



Practical implications for genetic modeling in the genomics era¹

P. M. VanRaden²

Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

Genetic models convert data into estimated breeding values and other information useful to breeders. The goal is to provide accurate and timely predictions of the future performance for each animal (or embryo). Modeling involves defining traits, editing raw data, removing environmental effects, including genetic by environmental interactions and correlations among traits, and accounting for nonadditive inheritance or nonnormal distributions. Data include phenotypes and pedigrees during the last century and genotypes within the last decade. The genomic data can include single nucleotide polymorphisms, quantitative trait loci, insertions, deletions, and haplotypes. Subsets must be selected to reduce computation because total numbers of variants that can be imputed have increased rapidly from thousands to millions. Current computation using 60,671 markers takes just a few days. Nonlinear models can account for the nonnormal distribution of genomic effects, but reliability is usually better than that of linear models only for traits influenced by major genes. Numbers of genotyped animals have also increased rapidly in the joint North American database from a few thousand in 2009 to over 1 million in 2015. Most are young females and will contribute to estimating allele effects in the future, but only about 150,000 have phenotypes so far. Genomic preselection can bias traditional animal models because Mendelian sampling of phenotyped progeny and mates is no longer expected to average zero; however, estimates of bias are small in current US data. Single-step models that combine pedigree and genomic relationships can account for preselection, but approximations are required for affordable computation. Traditional animal models may include all breeds and crossbreds, but most genomic evaluations are still computed within breed. Models that include inbreeding, heterosis, dominance, and interactions can

improve predictions for individual matings. Multitrait genomic models may be preferred for traits with many missing records or when foreign records are included as pseudo-observations, but most countries use multitrait traditional evaluations followed by single-trait genomic evaluations. Genomic reliabilities are about 70% for the more heritable traits. Researchers must choose from many available models and explain how the models work so that breeders can more confidently apply the predictions in their selection programs.

Key words: genetic evaluation, genomic selection, mixed models, multitrait

INTRODUCTION

Genetic markers are central to dairy cattle selection programs, allowing accurate and affordable prediction of each animal's merit using tens of thousands of genotypes. Many new issues arise when models use genotypes along with phenotypes and pedigrees, but many previous principles of modeling remain true. Models must still separate genetic from environmental effects on traits, and genetic effects are now further separated and tracked across each chromosome by markers. Since 2008, rapidly growing genomic data sets and changes in selection programs require constant updating of evaluation systems. More available choices and algorithms allow researchers to improve accuracy and control bias in genetic rankings while adding more data.

Modeling involves defining traits, editing raw data, adjusting for environmental effects, including genetic by environmental interactions and correlations among traits, and accounting for nonadditive inheritance or nonnormal distributions. Major previous advances in US evaluation models were use of daughters' average production adjusted for dams' production in 1937, use of herd-year-season groups and heritability in 1962, use of sire and maternal grandsire pedigrees in 1974, use of all relatives in 1989, use of foreign data from multi-trait, across-country evaluation (**MACE**) in 1995, and use of a multibreed model in 2007. Similar advances in modeling occurred in many other countries. The goal of modeling is to provide accurate and timely predictions of the future performance for each animal (or embryo).

Received June 30, 2015.

Accepted November 16, 2015.

¹Presented as part of the Relevance of Modeling in the Genomics Era Symposium at the ADSA-ASAS Joint Annual Meeting, Orlando, Florida, July 2015.

²Corresponding author: Paul.VanRaden@ars.usda.gov

Accurate estimates of the environmental effects may also be helpful in management decisions and benchmarking to compare nongenetic factors, but this report will focus on genetics.

Goals of this report were to (1) summarize the choices of data available for genomic modeling, (2) compare methods to test different models for predictive ability, and (3) describe how changes in breeding programs and available data may require changes in models.

GENETIC MODELING

Data

Genotypes can include SNP, QTL, insertions, deletions, and haplotypes; total numbers of known variants have increased rapidly since 2000 from hundreds to thousands to millions (Table 1). About 30 million total variants have been identified in cattle (Daetwyler et al., 2014) and could, in theory, be imputed for each genotyped animal. Thus, imputation can generate far more data than any routine analysis could affordably include. Research, therefore, focuses on estimating effects of many variants from progeny-tested bulls or phenotyped cows and then including the variants with larger estimated effects on future chips and in routine evaluations (Hayes et al., 2014; Wiggans et al., 2014; Brøndum et al., 2015).

The variants included are increasingly chosen from previous estimates or bioinformatics instead of additional random markers. Until 2013, patents prevented some variants from being used, but patents on naturally occurring variants are no longer valid in the United States (US Supreme Court, 2013). As costs of genotyping by sequencing decline, new variants and mutations could be detected with each new animal. Whole-genome sequencing can read all DNA at various depths of coverage, and exome sequencing can read just the sections of DNA that code for expressed genes (McClure et al., 2014).

Pedigrees were the basis of selection for many decades. Accuracy was excellent for bulls in AI service and for embryo transfer (ET) calves because of careful checking but was not so good for commercial cows. Ge-

netomic testing can discover or correct ancestry because nearly all AI sires and maternal grandsires are now genotyped for dairy cattle in many countries (Tooker et al., 2015). Evaluation models have not accounted for uncertain paternity and thus have underestimated the true genetic differences among such sires (Foulley et al., 1987), but paternity is certain when daughters are genotyped. Correct pedigrees are useful in imputation, avoiding inbreeding, and in quality control of the genotyping process, such as to avoid switched samples, but are no longer as important in modeling because pedigree relationships are being replaced by genomic relationships.

Phenotypes can be defined in many ways. A main goal is consistent definition across time and countries so that larger data sets can be used to estimate small effects of individual genes. Foreign data can improve reliability if traits have high correlations across countries (Lund et al., 2011), and this has led to widespread international genotype exchanges. Historical phenotypes can increase reliability if the recent reference population is small (Cooper et al., 2015) but could decrease reliability if generations of recombination have changed the linkage patterns or allele frequencies or if the traits have changed over time (Lourenco et al., 2014). Selection previously emphasized traits with high heritability, but large reference populations now allow progress from genomic selection even for traits with low heritability.

Environmental factors such as herd-year-season, age, parity, days in milk, and milking frequency were included in models or used to preadjust data for many decades. Data edits for phenotypes are important but are not affected by the addition of genotypes, and so previous methods should apply. Edits for genotypes are becoming more complex as new chips, new variants, and sequence data are added. Each of the 17 chips now included has different patterns of missing data and often different names for the same variant, requiring complex merges and edits (Nicolazzi et al., 2014) because error rates can also differ by chip or source of data. Imputation is now a key step in the US evaluation because only 25% of all animal genotypes are measured and 75% are missing and imputed. Imputation uses statistical methods and pedigrees to phase observed higher-

Table 1. Growth in number of variants and animals available to estimate genetic effects

Year	Variants	Reference animals ¹	Example reference
1991–2004	367	1,415	Ashwell et al., 2004
2007–2010	42,503	16,646	Wiggans et al., 2011
2011–2013	636,967	15,842	VanRaden et al., 2013
2014–future	28,300,000	234 + imputed	Daetwyler et al., 2014

¹Holstein reference animals in April 2015 included 27,464 progeny-tested bulls and 136,184 phenotyped cows.

density genotypes into haplotypes (segments of shared DNA) and then choose the 2 most likely haplotypes inherited by each animal genotyped at lower density.

Numbers of available genotyped animals are also increasing rapidly, nearly doubling each year over the past 6 yr from a few thousand in 2009 to >1 million in 2015 (Figure 1). The database includes genotypes from 48 countries, including about 90,000 animals from Europe, but most are from North America (Tooker et al., 2015). Most are young females and will contribute to estimating allele effects in the future, but only about 200,000 females have US phenotypes so far. Eventually most heifer calves could be genotyped, and the database may contain >10 million genotyped animals within a decade or two. Traditional genetic evaluations retained phenotypes and pedigrees for >50 yr, and genomic evaluation algorithms must be more efficient if all historical genotypes are retained and used. Currently, 4 breeds are evaluated, including 903,120 Holsteins, 116,042 Jerseys, 19,457 Brown Swiss, and 4,715 Ayrshires (including Scandinavian red dairy cattle), plus 2,281 Guernseys that could be evaluated in the near future. Data storage and processing requirements expand by number of animals \times variants genotyped, or number of animals \times depth of coverage (often about 10) \times length of genome (about 3 billion) for sequence data.

Models, Parameters, and Computation

Genomic predictions often model marker genotypes using multiple linear regressions and a normal distribution for the marker effects (genomic BLUP; **GBLUP**) but can also use many different prior distributions. Theory and results for a wide range of such models were reviewed by de los Campos et al. (2013). When estimating effects of sequence variants, the prior distri-

bution may also account for predicted functional effects of genes using bioinformatics (Hayes et al., 2014). Variants should always be treated as random rather than fixed because even if effects are known to be large for some traits, they may be very small for others. Because numbers of available variants now far exceed the ability to impute and use all in routine evaluations, genomic evaluations must now choose and use subsets of the known variants. Traditional evaluations generally included all data from all animals, subject only to edits on data quality, because the inverse of the pedigree relationship matrix required little computation (Henderson, 1976).

Multitrait genomic models are preferred for traits with many missing records, high correlations among traits, or when foreign records are included as pseudo-observations. Most countries use multitrait traditional evaluations (VanRaden et al., 2014), but few use multitrait genomic evaluations (Calus and Veerkamp, 2011; Tsuruta et al., 2011). Similarly, most countries use random regression models with additional genetic effects (Schaeffer, 2004) in the conventional evaluation, but use only 1 genetic effect per trait in the genomic model. A main reason is that foreign bulls have only 1 EBV per trait in MACE. Single-step test-day models are possible (Koivula et al., 2015), but methods to incorporate foreign data may be more difficult than those used for single-step lactation models (Přibyl et al., 2013).

Unknown-parent groups should be defined differently for traits with many decades of records than for traits recorded only recently. Young animals without records were often excluded from conventional models and then their parent averages calculated separately afterward. Single-step models may include these recent animals because some are genotyped and affect the accuracy of their nongenotyped parents, but separate unknown-parent groups should not be defined for the young animals if no phenotypes are available to estimate the effects. Misztal et al. (2013) compared several different grouping strategies for models including genomic information and recommended strategies to avoid biases when the genotyped animals have incomplete pedigrees.

Nonlinear models have been used to account for the nonnormal distribution of genomic effects, but precision is better than that of linear models only for traits influenced by major genes. Few routine genomic evaluations have attempted to account directly for nonnormality of categorical trait phenotypes such as stillbirth, calving ease, or longevity, because multistep genomic models have used deregressed breeding values. Nonparametric models and machine learning algorithms can be used for categorical traits (González-Recio and Forni, 2011), but the predictions may then be more difficult to

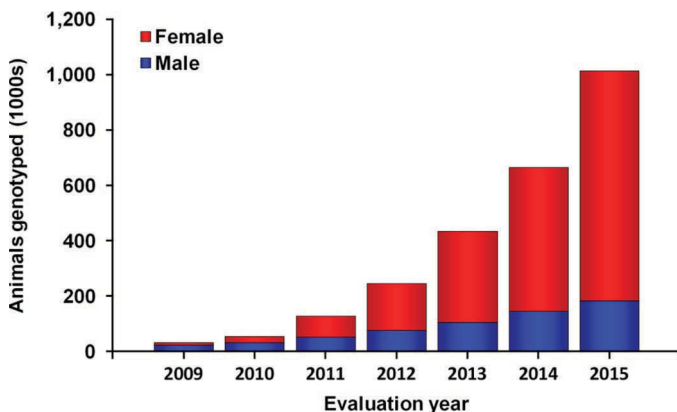


Figure 1. Cumulative numbers of genotyped males and females by year. Color version available online.

explain. Single-step models could, in theory, account for nonnormality of phenotypes, but most are animal models whereas categorical or survival models are often sire models.

Genomic preselection may bias estimates of environmental effects, genetic trend, and individual EBV from the conventional animal models used in multi-step methods (Patry and Ducrocq, 2011). Mendelian sampling of phenotyped progeny and herd mates is no longer expected to average zero, and conventional models do not adjust for genomic merit of mates. Single-step models that combine pedigree and genomic relationships (single-step GBLUP; Aguilar et al., 2010) can account for the preselection (Vitezica et al., 2011). Approximate formulas must be used in computing inverses when numbers of genotyped animals exceed 100,000 (Misztal et al., 2014), but such approximations may cause little or no loss of reliability (Fragomeni et al., 2015). Modified single-step GBLUP equations can use deregressed conventional breeding values instead of raw phenotypes to retain features of previous software more easily (Mäntysaari et al., 2011; Winkelman et al., 2015), or genomic pseudo-observations can be added as a correlated trait (Stoop et al., 2013), but those approaches may not account for selection as well.

Potential biases in US evaluations due to preferential treatment or genomic selection of females were estimated by comparing percentages of embryo transfer daughters, percentages of genotyped daughters, and merit of mates for bulls born 2010 and later that had sired >10 US daughters. The top 50 young genotyped Holstein bulls ranked on December 2014 net merit (NM\$) were compared. Future biases can be forecast for such bulls because genomic preselection of mates has already occurred and can be accurately measured when calves are born, even if those heifers do not have phenotypes yet. Potential bias from genomic preselection of a bull's mates was calculated as the average difference of genomic EBV (**GEBV**) – EBV from December 2014 data for the genotyped mates, multiplied by fraction of mates genotyped. Comparing preselected bulls primarily to each other can also bias estimates of genetic trend (Patry and Ducrocq, 2011), but such biases were not examined in this study.

Nonadditive genetic effects such as inbreeding depression and heterosis (VanRaden et al., 2014) or dominance (Sun et al., 2013) can improve accuracy compared with additive-only models. Traditional animal models may include all breeds and crossbreeds, but most genomic evaluations are still computed within breed. As more animals are genotyped, inbreeding depression and heterosis adjustments could use genomic instead of pedigree estimates. Scaling genomic relationships

to match pedigree relationships is more difficult with multiple breeds. Winkelman et al. (2015) reported a lack of convergence using genomic relationships among multiple breeds, but found that equations using a Euclidean distance matrix did converge. A better strategy may be to model marker effects as correlated instead of the same for different breeds, with additional math to account for crossbreeds (Makgahlela et al., 2013).

Mating programs using genomic rather than pedigree relationships can be very profitable (Sun et al., 2013), and dominance effects can be included with little extra computation but also perhaps little extra gain if dominance variance is small. Culling guides can also be a valuable product of the evaluation, using accumulated data for an individual to predict its own future phenotypes rather than the next generation (Kelleher et al., 2015). Currently few multistep evaluations include genomic information in predicted producing ability, but single-step evaluations could do this automatically.

Genotype-by-environment interactions can be modeled using discrete classifications such as sire by country or sire by herd or as random regressions on factors such as heat stress, herd level, or feeding system (pasture vs. intensive). Interactions may be important for some traits but were not accurately estimated from 100 progeny-test daughters because interaction variance is often much smaller than additive genetic variance (Zwald et al., 2003). However, genomic estimates of genotype-by-environment interaction might be fairly accurate even for young calves if large reference populations are available and sufficient animals that share genotypes have phenotypes measured across environments. Large reference populations allow predicting more terms in the model and more traits because genomic relationships measure actual proportions of alleles shared and thus capture information more efficiently than pedigree relationships, which measure only expected averages of alleles shared. The upper limit on reliability for predicting merit of calves is 50% with pedigree but is nearly 100% with genomic models.

The rapid processing needed for weekly genomic predictions may limit the complexity of models or require approximations. Although DNA samples can be obtained at birth and breeding decisions can wait until 1 yr of age, marketing and transfer of ownership happens quickly for top animals. Also, many herd owners now also use genomics as a culling tool, selling calves with low genomic predictions early in life instead of raising them for replacements (Weigel et al., 2012). High labor costs for calves during their first month created a demand for weekly evaluations, and these are computed only for animals with new genotypes (Wiggans et al., 2015). All genotypes are reprocessed each month, and

phenotypes are updated every 4 mo together with international phenotypes provided by MACE.

Prediction accuracy in dairy cattle is usually tested by predicting the latest 4 yr of progeny-tested bulls from data with the last 4 yr truncated (Mäntysaari et al., 2010). Often, this same test is used to choose models and adjust parameters to improve prediction accuracy and minimize biases. For example, ratios of polygenic to marker variance can be set using validation results instead of variance estimation (Liu et al., 2011). Two other common validation methods are using current evaluations with truncation of only recent bulls rather than recent data, or using cross-validation with random subsets of data rather than time truncation, but both methods tend to give inflated predictions of future success.

Standard validation tests the ability of genomic equations to predict 1 generation ahead, but many calves (or embryos) are now 2 or 3 generations away from the progeny-tested bulls used to compute the predictions. Also, only those bulls with the highest predicted merit now obtain daughters. This can greatly reduce the variance of true genetic merit and the correlations of predicted with true merit needed to estimate the reliability of the prediction system. Adjustments to account for this reduced variance are needed to accurately compute observed reliabilities.

PRACTICAL IMPLICATIONS

Growth of Modeling

Rapid growth of genomic databases can quickly cause previously tested statistical methods to become outdated. Computer programs also must be constantly revised to keep up with database sizes that include more genotyped animals and differing numbers of genotyped variants per animal. Conventional animal models adjust each animal's EBV for merit of parents, progeny, mates, and herd mates, whereas single-step GBLUP more precisely adjusts for their GEBV instead of EBV, thereby avoiding genomic preselection biases that can cause average GEBV to differ from EBV. No models may fully account for lack of random sampling, ET, and extreme prices that may cause preferential treatment. In the past, use of highly selected foreign bulls sometimes caused similar biases but with little effect on bulls tested in domestic programs.

Random-sample progeny testing previously ensured that every bull's first-crop daughters were part of a well-designed experiment to estimate genetic merit. Now, many of the best young bulls are used for elite matings and marketed individually instead of as a

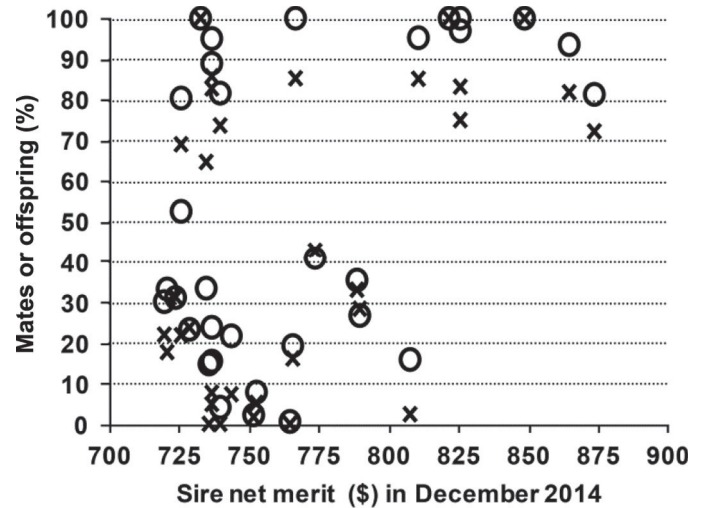


Figure 2. Percentages of mates with genotypes (O) or offspring from embryo transfer (X) for top young bulls based on December 2014 net merit.

group. Their first-crop evaluations are more similar to the second-crop evaluations of traditional selection. High percentages of the daughters of some bulls are from ET or from genotyped mates (Figure 2), whereas most others have low percentages. However, differences are very small between average GEBV and EBV for mates of the young bulls in Figure 2, with maximum potential biases in the bulls' published evaluations of only \$11 net merit, 1 kg of protein, 0.1 mo of productive life, and 0.1% daughter pregnancy rate due to mates' merit.

Adjustments for percentage of ET daughters have been implemented in Canada to account for preferential treatment (Beavers and Van Doormaal, 2014), but initial results from US yield traits indicate little current bias. Evaluations were stable across time for recent bulls that had a high percentage of ET for early daughters that then transitioned to lower percentage of ET daughters, with as many bulls increasing or decreasing when percentage of ET daughters decreased. However, other potential biases could emerge in the future as breeding programs continue to change. For example, use of the top genotyped heifers as embryo donors instead of for milk production could cause a negative bias in their sire's evaluation.

Computation using many conventional models is affordable with very large data sets because the pedigree relationship inverse is sparse. Genomic computation with multistep models is also affordable using iteration for marker effects, even with large reference populations. Completing a full evaluation currently takes about 5 d using <30 processors (Table 2). Single-step

Table 2. Current computation required for multistep reprocessing of all data every 4 mo and weekly updates for newly genotyped animals

Step	Time (h)	Processors	Memory (Gb)
Extract data and pedigree	10	7	1
Calculate inbreeding and heterosis	8	2	5
Compute conventional EBV for 12 traits	24	12	40
Impute genotypes (using prior haplotypes)	48	20	75
Compute genomic EBV (GEBV) for 27 traits	10	27	54
Compare genomic and pedigree relationships	6	20	75
Distribute final results	5	1	1
Total time for full evaluations	111		
Compute weekly GEBV for new animals	1.4	10	50

models were limited by the need to invert the genomic relationship matrix, but approximate methods to include that inverse should make single-step GBLUP affordable with many more genotyped animals (Misztal et al., 2014). Bayesian models that used Monte Carlo methods often required much more computation than GBLUP, but nonlinear iterative algorithms have given much of the same benefit in official US evaluations with very little extra computation (VanRaden, 2008). Global analysis of raw data is still not practical, and so national GEBV for young Holstein bulls from 11 populations are combined in a meta-analysis known as genomic MACE (**GMACE**; Sullivan et al., 2011), analogous to MACE (Schaeffer, 1994), that combines rankings across countries for bulls with daughter records. A preferred alternative to GMACE is to compare foreign and domestic young bulls directly on the same scale by exchanging genotypes.

Predictive Ability

Validation results can become outdated because of rapid growth of genomic databases and changes in selection programs and, therefore, are rechecked at least every 2 yr by Interbull (Uppsala, Sweden) member countries. For example, when genomic data sets were smaller, addition of genotypes from all bulls including very old bulls increased reliability. More recent research found that the oldest bulls could be dropped with little loss of accuracy because genotypes for many more recent bulls are available that contain the same DNA. In fact, for the more heritable traits, dropping all of the 27,229 progeny-tested bull genotypes causes little loss because the reference population now also includes >150,000 genotyped cows with records (Cooper et al., 2015). Early tests showed that adding highly selected females did not improve accuracy except with special adjustments to reduce bias (Wiggans et al., 2010), but most females now are genotyped as heifers before their phenotypes arrive, reducing preferential treatment bias.

The success of genomics has reduced our ability to test how well it works. The standard test predicts independent data such as daughter yield deviations or deregressed proofs from nongenomic models. If those statistics contain biases from preselection, regressions of later on earlier GEBV will be a less independent but sensible alternative test of success. An advantage of tests using only GEBV is that these use public information more easily understood by breeders.

Reliabilities of genomic predictions for young animal have steadily improved since their introduction in 2009. For most major Holstein populations, reliabilities now average 65 to 75% for more heritable traits and 45 to 65% for traits with lower heritability, and correspond fairly closely to a nonlinear function of the sum of reliabilities of reference bull EBV (Sullivan and Jakobsen, 2014). Reliabilities of other breeds such as Brown Swiss and Jersey are lower but have also improved with international exchanges. For all breeds, inclusion of more cows in the reference should continue this upward trend. Genomic modeling may be more difficult when many individuals each contribute only a little information, but many phenotypes are needed to accurately estimate effects from many genes.

CONCLUSIONS

Modeling is a key step in animal breeding because the model determines how the mixed model equations will convert the genomic, phenotypic, and pedigree files into predictions. Algorithms and computer resources are becoming limiting factors as numbers of animals and markers genotyped both expand rapidly. Results from validation tests are very helpful in choosing models and setting parameters. The best models use all data to predict future animal performance with the smallest errors and biases. As models become more complex, researchers must be able to explain how the models work so that breeders can more confidently apply the predictions in their selection programs. Breeders may

need additional supporting information and education to understand how well the theoretical methods work in practice.

ACKNOWLEDGMENTS

The author thanks the Council on Dairy Cattle Breeding (Bowie, MD) for supplying data under USDA Nonfunded Cooperative Agreement 58-1245-3-228N; S. M. Hubbard, J. R. Wright, and M. E. Tooker (Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD) for technical manuscript review and contributions to the research; and 3 anonymous reviewers for many helpful comments. The figure was created using Daniel's XL Toolbox add-in for Excel (version 6.60) by Daniel Kraus (Würzburg, Germany). This research was supported by appropriated project 1265-31000-096-00, "Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information," of the USDA's Agricultural Research Service. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

REFERENCES

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.
- Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassell, Y. Da, P. M. VanRaden, M. Ron, J. I. Weller, and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Dairy Sci.* 87:468–475.
- Beavers, L., and B. Van Doormaal. 2014. Steps to reduce bias in genetic and genomic evaluations. Accessed Jun. 10, 2015. <https://www.cdn.ca/document.php?id=362>.
- Brøndum, R. F., G. Su, L. Janss, G. Sahana, B. Guldbandsen, D. Boichard, and M. S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* 98:4107–4116.
- Calus, M. P. L., and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43:26.
- Cooper, T. A., G. R. Wiggans, and P. M. VanRaden. 2015. Short communication: Analysis of genomic predictor population for Holstein dairy cattle in the US—Effects of sex and age. *J. Dairy Sci.* 98:2785–2788.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. Van Tassell, I. Hulsegge, M. E. Goddard, B. Guldbandsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865.
- de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–345.
- Foulley, J. L., D. Gianola, and D. Planchenault. 1987. Sire evaluation with uncertain paternity. *Genet. Sel. Evol.* 19:83–102.
- Fragomeni, B. O., D. A. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Lawlor, 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.
- González-Recio, O., and S. Forni. 2011. Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. *Genet. Sel. Evol.* 43:7.
- Hayes, B. J., I. M. MacLeod, H. D. Daetwyler, P. J. Bowman, A. J. Chamberlain, C. J. Vander Jagt, A. Capitan, H. Pausch, P. Stothard, X. Liao, C. Schrooten, E. Mullaart, R. Fries, B. Guldbandsen, M. S. Lund, D. A. Boichard, R. F. Veerkamp, C. P. Van Tassell, B. Gredler, T. Druet, A. Bagnato, J. Vilkki, D. J. deKoning, E. Santus, and M. E. Goddard. 2014. Genomic prediction from whole genome sequence in livestock: The 1000 bull genomes project. Commun. 183 in Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, BC, Canada. Am. Soc. Anim. Sci., Champaign, IL.
- Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69–83.
- Kelleher, M. M., P. R. Amer, L. Shaloo, R. D. Evans, T. J. Byrne, F. Buckley, and D. P. Berry. 2015. Development of an index to rank dairy females on expected lifetime profit. *J. Dairy Sci.* 98:4225–4239.
- 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.
- 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.
- Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, T. J. Lawlor, S. Forni, and J. I. Weller. 2014. Are evaluations on young genotyped animals benefiting from the past generations? *J. Dairy Sci.* 97:3930–3942.
- Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbandsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol.* 43:43.
- Makgahlela, M. L., E. A. Mäntysaari, I. Strandén, M. Koivula, U. S. Nielsen, M. J. Sillanpää, and J. Juga. 2013. Across breed multitrait random regression genomic predictions in the Nordic Red dairy cattle. *J. Anim. Breed. Genet.* 130:10–19.
- Mäntysaari, E., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. *Interbull Bull.* 41:17–21.
- Mäntysaari, E. A., M. Koivula, I. Strandén, J. Pösö, and G. P. Aamand. 2011. Estimation of GEBVs using deregressed individual cow breeding values. *Interbull Bull.* 44:19–24.
- McClure, M. C., D. Bickhart, D. Null, P. VanRaden, L. Xu, G. Wiggans, G. Liu, S. Schroeder, J. Glasscock, J. Armstrong, J. B. Cole, C. P. Van Tassell, and T. S. Sonstegard. 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. *PLoS ONE* 9:e92769.
- Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97:3943–3952.
- Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. *J. Anim. Breed. Genet.* 130:252–258.
- Nicolazzi, E. L., M. Picciolini, F. Strozzi, R. D. Schnabel, C. Lawley, A. Pirani, F. Brew, and A. Stella. 2014. SNPchiMp: A database to disentangle the SNPchip jungle in bovine livestock. *BMC Genomics* 15:123.
- Patry, C., and V. Ducrocq. 2011. Accounting for genomic pre-selection in national BLUP evaluations in dairy cattle. *Genet. Sel. Evol.* 43:30.

- Přibyl, J., P. Madsen, J. Bauer, J. Přibyllová, M. Šimečková, L. Vostrý, and L. Zavadilová. 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. *J. Dairy Sci.* 96:1865–1873.
- Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678.
- Schaeffer, L. R. 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86:35–45.
- Stoop, W. M., H. Eding, M. L. van Pelt, L. C. M. de Haer, and G. de Jong. 2013. Using pseudo-observations to combine genomic and conventional data in the Dutch-Flemish national evaluation. *Interbull Bull.* 47:106–110.
- Sullivan, P. G., and J. H. Jakobsen. 2014. GMACE pilot #4: Adjusting the national reliability input data. *Interbull Bull.* 48:40–45.
- Sullivan, P. G., B. Zumbach, J. W. Dürr, and J. H. Jakobsen. 2011. International genomic evaluations for young bulls. *Interbull Bull.* 44:87–94.
- Sun, C., P. M. VanRaden, J. R. O'Connell, K. A. Weigel, and D. Gianola. 2013. Mating programs including genomic relationships and dominance effects. *J. Dairy Sci.* 96:8014–8023.
- Tooker, M. E., T. A. Cooper, J. Dürr, and P. M. VanRaden. 2015. Comparison of domestic and foreign genotypes by country and continent. *Interbull Bull.* 49:14–18.
- 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.
- United States Supreme Court. 2013. Ass'n for Molecular Pathology v. Myriad Genetics. Accessed June 10, 2015. http://www.supremecourt.gov/opinions/12pdf/12-398_1b7d.pdf.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* 96:668–678.
- VanRaden, P. M., M. E. Tooker, J. R. Wright, C. Sun, and J. L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J. Dairy Sci.* 97:7952–7962.
- 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.
- Weigel, K. A., P. C. Hoffman, W. Herring, and T. J. Lawlor Jr. 2012. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. *J. Dairy Sci.* 95:2215–2225.
- Wiggans, G. R., T. A. Cooper, D. J. Null, and P. M. VanRaden. 2014. Increasing the number of single nucleotide polymorphisms used in genomic evaluations of dairy cattle. Commun. 301 in Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, BC, Canada. Am. Soc. Anim. Sci., Champaign, IL.
- Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and J. B. Cole. 2010. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. *J. Dairy Sci.* 94:6188–6193.
- Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. *J. Dairy Sci.* 94:3202–3211.
- Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2015. Technical note: Rapid calculation of genomic evaluations for new animals. *J. Dairy Sci.* 98:2039–2042.
- Winkelman, A. M., D. L. Johnson, and B. L. Harris. 2015. Application of genomic evaluation to dairy cattle in New Zealand. *J. Dairy Sci.* 98:659–675.
- Zwald, N. R., K. A. Weigel, W. F. Fikse, and R. Rekaya. 2003. Identification of factors that cause genotype by environment interaction between herds of Holstein cattle in seventeen countries. *J. Dairy Sci.* 86:1009–1018.