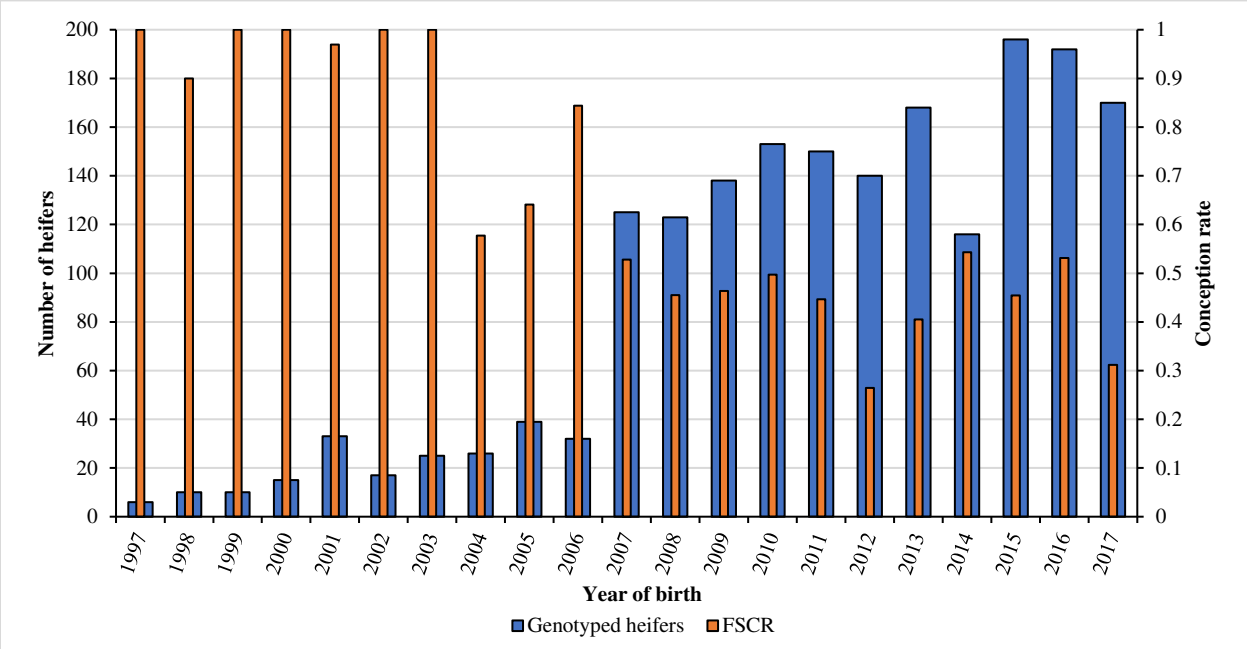


**Figure 5.9.** First-service conception rate (FSCR) of genotyped heifers grouped by year of birth at the Colorado State University Beef Improvement Center.

*5.2.6 Data structure exploration of STAY phenotypes*

Calving records for this study were available since 1990, therefore, not all females within the early years of data survived up to 2011 in order to be genotyped (Figure 5.10).

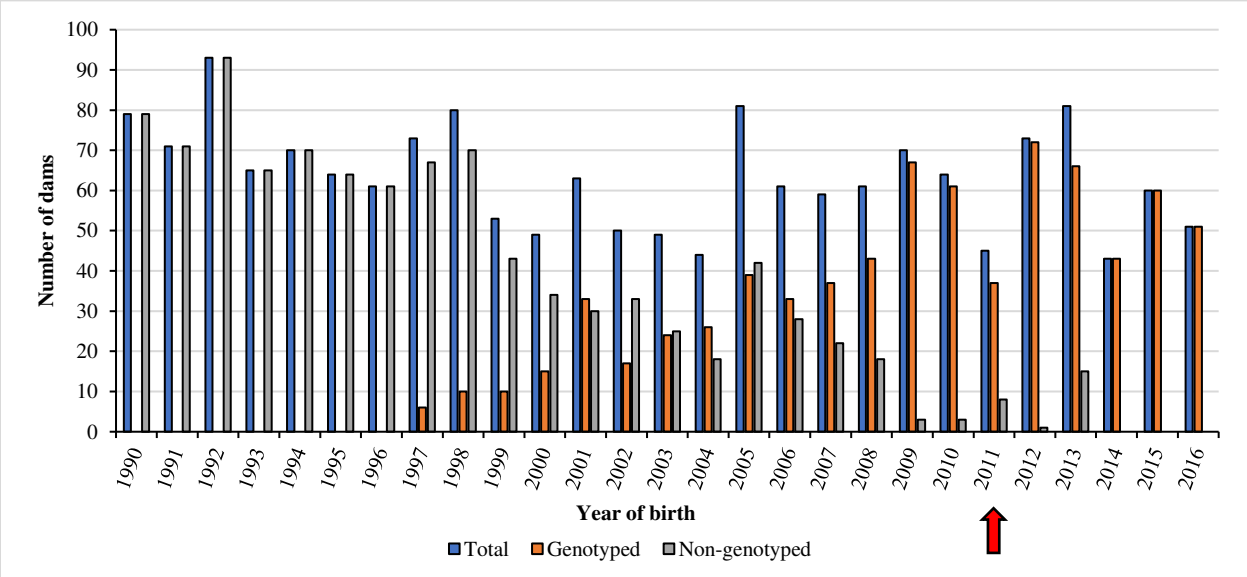


**Figure 5.10.** Summary of the genotyping status per year of birth of all dams with stayability records at the Colorado State University Beef Improvement Center (red arrow indicates the first genotyping year).

The first dataset implemented to test the viability of a ssRR-SHM genomic evaluation for STAY (DS-1_STAY) contained 8,907 records of all 1,713 multiparous cows. The final pedigree in DS-1_STAY consisted of a total of 3,569 individuals, with 660 unique sires and 2,026 dams. The average inbreeding coefficient of the final pedigree was 0.006. Among the animals included in the final pedigree file, a total of 882 individuals had genotypes available, while the remaining 2,687 animals were non-genotyped (Figure 5.11).

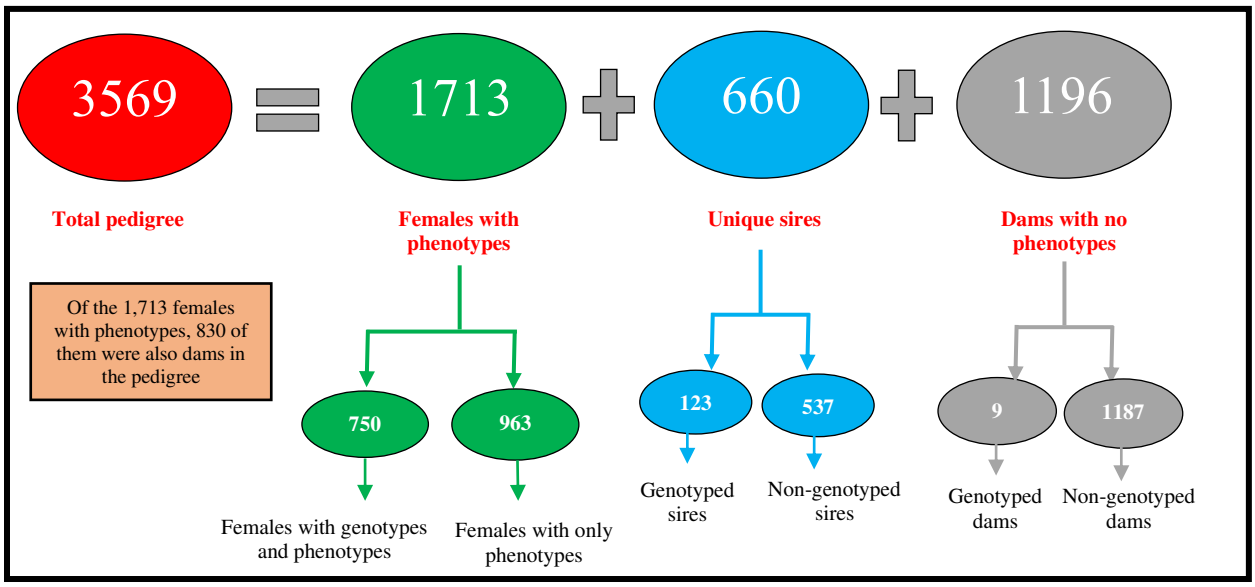


**Figure 5.11.** Data structure within the first dataset (DS-1_STAY) used in the random regression super-hybrid model genomic evaluation for stayability.

Given the cumulative nature of this fertility measure, all cows born in 2005 (and before) that were still present at the herd by 2011 had a 100% success for STAY06 observations (Figure 5.12). Consequently, the second dataset (DS-2_STAY) created to test the feasibility of a ssRR-SHM genomic prediction for STAY was formed by disregarding the phenotypic information of genotyped females born before 2006 (e.g., 180 cows were removed). This file contained 7,218 records on 1,533 cows. The final pedigree in DS-2_STAY included a total of 3,528 individuals, with 660 unique sires and 2,015 unique dams. The average inbreeding coefficient of the final

pedigree was 0.006. Among the animals included in the final pedigree file, 843 of them were genotyped and the remaining 2,685 were non-genotyped individuals (Figure 5.13).

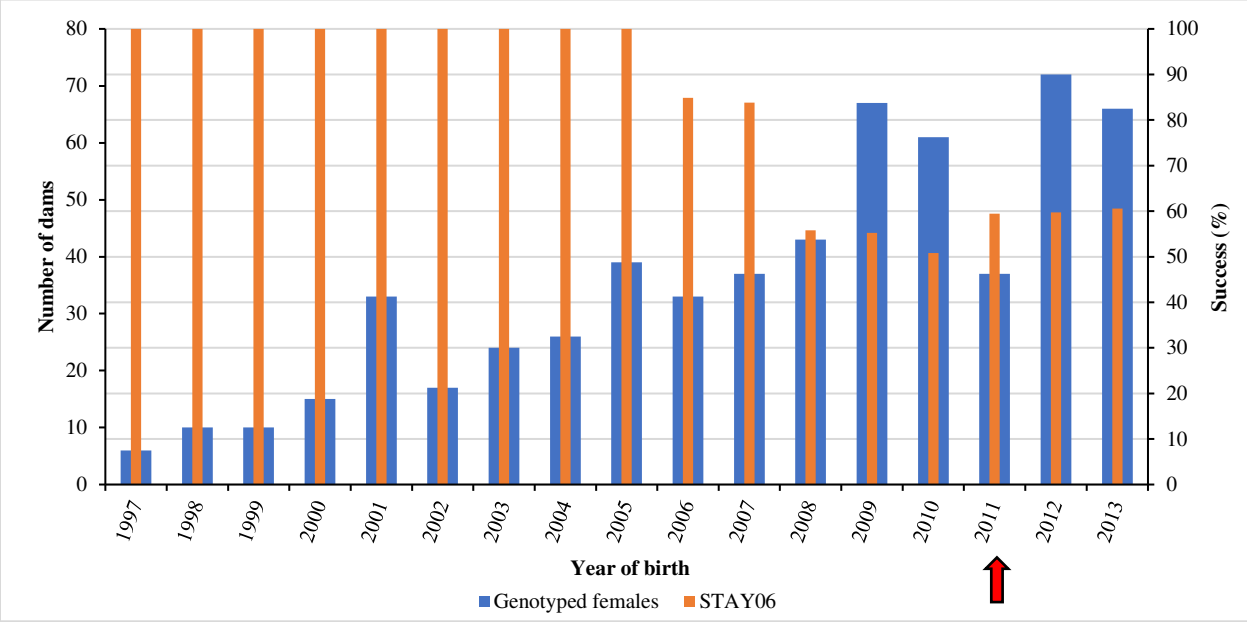


**Figure 5.12.** Success percentage of stayability at the age of 6 in genotyped dams grouped by year of birth at the Colorado State University Beef Improvement Center (red arrow indicates the first genotyping year).

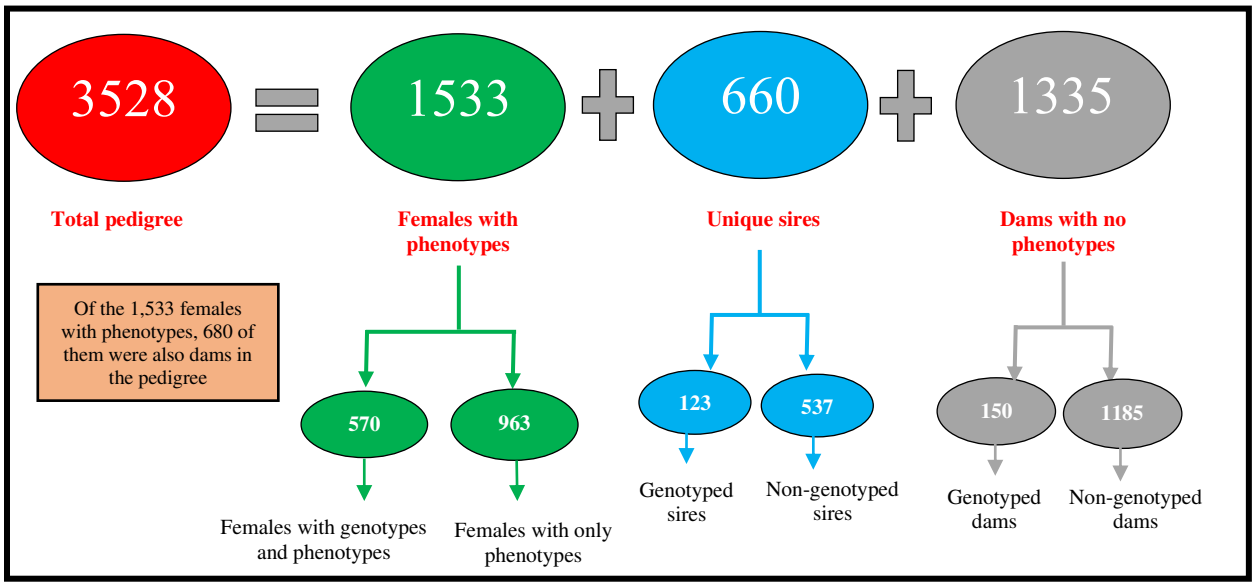


**Figure 5.13.** Data structure within the second dataset (DS-2_STAY) used in the random regression super-hybrid model genomic evaluation for stayability.

The third dataset generated to evaluate the feasibility of a ssRR-SHM genomic evaluation for STAY (DS-3_STAY) was generated by including only phenotypic information pertaining to all genotyped cows (e.g., 4,334 observations from 750 cows). For DS-3_STAY the final pedigree had an average inbreeding coefficient of 0.005 and included 2,450 individuals, as well as 595 unique sires and 1,422 dams. Within this pedigree, 881 animals had genotypes available and 1,569 were ungenotyped animals (Figure 5.14).

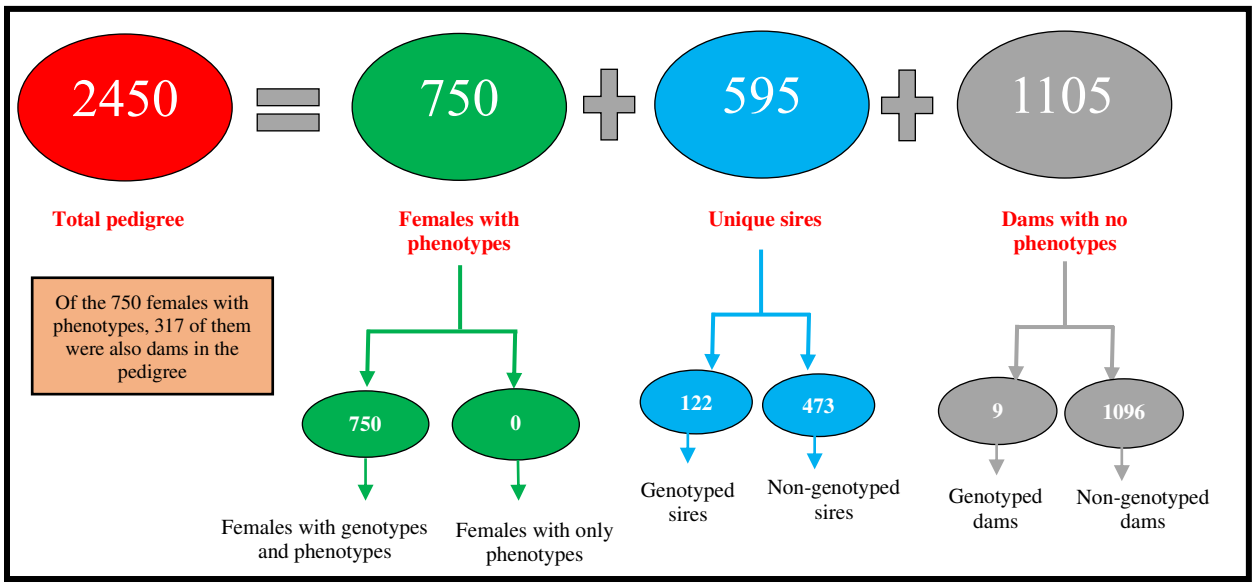


**Figure 5.14.** Data structure within the third dataset (DS-3_STAY) used in the random regression super-hybrid model genomic evaluation for stayability.

The last dataset generated to assess the viability of a ssRR-SHM genomic evaluation for STAY was constructed by including the information of all females born after 2005 (e.g., 570 genotyped cows and 98 ungenotyped cows). Hence, the fourth dataset (DS-4_STAY) included 2,279 observations from a total of 668 cows. The 3-generations pedigree file built for DS-4_STAY had an average inbreeding coefficient of 0.003 and consisted in a total of 2,340 individuals, with 570 unique sires and 1,330 unique dams. Within this pedigree, there were 837 genotyped and 1,503 ungenotyped animals (Figure 5.15).

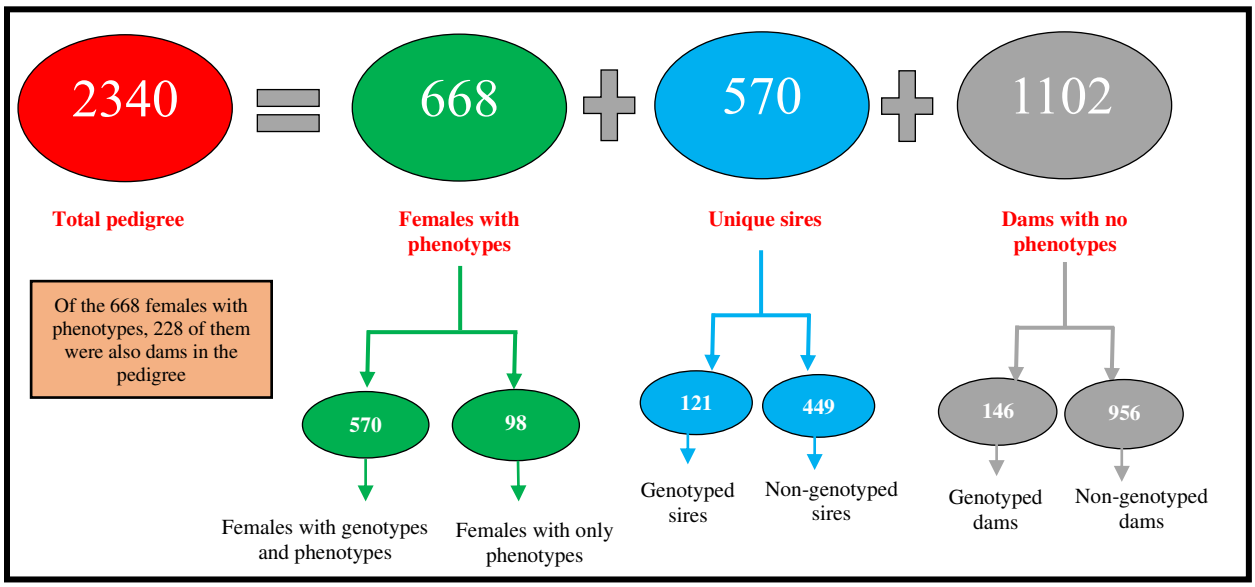


**Figure 5.15.** Data structure within the fourth dataset (DS-4_STAY) used in the random regression super-hybrid model genomic evaluation for stayability.

*5.2.7 Genetic and genomic evaluations for heifer pregnancy*

Two types of RRM were implemented for each heifer pregnancy dataset (DS-1_HPG through DS-4_HPG). The first type was a pedigree-based RRM such as the one described in equation 3.4 of the Materials and Methods section of Chapter 3. The purpose of this model was to generate base genetic predictions for HPG that were subsequently compared with genomic predictions generated by the single-step random regression super-hybrid models (ssRR-SHM; model type 2). Within the models implemented in all evaluations, HPG was regressed on the age at first exposure (AFE) using a linear RRM that utilized Legendre polynomials as the base function. The general model equation (5.1) for all ssRR-SHM is presented below:

$$\begin{bmatrix} \boldsymbol{y_n} \\ \boldsymbol{y_g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_n} \\ \boldsymbol{X_g} \end{bmatrix} \mathbf{b} + \begin{bmatrix} \boldsymbol{D^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D^1} \end{bmatrix} \begin{bmatrix} \boldsymbol{d^0} \\ \boldsymbol{d^1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_n^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_g^0 M_g} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_n^0} \\ \boldsymbol{\alpha^0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_n^1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_g^1 M_g} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_n^1} \\ \boldsymbol{\alpha^1} \end{bmatrix} + \mathbf{e} \qquad \text{Eq. 5.1}$$

where $\boldsymbol{y_n}$ and $\boldsymbol{y_g}$ corresponded to binary HPG observations on non-genotyped (**n**) and genotyped (**g**) individuals; $\boldsymbol{X_n}$ and $\boldsymbol{X_g}$ were appropriate incidence matrices that related fixed effects of breeding contemporary group (defined as a combination between breeding year and breeding

pasture), age of dam (according to the BIF recommendations; BIF, 2020) and a linear fixed regression of HPG on AFE contained in **b** to HPG observations in $\boldsymbol{y}$ (specifically sorted with **n** individuals first and **g** individuals after); $\boldsymbol{D^0}$ and $\boldsymbol{D^1}$ were appropriate age covariates (intercept and linear term) relating HPG observations in $\boldsymbol{y}$ to the additive random extra polygenic effects in $\boldsymbol{d^0}$ and $\boldsymbol{d^1}$; $\boldsymbol{Z_n^0}$ and $\boldsymbol{Z_n^1}$ corresponded to intercept and linear age covariates relating HPG observations in $\boldsymbol{y}$ to animal additive direct genetic effects accounted by imputed marker values in $\boldsymbol{u_n^0}$ and $\boldsymbol{u_n^1}$, respectively [with $\boldsymbol{u_n = M_n\alpha + \epsilon}$ (where $\boldsymbol{\epsilon}$ = imputation residual)]; $\boldsymbol{Z_g^0}$ and $\boldsymbol{Z_g^1}$ represented intercept and linear age covariates relating HPG observations in $\boldsymbol{y}$ to marker effects contained in $\boldsymbol{\alpha^0}$ and $\boldsymbol{\alpha^1}$, respectively; **M** denoted a matrix of centered marker values (coded as -1, 0 or 1, representing homozygous, heterozygous and opposite homozygous genotypes, respectively); and **e** represented a vector of random errors.

The extra polygenic additive genetic effects terms ($\boldsymbol{d}$) were fit assuming ½ of the additive genetic variance was not captured by markers and these effects were considered to be uncorrelated to other random effects (Golden et al., 2018a). For these models, variances of random effects we assumed to be equal to:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{d}\\ \boldsymbol{u_n}\\ \boldsymbol{\alpha}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G_d^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{G_u^{-1}} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{G_\alpha^{-1}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

With:

$$\boldsymbol{G_d^{-1}} = [\boldsymbol{G_0}(\mathbf{1}-\boldsymbol{c})]^{-1} \qquad \boldsymbol{G_u^{-1}} = [\boldsymbol{G_0 c}]^{-1}$$

$$\boldsymbol{G_\alpha^{-1}} = \left[\boldsymbol{G_0}\frac{c}{2_k\overline{pq}(1-\pi)}\right]^{-1} \qquad \boldsymbol{G_0} = \mathbf{Var}\begin{bmatrix}\boldsymbol{\beta_0}\\ \boldsymbol{\beta_1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_{\beta_0}^2} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{A\sigma_{\beta_1}^2}\end{bmatrix}$$

where $\boldsymbol{G_d^{-1}}$ denoted the variance of the extra polygenic effects, $\boldsymbol{G_u^{-1}}$ corresponded to the variance of the residual polygenic effects and $\boldsymbol{G_\alpha^{-1}}$ represented the variance of marker effects; $\boldsymbol{G_0}$ was a modified variance-covariance matrix of additive genetic random regression coefficients where the covariance between the intercept ($\boldsymbol{\beta_0}$) and the linear term ($\boldsymbol{\beta_1}$) was assumed to be zero, given no heifer had more than one observation for HPG (Speidel et al., 2018b); **A** corresponded to the additive numerator relationship matrix; $\boldsymbol{c}$ represented the proportion of genetic variance accounted for by the marker effects; $\boldsymbol{k}$ was the number of loci in the genotype matrix, $\boldsymbol{\overline{pq}}$ represented the average of the product of the p and q loci frequencies and; $\boldsymbol{\pi}$ represented the prior probability of a marker not affecting the trait being studied ($\pi$ = 0.99).

The mixed model equations (MME) of all these Bayes C$\pi$ ssRR-SHM were assembled using the BOLT software package (Release 1.2.7; http://www.thetasolutionsllc.com/bolt-software.html). These MME were solved in the first instance using a preconditioned conjugate gradient (PCG) method (Garrick et al., 2018). Afterwards, three parallel BayesC Gibb's single site samplers were seeded with the PCG solutions and a total of 300,000 effective samples were obtained using Markov chain Monte Carlo (MCMC) methods (each chain consisted in 115,000 iterations with the first 15,000 of them considered as burn-in). Linear functions of sampled model effects such as $\boldsymbol{\beta}$ (fixed effects), $\boldsymbol{\alpha}$ (marker effects), $\boldsymbol{\epsilon}$ (imputation residual effects for non-genotyped animals) and $\boldsymbol{d}$ (extra polygenic effects common to genotyped and non-genotyped animals) were computed to obtain estimated breeding values (EBV) specific to the average AFE on non-genotyped and genotyped animals, using equations 5.2 and 5.3, respectively:

$$\boldsymbol{EBV_n = [(M_n\alpha_m^0 + \epsilon) * \Phi_i^0] + [(M_n\alpha_m^1 + \epsilon) * \Phi_i^1] + (d_m^0 * \Phi_i^0) + (d_m^1 * \Phi_i^1) + (J_1{*}\ \beta) + (K_1{*}\ \beta)} \quad \text{Eq. 5.2}$$

$$\boldsymbol{EBV_g = (M_g\alpha_m^0 * \Phi_i^0) + (M_g\alpha_m^1 * \Phi_i^1) + (d_m^0 * \Phi_i^0) + (d_m^1 * \Phi_i^1) + (J_2{*}\ \beta) + (K_2{*}\ \beta)} \quad \text{Eq. 5.3}$$

where $\boldsymbol{EBV_n}$ ($\boldsymbol{EBV_g}$) corresponded to the estimated breeding value for non-genotyped and genotyped animals, respectively; $\Phi_i^0$ ($\Phi_i^1$) corresponded to intercept (linear) coefficients of Legendre polynomials standardized to the $\mathbf{i^{th}}$ age of interest (e.g., AFE = 422 d); $\boldsymbol{\alpha_m^0}$ ($\boldsymbol{\alpha_m^1}$) represented the intercept (linear term) marker effects random regression solutions for the $\mathbf{m^{th}}$ animal; $\boldsymbol{d_m^0}$ ($\boldsymbol{d_m^1}$) were the intercept (linear term) random regression solutions for the extra polygenic additive genetic effects of the $\mathbf{m^{th}}$ animal; $\boldsymbol{M_n}$ ($\boldsymbol{M_g}$) corresponded matrices of centered marker values for non-genotyped (genotyped) individuals; $\boldsymbol{J_1}$ and $\boldsymbol{J_2}$ were partitions of a fixed ($\boldsymbol{\beta}$) covariate effect that accounted for the difference in expected value of genetic merit between non-genotyped animals and genotyped animals, respectively; whereas $\boldsymbol{K_1}$ ($\boldsymbol{K_2}$) were partitions for non-genotyped (and genotyped) animals of an extra fixed ($\boldsymbol{\beta}$) covariate that accounted for the fact that the expectation of the $\boldsymbol{\alpha}$ (marker effects) was not zero (Fernando et al., 2014, 2016; Golden and Garrick, 2016). Posterior means and variances of breeding values for HPG at the average AFE were then averaged across the three chains and transformed to EPD expressed on a pseudo-probability scale as deviations from 50% as described by Speidel et al. (2018b).

Additionally, the number of times that each marker entered to the model during the sampling process was summed across the three chains in order to calculate its posterior probability of inclusion (PPI). The PPI was calculated through the division of the total number of times a marker was included in the model by the overall number of effective samples. Finally, according to the information contained in the manifest files of the original SNP arrays used to genotype the animals (UMD3.1.1 bovine assembly, Zimin et al., 2009), marker specific locations and their corresponding PPI were merged in order to conduct a genome-wide association analysis intended to identify relevant chromosomic regions related to HPG. Following the procedures reported by Pierce (2019), the five SNP with the highest PPI were considered as quantitative trait loci (QTL)

for which further exploration was performed. Specifically, once that the top five SNP were identified, their information was used to explore within the cattle QTL database (Cattle QTLdb; https://www.animalgenome.org/cgi-bin/QTLdb/BT/index) if similar QTL had been previously associated with beef cattle fertility traits. Furthermore, Ensembl genome database (Release 94, Zerbino et al., 2018) was utilized to search genes within one megabase of the QTL and the annotated gene located nearest to SNP was considered as a potential candidate gene.

*5.2.8 Genetic and genomic evaluations for heifer first-service conception rate*

The EPD estimations for heifer FSCR were performed using the same approach described for HPG. Two RRM evaluations were executed for each heifer FSCR dataset (DS-1_FSCR through DS-4_ FSCR). The first evaluation was a pedigree-based RRM like the one described in equation 3.9 of the materials and methods section of chapter 3. This evaluation was intended to obtain base genetic predictions for FSCR without including genomic information. Later, such base predictions were compared with the genomic predictions generated by the ssRR-SHM (second evaluation). In all models, FSCR was regressed on AFE applying a linear RRM that used Legendre polynomials as the base function. The model used in all the hybrid genomic evaluations is presented in matrix form in Equation 5.4 below:

$$\begin{bmatrix} \boldsymbol{y_n} \\ \boldsymbol{y_g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_n} \\ \boldsymbol{X_g} \end{bmatrix} \mathbf{b} + \begin{bmatrix} \boldsymbol{D^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D^1} \end{bmatrix} \begin{bmatrix} \boldsymbol{d^0} \\ \boldsymbol{d^1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_n^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_g^0 M_g} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_n^0} \\ \boldsymbol{\alpha^0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_n^1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_g^1 M_g} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_n^1} \\ \boldsymbol{\alpha^1} \end{bmatrix} + \boldsymbol{Q_m} + \boldsymbol{W_s} + \mathbf{e} \quad \text{Eq. 5.4}$$

where $\boldsymbol{y_n}$ and $\boldsymbol{y_g}$ corresponded to binary FSCR observations on non-genotyped (**n**) and genotyped (**g**) individuals; $\boldsymbol{X_n}$ and $\boldsymbol{X_g}$ were appropriate incidence matrices that related fixed effects of breeding contemporary group (defined as a combination between breeding year and semen type), AI technician, age of dam (BIF classes) and a linear fixed regression of FSCR on AFE contained in **b** to FSCR observations in ***y*** (specifically sorted with **n** individuals first and **g** individuals after); $\boldsymbol{D^0}$ and $\boldsymbol{D^1}$ were appropriate age covariates relating FSCR observations in ***y*** to the additive random

extra polygenic effects in $\boldsymbol{d^0}$ and $\boldsymbol{d^1}$; $\boldsymbol{Z_n^0}$ and $\boldsymbol{Z_n^1}$ corresponded to intercept and linear age covariates relating FSCR observations in $\boldsymbol{y}$ to animal additive direct genetic effects accounted by imputed marker values in $\boldsymbol{u_n^0}$ and $\boldsymbol{u_n^1}$, respectively [with $\boldsymbol{u_n = M_n\alpha + \epsilon}$ (where $\boldsymbol{\epsilon}$ = imputation residual)]; $\boldsymbol{Z_g^0}$ and $\boldsymbol{Z_g^1}$ represented intercept and linear age covariates relating FSCR observations in $\boldsymbol{y}$ to marker effects contained in $\boldsymbol{\alpha^0}$ and $\boldsymbol{\alpha^1}$, respectively; $\boldsymbol{Q}$ was an incidence matrix relating FSCR observations in $\boldsymbol{y}$ to a vector of unknown solutions of mating group (e.g., inseminated in heat or during a mass mate) random effects contained in $\boldsymbol{m}$; $\boldsymbol{W}$ was an incidence matrix relating FSCR observations in $\boldsymbol{y}$ to the vector of unknown solutions of service sire random effects contained in $\boldsymbol{s}$; **M** corresponded to a matrix of centered marker values (coded as -1, 0 or 1) and **e** represented a vector of random errors.

For these models, again the extra polygenic additive genetic effects terms ($\boldsymbol{d}$) were fit assuming ½ of the additive genetic variance was not captured by markers and it was assumed that these effects were uncorrelated to other random effects. Variances of random effects included in the ssRR-SHM for FSCR were assumed to be:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{d}\\ \boldsymbol{u_n}\\ \boldsymbol{\alpha}\\ \boldsymbol{m}\\ \boldsymbol{s}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G_d^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{G_u^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{G_\alpha^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_m\sigma_m^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_w\sigma_s^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

where $\boldsymbol{G_\alpha^{-1}}$, $\boldsymbol{G_u^{-1}}$ and $\boldsymbol{G_d^{-1}}$ remained as described for the ssRR-SHM for HPG, while $\boldsymbol{I_m}$, $\boldsymbol{I_w}$ and $\boldsymbol{I_n}$ were identity matrices whose orders were equal to the number of mating groups, AI sires and observations, respectively. The $\boldsymbol{\sigma_m^2}$, $\boldsymbol{\sigma_s^2}$ and $\boldsymbol{\sigma_e^2}$ were the mating group, AI sire and residual variances, respectively. Variance components used to feed all these parameters were the ones obtained with the heifer FSCR random regression analysis in chapter 3 (e.g., refer to Table 3.5).

Similar to the HPG evaluation, BOLT software was used to assemble the MME of this ssRR-SHM and the same strategy of first obtaining solutions with the PCG method and then execute an exhaustive MCMC-based sampling process (e.g., 315,000 iterations with 15,000 disregarded as a bun-in) was implemented to estimate EPD for FSCR at the average AFE (e.g., refer to equations 5.2 and 5.3). Correspondingly, the same approach taken to determine and explore QTL for HPG was applied to identify and investigate QTL for FSCR.

*5.2.9 Genetic and genomic evaluations for stayability*

In the case of STAY evaluations, base predictions for all datasets (DS-1_STAY through DS-4_STAY) were obtained by evaluating all STAY endpoints (STAY03 through STAY12) jointly using linear and pedigree-based RRM with Legendre polynomials as their base function. A detailed description of such models could be found in equation 4.2 of the Materials and Methods section of Chapter 4. In contrast, genomic predictions to which the base predictions were compared were obtained by implementing a ssRR-SHM similar to the one described by Golden et al. (2018a). The general model equation (Equation 5.5) used in all the hybrid genomic evaluations for STAY is shown in matrix form below:

$$\begin{bmatrix} \boldsymbol{y_n} \\ \boldsymbol{y_g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_n} \\ \boldsymbol{X_g} \end{bmatrix}\mathbf{b} + \begin{bmatrix} \boldsymbol{D^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D^1} \end{bmatrix}\begin{bmatrix} \boldsymbol{d^0} \\ \boldsymbol{d^1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_n^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_g^0 M_g} \end{bmatrix}\begin{bmatrix} \boldsymbol{u_n^0} \\ \boldsymbol{\alpha^0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_n^1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_g^1 M_g} \end{bmatrix}\begin{bmatrix} \boldsymbol{u_n^1} \\ \boldsymbol{\alpha^1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{W^0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W^1} \end{bmatrix}\begin{bmatrix} \boldsymbol{p^0} \\ \boldsymbol{p^1} \end{bmatrix} + \boldsymbol{I_{cg}} + \mathbf{e} \quad \text{Eq. 5.5}$$

where $\boldsymbol{y_n}$ and $\boldsymbol{y_g}$ corresponded to binary STAY observations on non-genotyped (**n**) and genotyped (**g**) individuals; $\boldsymbol{X_n}$ and $\boldsymbol{X_g}$ were incidence matrices relating STAY observations in $\boldsymbol{y}$ (explicitly sorted with **n** individuals first and **g** individuals after) to the fixed effects of age at first calving (defined in months and having 7 levels, e.g.; 21 to 27 mo), calving ease score of the immediate previous calving (4 levels), as well as a couple of fixed regression coefficients for breeding weight

and STAY observations on age at calving contained in **b**; $\mathbf{D^0}$ (intercept) and $\mathbf{D^1}$ (linear term) were age covariates relating STAY observations in $\boldsymbol{y}$ to the additive random extra polygenic effects contained in $\mathbf{d^0}$ and $\mathbf{d^1}$, respectively; $\boldsymbol{Z_n^0}$ and $\boldsymbol{Z_n^1}$ corresponded to intercept and linear age covariates relating STAY observations in $\boldsymbol{y}$ to animal additive direct genetic effects accounted by imputed marker values in $\boldsymbol{u_n^0}$ and $\boldsymbol{u_n^1}$, respectively [with $\boldsymbol{u_n = M_n\alpha + \epsilon}$ (where $\boldsymbol{\epsilon}$ = imputation residual)]; $\boldsymbol{Z_g^0}$ and $\boldsymbol{Z_g^1}$ represented intercept and linear age covariates relating STAY observations in $\boldsymbol{y}$ to marker effects contained in $\boldsymbol{\alpha^0}$ and $\boldsymbol{\alpha^1}$, respectively; $\mathbf{W^0}$ and $\mathbf{W^1}$ were intercept and linear age covariates relating STAY observations in $\boldsymbol{y}$ to the random permanent environmental effects contained in $\mathbf{p^0}$ and $\mathbf{p^1}$; $\boldsymbol{I}$ was a known incidence matrix relating STAY observations in $\boldsymbol{y}$ to their corresponding random contemporary group effects in **cg** (contemporary group was defined as a combination between breeding year and breeding pasture); **M** corresponded to a matrix of centered marker values (coded as -1, 0 or 1) and **e** represented a vector of random errors.

As described by Golden et al. (2018a), the extra polygenic additive genetic effects terms ($\boldsymbol{d}$) were fit assuming ½ of the additive genetic variance was not captured by markers and these effects were assumed to be uncorrelated to other random effects. Variances of the random effects included in all ssRR-SHM implemented for STAY evaluations were assumed to be:

$$\mathbf{Var}\begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{u_n} \\ \boldsymbol{\alpha} \\ \boldsymbol{p} \\ \boldsymbol{cg} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G_d^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G_u^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{G_\alpha^{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_p \otimes P_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_{cg}\sigma_{cg}^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{R} \end{bmatrix}$$

where $\boldsymbol{G_d^{-1}}$, $\boldsymbol{G_u^{-1}}$ and $\boldsymbol{G_\alpha^{-1}}$ remained as described for the previous ssRR-SHM applied for HPG and FSCR, however it is important to clarify that within STAY analyses, $\mathbf{G_0}$ corresponded to a (co)variance matrix of additive genetic random regression coefficients that did not required any

modification given the longitudinal nature of the trait under study; $\mathbf{P_0}$ was a (co)variance matrix of permanent environmental random regression coefficients; $\otimes$ was the Kronecker product; $\boldsymbol{I_p}$ and $\boldsymbol{I_{cg}}$ represented identity matrices whose order were equal to the number of observations and contemporary groups, respectively; $\boldsymbol{\sigma^2_{cg}}$ represented the variance associated to contemporary groups and, lastly, $\mathbf{R = diag\{\sigma^2_{e_k}\}}$ was a diagonal matrix of temporary environmental variances that themselves vary depending on the $k^{th}$ age endpoint. The same statistical software, solution and sampling methods as well as procedures to estimate EPD described for heifer fertility traits were implemented to obtain predictions for STAY at the age of 6 in all the STAY hybrid genomic evaluations. Concordantly, the same steps taken to determine and explore QTL for heifer fertility traits were adopted to identify and investigate QTL for STAY.

*5.3 Results and discussion*

*5.3.1 Comparison of heifer fertility traits genetic and genomic evaluations*

Phenotypic summary statistics for the final number of observations for each trait (e.g., HPG and/or FSCR) within each dataset tested in the heifer fertility analyses are shown in Table 5.1. In general, genotyped individuals showed slightly higher phenotypic averages for HPG and FSCR than their non-genotyped counterparts. This result was expected given the fact that the subset of genotyped animals represented the most recent generations of a population that has been subjected to selection pressure to improve female fertility for at least three decades (Doyle et al., 2000). Averages of raw phenotypes considering all the animals included in each dataset oscillated between 0.78 to 0.87 for HPG and between 0.28 to 0.49 for FSCR.

**Table 5.1.** Summary statistics per dataset and genotyping status of individuals within dataset for heifer pregnancy, first-service conception rate, age at first exposure and age of dam.

| Dataset | Subset | Item | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|---|
| DS-1_Trait[1] | Genotyped | HPG[5] | 1,884 | 0.87 | 0.3 | 0 | 1 |
| | | FSCR[6] | 1,884 | 0.49 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 1,884 | 418.3 | 21.9 | 347 | 467 |
| | | Age of dam (yr) | 1,879 | 5.2 | 3.1 | 2 | 13 |
| | Non-genotyped | HPG[5] | 2,450 | 0.82 | 0.4 | 0 | 1 |
| | | FSCR[6] | 2,450 | 0.44 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 2,450 | 425.0 | 20.0 | 349 | 479 |
| | | Age of dam (yr) | 2,257 | 4.6 | 2.5 | 2 | 13 |
| | All together | HPG[5] | 4,334 | 0.85 | 0.4 | 0 | 1 |
| | | FSCR[6] | 4,334 | 0.46 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 4,334 | 422.1 | 21.1 | 347 | 479 |
| | | Age of dam (yr) | 4,136 | 4.9 | 2.8 | 2 | 13 |
| DS-2_Trait[2] | Genotyped | HPG[5] | 1,264 | 0.82 | 0.4 | 0 | 1 |
| | | FSCR[6] | 1,264 | 0.31 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 1,264 | 417.8 | 22.6 | 347 | 466 |
| | | Age of dam (yr) | 1,259 | 5.3 | 3.1 | 2 | 13 |
| | Non-genotyped | HPG[5] | 1,773 | 0.76 | 0.4 | 0 | 1 |
| | | FSCR[6] | 1,773 | 0.26 | 0.4 | 0 | 1 |
| | | Age at first exposure (d) | 1,773 | 423.7 | 20.5 | 349 | 479 |
| | | Age of dam (yr) | 1,588 | 4.7 | 2.5 | 2 | 13 |
| | All together | HPG[5] | 3,037 | 0.78 | 0.4 | 0 | 1 |
| | | FSCR[6] | 3,037 | 0.28 | 0.4 | 0 | 1 |
| | | Age at first exposure (d) | 3,037 | 421.2 | 21.6 | 247 | 479 |
| | | Age of dam (yr) | 2,847 | 4.9 | 2.8 | 2 | 13 |
| DS-3_Trait[3] | Genotyped | HPG[5] | 1,671 | 0.86 | 0.3 | 0 | 1 |
| | | FSCR[6] | 1,671 | 0.44 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 1,671 | 417.8 | 22.2 | 347 | 466 |
| | | Age of dam (yr) | 1,666 | 5.2 | 3.0 | 2 | 13 |
| | Non-genotyped | HPG[5] | 2,450 | 0.82 | 0.4 | 0 | 1 |
| | | FSCR[6] | 2,450 | 0.44 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 2,450 | 425.0 | 20.0 | 349 | 479 |
| | | Age of dam (yr) | 2,257 | 4.6 | 2.5 | 2 | 13 |
| | All together | HPG[5] | 4,121 | 0.84 | 0.4 | 0 | 1 |
| | | FSCR[6] | 4,121 | 0.44 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 4,121 | 422.1 | 21.2 | 347 | 479 |
| | | Age of dam (yr) | 3,923 | 4.8 | 2.8 | 2 | 13 |
| DS-4_Trait[4] | Genotyped | HPG[5] | 1,884 | 0.87 | 0.3 | 0 | 1 |
| | | FSCR[6] | 1,884 | 0.49 | 0.5 | 0 | 1 |
| | | Age at first exposure (d) | 1,884 | 418.3 | 21.9 | 347 | 467 |
| | | Age of dam (yr) | 1,879 | 5.2 | 3.1 | 2 | 13 |

[1]DS-1_Trait = dataset 1 for HPG or FSCR; [2]DS-2_Trait = dataset 2 for HPG or FSCR; [3]DS-3_Trait = dataset 3 for HPG or FSCR; [4]DS-4_Trait = dataset 4 for HPG or FSCR; [5]HPG = Heifer pregnancy; [6]FSCR = First-service conception rate.

Smaller averages were found in the dataset where all phenotypic records belonging to replacement heifers were removed (DS-2_Trait). Naturally, since replacement heifers were selected because of their superior performance, the removal of their records caused a general decline in phenotypic averages for both HPG and FSCR. Conversely, the highest averages for both traits were found in the dataset that only included genotyped heifers (e.g., DS-4_Trait). Considering that genotyped heifers belong to the most recent generations within the CSU-BIC, the previous could be a reflection of the selection procedures applied in the Rouse Angus herd for several years.

Despite the differences observed in the average performance for HPG and FSCR among all datasets and between different subsets of animals within datasets (e.g., genotyped *vs* non-genotyped), mean performance of both traits studied were within the respective range of values reported in literature. For instance, Doyle et al. (2000) reported a range of HPG percentage averages between 76.6 to 95.7% within the same experimental Angus cattle population at the CSU-BIC when analyzing data collected between 1985 to 1993. More recently, average HPG rates of 0.76 and 0.77 were reported for Red Angus cattle by Speidel et al. (2018b) and Boldt et al. (2018), respectively. Also, Azzam et al. (1989) reported a range between 0.26 and 0.81 in FSCR for Simmental influenced heifers exposed to their first artificial insemination event at 1 or 1.5 yr of age. A more recent report in Angus heifers documented a 60% FSCR average when analyzing information from 6 different herds distributed across 5 states within the US (Bormann et al., 2006). No substantial differences were found among datasets or animal subsets with respect to relevant systematic effects such as age of dam and age at first exposure. The EPD summary statistics of the pedigree-based and the genomic evaluations for HPG are shown in Tables 5.2 and 5.3, respectively.

**Table 5.2.** Heifer pregnancy expected progeny differences (EPD; at 422 d of age) summary statistics obtained with pedigree-based random regression models.

| Dataset | Subset | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| DS-1_HPG[1] | Genotyped | 2,037 | 0.594 | 1.042 | -2.849 | 4.048 |
| | Non-genotyped | 4,736 | 0.153 | 0.944 | -4.301 | 4.361 |
| | All animals | 6,773 | 0.286 | 0.995 | -4.301 | 4.361 |
| DS-2_HPG[2] | Genotyped | 2,037 | 0.112 | 0.901 | -2.972 | 2.699 |
| | Non-genotyped | 4,701 | -0.040 | 0.827 | -3.519 | 3.217 |
| | All animals | 6,738 | 0.006 | 0.853 | -3.519 | 3.217 |
| DS-3_HPG[3] | Genotyped | 2,029 | 0.417 | 1.016 | -2.951 | 3.811 |
| | Non-genotyped | 4,735 | 0.167 | 0.927 | -4.257 | 4.167 |
| | All animals | 6,764 | 0.242 | 0.961 | -4.257 | 4.167 |
| DS-4_HPG[4] | Genotyped | 2,032 | 0.038 | 1.000 | -3.297 | 3.160 |
| | Non-genotyped | 1,902 | -0.037 | 0.435 | -3.011 | 1.842 |
| | All animals | 3,934 | 0.002 | 0.781 | -3.297 | 3.160 |

[1]DS-1_HPG = dataset 1 (4,334 heifers); [2]DS-2_HPG = dataset 2 (3,037 heifers); [3]DS-3_HPG = dataset 3 (4,121 heifers); [4]DS-4_HPG = dataset 4 (1,884 heifers).

**Table 5.3.** Heifer pregnancy expected progeny differences (EPD; at 422 d of age) summary statistics obtained with single-step random regression super-hybrid models.

| Dataset | Subset | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| DS-1_HPG[1] | Genotyped | 2,037 | 9.977 | 1.114 | -1.951 | 14.765 |
| | Non-genotyped | 4,736 | 5.683 | 3.666 | -7.126 | 15.770 |
| | All animals | 6,773 | 6.974 | 3.694 | -7.126 | 15.770 |
| DS-2_HPG[2] | Genotyped | 2,037 | 0.828 | 0.921 | -11.390 | 5.165 |
| | Non-genotyped | 4,701 | 0.436 | 1.160 | -4.896 | 5.095 |
| | All animals | 6,738 | 0.554 | 1.108 | -11.390 | 5.165 |
| DS-3_HPG[3] | Genotyped | 2,029 | 6.504 | 1.135 | -6.729 | 11.658 |
| | Non-genotyped | 4,735 | 3.804 | 2.676 | -7.136 | 11.255 |
| | All animals | 6,764 | 4.614 | 2.632 | -7.136 | 11.658 |
| DS-4_HPG[4] | Genotyped | 2,032 | 47.151 | 0.172 | 45.659 | 47.704 |
| | Non-genotyped | 1,902 | 28.718 | 17.646 | -0.496 | 49.894 |
| | All animals | 3,934 | 38.239 | 15.343 | -0.496 | 49.894 |

[1]DS-1_HPG = dataset 1 (4,334 heifers); [2]DS-2_HPG = dataset 2 (3,037 heifers); [3]DS-3_HPG = dataset 3 (4,121 heifers); [4]DS-4_HPG = dataset 4 (1,884 heifers).

Genetic predictions of HPG obtained with the pedigree-based RRM showed a stable and consistent behavior in all datasets. Specifically, the average EPD was in the middle of the respective range of EPD values and all ranges were similar regardless of the dataset or the subset

of animals being analyzed (Table 5.2). In every data structure scenario (DS-1_HPG through DS-4_HPG), EPD averages of genotyped animals were higher than the EPD of non-genotyped individuals; although such superiority was more marked in DS-1_HPG and DS-3_HPG due to the inclusion of phenotypic records of genotyped and non-genotyped heifer replacements in these two datasets. A slightly higher EPD average (e.g., 2.139) and spread in EPD values (e.g., -7.06 to 9.74) was reported by Speidel et al. (2018b) when analyzing HPG in Red Angus cattle through the application of random regression techniques. Differences among studies could be explained by differences in population size, since in Speidel et al. (2018b) a total of 2,625,287 animals were included in the analysis, whereas in the current study pedigrees sizes ranged between 3,934 to 6,773 individuals depending on the dataset used for the evaluation.

Genomic predictions for HPG obtained with the ssRR-SHM showed considerable higher means and larger ranges in EPD values (Table 5.3) in comparison to the corresponding base genetic predictions that did not included genomic information (Table 5.2). For all datasets and regardless of the genotyping status of the animals, average EPD were always positive, which suggested a general tendency of overprediction within the single-step hybrid genomic evaluations. In this regard, single-step genomic evaluations have been previously reported to over-predict differences in breeding values (Koivula et al., 2015; Mäntysaari et al., 2020). According to Tsuruta et al. (2019), possible reasons for inflations in genomic predictions of young genotyped animals are preselection and incompatibilities between pedigree-based (**A**) and genomic relationship matrices (**G**). Although the ssRR-SHM used in the present study completely avoided issues associated to the usage of a **G** matrix, the bias introduced by preselection of genotyped animals still inflated resulting genomic predictions.

Nordbø et al. (2019) described two types of bias arising in single-step genomic best linear unbiased prediction methods: (1) a general inflation of genomic breeding values that consequently yields a wider spread of the breeding value estimates; and (2) level-bias of breeding values, which influences the predicted genetic levels for groups of animals (e.g., overprediction of genotyped animals vs. non-genotyped animals). Both phenomena manifested in the current investigation; however, the magnitude in which they affected the resulting genomic predictions varied depending on the data structure of the phenotypic file. The most severe upward distortion in EPD occurred in DS-4_HPG which was the dataset that included phenotypic information of only the subset of genotyped heifers at the CSU-BIC. Within this dataset, the average HPG of the 1,884 heifers contributing phenotypes to the evaluation was considerably high (0.87). Furthermore, 589 heifers with successful observations for HPG were selected as replacements within the herd and they represented 30% of all dams in the pedigree (1,944 total dams). In this context, it has been reported that after selection it is expected to find a reduction in the genetic variance of the selected pool of animals, which could potentially lead to important overestimations of their genomic estimated breeding values (Schaeffer, 2014; Dehnavi et al., 2018). Although less marked, overestimations of the genetic merit of individuals for HPG also occurred in DS-1_HPG and DS-3_HPG, presumably due to the same reasons that in DS-1_HPG.

A much more reasonable and expected result was obtained for the genomic predictions yielded with the evaluation that used as phenotypic input the second dataset (DS-2_HPG). The removal of the phenotypes belonging to all heifers (genotyped and non-genotyped) that eventually became replacements in the herd, also removed the issues associated with the pre-selection bias. A similar result was reported by Koivula et al. (2016) in a study that evaluated the effects of including different sets of phenotypes and genotypes of elite females (e.g., dams previously

selected due to superior genetic merit) into the reference population in single-step evaluations. In this work, authors reported that the exclusion of information from elite dams was enough to overcome bias. For DS-2_HPG, results for the average and range in EPD were comparable to values reported in the literature. For instance, the mean genomic merit for conception rate in a group of Holstein heifers classified as highly fertile was 2.75 (ranging between 1.5 and 5.5), conversely, the average genomic merit for the same trait in another group of Holstein heifers classified as lowly fertile was 0.06, with values that oscillated between -2.1 and 1.2 (Veronese et al., 2019a, 2019b).

Pearson correlations, rank correlations and regression coefficients of HPG genomic predictions on HPG pedigree-based predictions are shown in Table 5.4. In all data structure scenarios, predictions of the subset of genotyped animals were highly correlated; however, similarities in predictions for non-genotyped animals were highly variable depending on the dataset studied. The most severe disparity among predictions for non-genotyped individuals was evident for DS-4_HPG, which could be attributed to the previously discussed problems associated to the pre-selection bias. Conversely, correlations among predictions for the genotyped animals within the same dataset resulted in the highest from all analyses. In this regard, it has been reported that although level-bias affects the average level of EBV between groups of animals, even when average level between groups could be wrong, the average ranking within groups might be correct (Nordbø et al., 2019).

Examining the results of all subsets of animals comprehensively, it became evident that genomic predictions obtained with DS-2_HPG were the most stable, since their level of concordance with its corresponding pedigree-based prediction excelled any other comparison from another dataset. Pearson correlations between base and genomic predictions obtained with DS-

2_HPG were even higher than those reported by Forni et al. (2011) when comparing pedigree-based EBV to genomic EBV for the total number of piglets born per litter using different genomic relationship matrices (such correlations ranged 0.791 to 0.891). Saatchi et al. (2018) also reported high correlations (e.g., 0.95) and close to 1 regression coefficients between single-step Bayesian regression EPD and conventional pedigree-based EPD for birth weight in a multi-breed beef cattle population. Such result agreed with the correlations and regression coefficients obtained for all subset of animals within DS-2_HPG in the current study and represented the only published work comparing pedigree-based predictions to genomic predictions using a hybrid Bayesian marker effects models. Given the notable superiority of the results obtained with the second dataset, the genome-wide association study for HPG was performed using results of this particular evaluation only. Additionally, the heifers that contributed with phenotypes in DS-2_HPG were ranked in quartiles according their genomic EPD and it was explored if such classification effectively translated into expressed differences in HPG phenotypic performance (Appendix B, Table B-2 and Figure B-1).

**Table 5.4.** Pearson correlation, rank correlation and regression coefficients of genomic predictions on pedigree-based predictions for heifer pregnancy.

| Dataset | Subset | Pearson correlation | Rank Correlation | Regression coefficient |
|---|---|---|---|---|
| DS-1_HPG[1] | Genotyped | 0.867 | 0.892 | 0.926 |
| | Non-genotyped | 0.490 | 0.527 | 1.905 |
| | All animals | 0.513 | 0.626 | 1.905 |
| DS-2_HPG[2] | Genotyped | 0.827 | 0.910 | 0.845 |
| | Non-genotyped | 0.927 | 0.893 | 1.300 |
| | All animals | 0.889 | 0.892 | 1.155 |
| DS-3_HPG[3] | Genotyped | 0.861 | 0.908 | 0.961 |
| | Non-genotyped | 0.611 | 0.615 | 1.763 |
| | All animals | 0.593 | 0.673 | 1.623 |
| DS-4_HPG[4] | Genotyped | 0.944 | 0.959 | 0.162 |
| | Non-genotyped | -0.103 | -0.101 | -4.205 |
| | All animals | 0.004 | 0.319 | 0.081 |

[1]DS-1_HPG = dataset 1 (4,334 heifers); [2]DS-2_HPG = dataset 2 (3,037 heifers); [3]DS-3_HPG = dataset 3 (4,121 heifers); [4]DS-4_HPG = dataset 4 (1,884 heifers).

With respect to the results of FSCR analyses, EPD summary statistics of the pedigree-based and the genomic evaluations for this trait are shown in Tables 5.5 and 5.6, respectively. Similar to HPG results, base genetic predictions for FSCR yielded by the pedigree-based RRM showed a consistent behavior regardless of the dataset implemented. Means EPD for all datasets were close to zero and all ranges were considerable small in amplitude, which could be considered normal and expected due to the low genetic variability of FSCR (Mu et al., 2016). Although no mean EPD for FSCR was found in literature, Bormann et al. (2006) reported a range of breeding values (in an observed scale) between -0.01 to 0.02 for this trait for Angus sires, a result that resembles to the results of the present study even when they are expressed in a pseudo-probability scale.

**Table 5.5.** First-service conception rate expected progeny differences (EPD; at 422 d of age) summary statistics obtained with pedigree-based random regression models.

| Dataset | Subset | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| DS-1_FSCR[1] | Genotyped | 2,037 | 0.044 | 0.321 | -0.982 | 1.071 |
| | Non-genotyped | 4,736 | -0.008 | 0.253 | -1.108 | 1.442 |
| | All animals | 6,773 | 0.007 | 0.276 | -1.108 | 1.442 |
| DS-2_FSCR[2] | Genotyped | 2,037 | -0.151 | 0.260 | -1.002 | 0.876 |
| | Non-genotyped | 4,701 | -0.116 | 0.210 | -0.975 | 0.979 |
| | All animals | 6,738 | -0.126 | 0.227 | -1.002 | 0.979 |
| DS-3_FSCR[3] | Genotyped | 2,029 | -0.011 | 0.390 | -1.154 | 0.959 |
| | Non-genotyped | 4,735 | -0.006 | 0.242 | -1.078 | 1.424 |
| | All animals | 6,764 | -0.008 | 0.268 | -1.154 | 1.424 |
| DS-4_FSCR[4] | Genotyped | 2,032 | 0.020 | 0.322 | -0.770 | 1.310 |
| | Non-genotyped | 1,902 | -0.018 | 0.119 | -0.793 | 0.750 |
| | All animals | 3,934 | 0.002 | 0.246 | -0.793 | 1.310 |

[1]DS-1_FSCR = dataset 1 (4,334 heifers); [2]DS-2_FSCR = dataset 2 (3,037 heifers); [3]DS-3_FSCR = dataset 3 (4,121 heifers); [4]DS-4_FSCR = dataset 4 (1,884 heifers).

**Table 5.6.** First-service conception rate expected progeny differences (EPD; at 422 d of age) summary statistics obtained with single-step random regression super-hybrid models.

| Dataset | Subset | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| DS-1_FSCR[1] | Genotyped | 2,037 | 6.727 | 0.395 | -0.536 | 8.750 |
| | Non-genotyped | 4,736 | 4.124 | 2.505 | -0.908 | 10.397 |
| | All animals | 6,773 | 4.907 | 2.421 | -0.908 | 10.397 |
| DS-2_FSCR[2] | Genotyped | 2,037 | -4.155 | 0.322 | -9.742 | 2.519 |
| | Non-genotyped | 4,701 | -2.593 | 1.571 | -6.569 | 0.933 |
| | All animals | 6,738 | -3.065 | 1.506 | -9.742 | 2.519 |
| DS-3_FSCR[3] | Genotyped | 2,029 | 2.259 | 0.301 | -1.659 | 3.515 |
| | Non-genotyped | 4,735 | 1.387 | 0.871 | -0.860 | 3.479 |
| | All animals | 6,764 | 1.648 | 0.847 | -1.659 | 3.515 |
| DS-4_FSCR[4] | Genotyped | 2,032 | -24.47 | 0.443 | -32.895 | -21.734 |
| | Non-genotyped | 1,902 | -13.16 | 9.229 | -35.760 | 0.164 |
| | All animals | 3,934 | -19.00 | 8.558 | -35.760 | 0.164 |

[1]DS-1_FSCR = dataset 1 (4,334 heifers); [2]DS-2_FSCR = dataset 2 (3,037 heifers); [3]DS-3_FSCR = dataset 3 (4,121 heifers); [4]DS-4_FSCR = dataset 4 (1,884 heifers).

Predictions obtained with the random regression hybrid marker effects models (Table 5.6) were much more variable than their counterparts yielded with pedigree-based RRM (Table 5.5). Apparently, an overvaluation of genomic information also occurred with the ssRR-SHM implemented to analyze FSCR, a phenomenon that seems to be a challenge of all single-step genomic evaluation procedures (Mäntysaari et al., 2020). Some reports have suggested that genomic breeding values of juvenile dairy bulls were inflated when compared to their actual daughter performance once phenotypic information was generated (Mäntysaari et al., 2010). The vast majority of genomic evaluations that have been published have been based on the usage of single-step procedures relying on the blending between pedigree (**A**) and genomic relationship (**G**) matrices (Legarra et al., 2009; Aguilar et al., 2010). Within this procedure, it has been acknowledged that incompatibilities between **A** and **G** matrices represented a critical source of inflation of resulting predictions (Misztal et al., 2017; 2020). Consequently, important research efforts have been undertaken in order to identify the most adequate scaling factors contributing to

improve the matching between both types of relationship matrices (Misztal et al., 2010; Harris et al., 2011; Tsuruta et al., 2011; Martini et al., 2018).

The genomic evaluation procedure implemented in the current study was not dependent upon the utilization of a genomic relationship matrix. In our single-step Bayesian regression model, the congruity between genomic and pedigree information was partially reached by fitting an additional fixed covariate (e.g., J equation) that accounted the possible difference in expected value of genetic merit between non-genotyped animals and genotyped animals (Fernando et al., 2014; Hsu et al., 2017; Misztal et al., 2020). Furthermore, an extra fixed covariate (e.g., K equation) was also included into all the hybrid marker effects models in order to prevent the Markov chain Monte Carlo procedure from diverging away from realistic values during the sampling process (BOLT software package, Release 1.2.7; http://www.thetasolutionsllc.com/bolt-software.html). Additionally, two extra random effects were also included in all Bayesian regression models to avoid the occurrence of biased results. First, given the imperfect nature of the SNP imputation process performed with non-genotyped animals, an imputation error term was explicitly fitted to the model. Also, recognizing that incomplete linkage disequilibrium exists between SNP markers and causal mutations responsible for variations in quantitative traits, an extra polygenic effect was included in the model to account for the variation not captured by markers (Liu et al., 2016; Golden et al., 2018a). Inclusion of all these effects into single-step genomic evaluations models have proved to reduce the inflation of genomic predictions (Liu et al., 2011; Gao et al., 2012; Su et al., 2014).

In spite of the inclusion of all the appropriated effects into the Bayesian regression models implemented in the present study to control the inflation of genomic predictions, such a problem still arose for FSCR (Table 5.7). Particularly for the subsets of non-genotyped animals and for all

animals considered together, the majority of the regression coefficients of genomic predictions on pedigree-based predictions were above 1. A similar situation was reported by Gao et al. (2018) whose application of single-step Bayesian regression models (ssBR) overestimated the genomic breeding value of three milk traits in dairy cattle; regardless of the fact that the ssBR did included the extra polygenic effect to account for the variance not explained by markers. Misztal et al. (2020) explained that a common problem leading to overestimation of genomic breeding values in all single-step methodologies, was an incomplete accounting of inbreeding due to missing pedigree connections or pedigree errors. Vitezica et al. (2011) suggested that the reason of the previous was that the knowledge of the genetic merit of some animals (e.g., parents) decreased the uncertainty (variance) of the genetic merit of their relatives (e.g., progeny). However, when such relationships exist but they are not correctly captured in the pedigree, the opposite occurs and an inflation of breeding values occurs in young animals. Although worth considering this source of overestimation, is unlikely that this problem originated the results of our study given the constant and cautious monitoring of the pedigree recording within the CSU-BIC (Crawford et al., 2016).

**Table 5.7.** Pearson correlation, rank correlation and regression coefficients of genomic predictions on pedigree-based predictions for heifer first-service conception rate.

| Dataset | Subset | Pearson correlation | Rank Correlation | Regression coefficient |
|---|---|---|---|---|
| DS-1_FSCR[1] | Genotyped | 0.570 | 0.699 | 0.701 |
| | Non-genotyped | 0.076 | 0.086 | 0.749 |
| | All animals | 0.125 | 0.250 | 1.099 |
| DS-2_FSCR[2] | Genotyped | 0.570 | 0.680 | 0.706 |
| | Non-genotyped | 0.452 | 0.517 | 3.380 |
| | All animals | 0.381 | 0.490 | 2.526 |
| DS-3_FSCR[3] | Genotyped | 0.733 | 0.782 | 0.691 |
| | Non-genotyped | 0.229 | 0.246 | 0.823 |
| | All animals | 0.238 | 0.360 | 0.755 |
| DS-4_FSCR[4] | Genotyped | 0.408 | 0.578 | 0.561 |
| | Non-genotyped | 0.186 | 0.266 | 14.445 |
| | All animals | 0.011 | 0.242 | 0.337 |

[1]DS-1_FSCR = dataset 1 (4,334 heifers); [2]DS-2_FSCR = dataset 2 (3,037 heifers); [3]DS-3_FSCR = dataset 3 (4,121 heifers); [4]DS-4_FSCR = dataset 4 (1,884 heifers).

Similar to the previously discussed for HPG, the most plausible reason for the overestimation of the genetic merit for FSCR seems to be the preselection bias of some of the genotyped females with phenotypic records. This became apparent when considering that the most reasonable correlations and regression coefficients were obtained for DS-3_FSCR; the dataset in where phenotypic records of heifers born before 2007 were removed due to the artificial inflation of FSCR averages caused by the reduced number of females represented in the data (e.g., <40 per year; Figure 5.9). Within this dataset, Pearson and rank correlations between pedigree-based RRM and ssRR-SHM genomic predictions were positive and strong (e.g., > 0.7) for the subpopulation of genotyped animals. This result agrees with a report by Wei et al. (2020) about correlations between regular breeding values and genomic breeding values ranging between 0.525 and 0.769 for wool traits in a subset of genotyped Merino sheep. In the case of the regression coefficients, the results for genotyped animals suggested that genomic predictions slightly underestimated EPD for FSCR in comparison to the pedigree-based predictions. Nonetheless, the strong correlations obtained for this group of animals allows to consider that even when the models produced numerically different predictions, the ranking of animals remained similar.

Pearson and rank correlations among predictions for non-genotyped animals obtained with the DS-3_FSCR were also positive, but they were considerably weak (e.g., <0.25). This result is perhaps a reflection of inaccuracies during the imputation process since the linear regression method used to calculate the number of copies of a particular allele in non-genotyped individuals, depends only on observed genotypes of close relatives (parents, offspring, siblings) and mates; but it is independent from the rest of the pedigree (Gengler et al., 2007). This implies that for the vast majority of non-genotyped heifers with phenotypic records (n = 2,450), their imputed genotypes were calculated using a limited amount of genotypic information from the small proportion of

genotyped animals born between 1997 and 2006. Additionally, heifers within the early years of data (1991 to 1996) may not had a close genotyped relative from which to obtain genotypic information for the imputation. Regression coefficients obtained with DS-3_FSCR suggested that the removal of phenotypic records from the 213 heifers with genotypes and phenotypes born between 1997 and 2006 contributed to preventing an overestimation of genomic breeding values for all subset of animals (only dataset with regression coefficients <1 for all groups). Consequently, the genome-wide association study for FSCR was performed using results of this particular evaluation only. Additionally, heifers contributing phenotypes within DS-3_FSCR dataset were ranked in quartiles according their genomic EPD to explore if such classification translated into expressed differences in FSCR phenotypic performance (Appendix B, Table B-3 and Figure B-2).

*5.3.2 Genome-wide association study for heifer pregnancy*

The location information of all the genetic markers utilized in this investigation was based on the UMD3.1.1 bovine assembly. Such bovine genome assembly has been criticized for having a variety of assembly errors, genome segmental inversions and not accurate chromosomal placements (Medrano, 2017). Therefore, it was opted to perform all GWAS analyzes using single SNP instead of marker windows. According to Speidel et al. (2018a), results for single SNP associations obtained from GWAS procedures should not change even with the advent of a new assembly, other than a possible refinement of the location of associated SNP within its respective chromosome. The five SNP that resulted with the highest PPI in the GWAS for HPG are shown in Table 5.8. Two of the five SNP were located intronic regions of the same gene (TMEM117) in chromosome 5, whereas the remaining three SNP were located on intergenic regions of chromosomes 3, 7 and 13. The corresponding Manhattan plot of the genome-wide screening for HPG is depicted in Figure 5.16.

**Table 5.8.** Single nucleotide polymorphisms (SNP) associated to heifer pregnancy in Angus cattle.

| SNP ID[a] | Chr[b] | Location[c] | PPI[d] | Gene | Gene location[e] |
|---|---|---|---|---|---|
| rs41585874 | 3 | 107,542,688 | 0.0156 | NDUFS5 | 21.51 kb |
| rs43119961 | 5 | 36,553,084 | 0.0165 | TMEM117 | Intron |
| rs43435407 | 5 | 36,578,127 | 0.0168 | TMEM117 | Intron |
| rs110232154 | 7 | 85,619,989 | 0.0161 | XRCC4 | 63.05 kb |
| rs109797421 | 13 | 24,860,333 | 0.0168 | OTUD1 | 203.67 kb |

[a]Reference SNP cluster identification assigned by the National Center for Biotechnology Information (NCBI).
[b]Chromosome in which the SNP was located within the *Bos taurus* UMD3.1.1 assembly.
[c]Chromosome position in *Bos taurus* UMD3.1.1.
[d]Posterior probability of inclusion in the model.
[e]Location of the SNP within a gene or distance to the closest annotated gene (in kilobases).

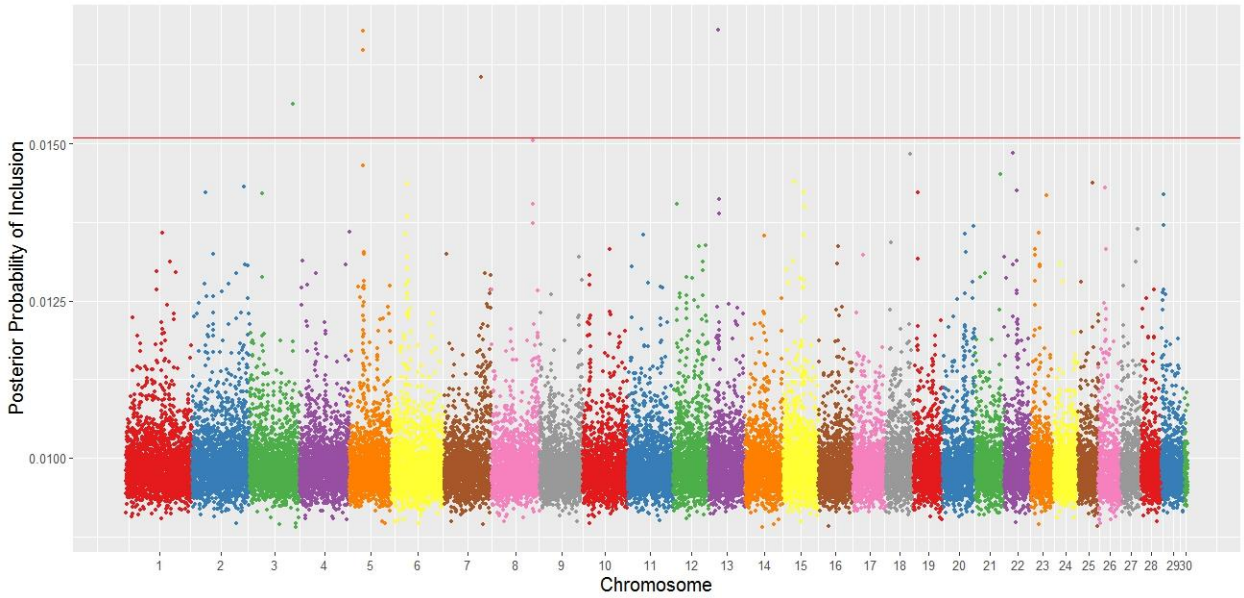


**Figure 5.16.** Manhattan plot identifying SNP associated with heifer pregnancy in Angus cattle (red horizontal line denotes a 1.51% posterior probability of inclusion).

In general, the PPI for all markers within the GWAS for HPG was low (e.g., <0.02). Similar results were found in a GWAS for the same trait performed in Red Angus cattle by Speidel et al. (2018a). Specifically, authors reported markers PPI ranging between 0.03 to 0.06 and suggested that the scarcity of HPG phenotypic records on genotyped heifers (e.g., 567 animals), as well as the low heritability assumed for the trait ($h^2$ = 0.12), were the main drivers of the results. In

accordance with these arguments, it has been suggested that the ability of Bayesian multiple-SNP regression models to detect strong associations between individual markers and the trait of interest depend on several factors, among which imperfect linkage disequilibrium between the marker and the causal mutation as well as the proportion of the variance in the trait explained by the real QTL stand out (Wolc et al., 2012, Garrick and Fernando, 2013; Pierce et al., 2020).

The SNP rs41585874 located in chromosome 3 was located within a distance of 24.51 kb from NDUFS5, a gene that codes for the subunit S5 of a proton-pumping enzyme named NADH-ubiquinone oxidoreductase that plays an important role within the mitochondrial respiratory chain (Murai et al., 2009). The gene NDUFS5 has been previously reported to be a member of a network generated through an Ingenuity Pathway Analysis (IPA) associated to endometrial-related conception rate of beef heifers (Killen et al., 2016). Interestingly, using the same molecular approach (IPA), this gene was found to be differentially expressed between days 7 and 13 of the estrous cycle in the endometrium of cross-bred beef heifers with normal circulating progesterone concentrations in comparison to heifers with low concentrations of progesterone (Forde et al., 2012). Additionally, it has been reported that NDUFS5 forms part of a group of genes that participate in biological processes related to energy pathways and mitochondrion organization that decreased their abundance during an *in vitro* oocyte maturation process (Reyes et al., 2015).

The two genetic markers that were located within intronic regions of the TMEM117 gene were rs43119961 and rs43435407. This gene is located at BTA5 and encodes for a multi-pass transmembrane protein (transmembrane protein 117) that has recently been identified as a mediator of an endoplasmic reticulum stress-induced cell death pathway (Tamaki et al., 2017). Previous research efforts have reported associations between TMEM117 and important traits in cattle; for instance, Veerkamp et al. (2012) reported that a SNP within this gene had a large effect

on the body condition scores of Holstein cows. Waters et al. (2014) also informed that TMEM117 was a differentially expressed gene involved in biological processes influenced by n-3-polyunsaturated fatty acids supplementation in the bovine endometrium. Furthermore, a GWAS identified that polymorphisms within this gene were associated to the muscle fatty acid composition in Simmental cattle (Zhu et al., 2017). More recently, TMEM117 was identified as a gene that has responded to the selection pressure applied in Gir cattle to improve its milk production ability (Maiorano et al., 2018) and in Nelore cattle to increase its reproductive performance (Montes et al., 2019). Considering all these reports, it is possible to hypothesize that the relationship between TMEM117 and HPG may be given by an involvement of this gene in the correct growth and development of beef heifers that allow them to reach an appropriate level of adiposity crucial for a reproductive success.

Furthermore, possibilities of indirect associations via linkage disequilibrium of TMEM117 with other genes also exist since this gene is located within a large QTL (between 5 and 80 Mb) on BTA5 that has been extensively studied and linked to various fertility traits in cattle. For instance, genes located between 1 to 11 Mb downstream from TMEM117 in BTA5 have been reported to be associated with twinning and ovulation rate (Kappes et al., 2000; Allan et al., 2009; Kim et al., 2009). Luna-Nevárez et al. (2011) reported that a SNP in the STAT2 gene (located ~20.4 Mb upstream from TMEM117) was associated to rebreeding traits such as calving interval and days to calving in beef heifers. Leyva-Corona et al. (2018) informed about the association of polymorphisms within the IGF1 and PMCH genes (located between 29.4 and 29.6 Mb upstream from TMEM117) with the number of services per conception in heat-stressed Holstein cows. Additionally, it has been recently suggested that genes within this genomic region in BTA5 possess pleiotropic effects on several traits affecting reproduction in cattle (Fernández et al., 2019).

Although positioned in a non-coding region of BTA7, the SNP rs110232154 was located within a distance of 63.05 kb from the gene XRCC4 that codes a DNA repair protein (Li et al., 1995). According to Barreta et al. (2012), DNA damage during early phases of embryo development represent one of the most powerful blockers of the cellular division process that culminates triggering the conceptus apoptosis. In this regard, XRCC4 has been reported to be within the main genes responsible for controlling the DNA repairing actions in human oocytes and blastocysts (Jaroudi et al., 2009). In cattle, variants within this gene have been linked with an increased resistance to paratuberculosis, a disease characterized by an overall reduction in productive and reproductive performance of infected animals (Pant et al., 2010; Brito et al., 2017). Additionally, a whole-genome association study implemented with a single-step methodology performed in Nelore cattle, identified XRCC4 as a putative candidate gene related to the fatty acid profile of the *longissimus thoracis* muscle (Lemos et al., 2016).

The last SNP associated with HPG in the current study was rs109797421 and was located at BTA13 at approximately 203.67 kb of distance from the gene OTUD1. The enzyme coded by this gene belongs to the ovarian tumor subfamily of deubiquitinases whose primary job is to remove posttranslational modifications (mostly lysine residues), that regulate cellular processes like transcription, translation and DNA damage response (Komander and Rape, 2012; Mevissen et al., 2013). From a reproductive standpoint, Sbardella (2020) reported that the chromosomic region in which the gene OTUD1 lies was associated with the ability of Nelore heifers to have early calvings ($\leq$ 30 mo). In addition, this gene has been found to be differentially expressed in Nelore steers with divergent residual feed intakes (Tizioto et al., 2016) and also in muscle samples collected from Nelore animals with dissimilar marbling scores (Fonseca et al., 2020).

*5.3.3 Genome-wide association study for heifer first-service conception rate*

The five SNP with the highest PPI in the GWAS for heifer FSCR are shown in Table 5.9. For this trait, all SNP were located on non-coding regions of BTA6, 7, 12 and 13. The corresponding Manhattan plot of the whole-genome study for FSCR is in Figure 5.17.

**Table 5.9.** Single nucleotide polymorphisms (SNP) associated to Heifer first-service conception rate in Angus cattle.

| SNP ID[a] | Chr[b] | Location[c] | PPI[d] | Gene | Gene location[e] |
|---|---|---|---|---|---|
| rs41615514 | 6 | 12,412,107 | 0.0123 | UGT8 | 210.63 kb |
| rs109154069 | 6 | 17,282,916 | 0.0122 | SEC24B | 224.10 kb |
| rs110596313 | 7 | 49,715,020 | 0.0117 | SPOCK1 | 246.39 kb |
| rs110013823 | 12 | 860,453 | 0.0118 | TDRD3 | 666.55 kb |
| rs110788468 | 18 | 65,604,707 | 0.0119 | A1BG | 221.91 kb |

[a]Reference SNP cluster identification assigned by the National Center for Biotechnology Information (NCBI).
[b]Chromosome in which the SNP was located within the *Bos taurus* UMD3.1.1 assembly.
[c]Chromosome position in *Bos taurus* UMD3.1.1.
[d]Posterior probability of inclusion in the model.
[e]Location of the SNP within a gene or distance to the closest annotated gene (in kilobases).

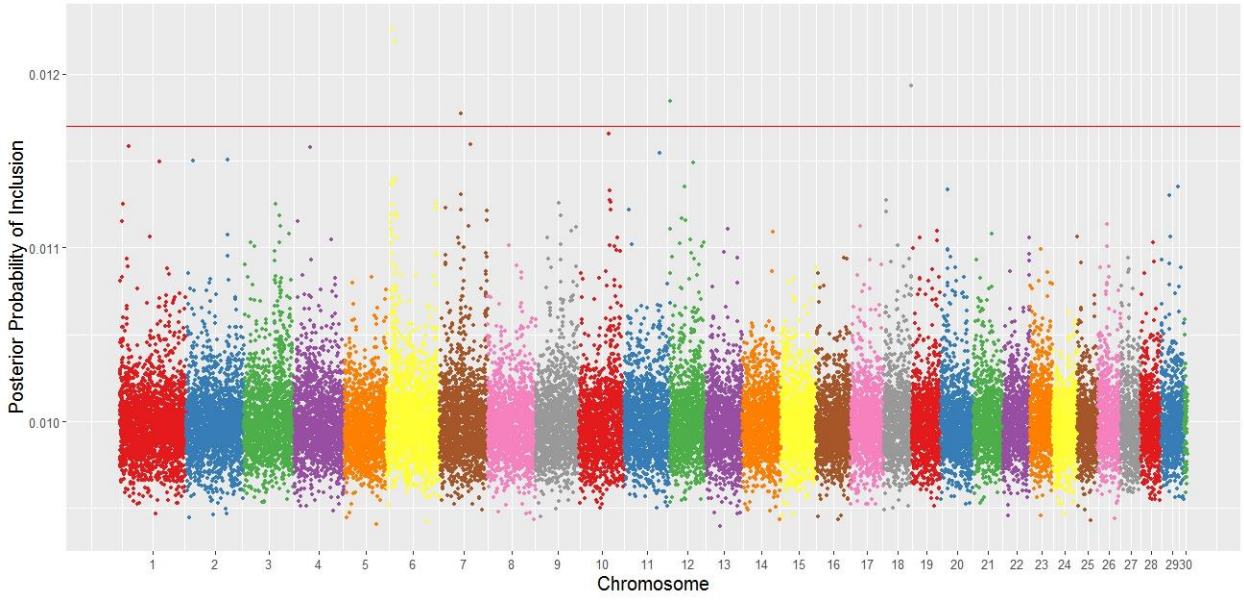


**Figure 5.17.** Manhattan plot identifying SNP associated with heifer first-service conception rate in Angus cattle (red horizontal line denotes a 1.16% posterior probability of inclusion).

In BTA6 there were two SNP representing the maximum peaks within the genome-wide inspection of markers associated to FSCR, such markers were rs41615514 and rs109154069. The first marker was located in an intergenic position at approximately 210.63 kb from the gene UGT8 that encodes for an enzyme known as UDP glycosyltransferase 8. The UGT family of enzymes has been proposed as an important modulator of the levels and the actions steroid hormones like estrogens, androgens, and progesterone (Hum et al., 1999; Meech et al., 2019). The previous could help to understand the relationship between variants within UGT8 and differences in phenotypic performance in FSCR, since this trait is heavily dependent on the action of steroid hormones. Interestingly, a study that investigated the transcriptomic profiles of peripheral white blood cells at the time of artificial insemination in beef heifers, found that UGT8 was a differentially expressed gene associated to pregnancies originated from AI or natural mating. Particularly, this gene was overexpressed in pregnancies originated thru natural service in comparison to AI-originated pregnancies (Dickinson et al., 2018).

The second SNP within BTA6 was rs109154069 and it was positioned 224.10 kb upstream from the gene SEC24B (SEC24 homolog B, COPII coat complex component). According to Zapaterra (2017), this gene was involved in the regulation of cholesterol biosynthesis and with the metabolism of lipids and lipoproteins; therefore, similar to UGT8, its relationship with phenotypic variations in FSCR may be given through its connection with steroid hormones. Furthermore, Reyes et al. (2015) reported that SEC24B was part of a conglomerate of genes involved in intracellular protein transport activities whose transcripts decreased in abundance during an *in vitro* maturation process of bovine oocytes. Additionally, a GWAS executed to identify genetic variants associated to the sperm membrane integrity in frozen-thawed semen of dairy bulls,

showed that SEC24B was one of the most promising candidate genes associated to variations sperm quality (Kamiński et al., 2016).

The SNP rs110596313 was located at approximately 246.39 kb from the gene SPOCK1 in BTA7. This gene codes for a proteoglycan protein isolated from testes tissue samples; however, GWAS in human-based investigations have linked SPOCK1 as a key gene underlying age at menarche (Liu et al., 2009). This is biologically relevant since FSCR is a trait that greatly depends on age at puberty and the initiation of estrous cycles of heifers by the moment when they are subjected to their first AI event. In support of the previous, Fortes and colleagues identified SPOCK1 as an important candidate gene related to age at puberty in beef cattle using a systems biology approach known as association weight matrix (Fortes et al., 2010). Additionally, this gene has also been contemplated as possibly related to an increased resistance to paratuberculosis, which as was previously discussed, is a disease that cause a severe reproductive underperformance (Brito et al., 2017).

The genetic marker found in the non-coding region of BTA12 (rs110013823) was placed at 666.55 kb from the gene TDRD3 (tudor domain containing 3). Tudor domain-containing proteins are believed to function as RNA binding proteins required for embryonic development. In mature animals, these proteins seem to function in only a restricted range of secretory exocrine and endocrine organs as the mammary gland, anterior pituitary, corpus luteum, ovaries and placenta (Broadhurst et al., 2005). Valour et al. (2014) reported that TDRD3 was a differentially expressed gene related to lipid metabolism activities of 18-d-old bovine embryos. Interestingly, authors of such study identified that transcripts of this gene were higher in embryos produced by growing heifers than in embryos produced by cows. Mahdipour (2015) also reported that this gene was constantly expressed at various maturation stages of bovine oocytes and that its expression

even increased in 4- and 8-cell embryos. Moreover, in humans it has been reported that genetic variants within TDRD3 were associated with the number of punctured ovarian follicles and the oocytes retrieved from women undergoing in vitro fertilization treatments (Laisk-Podar et al., 2015).

The last polymorphism identified in this study as potentially associated with heifer FSCR was rs110788468. This SNP was positioned at 221.91 kb of distance from the gen A1BG (alpha-1B glycoprotein) in BTA18. Although the function of the alpha-1B-glycoprotein has not been fully described, it is known that it belongs to an immunoglobulin family that commonly circulates in the blood plasma (Chalupnik et al., 2016). Recently and thru a multi-OMICS approach, the A1GB protein was described as an endocrine biomarker associated to the response to high-altitude hypoxia in cattle, since it was found to be differentially expressed Holstein cows managed at 3,000 m above the sea level (Kong et al., 2019). This is relevant to our study, since all heifers of this investigation were born and raised at a research herd located at an elevation ranging from 2,150 to 2,411 m that has a breeding program focused on fertility, maternal ability, early growth and adaptability of high altitudes (Doyle et al., 2000; Crawford et al., 2016; Pierce et al., 2020). Possibly, the combined selection pressures for fertility and high-altitude adaptability applied to this Angus population have led to the development of heifers with a better adaptability to hypoxic environments in which their fertility excel. Further support to such hypothesis was based on studies reporting that the A1BG protein is commonly expressed in bovine conceptus fluids since the A1GB gene is a major regulator of the gene networks existent between the oocyte and the surrounding cumulus cells (Riding et al., 2008; Biase and Kimble et al., 2018).

*5.3.4 Comparison of genetic and genomic evaluations for stayability*

Considering that only females that calve at the age of 2 years are retained at the CSU-BIC, all cows kept in the breeding herd were grouped by their year of birth and then according to their calving records, it was determined how many of them received a successful observation at each of the age endpoints contemplated in this study (Table 5.10). The same procedure was followed for genotyped (Table 5.11) and non-genotyped (Table 5.12) cows. Naturally, the number of cows that was able to remain in the herd decreased as the specific age endpoint increased. Considering all females (regardless of their genotyping status), 13.89% of the cows within the data were able to reach 12 yr of age. Nonetheless, splitting the animals according to their genotyping status, the percentage of females reaching the maximum age in the study was higher for genotyped (16.53%) than for non-genotyped (11.84%) cows. Given the age of 6 has been considered as a financial breakeven cow age within the US beef industry (Snelling et al., 1995; Brigham et al., 2006), a special emphasis was placed in this age endpoint (STAY06). Phenotypic summary statistics specific for STAY06 and according to the final number of observations available within each dataset tested in the stayability analyses (DS-1_STAY through DS-4_STAY) are shown in Table 5.13.

**Table 5.10.** Number of cows from the Colorado State University Beef Improvement Center that received a successful observation at each of the age endpoints contemplated in the study.

| | | Age endpoint | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Year of birth** | **Number of cows (2-yr-old)** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| 1990 | 79 | 79 | 76 | 67 | 59 | 47 | 41 | 40 | 32 | 20 | 16 |
| 1991 | 71 | 63 | 54 | 46 | 39 | 34 | 30 | 27 | 22 | 16 | 7 |
| 1992 | 93 | 75 | 66 | 46 | 43 | 31 | 29 | 28 | 23 | 21 | 18 |
| 1993 | 65 | 57 | 38 | 33 | 30 | 30 | 25 | 21 | 19 | 17 | 14 |
| 1994 | 70 | 47 | 43 | 39 | 35 | 30 | 27 | 26 | 23 | 19 | 14 |
| 1995 | 64 | 56 | 48 | 38 | 33 | 30 | 27 | 23 | 18 | 15 | 10 |
| 1996 | 61 | 55 | 44 | 38 | 33 | 31 | 29 | 26 | 21 | 15 | 9 |
| 1997 | 73 | 61 | 56 | 49 | 42 | 36 | 31 | 28 | 24 | 18 | 13 |
| 1998 | 80 | 67 | 61 | 56 | 52 | 43 | 41 | 33 | 29 | 23 | 18 |
| 1999 | 53 | 47 | 43 | 38 | 31 | 26 | 22 | 22 | 19 | 14 | 9 |
| 2000 | 49 | 44 | 43 | 39 | 36 | 33 | 30 | 28 | 22 | 16 | 11 |
| 2001 | 63 | 59 | 53 | 50 | 49 | 45 | 42 | 41 | 37 | 33 | 24 |
| 2002 | 50 | 41 | 33 | 32 | 29 | 23 | 20 | 18 | 15 | 14 | 10 |
| 2003 | 49 | 40 | 38 | 36 | 33 | 29 | 28 | 25 | 21 | 18 | 13 |
| 2004 | 44 | 42 | 38 | 38 | 34 | 27 | 25 | 22 | 18 | 15 | 13 |
| 2005 | 81 | 67 | 58 | 51 | 45 | 40 | 39 | 30 | 26 | 19 | 13 |
| 2006 | 61 | 54 | 50 | 43 | 36 | 32 | 23 | 22 | 17 | 16 | 14 |
| 2007 | 59 | 56 | 48 | 42 | 40 | 30 | 29 | 29 | 23 | 18 | 12 |
| 2008 | 61 | 49 | 40 | 37 | 30 | 30 | 28 | 26 | 22 | 14 | . |
| 2009 | 70 | 59 | 58 | 41 | 38 | 35 | 30 | 27 | 20 | . | . |
| 2010 | 61 | 57 | 41 | 39 | 31 | 29 | 26 | 22 | . | . | . |
| 2011 | 40 | 31 | 26 | 25 | 23 | 21 | 18 | . | . | . | . |
| 2012 | 80 | 74 | 67 | 58 | 51 | 37 | . | . | . | . | . |
| 2013 | 67 | 64 | 53 | 52 | 40 | . | . | . | . | . | . |
| 2014 | 58 | 48 | 41 | 33 | . | . | . | . | . | . | . |
| 2015 | 60 | 47 | 39 | . | . | . | . | . | . | . | . |
| 2016 | 51 | 41 | . | . | . | . | . | . | . | . | . |
| Total | 1713 | 1480 | 1255 | 1066 | 912 | 749 | 640 | 564 | 451 | 341 | 238 |

**Table 5.11.** Number of genotyped cows from the Colorado State University Beef Improvement Center that received a successful observation at each of the age endpoints contemplated in the study.

| | | Age endpoint | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Year of birth** | **Number of cows (2-yr-old)** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| 1990 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1991 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1992 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1993 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1994 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1995 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1996 | 0 | . | . | . | . | . | . | . | . | . | . |
| 1997 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 1998 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| 1999 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 9 |
| 2000 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 11 |
| 2001 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 30 | 24 |
| 2002 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 16 | 14 | 13 | 9 |
| 2003 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 21 | 19 | 17 | 12 |
| 2004 | 26 | 26 | 26 | 26 | 26 | 25 | 24 | 21 | 17 | 14 | 12 |
| 2005 | 39 | 39 | 39 | 39 | 39 | 35 | 34 | 27 | 25 | 18 | 12 |
| 2006 | 33 | 33 | 33 | 33 | 28 | 26 | 17 | 16 | 12 | 12 | 10 |
| 2007 | 37 | 36 | 36 | 31 | 31 | 23 | 22 | 22 | 18 | 14 | 9 |
| 2008 | 43 | 42 | 34 | 31 | 24 | 24 | 22 | 20 | 16 | 10 | . |
| 2009 | 67 | 57 | 56 | 40 | 37 | 34 | 29 | 26 | 19 | . | . |
| 2010 | 61 | 57 | 41 | 39 | 31 | 29 | 26 | 22 | . | . | . |
| 2011 | 37 | 30 | 25 | 24 | 22 | 20 | 17 | . | . | . | . |
| 2012 | 72 | 66 | 59 | 50 | 43 | 32 | . | . | . | . | . |
| 2013 | 66 | 63 | 52 | 51 | 40 | . | . | . | . | . | . |
| 2014 | 43 | 35 | 31 | 24 | . | . | . | . | . | . | . |
| 2015 | 60 | 47 | 39 | . | . | . | . | . | . | . | . |
| 2016 | 51 | 41 | . | . | . | . | . | . | . | . | . |
| Total | 750 | 687 | 586 | 503 | 436 | 363 | 306 | 265 | 214 | 169 | 124 |

**Table 5.12.** Number of non-genotyped cows from the Colorado State University Beef Improvement Center that received a successful observation at each of the age endpoints contemplated in the study.

| | | Age endpoint | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Year of birth** | **Number of cows (2-yr-old)** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| 1990 | 79 | 79 | 76 | 67 | 59 | 47 | 41 | 40 | 32 | 20 | 16 |
| 1991 | 71 | 63 | 54 | 46 | 39 | 34 | 30 | 27 | 22 | 16 | 7 |
| 1992 | 93 | 75 | 66 | 46 | 43 | 31 | 29 | 28 | 23 | 21 | 18 |
| 1993 | 65 | 57 | 38 | 33 | 30 | 30 | 25 | 21 | 19 | 17 | 14 |
| 1994 | 70 | 47 | 43 | 39 | 35 | 30 | 27 | 26 | 23 | 19 | 14 |
| 1995 | 64 | 56 | 48 | 38 | 33 | 30 | 27 | 23 | 18 | 15 | 10 |
| 1996 | 61 | 55 | 44 | 38 | 33 | 31 | 29 | 26 | 21 | 15 | 9 |
| 1997 | 67 | 55 | 50 | 43 | 36 | 30 | 25 | 22 | 18 | 12 | 7 |
| 1998 | 70 | 57 | 51 | 46 | 42 | 33 | 31 | 23 | 19 | 13 | 8 |
| 1999 | 43 | 37 | 33 | 28 | 21 | 16 | 12 | 12 | 9 | 4 | 0 |
| 2000 | 34 | 29 | 28 | 24 | 21 | 18 | 15 | 13 | 7 | 1 | 0 |
| 2001 | 30 | 26 | 20 | 17 | 16 | 12 | 9 | 8 | 4 | 3 | 0 |
| 2002 | 33 | 24 | 16 | 15 | 12 | 6 | 3 | 2 | 1 | 1 | 1 |
| 2003 | 25 | 16 | 14 | 12 | 9 | 5 | 4 | 4 | 2 | 1 | 1 |
| 2004 | 18 | 16 | 12 | 12 | 8 | 2 | 1 | 1 | 1 | 1 | 1 |
| 2005 | 42 | 28 | 19 | 12 | 6 | 5 | 5 | 3 | 1 | 1 | 1 |
| 2006 | 28 | 21 | 17 | 10 | 8 | 6 | 6 | 6 | 5 | 4 | 4 |
| 2007 | 22 | 20 | 12 | 11 | 9 | 7 | 7 | 7 | 5 | 4 | 3 |
| 2008 | 18 | 7 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 4 | . |
| 2009 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | . | . |
| 2010 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | . | . |
| 2011 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | . | . | . | . |
| 2012 | 8 | 8 | 8 | 8 | 8 | 5 | . | . | . | . | . |
| 2013 | 1 | 1 | 1 | 1 | 0 | . | . | . | . | . | . |
| 2014 | 15 | 13 | 10 | 9 | . | . | . | . | . | . | . |
| 2015 | 0 | . | . | . | . | . | . | . | . | . | . |
| 2016 | 0 | . | . | . | . | . | . | . | . | . | . |
| Total | 963 | 793 | 669 | 563 | 476 | 386 | 334 | 299 | 237 | 172 | 114 |

**Table 5.13.** Summary statistics per dataset and genotyping status of individuals within dataset that reached the stayability at the age of 6 endpoint.

| Dataset | Subset | Item | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|---|
| DS-1_STAY | Genotyped | STAY06 (%) | 479 | 91.02 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 479 | 23.7 | 0.9 | 21 | 27 |
| | | Breeding weight (lbs) | 38 | 1,244.82 | 132.9 | 1,050 | 1,630 |
| | Non-genotyped | STAY06 (%) | 554 | 85.92 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 554 | 23.9 | 0.7 | 22 | 26 |
| | | Breeding weight (lbs) | 335 | 1,248.6 | 127.1 | 851 | 1,677 |
| | All together | STAY06 (%) | 1,033 | 88.29 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 1,033 | 23.8 | 0.8 | 21 | 27 |
| | | Breeding weight (lbs) | 373 | 1,248.2 | 127.5 | 851 | 1,677 |
| DS-2_STAY | Genotyped | STAY06 (%) | 299 | 85.62 | 0.4 | 0 | 1 |
| | | Age at first calving (mo) | 299 | 23.7 | 0.9 | 21 | 27 |
| | | Breeding weight (lbs) | 23 | 1,216.5 | 115.2 | 1,050 | 1,488 |
| | Non-genotyped | STAY06 (%) | 554 | 85.92 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 554 | 23.9 | 0.7 | 22 | 26 |
| | | Breeding weight (lbs) | 335 | 1,248.6 | 127.1 | 851 | 1,677 |
| | All together | STAY06 (%) | 853 | 85.81 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 853 | 23.8 | 0.8 | 21 | 27 |
| | | Breeding weight (lbs) | 358 | 1,246.5 | 126.4 | 851 | 1,677 |
| DS-3_STAY | Genotyped | STAY06 (%) | 479 | 91.02 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 479 | 23.7 | 0.9 | 21 | 27 |
| | | Breeding weight (lbs) | 38 | 1,244.8 | 132.9 | 1,050 | 1,630 |
| DS-4_STAY | Genotyped | STAY06 (%) | 299 | 85.62 | 0.4 | 0 | 1 |
| | | Age at first calving (mo) | 299 | 23.7 | 0.9 | 21 | 27 |
| | | Breeding weight (lbs) | 23 | 1,216.5 | 115.2 | 1,050 | 1,488 |
| | Non-genotyped | STAY06 (%) | 38 | 86.84 | 0.3 | 0 | 1 |
| | | Age at first calving (mo) | 38 | 23.9 | 0.9 | 22 | 26 |
| | | Breeding weight (lbs) | 1 | 1,172.0 | - | - | - |
| | All together | STAY06 (%) | 337 | 85.76 | 0.4 | 0 | 1 |
| | | Age at first calving (mo) | 337 | 23.7 | 0.9 | 21 | 27 |
| | | Breeding weight (lbs) | 24 | 1,214.6 | 113.0 | 1,050 | 1,488 |

[1]DS-1_STAY = dataset 1 (1,713 cows); [2]DS-2_STAY = dataset 2 (1,533 cows); [3]DS-3_STAY = dataset 3 (750 cows); [4]DS-4_STAY = dataset 4 (668 cows).

Depending on the dataset used for the analysis, the number of females that produced 4 consecutive calves (successful observations from STAY02 through STAY05) and therefore had an opportunity to manifest a fifth calving record associated to STAY06 varied considerably. For

instance, in DS-1_STAY a total of 1,033 cows (479 genotyped and 554 non-genotyped) were still present at the herd at the age of 6, whereas, the number of animals reaching the STAY06 endpoint decreased to 853, 479 and 337 for DS-2_STAY, DS-3_STAY and DS-4_STAY, respectively. Although the averages for STAY06 shown in Table 5.13 seem high (>85%) for all datasets and subsets of animals, it should be noted that these values were calculated with the phenotypic information of females that were still present in the herd at the age of 6. In other words, these averages do not reflect the percentage of females that were culled before this age endpoint. When considering females that failed to produce a calf before the age of 6, the overall percentage of females that remained in the herd up to the STAY06 endpoint was 53.2%. Interestingly, when cows were classified according to their genotyping status, the percentage of cows remaining in the herd until the age of 6 was higher for genotyped (58.1%) than for non-genotyped (49.4%) cows (Figure 5.18). All these values, fall within the range of 38 to 60% success for STAY06 reported in literature for various beef cattle breeds (Snelling et al., 1995; Brigham et al., 2007; Engle et al., 2018).

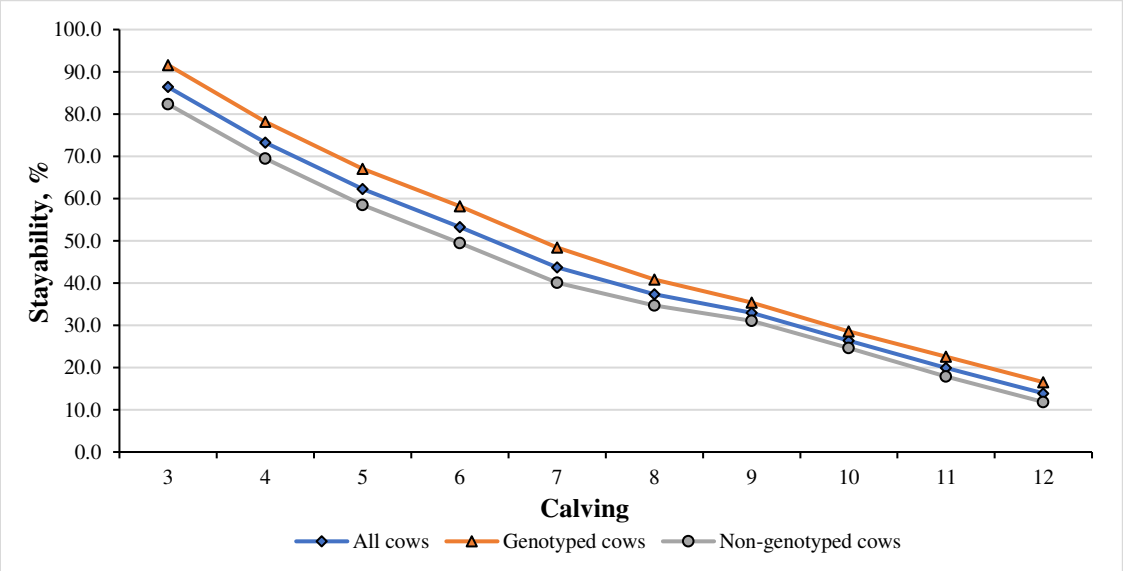


**Figure 5.18**. Average stayability (%) to consecutive calvings for genotyped, non-genotyped and overall cows from the Colorado State University Beef Improvement Center.

Comparing the phenotypic trends for STAY at consecutive calvings for genotyped and non-genotyped animals, the subset of cows with genotypic information had greater percentages of individuals remaining in the herd in all age endpoints (Figure 5.18). Given the subpopulation of genotyped cows was formed by females from the last generations within the herd, these results could be considered as further evidence of a successful selection pressure imposed to the Angus population of the CSU-BIC intended to increase female fertility (Snelling et al., 1995; Doyle et al., 2000). The EPD summary statistics of the pedigree-based genetic evaluations for STAY06 are shown in Table 5.14.

**Table 5.14.** Stayability at the age of 6 expected progeny differences (EPD) summary statistics obtained with pedigree-based random regression models.

| Dataset | Subset | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| DS-1_STAY[1] | Genotyped | 882 | 3.389 | 3.841 | -8.077 | 12.116 |
| | Non-genotyped | 2,687 | -2.528 | 4.764 | -18.104 | 13.435 |
| | All animals | 3,569 | -1.066 | 5.219 | -18.104 | 13.435 |
| DS-2_STAY[2] | Genotyped | 843 | 3.303 | 4.630 | -12.070 | 12.187 |
| | Non-genotyped | 2,685 | -2.949 | 5.079 | -18.262 | 14.466 |
| | All animals | 3,528 | -1.455 | 5.644 | -18.262 | 14.466 |
| DS-3_STAY[3] | Genotyped | 881 | -0.762 | 3.472 | -10.421 | 6.607 |
| | Non-genotyped | 1,569 | 0.323 | 1.248 | -8.440 | 5.642 |
| | All animals | 2,450 | -0.067 | 2.366 | -10.421 | 6.607 |
| DS-4_STAY[4] | Genotyped | 837 | -0.392 | 3.120 | -7.928 | 7.046 |
| | Non-genotyped | 1,503 | -0.313 | 1.821 | -9.095 | 6.591 |
| | All animals | 2,340 | -0.341 | 2.369 | -9.095 | 7.046 |

[1]DS-1_STAY = dataset 1 (1,713 cows); [2]DS-2_STAY = dataset 2 (1,533 cows); [3]DS-3_STAY = dataset 3 (750 cows); [4]DS-4_STAY = dataset 4 (668 cows).

The RRM pedigree-based genetic predictions for STAY06 behaved slightly different between datasets. For DS-1_STAY and DS-2_STAY, average EPD appeared similar for all subsets of animals, although a slightly higher variability in predictions was noted for the DS-2_STAY (higher standard deviations and wider ranges for all groups of animals). In the same datasets, EPD averages of genotyped cows were higher than those obtained for non-genotyped individuals. This

could be explained by the inflated averages of success for STAY06 within the genotyped groups of cows contained in these data structures (180 cows with a 100% success for STAY06 in DS-1_STAY and 70 cows with >85% success for STAY06 in DS-2_STAY; Figure 5.12). In the case of DS-3_STAY and DS-4_STAY, average EPD for STAY06 were closer to zero and their respective ranges were narrower in comparison with the first two data files. Such changes in predictions could have resulted from the restriction of the usage of phenotypic records coming only from genotyped animals, in which the genetic variability for the trait may have decreased over time due to artificial selection (Bulmer, 1971). Resulting predictions using DS-4_STAY showed a greater stability than when using any other data file, since EPD averages, standard deviations and ranges were more concordant between subsets of animals within this particular data structure. Regarding genomic predictions for STAY06, their summary statistics are presented in Table 5.15.

**Table 5.15.** Stayability at the age of 6 expected progeny differences (EPD) summary statistics obtained with the single-step random regression super-hybrid models.

| Dataset | Subset | N | Average | SD | Min | Max |
|---|---|---|---|---|---|---|
| DS-1_STAY[1] | Genotyped | 882 | 32.886 | 1.754 | 27.398 | 44.062 |
| | Non-genotyped | 2,687 | 15.966 | 10.223 | -4.236 | 41.726 |
| | All animals | 3,569 | 20.146 | 11.519 | -4.236 | 44.062 |
| DS-2_STAY[2] | Genotyped | 843 | 16.471 | 1.612 | 11.137 | 36.527 |
| | Non-genotyped | 2,685 | 7.021 | 4.806 | -2.248 | 23.883 |
| | All animals | 3,528 | 9.279 | 5.869 | -2.248 | 36.527 |
| DS-3_STAY[3] | Genotyped | 881 | 25.937 | 1.626 | 20.879 | 38.063 |
| | Non-genotyped | 1,569 | 13.814 | 9.831 | -0.377 | 37.558 |
| | All animals | 2,450 | 18.173 | 9.833 | -0.377 | 38.063 |
| DS-4_STAY[4] | Genotyped | 837 | 6.113 | 2.096 | -1.509 | 32.722 |
| | Non-genotyped | 1,503 | 2.794 | 2.305 | -1.207 | 15.699 |
| | All animals | 2,340 | 3.981 | 2.741 | -1.509 | 32.722 |

[1]DS-1_STAY = dataset 1 (1,713 cows); [2]DS-2_STAY = dataset 2 (1,533 cows); [3]DS-3_STAY = dataset 3 (750 cows); [4]DS-4_STAY = dataset 4 (668 cows).

Genomic predictions for STAY06 obtained with the ssRR-SHM were much more variable than the pedigree-based predictions for the same trait. Due to the longitudinal nature of this trait, fewer animals, but with multiple observations, were present in the evaluation. This situation may have caused an exacerbation of the issues associated with preselection bias since 14.6% of animals (all cows born before 2007) contributing phenotypes to the evaluation possessed clearly inflated phenotypic records for STAY06 (Figure 5.12). In every data structure scenario, the EPD averages of genotyped animals were higher than the EPD of non-genotyped individuals. This type of result was more evident in DS-1_STAY and DS-3_STAY due to the inclusion of observations from the 180 cows with a 100% success for STAY06. For DS-2_STAY, the removal of observations from cows with only successful records for STAY06 reduced the overdispersion of EPD for this trait (narrower ranges of EPD values). Although in this dataset, the 70 cows with an unusually high STAY06 percentage of success (>85%) may have also caused an overestimation of breeding values (e.g., average EPD of 16.471). The most reasonable set of results were obtained when using DS-4_STAY since averages, standard deviations and ranges were more similar between all subsets of animals. Furthermore, the vast majority of the prediction values fell within the range of values (-21.1 to 25.3) that has been reported for this trait in various beef breeds (Snelling et al., 1994; Brigham et al., 2006).

Pearson correlations, rank correlations and regression coefficients of STAY06 genomic predictions on STAY06 pedigree-based predictions are shown in Table 5.16. In all data files, predictions for genotyped animals were highly correlated, nonetheless, similarities in predictions for non-genotyped animals were considerably lower or almost non-existent depending on the dataset studied. Greater degrees of discrepancies for non-genotyped individuals were found in the first couple of data files analyzed, in both cases, Pearson and Spearman's correlations were close

to zero, which suggested that the introduction of imputed genomic information on these animals created a distortion in their predictions. It is probable that the imputation process performed on non-genotyped animals was inaccurate since it depended only on observed genotypes from first degree relatives such as parents, offspring or siblings, being independent from the rest of the pedigree (Gengler et al., 2007). In this sense, the majority of the genotyped animals belonged to the latest generations within the CSU-BIC; however, more than half of the females contributing with phenotypes belonged to previous generations (e.g., 1900's decade). Consequently, even when females of the early years of data were present at the pedigree of the genotyped animals, such pedigree relationship was not close enough to ensure an accurate imputation of their genotypes.

**Table 5.16.** Pearson correlation, rank correlation and regression coefficients of genomic predictions on pedigree-based predictions for stayability at the age of 6.

| Dataset | Subset | Pearson correlation | Rank correlation | Regression coefficient |
|---|---|---|---|---|
| DS-1_STAY[1] | Genotyped | 0.863 | 0.857 | 0.394 |
| | Non-genotyped | 0.013 | 0.011 | 0.028 |
| | All animals | 0.341 | 0.450 | 0.754 |
| DS-2_STAY[2] | Genotyped | 0.797 | 0.860 | 0.278 |
| | Non-genotyped | 0.060 | -0.006 | 0.056 |
| | All animals | 0.401 | 0.420 | 0.417 |
| DS-3_STAY[3] | Genotyped | 0.944 | 0.962 | 0.442 |
| | Non-genotyped | 0.282 | 0.349 | 2.221 |
| | All animals | 0.047 | 0.323 | 0.196 |
| DS-4_STAY[4] | Genotyped | 0.844 | 0.925 | 0.567 |
| | Non-genotyped | 0.349 | 0.215 | 0.441 |
| | All animals | 0.439 | 0.372 | 0.508 |

[1]DS-1_STAY = dataset 1 (1,713 cows); [2]DS-2_STAY = dataset 2 (1,533 cows); [3]DS-3_STAY = dataset 3 (750 cows); [4]DS-4_STAY = dataset 4 (668 cows).

For DS-3_STAY and DS-4_STAY, the similarities between pedigree-based and genomic predictions for STAY06 were higher for all subsets of animals (Table 5.16). Since the last couple of data files analyzed contained information predominantly from genotyped animals, correlations for this subgroup of animals were expected to increase. Conversely, what was more interesting

was to explore the changes in the concordance degree between predictions for non-genotyped individuals. In DS-3_STAY, the only source of phenotypic information was the genotyped subgroup of cows, this may explain why for this subset of animals it was reached the highest degree of congruence among predictions. Albeit, within the same data file, a third of the females (250 out of the 750) contributing phenotypes for the evaluation had atypical high success percentages for STAY06 (females born before 2007), which apparently originated an overestimation of the genetic merit for non-genotyped individuals (e.g., regression coefficient >1). In DS-4_STAY, the removal of phenotypes from the 180 genotyped cows with a 100% success for STAY06, prevented the overestimation of genomic breeding values for non-genotyped animals. Considering all results together, it was concluded that genomic predictions obtained with DS-4_STAY were the most reliable; therefore, the GWAS for STAY06 was based on the results of this specific evaluation. Cows with phenotypes within DS-4_STAY were ranked in quartiles based on their genomic EPD and it was explored if such classification effectively translated into expressed differences in STAY06 phenotypic performance (Appendix B, Table B-4 and Figure B-3).

*5.3.5 Genome-wide association study for stayability*

Among the five SNP associated to STAY06 (Table 5.17), one was located within a coding region and the four remaining were located at non-coding chromosomic segments. The corresponding Manhattan plot of the whole-genome study for STAY06 is in Figure 5.19.

**Table 5.17.** Single nucleotide polymorphisms (SNP) associated stayability in Angus cattle.

| SNP ID[a] | Chr[b] | Location[c] | PPI[d] | Gene | Gene location[e] |
|---|---|---|---|---|---|
| rs41256934 | 2 | 125,653,839 | 0.0086 | MED18 | Intron |
| rs41607880 | 4 | 89,380,482 | 0.0082 | GPR37 | 219.11 kb |
| rs43426517 | 5 | 11,176,710 | 0.0089 | ACSS3 | 271.22 kb |
| rs41636773 | 18 | 53,970,861 | 0.0082 | IGFL1 | 6.30 kb |
| rs110175546 | 20 | 17,240,999 | 0.0084 | KIF2A | 231.45 kb |

[a]Reference SNP cluster identification assigned by the National Center for Biotechnology Information (NCBI).
[b]Chromosome in which the SNP was located within the *Bos taurus* UMD3.1.1 assembly.
[c]Chromosome position in *Bos taurus* UMD3.1.1.
[d]Posterior probability of inclusion in the model.
[e]Location of the SNP within a gene or distance to the closest annotated gene (in kilobases).

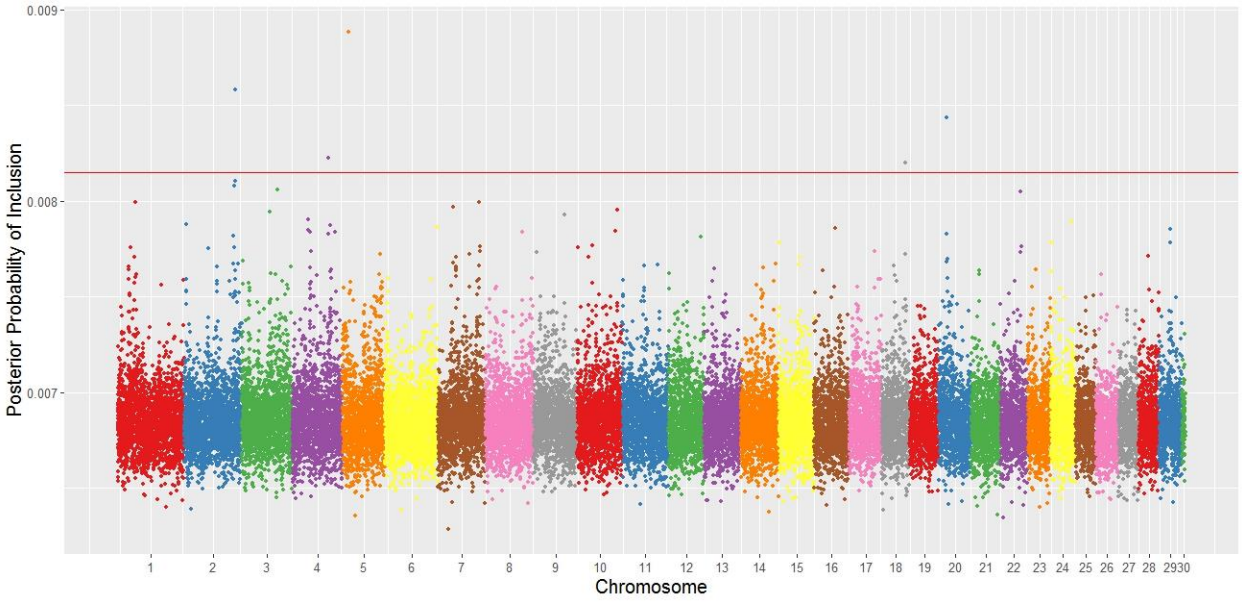


**Figure 5.19.** Manhattan plot identifying SNP associated with stayability in Angus cattle (red horizontal line denotes a 0.82% posterior probability of inclusion).

According to the UMD3.1.1 bovine genome assembly, SNP rs41256934 was a synonymous variant positioned at exon 3 of gene MED18 (mediator complex subunit 18) located in BTA2 (Zimin et al., 2009). The mediator complex is formed by multiple proteins whose role is to regulate the gene transcription process through the modulation of the activity of the RNA

polymerase II enzyme (Malik and Roeder, 2005). Particularly, the protein corresponding to subunit 18 of this mediator complex (specific product of MED18 gene) has been suggested to promote the repression of the transcriptional process (Kamafuji et al., 2014). A study focused on elucidating the transcriptome dynamics and molecular cross-talking between bovine oocyte and its companion cumulus cells revealed that MED18 was exclusively expressed in the germinal vesicle of oocytes subjected to an in vitro maturation procedure (Regassa et al., 2011). Interestingly, the transcriptomic profile of oocytes that were directly aspired from ovarian follicles of synchronized cows was also investigated and, results from such investigation, indicated that transcripts of MED18 were higher in oocytes collected from cows showing estrus signs (Dickinson, 2016). These reports suggest that this gene was involved in controlling the transcriptional processes that need to occur during the development of competent oocytes, which may explain its relationship with a trait related to the sustained fertility of beef cows.

In BTA4, the SNP associated with STAY06 was rs41607880. This polymorphism was located in an intergenic region that was 219.11 kb downstream from the gene GPR37 which encodes for the G protein-coupled receptor 37. The G protein-coupled receptors are important membrane proteins that detect signaling molecules such as hormones and neurotransmitters (Venkatakrishnan et al., 2013). A report in humans suggested that this gene (along with other five members of the G protein-coupled receptors family) was overexpressed in mature oocytes when compared to immature oocytes (Assou et al., 2006). In cattle, the GPR37 was found to be differentially expressed in artificially maturated oocytes (Regassa et al., 2011); as well as on the endometrium of beef heifers supplemented with a rumen protected source of n-3-polyunsaturated fatty acid that were slaughtered on day 17 of their estrus cycle (Waters et al., 2014). Oliveira Júnior et al. (2017) informed about a genomic window on BTA16 that that contained the gene GPR37L1

(G protein-coupled receptor 37 like-1) that explained more than 1% of the additive genetic variance for number of antral follicles in Nelore heifers. Even though such investigation informed of a different gene in a different chromosome, it is interesting to note that both, GPR37 and GPR37L1 have been related to fertility traits and that the protein products encoded by these two genes share more than 60% amino acid similarity (Hertz et al., 2019).

The genetic marker found on BTA5 (rs43426517) was positioned at 271.22 kb upstream from the gene ACSS3 which encodes for one of the three acyl-CoA synthetase short chain family members that exist (specifically, member 3). This family of enzymes is responsible for ligate the acetate produced in the ruminal fermentation to the coenzyme A in order to form acetyl-coA, a molecule needed for lipogenesis and histone acetylation (Xu et al., 2017). Lipids are known regulators of conceptus development in cattle since they are required for its elongation, a critical phase that leads maternal recognition of pregnancy, implantation, and onset of placentation (Ribeiro, 2018). A higher expression level of ACSS3 have been reported in bovine embryos of 18 d of age collected from growing dairy heifers when compared to embryos of the same age obtained from multiparous dairy cows at different lactation stages (Valour et al., 2014). Recently, Mota et al. (2020) identified through a GWAS that the ACSS3 gene was potentially associated to the age at first calving in Nelore heifers. Vineeth et al. (2020) reported that SNP within this gene resulted associated to productive traits in Sahiwal cattle.

Among the SNP located at intergenic regions, the SNP rs41636773 was the more closely positioned to a coding sequence since it was located 6.30 kb downstream from the gene IGFL1 (IGF like family member 1). A study performed in bovines intended to characterize the transcriptome of the conceptus-endometrium interactions during maternal recognition of pregnancy, indicated that IGFL1 gene was a highly expressed growth factor in embryos retrieved

from pregnant heifers slaughtered 16 d post insemination (Mamo et al., 2012). Moreover, Cole et al. (2011) identified that the chromosomic region (15 Mb long) where the IGF1L gene is located was associated with a variety of traits in dairy cattle among which productive life, daughter pregnancy rate and calving ease may be the more related to STAY in beef cattle. Speidel et al. (2018a) reported that a SNP within the same chromosomic region was precisely associated with STAY in Red Angus cattle. Furthermore, research in humans and mice suggested that the IGF-like (IGFL) family of genes share structural homology to the IGF family (e.g., IGF-1 and IGF-2) and that both genes encode short-length proteins (around 100 amino acids) that seem to be involved in biological processes like regulation of metabolism, growth and reproduction (Emtage et al., 2006). A plethora of studies have been conducted to investigate the implications of the insulin-like growth factor 1 (IGF1) gene with reproductive performance in cattle. In general, it has been documented that the protein product of the IGF1 gene influences ovarian activity by regulating the action of gonadotropins on follicular growth, steroidogenesis and the establishment of follicular dominance (Werner and Le Roith, 2000; Rivera et al., 2001; Monget et al., 2002; Grossi et al., 2015). At a genetic level, polymorphisms within the IGF1 gene have been associated with various fertility-related traits in cattle such as body condition score at calving (Mullen et al., 2011), postpartum resumption of ovarian cyclicity (Nicolini et al., 2013), calving to conception interval (Silveira et al., 2015), and number of services per conception (Leyva-Corona et al., 2018).

The last of the five polymorphisms that showed the highest PPI for STAY06 in this study was rs41636773, this SNP was located at 231.45 kb of distance from the kinesin family member 2A gene (KIF2A). This member of the kinesin family of proteins has been identified as key regulator of microtubule dynamics during mitosis due to its participation in processes like intracellular transport, cell division, and bipolar spindle assembly (Manning et al., 2007; Wang et

al., 2019). From a reproductive standpoint, it has been reported that kinesins are involved in a crucial fertilization event such as the acrosome reaction of bovine spermatozoa (Oikonomopoulou et al., 2009). Furthermore, an increased expression of KIF2A gene was found in the endometrium of crossbred beef heifers that were supplemented with n-3-polyunsaturated fatty acids and slaughtered at day 17 of their estrus cycle (Waters et al., 2012; Waters et al., 2014).

*5.4 Conclusion*

The implementation of random regression super-hybrid models for the genomic evaluation of singly-observed binary fertility traits like HPG and FSCR, as well as for the evaluation of a longitudinally recorded binary trait such as STAY was feasible in a single-herd purebred Angus population. Nonetheless, genomic predictions yielded by ssRR-SHM were highly dependent on the specific data structures relative to each one of the traits analyzed. In all cases, the presence of preselection bias on the subset of genotyped individuals that contributed with phenotypes for the evaluation was the main reason of overestimations of genomic predictions for all animals (including non-genotyped individuals). Furthermore, inaccurate imputation of genotypes for the non-genotyped subset of animals also impacted resulting genomic predictions, although this issue was restricted to this subgroup of animals only. Removal of phenotypic records from preselected animals ameliorated problems associated with overestimation of genomic predictions and improved correlations among genomically-enhanced and pedigree-based EPD for all traits.

Regarding GWAS analyses, although the PPI obtained for all traits were considerably low, all SNP identified as QTL after the application of ssRR-SHM resulted located either within or relatively close to genes that have been previously associated with important reproductive processes and fertility traits in cattle. The previous imply that in spite of yielding small signals for QTL detection, the models implemented in this investigation identified important chromosomal

regions influencing the traits under study. Furthermore, considering the low heritability and the high biological complexity of all traits studied, the obtention of low PPI should not be seen as a statistical modelling problem but rather as a reflection of the reality about the genetic component of fertility in cattle.

*Literature cited*

Allan, M. F., L. A. Kuehn, R. A. Cushman, W. M. Snelling, S. E. Echternkamp, and R. M. Thallman. 2009. Confirmation of quantitative trait loci using a low-density single nucleotide polymorphism map for twinning and ovulation rate on bovine chromosome 5. J. Anim. Sci. 87:46-56.

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. J. Dairy Sci. 93:743-752.

Assou, S., T. Anahory, V. Pantesco, T. Le Carrour, F. Pellestor, B. Klein, L. Reyftmann, H. Dechaud, J. De Vos, and S. Hamamah. 2006. The human cumulus–oocyte complex gene-expression profile. Hum. Reprod. 21:1705-1719.

Averill, T., R. Rekaya, and K. Weigel. 2006. Random regression models for male and female fertility evaluation using longitudinal binary data. J. Dairy Sci. 89:3681-3689.

Azzam, S. M., J. E. Kinder, and M. K. Nielsen. 1989. Conception rate at first insemination in beef cattle: effects of season, age and previous reproductive performance. J. Anim. Sci. 67:1405-1410.

Barreta, M.H., B. G. Gasperin, V. B. Rissi, M. P. de Cesaro, R. Ferreira, J. F. de Oliveira, P. B. D. Gonçalves, and V. Bordignon. 2012. Homologous recombination and non-homologous end-joining repair pathways in bovine embryos with different developmental competence. Exp. Cell Res. 318:2049-2058.

Beef Improvement Federation (BIF). 2020. Guidelines for Uniform Beef Improvement Programs: New, Improved and Wiki. http://guidelines.beefimprovement.org/index.php/Special:AllPages Accessed 29 October 2020).

Biase, F. H., and K. M. Kimble. 2018. Functional signaling and gene regulatory networks between the oocyte and the surrounding cumulus cells. BMC Genom. 19:1-13.

Boldt, R. J. 2017. Genetic parameters for fertility and production traits in Red Angus cattle. Master's Thesis, Colorado State University, Fort Collins.

Boldt, R. J., S. E. Speidel, M. G. Thomas, and R. M. Enns. 2018. Genetic parameters for fertility and production traits in Red Angus cattle. J. Anim. Sci. 96:4100-4111.

Bormann, J. M., L. R. Totir, S. D. Kachman, R. L. Fernando, and D. E. Wilson. 2006. Pregnancy rate and first-service conception rate in Angus heifers. J. Animal Sci. 84:2022-2025.

Brigham, B. W., S. E. Speidel, R. M. Enns, and D. J. Garrick. 2007. Stayability to alternate ages. Proc. West. Sec. Amer. Soc. Anim. Sci. 58:27-30.

Brito, L.F., S. Mallikarjunappa, M. Sargolzaei, A. Koeck, J. Chesnais, F. S. Schenkel, K. G. Meade, F. Miglior, and N. A. Karrow. 2018. The genetic architecture of milk ELISA scores as an indicator of Johne's disease (paratuberculosis) in dairy cattle. J. Dairy Sci. 101:10062-10075.

Broadhurst, M. K., R. S. F. Lee, S. Hawkins, and T. T. Wheeler. 2005. The p100 EBNA-2 coactivator: a highly conserved protein found in a range of exocrine and endocrine cells and tissues in cattle. BBA-Gene Struct. Expr. 1681:126-133.

Bulmer, M. 1971. The effect of selection on genetic variability. Amer. Naturalist. 105:201-211.

Cammack, K. M., M. G. Thomas, and R. M. Enns. 2009. Reproductive traits and their heritabilities in beef cattle. Prof. Anim. Sci. 25:517-528.

Chalupnik, A. D., M. Ozgo, A. Herosimczyk, A. Lepczynski, and K. Michalek. 2016. Lactose-rich milk replacer modifies the proteome of bloodplasma in 2-week-old calves. Turk. J. Vet. Anim. Sci. 40:21-27.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Secondgeneration PLINK: rising to the challenge of larger and richer datasets. GigaSci 4, 7.

Chudleigh F., M. Bowen and B. Holmes. 2019. Farm economic thinking and the genetic improvement of fertility in northern beef herds. In: Proc. 63rd Australian Agricultural and Resource Economics Society, Melbourne, Australia, pp. 12-37.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC Genom. 12:408.

Crawford, N. F., M. G. Thomas, T. N. Holt, S. E. Speidel, and R. M. Enns. 2016. Heritabilities and genetic correlations of pulmonary arterial pressure and performance traits in Angus cattle at high altitude. J. Anim. Sci. 94:4483-4490.

Dehnavi, E., S. A. Mahyari, F. S. Schenkel, and M. Sargolzaei. 2018. The effect of using cow genomic information on accuracy and bias of genomic breeding values in a simulated Holstein dairy cattle population. J. Dairy Sci. 101:5166-5176.

Dickinson, S. E. 2016. Effect of pre-ovulatory follicle size on oocyte transcript abundance in beef cows. Master's Thesis, University of Missouri, Columbia, USA.

Dickinson, S.E., B. A. Griffin, M. F. Elmore, L. Kriese-Anderson, J. B. Elmore, P. W. Dyce, S. P. Rodning, and F. H. Biase. 2018. Transcriptome profiles in peripheral white blood cells at the time of artificial insemination discriminate beef heifers with different fertility potential. BMC Genom. 19:1-11.

Doyle, S. P., B. L. Golden, R. D. Green, and J. S. Brinks. 2000. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. J. Anim. Sci. 78:2091-2098.

Dekkers, J. C. M. 2010. Use of high-density marker genotyping for genetic improvement of livestock by genomic selection. CABI Anim. Sci. Rev. 2010:197–210.

Emtage, P., P. Vatta, M. Arterburn, M. W. Muller, E. Park, B. Boyle, S. Hazell, R. Polizotto, W. D. Funk, and Y. T. Tang., 2006. IGFL: A secreted family with conserved cysteine residues and similarities to the IGF superfamily. Genomics. 88:513-520.

Engle, B. N., A. D. Herring, J. E. Sawyer, D. G. Riley, J. O. Sanders, and C. A. Gill. 2016. Genome-wide association study for stayability measures in Nellore-Angus crossbred cows. J. Anim. Sci. 94:142-142.

Fernández, J. C., J. E. Pérez, N. Herrera, R. Martínez, D. Bejarano, and J. F. Rocha. 2019. Genomic association study for age at first calving and calving interval in Romosinuano and Costeño con Cuernos cattle. Genet. Mol. Res. 18:1-13.

Fernando, R. L., J. C. M. Dekkers and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50

Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96.

Fonseca, L.F.S., D. B. dos Santos Silva, D. F. J. Gimenez, F. Baldi, J. A. Ferro, L. A. L. Chardulo, and L. G. de Albuquerque. 2020. Gene expression profiling and identification of hub genes in Nellore cattle with different marbling score levels. Genomics. 112:873-879.

Forde, N., J. P. Mehta, M. Minten, M. A. Crowe, J. F. Roche, T. E. Spencer, and P. Lonergan. 2012. Effects of low progesterone on the endometrial transcriptome in cattle. Biol. Reprod. 87(5), pp.124:1-11.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43:1.

Fortes, M. R., A. Reverter, Y. Zhang, E. Collis, S. H. Nagaraj, N. N. Jonsson, K. C. Prayaga, W. Barris, and R. J. Hawken. 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. Proc. Natl. Acad. Sci. 107:13642-13647.

Gao, H., O. F. Christensen, P. Madsen, U. S. Nielsen, Y. Zhang, M. S. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genet. Sel. Evol. 44:8.

Gao, H., M. Koivula, J. Jensen, I. Strandén, P. Madsen, T. Pitkänen, G. P. Aamand, and E. A. Mäntysaari. 2018. Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population. J. Dairy Sci. 101:10082-10088.

Garrick, D.J., Fernando, R.L., 2013. Implementing a QTL detection study (GWAS) using genomic prediction methodology. In: Gondro, C., van der Werf, J., Hayes, B. (Eds.), Genome-wide Association Studies and Genomic Prediction. Humana Press, Totowa, NJ, pp. 275 –298.

Garrick, D. J., D. P. Garrick, and B. L. Golden. 2018. An introduction to BOLT software for genetic and genomic evaluations. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/introduction-bolt-software-genetic-and-genomic-evaluations.pdf

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal. 1:21-28.

Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits, a framework for the next generation of EPDs. In: Proc. Beef Improvement Federation, Wichita, KS. p. 2–13.

Golden, B. L. and D. J. Garrick. 2016. An introduction to BOLT. Theta Solutions LLC. http://manual.thetasolutionsllc.com/IntroBolt.pdf

Golden, B. L., S. Weerasinghe, B. Crook, S. Sanders, and D. J. Garrick. 2018a. A Single-Step Hybrid Marker Effects Model Using Random Regression for Stayability in Hereford Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/single-step-hybrid-marker-effects-model-using-random-regression-stayability-hereford-cattle.pdf

Golden, B. L., M. L. Spangler, W. M. Snelling, and D. J. Garrick. 2018. Current single-step national beef cattle evaluation models used by the American Hereford Association and International Genetic Solutions, computational aspects, and implications of marker selection. In: Proc. Beef Imp. Fed. 11th Genetic Prediction Workshop, Refining Genomic Evaluation and Selection Indices; December 12 to 13, 2018; Kansas City (MO); p. 14–22.

Grossi, D. do A., M. E. Buzanskas, N. V. Grupioni, C. C. P. de Paz, L. C. Regitano, L.C. de A. Regitano, M. M. de Alencar, F. S. Schenkel, and D. P. Munari. 2015. Effect of IGF1, GH, and PIT1 markers on the genetic parameters of growth and reproduction traits in Canchim cattle. Mol. Biol. Rep. 42, 245–251.

Harris, B. L., D. L. Johnson, and R. J. Spelman. 2011. Choice of parameters for removal of inflation in genomic breeding values for dairy cattle. Proc. Ass. Adv. Anim. Breed. Genet. 19:359-62.

Hayes, B. J., N. J. Corbet, J. M. Allen, A. R. Laing, G. Fordyce, R. Lyons, M. R. McGowan, and B. M. Burns. 2019. Towards multi-breed genomic evaluations for female fertility of tropical beef cattle. J. Anim. Sci. 97:55-62.

Hertz, E., L. Terenius, V. Vukojević, and P. Svenningsson. 2019. GPR37 and GPR37L1 differently interact with dopamine 2 receptors in live cells. Neuropharmacology. 152:51-57.

Hsu, W.L., D. J. Garrick, and R. L. Fernando. 2017. The accuracy and bias of single-step genomic prediction for populations under selection. G3-Gen. Genom. Genet. 7:2685-2694.

Hum, D. W., A. Bélanger, É. Lévesque, O. Barbier, M. Beaulieu, C. Albert, M. Vallée, C. Guillemette, A. Tchernof, D. Turgeon, and S. Dubois. 1999. Characterization of UDP-glucuronosyltransferases active on steroid hormones. J. Steroid Biochemistry Mol. Biol. 69:413-423.

Jamrozik, J., S. McGrath, R. A. Kemp, and S. P. Miller. 2013. Estimates of genetic parameters for stayability to consecutive calvings of Canadian Simmentals by random regression models. J. Anim. Sci. 91:3634-3643.

Jaroudi, S., G. Kakourou, S. Cawood, A. Doshi, D. M. Ranieri, P. Serhal, J. C. Harper, and S. B. SenGupta. 2009. Expression profiling of DNA repair genes in human oocytes and blastocysts using microarrays. Human Reprod. 24:2649-2655.

Kumafuji, M., H. Umemura, T. Furumoto, R. Fukasawa, A. Tanaka, and Y. Ohkuma. 2014. Mediator MED 18 subunit plays a negative role in transcription via the CDK/cyclin module. Genes Cells. 19:582-593.

Kamiński, S., D. M. Hering, K. Oleński, M. Lecewicz, and W. Kordan. 2016. Genome-wide association study for sperm membrane integrity in frozen-thawed semen of Holstein-Friesian bulls. Anim. Reprod. Sci. 170:135-140.

Kappes, S. M., G. L. Bennett, J. W. Keele, S. E. Echternkamp, K. E. Gregory, and R. M. Thallman. 2000. Initial results of genomic scans for ovulation rate in a cattle population selected for increased twinning rate. J. Anim. Sci. 78:3053-3059.

Killeen, A. P., M. G. Diskin, D. G. Morris, D. A. Kenny, and S. M. Waters. 2016. Endometrial gene expression in high-and low-fertility heifers in the late luteal phase of the estrous cycle and a comparison with midluteal gene expression. Physiol. Genomics. 48:306-319.

Kim, E. S., X. Shi, O. Cobanoglu, K. Weigel, P. J. Berger, and B. W. Kirkpatrick. 2009. Refined mapping of twinning-rate quantitative trait loci on bovine chromosome 5 and analysis of insulin-like growth factor-1 as a positional candidate gene. J. Anim. Sci. 87:835-843.

Kluska, S., B. F. Olivieri, M. Bonamy, H. L. J. Chiaia, F. L. B. Feitosa, M. P. Berton, E. Peripolli, M. V. A. Lemos, R. L. Tonussi, R. B. Lôbo, C. Ulhoa Magnabosco, F. Di Croce, J. Osterstock, A. S. C. Pereira, D. P. Munari, L. A. Bezerra, F. B. Lopes and F. Baldi. 2018. Estimates of genetic parameters for growth, reproductive, and carcass traits in Nelore cattle using the single step genomic BLUP procedure. Livest. Sci. 216:203–209.

Koivula, M., I. Strandén, J. Pösö, G. P. Aamand, and E. A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775-2784.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 10:1061-1066.

Komander, D., and M. Rape. 2012. The ubiquitin code. Annu. Rev. Biochem. 81, 203–229.

Kong, Z., C. Zhou, L. Chen, A. Ren, D. Zhang, Z. Basang, Z. Tan, J. Kang, and B. Li. 2019. Multi-Omics Analysis Reveals Up-Regulation of APR Signaling, LXR/RXR and FXR/RXR Activation Pathways in Holstein Dairy Cows Exposed to High-Altitude Hypoxia. Animals. 9:406.

Laisk-Podar T., T. Kaart M. Peters, and A. Salumets. 2015. Genetic variants associated with female reproductive ageing–potential markers for assessing ovarian function and ovarian stimulation outcome. Reprod. Biomed. Online. 1:199-209.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Lemos, M. V., H. L. J. Chiaia, M. P. Berton, F. L. Feitosa, C. Aboujaoud, G. M. Camargo, A. S. Pereira, L. G. Albuquerque, A. M. Ferrinho, L. F. Mueller, and M. R. Mazalli. 2016. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. BMC Genom. 17:213.

Leyva-Corona, J. C., J. R. Reyna-Granados, R. Zamorano-Algandar, M. A. Sánchez-Castro, M. G. Thomas, R. M. Enns, S. E. Speidel, J. F. Medrano, G. Rincon, and P. Luna-Nevárez. 2018. Polymorphisms within the prolactin and growth hormone/insulin-like growth factor-1 functional pathways associated with fertility traits in Holstein cows raised in a hot-humid climate. Trop. Anim. Health Prod. 50:1913-1920.

Li, Z., T. Otevrel, Y. Gao, H. L. Cheng, B. Seed, T. D. Stamato, G. E. Taccioli, and F. W. Alt. 1995. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V (D) J recombination. Cell. 83:1079-1089.

Liu, Y. Z., Y. F. Guo, L. Wang, L. J. Tan, X. G. Liu, Y. F. Pei, H. Yan, D. H. Xiong, F. Y. Deng, N. Yu, and Y. P. Zhang. 2009. Genome-wide association analyses identify SPOCK as a key novel gene underlying age at menarche. PLoS Genet. 5:e1000420.

Liu, Z., F. R. Seefried, F. Reinhardt, S. Rensing, G. Thaller, and R. Reents. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. 43:19.

Liu, Z., M. E. Goddard, B. J. Hayes, F. Reinhardt, and R. Reents. 2016. Equivalent genomic models with a residual polygenic effect. J. Dairy Sci. 99:2016-2025.

Luna-Nevarez, P., G. Rincon, J. F. Medrano, D. G. Riley, C. C. Chase Jr., S. W. Coleman, D. M. VanLeeuwen, K. L. DeAtley, A. Islas-Trejo, G. A. Silver, and M. G. Thomas. 2011. Single nucleotide polymorphisms in the growth hormone–insulin-like growth factor axis in straightbred and crossbred Angus, Brahman, and Romosinuano heifers: population genetic analyses and association of genotypes with reproductive phenotypes. J. Anim. Sci. 89:926-934.

Mahdipour, M. 2015. Maintenance of genomic integrity and developmental competence in the mammalian oocyte. Ph.D diss., Utrecht University, Iran.

Maiorano, A. M., D. L. Lourenco, S. Tsuruta, A. M. T. Ospina, N. B. Stafuzza, Y. Masuda, A. E. V. Filho, J. N. D. S. G. Cyrillo, R. A. Curi, and J. A. I. D .V. Silva. 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One. 13:p.e0200694.

Malik, S. and R. G. Roeder. 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem. Sci. 30:256-263.

Mamo, S., J. P. Mehta, N. Forde, P. McGettigan, and P. Lonergan. 2012. Conceptus-endometrium crosstalk during maternal recognition of pregnancy in cattle. Biol. Reprod. 87:1-9.

Manning, A. L., N. J. Ganem, S. F. Bakhoum, M. Wagenbach, L. Wordeman, and D. A. Compton. 2007. The kinesin-13 proteins Kif2a, Kif2b, and Kif2c/MCAK have distinct roles during mitosis in human cells. Mol. Biol. Cell. 18:2970-2979.

Mäntysaari, E.A., Z. Liu, and P. Van Raden. 2010. Interbull validation test for genomic evaluations. In: Proc. Interbull International Workshop March 4-5, 2010, Paris, France. Interbull Bulletin. 41:17-21.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. J. Dairy Sci. 103:5314-5326.

Martini, J. W., M. F. Schrauf, C. A. Garcia-Baccino, E. C. Pimentel, S. Munilla, A. Rogberg-Muñoz, R. J. Cantet, C. Reimer, N. Gao, V. Wimmer, and H. Simianer. 2018. The effect of the $H^{-1}$ scaling factors $\tau$ and $\omega$ on the structure of H in the single-step procedure. Genet. Sel. Evol. 50:1-9.

Meech, R., D. G. Hu, R. A. McKinnon, S. N. Mubarokah, A. Z. Haines, P. C. Nair, A. Rowland, and P. I. Mackenzie. 2019. The UDP-glycosyltransferase (UGT) superfamily: new members, new functions, and novel paradigms. Physiol. Rev. 99:1153-1222.

Medrano, J. F. 2017. The new bovine reference assembly and its value for genomic research. In: Proc. 22nd Assoc. Adv. Anim. Breeding Genet. Conf. 22:37.

Melton, B. E. 1995. Conception to consumption: The economics of genetic improvement. In: Proc. Beef Improvement Federation 27th Research Symposium and Annual Meeting, Sheridan, Wyoming. pp. 40−47.

Mevissen, T.E., M. K. Hospenthal, P. P. Geurink, P. R. Elliott, M. Akutsu, N. Arnaudo, R. Ekkebus, Y. Kulathu, T. Wauer, F. El Oualid, and S. M. Freund. 2013. OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. Cell. 154:169-184.

Misztal, I., I. Aguilar, A. Legarra, and T. J. Lawlor. 2010. Choice of parameters for single-step genomic evaluation for type. J. Dairy Sci. 93(Suppl. 1):533. (Abstr.).

Misztal, I., H. L. Bradford, D. A. L. Lourenco, S. Tsuruta, Y. Masuda, A. Legarra and T. J. Lawlor. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bulletin 51:38-42.

Misztal, I. and D. Lourenco. 2018. Current Research in Unweighted and Weighted ssGBLUP. In: Proc. Beef Imp. Fed. 11th Genetic Prediction Workshop, Refining Genomic Evaluation and Selection Indices; December 12 to 13, 2018; Kansas City (MO); p. 6–13.

Misztal, I., D. Lourenco, and A. Legarra. 2020. Current status of genomic evaluation. J. Anim. Sci. 98:1-14.

Monget P., S. Fabre, P. Mulsant, F. Lecerf, J. M. Elsen, S. Mazerbourg, C. Pisselet, and D. Monniaux. 2002. Regulation of ovarian folliculogenesis by IGF and BMP system in domestic animals. Domest. Anim. Endocrinol. 23:139–154.

Montes, D. E., C. U. Braz, A. M. Ribeiro, L. Cavani, M. M. D. Barbero, L. G. Albuquerque, R. A. Curi, and H. N. Oliveira. 2019. Selection signatures in candidate genes and QTL for reproductive traits in Nellore heifers. Anim. Reprod. Sci. 207:1-8.

Mota, L. F., F. B. Lopes, G. A. F. Júnior, G. J. Rosa, A. F. Magalhães, R. Carvalheiro, R. and L. G. Albuquerque. 2020. Genome-wide scan highlights the role of candidate genes on phenotypic plasticity for age at first calving in Nellore heifers. Sci. Rep. 10:1-13.

Mu, Y., G. Vander Voort, M. K. Abo-Ismail, R. Ventura, J. Jamrozik, and S. P. Miller. 2016. Genetic correlations between female fertility and postweaning growth and feed efficiency traits in multibreed beef cattle. Can. J. Anim. Sci. 96:448-455.

Mullen, M. P., C. O. Lynch, S. M. Waters, D. J. Howard, P. O'boyle, D. A. Kenny, F. Bluckey, B. Horam, and M. G. Diskin. 2011. Single nucleotide polymorphisms in the growth hormone and insulin-like growth factor-1 genes are associated with milk production, body condition score and fertility traits in dairy cows. Genet. Mol. Res. 10:1819-1830.

Murai, M., Sekiguchi, K., Nishioka, T. and Miyoshi, H., 2009. Characterization of the Inhibitor Binding Site in Mitochondrial NADH− Ubiquinone Oxidoreductase by Photoaffinity Labeling Using a Quinazoline-Type Inhibitor. Biochemistry. 48:688-698.

Nicolini, P., M. Carriquiry, and A. Meikle. 2013. A polymorphism in the insulin-like growth factor 1 gene is associated with postpartum resumption of ovarian cyclicity in Holstein-Friesian cows under grazing conditions. Acta Vet. Scand. 55:1-8.

Nordbø, Ø., A. B. Gjuvsland, L. S. Eikje, and T. Meuwissen. 2019. Level-biases in estimated breeding values due to the use of different SNP panels over time in ssGBLUP. Genet. Sel. Evol. 51:76.

Oikonomopoulou, I., H. Patel, P. F. Watson, and P. D. Chantler. 2009. Relocation of myosin and actin, kinesin and tubulin in the acrosome reaction of bovine spermatozoa. Reprod. Fertil. Dev. 21:364-377.

Oliveira Júnior, G. A., B. C. Perez, J. B. Cole, M. H. A. Santana, J. Silveira, G. Mazzoni, R. V. Ventura, M. L. Santana Júnior, H. N. Kadarmideen, D. J. Garrick, and J. B. S. Ferraz. 2017. Genomic study and medical subject headings enrichment analysis of early pregnancy rate and antral follicle numbers in Nelore heifers. J. Anim. Sci. 95:4796-4812.

Pant, S. D., F. S. Schenkel, C. P. Verschoor, Q. You, D. F. Kelton, S. S. Moore, and N. A. Karrow. 2010. A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle. Genomics. 95:176-182.

Pierce, C. F. 2019. Identifying Single Nucleotide Polymorphisms Associated with Beef Cattle Terrain-use in the Western United States. Master's Thesis, Colorado State University, Fort Collins.

Pierce, C.F., S. E. Speidel, S. J. Coleman, R. M. Enns, D. W. Bailey, J. F. Medrano, A. Cánovas, P. J. Meiman, L. D. Howery, W. F. Mandeville, and M. G. Thomas. 2020. Genome-wide association studies of beef cow terrain-use traits using Bayesian multiple-SNP regression. Livest. Sci. 232:103900.

Ramirez-Valverde, R., I. Misztal, and J. K. Bertrand. 2001. Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. J. Anim. Sci. 79:333-338.

Regassa, A., F. Rings, M. Hoelker, U. Cinar, E. Tholen, C. Looft, K. Schellander, and D. Tesfaye. 2011. Transcriptome dynamics and molecular cross-talk between bovine oocyte and its companion cumulus cells. BMC Genom. 12:57.

Reyes, J. M., J. L. Chitwood, and P. J. Ross. 2015. RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. Mol. Reprod. Dev. 82:103-114.

Ribeiro, E. S. 2018. Symposium review: Lipids as regulators of conceptus development: Implications for metabolic regulation of reproduction in dairy cattle. J. Dairy Sci. 101:3630-3641.

Riding, G. A., A. Jones, M. K. Holland, J. R. Hill, and S. A. Lehnert. 2008. Proteomic analysis of bovine conceptus fluids during early pregnancy. Proteomics. 8:160-177.

Rivera G. M., Y. A. Chandrasekher, A. C. Evans, L. C. Giudice, and J. E. Fortune. 2001. A potential role for insulin-like growth factor binding protein-4 proteolysis in the establishment of ovarian follicular dominance in cattle. Biol. Reprod. 65:102–111.

Saatchi, M., L. Hyde, W. Shafer, D. Garrick, and B. Golden. 2018. Incorporation of external information for international multi-breed beef cattle genetic evaluations using multi-trait single-step Bayesian regression model. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand.

Sbardella, A. P. 2020. Estudo genômico de características de desempenho reprodutivo em bovinos da raça Nelore. Ph.D. diss., Universidade Estadual Paulista, Jaboticabal, Brazil.

Schaeffer, L. R. 2014. Is The Animal Model Obsolete? Center for Genetic Improvement of livestock (CGIL), Department of Animal Bioscience, University of Guelph, Guelph, Ontario, Canada, Sep 19 - Dec 3. 2014. http://animalbiosciences.uoguelph.ca/sites/default/files/wrkshop_0.pdf.

Silveira, P. A. S., L. M. C. Pegoraro, W. R. Butler, M. N. Correa, and A. Schneider. 2015. Effect of IGF-1 SnaBI polymorphism on reproductive parameters and metabolic parameters in dairy cows. In: Embrapa Clima Temperado-Resumo em anais de congresso (ALICE). Anim. Reprod. Page 582.

Snelling, W. M., B. L. Golden, and R. M. Bourdon. 1994. An EPD for stayability in beef cows. Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Ontario, Canada. 17:169–172.

Snelling, W. M., B. L. Golden, and R. M. Bourdon. 1995. Within-herd genetic analyses of stayability of beef females. J. Anim. Sci. 73:993-1001.

Speidel, S. E., B. A. Buckley, R. J. Boldt, R. M. Enns, J. Lee, M. L. Spangler, and M. G. Thomas. 2018a. Genome-wide association study of Stayability and Heifer Pregnancy in Red Angus cattle. J. Anim. Sci. 96:846-853.

Speidel, S. E., R. M. Enns, and B. L. Golden. 2018b. Use of a Random Regression model for the Evaluation of Heifer Pregnancy in Red Angus Cattle. In: Proc. 11th World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/system/files/proceedings/2018/use-random-regression-model-evaluation-heifer-pregnancy-red-angus-cattle.pdf

Su, G., O. F. Christensen, L. Janss, and M. S. Lund. 2014. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. J. Dairy Sci. 97:6547-6559.

Tamaki, T., K. Kamatsuka, T. Sato, S. Morooka, K. Otsuka, M. Hattori, and T. Sugiyama. 2017. A novel transmembrane protein defines the endoplasmic reticulum stress-induced cell death pathway. Biochem. Bioph. Res. Co. 486:149-155.

Tizioto, P.C., L. L. Coutinho, P. S. Oliveira, A. S. Cesar, W. J. Diniz, A. O. Lima, M. I. Rocha, J. E. Decker, R. D. Schnabel, G. B. Mourão, and R. R. Tullio. 2016. Gene expression differences in Longissimus muscle of Nelore steers genetically divergent for residual feed intake. Sci. Rep. 6:39493.

Toghiani, S., E. Hay, P. Sumreddee, T. W. Geary, R. Rekaya, and A. J. Roberts. 2017. Genomic prediction of continuous and binary fertility traits of females in a composite beef cattle breed. J. Anim. Sci. 95:4787-4795.

Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198-4204.

Tsuruta, S., D. A. L. Lourenco, Y. Masuda, I. Misztal, and T. J. Lawlor. 2019. Controlling bias in genomic breeding values for young genotyped bulls. J. Dairy Sci. 102:9956-9970.

Valour, D., S. A. Degrelle, A. A. Ponter, C. Giraud-Delville, E. Campion, C. Guyader-Joly, C. Richard, F. Constant, P. Humblot, C. Ponsart, and I. Hue. 2014. Energy and lipid metabolism gene expression of D18 embryos in dairy cows is related to dam physiological status. Physiol. Genomics. 46:39-56.

VanRaden, P. 2011. Findhap.f90. Accessed Jul. 15, 2020. http://aipl.arsusda.gov/software/findhap

Veerkamp, R.F., M. P. Coffey, D. P. Berry, Y. De Haas, E. Strandberg, H. Bovenhuis, M. P. L. Calus, and E. Wall. 2012. Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. Animal. 6:1738-1749.

Venkatakrishnan, A. J., X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler, and M. M. Babu. 2013. Molecular signatures of G-protein-coupled receptors. Nature. 494:185-194.

Veronese, A., O. Marques, R. Moreira, A. L. Belli, R. S. Bisinotto, T. R. Bilby, F. Peñagaricano, and R. C. Chebel. 2019a. Genomic merit for reproductive traits. I: Estrous characteristics and fertility in Holstein heifers. J. Dairy Sci. 102:6624-6638.

Veronese, A., O. Marques, F. Peñagaricano, R. S. Bisinotto, K. G. Pohler, T. R. Bilby, and R. C. Chebel. 2019b. Genomic merit for reproductive traits. II: Physiological responses of Holstein heifers. J. Dairy Sci. 102:6639-6648.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. Camb. 93:357-366.

Vineeth, M. R., T. Surya, J. Sivalingam, A. Kumar, S. K. Niranjan, S. P. Dixit, K. Singh, M. S. Tantia, and I. D. Gupta. 2020. Genome-wide discovery of SNPs in candidate genes related to production and fertility traits in Sahiwal cattle. Trop. Anim. Health Prod. 52:1707-1715.

Vostrý, L., Z. Veselá, A. Svitáková, and H. Vostrá Vydrová. 2014. Comparison of models for estimating genetic parameters and predicting breeding values for birth weight and calving ease in Czech Charolais cattle. Czech J. Anim. Sci. 59:302-309.

Walmsley, B. J., S. J. Lee, P. F. Parnell, and W. S. Pitchford. 2018. A review of factors influencing key biological components of maternal productivity in temperate beef cattle. Anim. Prod. Sci. 58:1-19.

Waters, S. M., G. S. Coyne, D. A. Kenny, D. E. MacHugh, and D. G. Morris. 2012. Dietary n-3 polyunsaturated fatty acid supplementation alters the expression of genes involved in the control of fertility in the bovine uterine endometrium. Physiol. Genom. 44:878-888.

Waters, S. M., G. S. Coyne, D. A. Kenny, and D. G. Morris. 2014. Effect of dietary n-3 polyunsaturated fatty acids on transcription factor regulation in the bovine endometrium. Mol. Biol. Rep. 41:2745-2755.

Wang, G., Z. Wang, and H. Yu. 2020. Kinesin family member 2A high expression correlates with advanced tumor stages and worse prognosis in non-small cell lung cancer patients. J. Clin. Lab. Anal. 34:e23135.

Werner H. and D. Le Roith. 2000. New concepts in regulation and function of the insulin-like growth factors: implications for understanding normal growth and neoplasia. Cell Mol. Life Sci. 57:932–942.

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, W. G. Hill, and J. C. M. Dekkers. 2012. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Animal Genetics, 43, pp.87-96.

Wei, C., H. Luo, B. Zhao, K. Tian, X. Huang, Y. Wang, X. Fu, Y. Tian, J. Di, X. Xu, W. Wu, H. Tulafu, M. Yasen, Y., Zhang, and W. Zhao. 2020. The Effect of Integrating Genomic Information into Genetic Evaluations of Chinese Merino Sheep. Animals. 10:569.

Xu, H., J. Luo, G. Ma, X. Zhang, D. Yao, M. Li, and J. J. Loor. 2018. Acyl-CoA synthetase short-chain family member 2 (ACSS2) is regulated by SREBP-1 and plays a role in fatty acid synthesis in caprine mammary epithelial cells. J. Cell. Physiol. 233:1005-1016.

Zappaterra, M. 2017. Genomics and New Approaches to Study Complex Traits in Pigs and Other Livestock Species. A Focus on the Investigation of Gene Networks Related to Fat Quality and Deposition in Pigs and Preliminary Research to Study Factors Related to Performances in Piglets and Poultry. Ph.D. diss., Alma Mater Studiorum – Università di Bologna, Bologna, Italy.

Zerbino, D. R., P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, and L. Gil. 2018. Ensembl 2018. Nucleic Acids Research. 46:D754-D761.

Zhu, B., H. Niu, W. Zhang, Z. Wang, Y. Liang, L. Guan, P. Guo, Y. Chen, L. Zhang, Y. Guo, and H. Ni. 2017. Genome wide association study and genomic prediction for fatty acid composition in Chinese Simmental beef cattle using high density SNP array. BMC Genom. 18:464.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10:R42.

# APPENDIX A

Below are the results of an alternative RRM genetic evaluation for STAY at consecutive ages where the only change performed in relationship to the RRM used to evaluate STAY in chapter 4 is the grouping of observations of 11 and 12-yr-old cows. In summary, changes in the $h^2$ estimates obtained with this alternative RRM are depicted in Figure A-1. Estimations of phenotypic and genetic correlations are shown in Table A-1. Finally, estimations of permanent environmental correlations between consecutive STAY endpoints are shown in Table A-2.

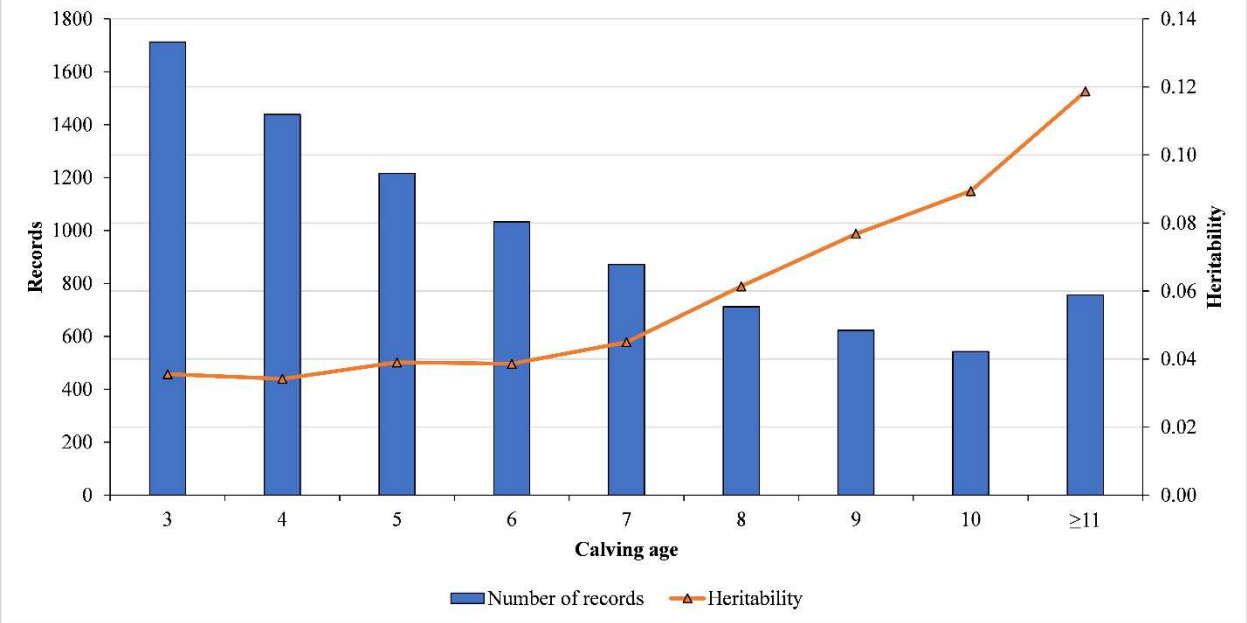


**Figure A-1.** Changes in heritability estimates for stayability and their relationship with the number of records at each endpoint in Angus cows when lumping ages of 11 and 12 yr together.

**Table A-1.** Genetic (above diagonal) and phenotypic (below diagonal) correlations for stayabilities to consecutive calvings (lumping ages of 11 and 12 yr together).

| Calving No | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | ≥11 |
|---|---|---|---|---|---|---|---|---|---|
| 3 | | 0.99 | 0.96 | 0.93 | 0.90 | 0.87 | 0.84 | 0.82 | 0.79 |
| 4 | 0.66 | | 0.99 | 0.98 | 0.96 | 0.93 | 0.91 | 0.89 | 0.87 |
| 5 | 0.51 | 0.65 | | 0.99 | 0.98 | 0.97 | 0.96 | 0.94 | 0.93 |
| 6 | 0.42 | 0.50 | 0.65 | | 0.99 | 0.99 | 0.98 | 0.97 | 0.96 |
| 7 | 0.35 | 0.40 | 0.48 | 0.59 | | 0.99 | 0.99 | 0.99 | 0.98 |
| 8 | 0.31 | 0.34 | 0.40 | 0.47 | 0.68 | | 0.99 | 0.99 | 0.99 |
| 9 | 0.28 | 0.31 | 0.35 | 0.40 | 0.55 | 0.65 | | 0.99 | 0.99 |
| 10 | 0.24 | 0.26 | 0.29 | 0.32 | 0.42 | 0.44 | 0.52 | | 0.99 |
| ≥11 | 0.23 | 0.24 | 0.25 | 0.27 | 0.33 | 0.30 | 0.32 | 0.48 | |

**Table A-2.** Permanent environmental correlations for stayabilities to consecutive calvings (lumping ages of 11 and 12 yr together).

| Calving No | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| 4 | 0.99 | | | | | | | |
| 5 | 0.97 | 0.99 | | | | | | |
| 6 | 0.92 | 0.96 | 0.99 | | | | | |
| 7 | 0.83 | 0.88 | 0.94 | 0.98 | | | | |
| 8 | 0.69 | 0.77 | 0.85 | 0.92 | 0.98 | | | |
| 9 | 0.53 | 0.63 | 0.73 | 0.83 | 0.92 | 0.98 | | |
| 10 | 0.37 | 0.47 | 0.59 | 0.71 | 0.83 | 0.93 | 0.98 | |
| ≥11 | 0.21 | 0.32 | 0.48 | 0.59 | 0.73 | 0.85 | 0.94 | 0.99 |

Below is described an alternative EPD calculation for STAY performed using a univariate repeatability threshold model (REP) along with a probit link function that transformed binary observations to an underlying normal distribution. The model Equation (A-1) was:

$$\boldsymbol{y^* = Xb + Z_1u + Z_2p + Qcg + e} \qquad \text{Eq. A-1}$$

where $\boldsymbol{y^*}$ corresponded to a vector of transformed observations of STAY on the underlying scale; $\boldsymbol{b}$ was a vector of unknown solutions for fixed effects, which included AFC, CE and the individual's breeding weight as a linear covariate; $\boldsymbol{u}$ corresponded to a vector of unknown solutions of animal random effects; $\boldsymbol{p}$ corresponded to a vector of unknown solutions of permanent environmental random effects; $\boldsymbol{cg}$ represented a vector of unknown solutions of contemporary group random effects; $\boldsymbol{X}$, $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$ and $\boldsymbol{Q}$ were known incidence matrices relating observations in $\boldsymbol{y^*}$ to fixed ($\boldsymbol{b}$), animal random ($\boldsymbol{u}$), permanent environment ($\boldsymbol{p}$) and contemporary group random ($\boldsymbol{cg}$) effects; and $\boldsymbol{e}$ was the vector of unknown residual errors. The mean for random effects was assumed to be 0 while variances were assumed to be distributed as:

$$\mathbf{Var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{p}\\ \boldsymbol{cg}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A\sigma_a^2} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_p\sigma_p^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_{cg}\sigma_{cg}^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I_n\sigma_e^2}\end{bmatrix}$$

where **A** represented the additive numerator relationship matrix amongst animals included in the pedigree; $\boldsymbol{I_p}$, $\boldsymbol{I_{cg}}$ and $\boldsymbol{I_n}$ were identity matrices with orders equal to the number of individuals, contemporary groups and observations, respectively. The $\boldsymbol{\sigma_a^2}$, $\boldsymbol{\sigma_p^2}$, $\boldsymbol{\sigma_{cg}^2}$ and $\boldsymbol{\sigma_e^2}$ denoted the additive, permanent environmental, contemporary group and residual variances, respectively. In this model, the residual variance ($\boldsymbol{\sigma_e^2}$) was constrained to be equal to 1. The prediction error variance of the i$^{th}$ animal (*PEVi*) was obtained by squaring the standard error reported next to the BLUP of each individual evaluated on the ASREML output solutions file (Gilmour et al., 2009).

These values represented approximations of the diagonal elements of the inverse of the coefficient matrix assembled in the final iteration round performed by the statistical software package. Calculation of the accuracy (ACC) of prediction was performed following the guidelines of the Beef Improvement Federation (2020) using Equation 3.7 and then, the resulting mean ACC of this model was compared to the mean accuracies of the evaluations presented in chapter 4. Mean accuracies of prediction obtained with the traditional threshold model (TM), the repeatability threshold model (REP) and the random regression model (RRM) are presented in Figure A-2 (A-D).

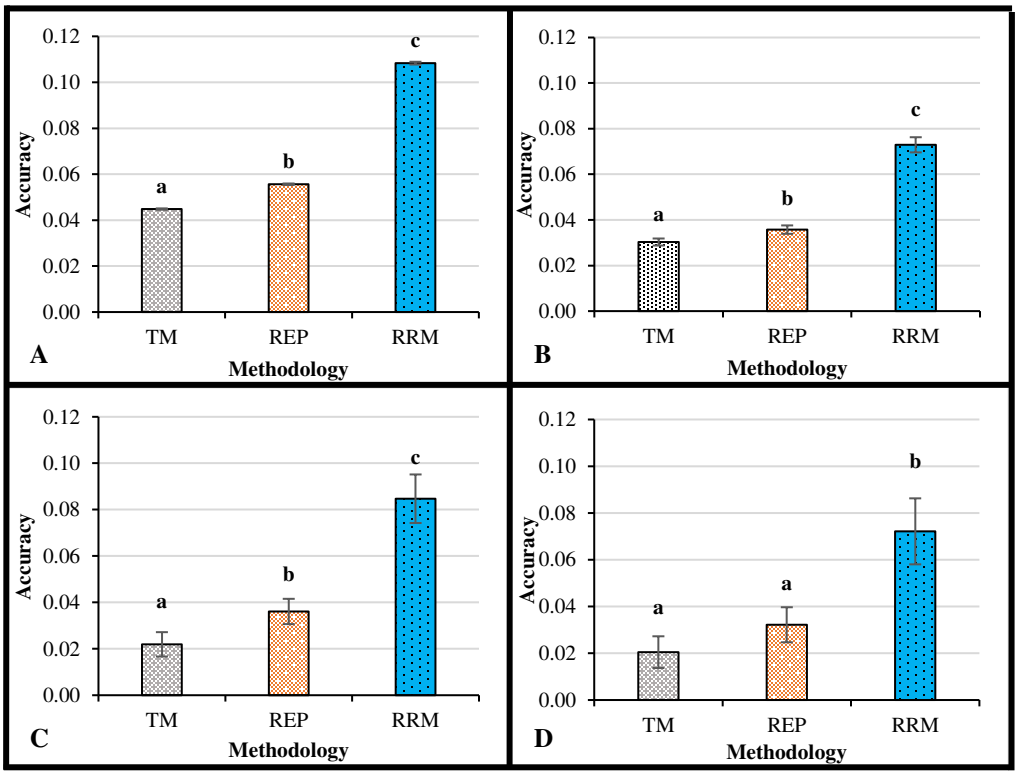


**Figure A-2.** Mean accuracies for stayability predictions at the age of 6 yr obtained with each statistical methodology. **A)** Mean accuracy for all animals in the pedigree (n = 14,140), **B)** Mean accuracies for all sires in pedigree (n = 971), **C)** Mean accuracies for sires that have produced progeny in the last five yr (n = 85), **D)** Mean accuracies for sires that have produced progeny in the last three yr (n = 51). Different letters indicate a statistical difference at the P <0.05 level among methodologies according to the least significant difference test.

Considering all animals in the pedigree (Figure A-2 A), the mean accuracy for STAY predictions obtained with the REP model was 0.056 with a minimum of 0.00004 and a maximum of 0.316. Similar increments in accuracies of predictions for all the sires in pedigree (Figure A-2 B) were obtained with REP, where the mean, minimum and maximum accuracy values were 0.036, 0.008 and 0.316. In the case of sires that have produced progeny in the last 5 yr within the CSU-BIC (Figure A-2 C), the average accuracy was 0.036, with values that ranged between 0.008 to 0.303. The last group animals whose mean accuracy values obtained by each method were compared was the sires that have produced progeny within the last 3 yr within the CSU-BIC (Figure A-2 D). For this group of animals, the mean, minimum and maximum accuracy values were 0.032, 0.008 and 0.303.

APPENDIX B

**Table B-1.** Number of genotyped animals per birth year at the Colorado State University – Beef Improvement Center.

| Year of birth | Genotyped animals | Year of genotyping |
|---|---|---|
| 1997 | 6 | 2011 |
| 1998 | 10 | 2011 |
| 1999 | 10 | 2011 |
| 2000 | 16 | 2011 |
| 2001 | 33 | 2011 |
| 2002 | 17 | 2011 |
| 2003 | 27 | 2011 |
| 2004 | 38 | 2011 |
| 2005 | 43 | 2011 |
| 2006 | 34 | 2011 |
| 2007 | 130 | 2011 |
| 2008 | 125 | 2011 |
| 2009 | 148 | 2011 |
| 2010 | 311 | 2011 |
| 2011 | 372 | 2011 |
| 2012 | 362 | 2012 |
| 2013 | 367 | 2013 |
| 2014 | 297 | 2014 |
| 2015 | 390 | 2015 |
| 2016 | 400 | 2016 |
| 2017 | 380 | 2017 |
| External sires | 105 | - |

*Genomic EPD quartile classification of heifers contributing phenotypes to DS-2_HPG*

After the genomic expected progeny differences (GEPD) were obtained for all animals contained in the final pedigree file of DS-2_HPG (n = 6,738) using the single-step random regression super-hybrid model (ssRR-SHM), a special emphasis was placed on the subset of 3,037 heifers that had phenotypes for HPG. Predictions for this group of animals were extracted with the purpose of matching them with their registered observation for HPG. Afterwards, heifers were ranked in quartiles according to their GEPD (Q1: ≥75%, Q2: ≥50% <75%, Q3: ≥25 to <50% and Q4: 0 to <25%) with the ultimate goal of exploring if the quartile classification effectively translated into expressed differences in phenotypic performance. A general linear model was executed to compare the adjusted success rate for HPG within each quartile-genetic group (SAS 9.3; SAS Inst.Inc., Cary, NC). The equation of such model (Eq. B-1) was the following:

$$\mathbf{y_{ijklm} = QG_i + CG_j + AOD_k + AFE_l + e_{ijklm}} \qquad \text{Eq. B-1}$$

where $\mathbf{y_{ijkl}}$ corresponded to the *m*th HPG phenotypic value associated to the *l*th age at first exposure, the *k*th age of dam category, the *j*th contemporary group and the *i*th quartile group; $\mathbf{QG_j}$ represented the *i*th quartile group (e.g., 4 classes: Q1 through Q4), $\mathbf{CG_j}$ denoted the *j*th contemporary group (e.g., 19 groups); $\mathbf{AOD_k}$ corresponded to the *k*th age of dam (e.g., 8 categories) and $\mathbf{AFE_l}$ was the *l*th covariate value of age at first exposure (ages ranging from 350 to 465 d); $\mathbf{e_{ijklm}}$ represented the residual term. Comparisons of the adjusted success rate for HPG were performed using the Tukey test. Average values for HPG genomic predictions, adjusted phenotypic success rate for HPG, age at first exposure and age of dam according to the quartile classification generated by the ssRR-SHM evaluation are shown in **Table B-2.** The adjusted average phenotypic performance for HPG of each quartile subgroup generated from this genomic evaluation is shown in **Figure B-1**.

**Table B-2.** Summary statistics according to the quartile classification derived from the single-step random regression super-hybrid model for heifer pregnancy (HPG).

| | Quartile classification for the 3,037 heifers that contributed phenotypic information for DS-2_HPG* | | | |
|---|---|---|---|---|
| Variable | Q1ᵠ | Q2ᵠ | Q3ᵠ | Q4ᵠ |
| N | 760 | 759 | 759 | 759 |
| Average GEPD for HPG | 2.10 | 1.11 | 0.44 | -1.00 |
| Average success rate for HPG | 0.98 | 0.97 | 0.83 | 0.35 |
| Averages AFE‡, d | 424.64 | 420.59 | 418.41 | 421.25 |
| Average AOD§, (yr) | 4.42 | 5.30 | 5.07 | 4.93 |

*Second dataset used in the random regression super-hybrid model genomic evaluation for heifer pregnancy, ᵠQ1: >75%, Q2: >50% to 75%, Q3: >25 to 50% and Q4: 0 to 25%, ‡AFE = age at first exposure, §AOD = age of dam.

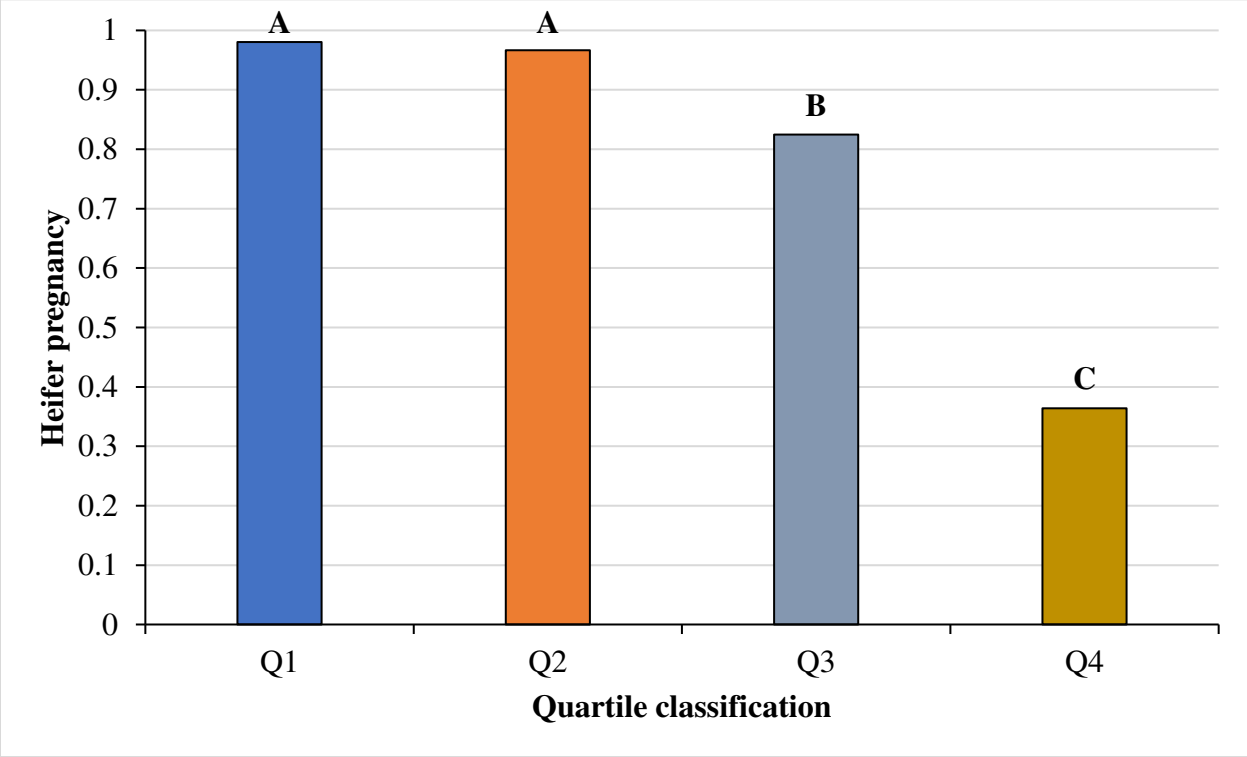


**Figure B-1.** Heifer pregnancy rate adjusted for non-genetic effects according to the quartile classification derived from the single-step random regression super-hybrid model for the 3,037 heifers that had genotypes and phenotypes within DS-2_HPG. Different letters indicate a statistical difference at the $P < 0.05$ level among quartiles according to the Tukey test.

*Genomic EPD quartile classification of heifers contributing phenotypes to DS-3_FSCR*

After the GEPD for heifer first-service conception rate (FSCR) were obtained for all animals contained in the final pedigree file of DS-3_FSCR (n = 6,764), a special emphasis was placed on the subset of 4,121 heifers that had phenotypic records for FSCR. Genomic predictions for this group of animals were extracted with the purpose of matching them with their registered phenotype for FSCR. Subsequently, heifers were ranked in quartiles according to their GEPD (Q1: ≥75%, Q2: ≥50% <75%, Q3: ≥25 to <50% and Q4: 0 to <25%) with the objective of exploring if the quartile classification effectively translated into expressed differences in phenotypic performance. Using SAS software, a general linear model was executed to compare the adjusted success rate for FSCR among quartile-genetic groups. The model equation (Eq. B-2) was as follows:

$$\mathbf{y_{ijklmnop} = QG_i + CG_j + AOD_k + MG_l + AIT_m + SS_n + AFE_o + e_{ijklmnop}} \quad \text{Eq. B-2}$$

where $\mathbf{y_{ijklmnop}}$ corresponded to the *p*th FSCR phenotypic value associated to the *o*th age at first exposure, the *n*th service sire, the *m*th artificial insemination technician, the *l*th mating group, the *k*th age of dam category, the *j*th contemporary group and the *i*th quartile group; $\mathbf{QG_j}$ represented the *i*th quartile group (e.g., 4 classes: Q1 through Q4), $\mathbf{CG_j}$ denoted the *j*th contemporary group (e.g., 26 groups); $\mathbf{AOD_k}$ corresponded to the *k*th age of dam (e.g., 8 categories), $\mathbf{MG_l}$ was the *l*th mating group (e.g., 2 classes), $\mathbf{AIT_m}$ corresponded to the *m*th AI technician (e.g., 51 technicians), $\mathbf{SS_n}$ represented the *n*th service sire (e.g., 44 sires) and $\mathbf{AFE_o}$ was the *o*th covariate value of age at first exposure (ages ranging from 350 to 465 d); $\mathbf{e_{ijklmnop}}$ represented the residual term. Comparisons of the adjusted mean phenotypic performance for FSCR were performed using the Tukey test. The average values for FSCR genomic predictions, adjusted FSCR phenotypic performance, age at first exposure and age of dam according to the quartile classification generated

by the ssRR-SHM evaluation are shown in **Table B-3.** The adjusted average phenotypic performance for FSCR of each quartile subgroup generated from this genomic evaluation is shown in **Figure B-2**.

**Table B-3.** Summary statistics according to the quartile classification derived from the single-step random regression super-hybrid model for heifer first-service conception rate (FSCR).

| | Quartile classification for the 4,121 heifers that contributed phenotypic information for DS-3_FSCR* | | | |
|---|---|---|---|---|
| Variable | Q1ᴪ | Q2ᴪ | Q3ᴪ | Q4ᴪ |
| N | 1031 | 1030 | 1030 | 1030 |
| Average GEPD for FSCR | 2.55 | 2.24 | 1.99 | 1.33 |
| Average FSCR rate | 0.66 | 0.46 | 0.33 | 0.22 |
| Averages AFE‡, d | 419.66 | 420.25 | 421.79 | 426.64 |
| Average AOD§, (yr) | 4.69 | 4.93 | 4.85 | 4.86 |

*Third dataset used in the random regression super-hybrid model genomic evaluation for heifer firs-service conception rate, ᴪQ1: >75%, Q2: >50% to 75%, Q3: >25 to 50% and Q4: 0 to 25%, ‡AFE = age at first exposure, §AOD = age of dam.

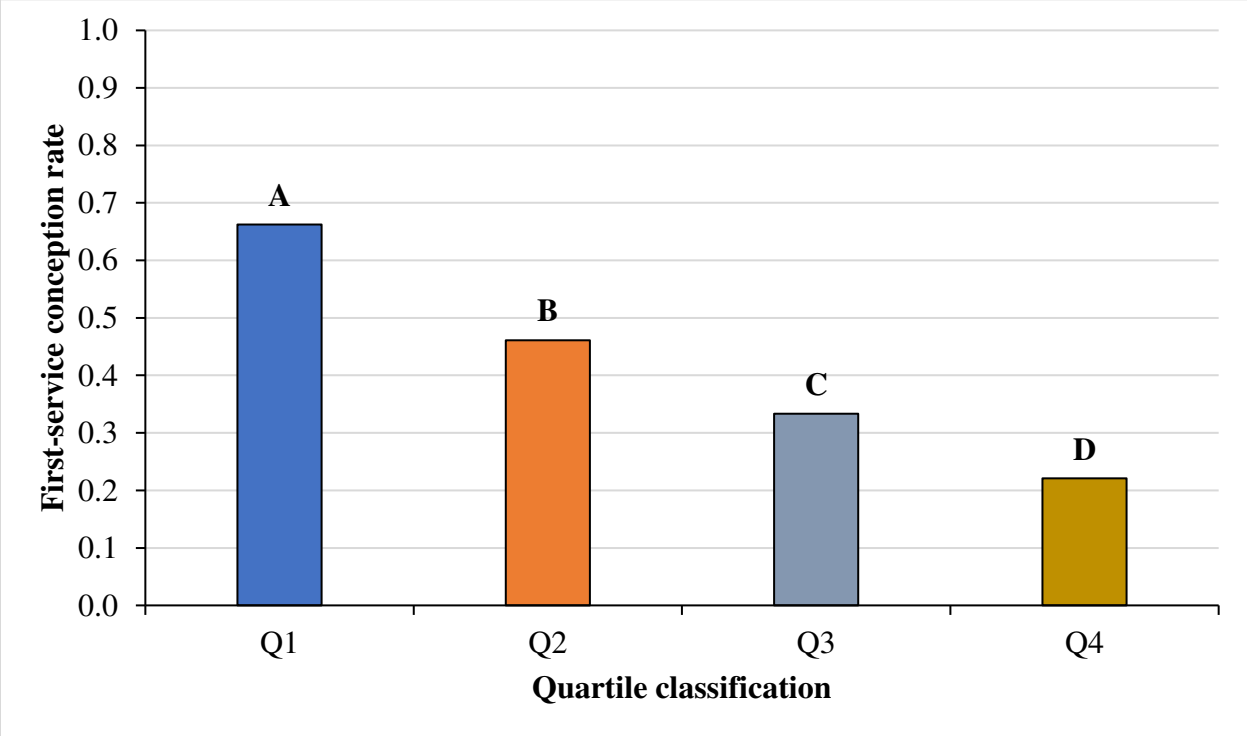


**Figure B-2.** Heifer first-service conception rate adjusted for non-genetic effects according to the quartile classification derived from the single-step random regression super-hybrid model for the 4,121 heifers that had genotypes and phenotypes within DS-3_FSCR. Different letters indicate a statistical difference at the $P < 0.05$ level among quartiles according to the Tukey test.

*Genomic EPD quartile classification of heifers contributing phenotypes to DS-4_STAY*

After the GEPD were obtained for all animals contained in the final pedigree file of DS-4_STAY (n = 2,340), a special emphasis was placed on a subset of 497 cows that had phenotypic records for stayability at the age of 6 (STAY06). Importantly, even when in this dataset there was a total of 668 cows with phenotypic records, only the subset of 497 cows selected for this analysis were old enough to express STAY06 phenotypes. Predictions for this group of animals were extracted and matched with their registered observation of the age endpoint of interest (e.g., 6 yr). Subsequently, cows were ranked in quartiles according to their GEPD for STAY06 (Q1: ≥75%, Q2: ≥50% <75%, Q3: ≥25 to <50% and Q4: 0 to <25%) to investigate if such quartile classification actually translated into expressed differences in phenotypic performance. A general linear model was executed to compare the proportion of cows that actually remained productive in the herd until the age of 6. The model equation (Eq. B-3) was the following:

$$\mathbf{y_{ijkl} = QG_i + CG_j + AFC_k + e_{ijkl}} \qquad \text{Eq. B-3}$$

where $\mathbf{y_{ijkl}}$ corresponded to the *l*th STAY phenotypic value associated to the *k*th age at first calving registered, the *j*th contemporary group and the *i*th quartile genetic group; $\mathbf{QG_i}$ represented the *i*th quartile genetic group (e.g., 4 classes: Q1 through Q4), $\mathbf{CG_j}$ corresponded to the *j*th contemporary group (e.g., 8 groups) and $\mathbf{AFC_k}$ denoted the *k*th age at first calving (e.g., 6 categories); $\mathbf{e_{ijkl}}$ represented the residual term. Comparisons of the adjusted mean phenotypic performance for STAY06 were performed using the Tukey test. As important notes, breeding weight values were not included as explanatory variable in this model because only 25 animals had observations associated to the age endpoint of interest (e.g., 6 yr of age). In the case of calving ease scores, they were not included in the model because there was no variability in the observations of this variable related to STAY06 (e.g., all scores were 1)**.**

The average values for STAY06 genomic predictions, the adjusted phenotypic success rate for STAY06 and the average age at first calving according to the quartile classification generated by the ssRR-SHM evaluation are shown in **Table B-4** and **Figure B-3**.

**Table B-4.** Summary statistics according to the quartile classification derived the single-step random regression super-hybrid model for stayability at the age of 6 (STAY06).

| | Quartile classification for the 497 cows that contributed phenotypic information for DS-4_STAY* | | | |
|---|---|---|---|---|
| Variable | Q1ᴪ | Q2ᴪ | Q3ᴪ | Q4ᴪ |
| N | 125 | 124 | 124 | 124 |
| Average GEPD for STAY06 | 8.10 | 6.24 | 4.80 | 2.22 |
| Phenotypic success rate for STAY06 | 0.91 | 0.82 | 0.47 | 0.10 |
| Averages AFC‡, mo | 23.81 | 23.72 | 23.70 | 23.63 |

*Fourth dataset used in the random regression super-hybrid model genomic evaluation for stayability, ᴪQ1: >75%, Q2: >50% to 75%, Q3: >25 to 50% and Q4: 0 to 25%, ‡AFC = age at first calving.

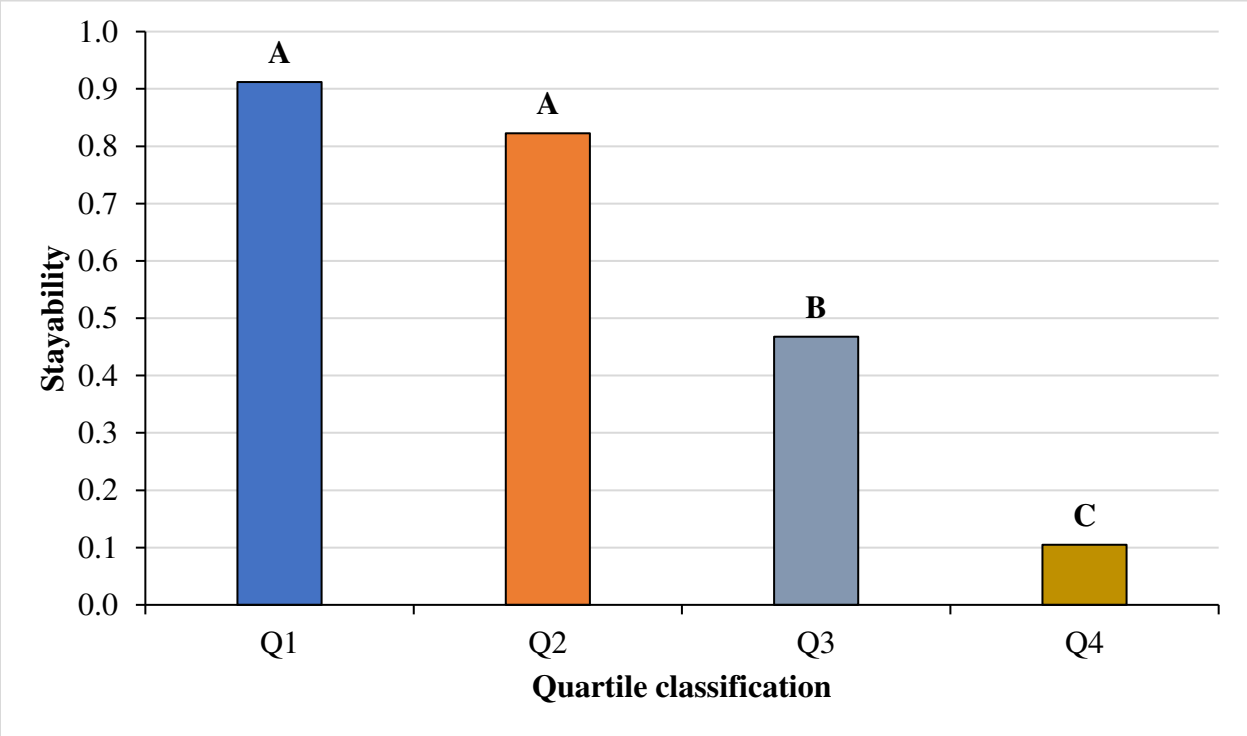


**Figure B-3.** Proportion of cows remaining productive in the herd until the age of 6 (stayability at the age of 6) adjusted for non-genetic effects according to the quartile classification derived from the single-step random regression super-hybrid model for the 497 cows that had genotypes and phenotypes within DS-4_STAY. Different letters indicate a statistical difference at the $P < 0.05$ level among quartiles according to the Tukey test.