Multiple trait genomic evaluation of conception rate in Holsteins

Journal:	*Journal of Dairy Science*
Manuscript ID:	JDS-10-3893.R1
Article Type:	Research
Date Submitted by the Author:	27-Dec-2010
Complete List of Authors:	Aguilar, Ignacio; University of Georgia, Animal and Dairy Science Misztal, Ignacy; University of Georgia Tsuruta, Shogo; University of Georgia, Animal and Dairy Science Wiggans, George; Animal Improvement Programs Lab Lawlor, Thomas; Holstein Association Research & Development
Key Words:	BLUP, genomic selection, fertility, genetic evaluation

SCHOLARONE™
Manuscripts

INTERPRETIVE SUMMARY

Multiple trait genomic evaluation of conception rate in Holsteins. *By Aguilar et al., page 000.* The goal of this study was to assess the feasibility and accuracy of a multiple trait single-step genomic evaluation for a lowly heritable trait such as the outcome of artificial insemination. Multiple trait genetic evaluation for conception rate using outcomes of all available inseminations is technically possible. Large improvement in accuracy is possible using the genomic information and computation with the single-step approach is straightforward.

For Peer Review

Conception Rate Genomic Evaluation

Multiple trait genomic evaluation of conception rate in Holsteins

I. Aguilar[*†1], I. Misztal[*], S. Tsuruta[*], G. R. Wiggans[‡] and T. J. Lawlor[§]

* Animal and Dairy Science Department, University of Georgia, Athens 30602

† Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay

‡ Agricultural Research Service, USDA, Beltsville, MD 20705-2350

§ Holstein Association USA Inc., Brattleboro, VT 05302-0808

[1] corresponding author: Ignacio Aguilar, Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, CP 90200, Uruguay, tel. +598-23677641, fax +598-23677609, e-mail: iaguilar@inia.org.uy

For Peer Review

ABSTRACT

A national data set of artificial inseminations in US Holsteins was used to obtain genetic evaluations for conception rate (**CR**). This objective of this study was to investigate the feasibility and resulting accuracy from using all available phenotypic, pedigree and genomic information. Evaluations were performed by regular BLUP or by BLUP with the traditional pedigree and genomic relationships combined in a unified single-step-procedure (SSP). Genetic parameters of CR in the first three parities were estimated with data from New York State only. Heritability estimates were around 2% and genetic correlations between CR in different parities were > 0.73. The R^2 obtained with the single-step approach were almost twice as large as those achieved with regular BLUP. Computing the single-step approach took 2 hrs, and it was 33% slower than a regular BLUP. A multiple trait evaluation of conception rate using the SSP is both possible and advantageous.

(**Key words**: BLUP, genomic selection, fertility, genetic evaluation)

INTRODUCTION

A worldwide decline of fertility in Holsteins has created a need for more accurate evaluation of fertility traits. The fertility complex in dairy cattle can be broken down into a number of traits (Gonzalez-Recio et al., 2005; Jamrozik et al., 2005). Typical fertility traits such as non-return rate and days open have their advantages and disadvantages (Huang et al., 2007). An important trait in the complex of fertility traits is conception rate (**CR**), which is defined as the probability of a successful outcome of individual breeding services (Averill et al., 2004; Huang et al., 2007). Treating each

For Peer Review

service separately allows for the specific effects influencing each breeding service to be accounted for.

Because of low heritability of CR (Gonzalez-Recio et al., 2005; González-Recio et al., 2006; Tsuruta et al., 2009), accuracies of bull estimated breeding values (**EBV**) for CR are usually low. However, accuracy can be improved using all available services from each parity. Records from later parities will lead to increased accuracy on older animals, which will contribute to higher accuracies of the younger relatives. Furthermore, accuracy can be boosted by utilizing genomic information (Veerkamp and Beerda, 2007).

The simplest and most efficient way to utilize genomic information is with a single-step approach (Aguilar et al., 2010). In this approach, genomic information is used to enhance relationship information and no changes to the model are required. The goals of this study were to estimate genetic parameters of CR in the first three parities, run a national evaluation with and without genomic information, and estimate gains in accuracy from the inclusion of genomic information.

MATERIAL AND METHODS

Data

Holstein service records in the first, second and third parities were obtained from the Animal Improvement Programs Laboratory, ARS, USDA (Beltsville, MD). Records from breedings between 2002 and 2008 were used. Breeding information prior to 2002 was scarce. Data editing followed criteria presented by (Kuhn et al., 2008). Only AI services were used and DIM at insemination was required to be between 30 and 365 d.

For Peer Review

Success of insemination was determined via several reproductive criteria (heat detection, natural service, AI and pregnancy diagnosis), as well as the presence of a subsequent calving. Service sires were restricted to Holstein bulls. Variance components were estimated using a subset of the data from only New York State. All available insemination records were used in the national genetic evaluation. A summary of both data sets is in Table 1.

Model

Conception rates in the first three parities were considered as correlated traits. Parameter estimates and predictions of EBV were obtained using a multiple-trait linear model. To quantify the benefit of using all three parities in a multiple-trait linear model, first parity service records were also analyzed with a single-trait linear model. Fixed effects in the model included contemporary group defined by herd-year of calving, month of service, age at calving, and days to service after calving. Random effects were service sire (s), additive genetic (a), permanent environmental (p) and residual (e). The (co)variance structure was:

$$\mathrm{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{s}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G}\otimes\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}\otimes\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{S}\otimes\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}\otimes\mathbf{I}\end{bmatrix}$$

where **A** is the numerator relationship matrix; **G**, **P** and **S** are 3 x 3 (co)variance matrices for additive genetic, permanent and service sire effects respectively; **I** are identity matrices, and **R** is a 3 x 3 diagonal matrix of residual variances.

Variance components

For Peer Review

Parameters were estimated by the GIBBS2F90 program (Misztal et al., 2002) via a Bayesian approach using Gibbs sampling. Genomic data were not included for variance component estimation. Of a total of 100,000 samples, the first 10,000 were discarded as a burn-in, and every 10th sample was retained to calculate posterior means and standard deviations of heritability and correlation estimates.

Genetic evaluation

Genetic evaluations were computed using a modified version of BLUP90IOD (Tsuruta et al., 2001; Aguilar et al., 2010). Approximate accuracies were calculated using ACCF90 (Misztal et al., 2002). Deregressed evaluations (**DD**) were obtained from EBV and approximate accuracies (VanRaden et al., 2009). A subset of records up to 2005 was used to assess the accuracy of prediction of breeding values. Two sets of EBV were obtained. The first set used a genetic evaluation with the regular numerator relationship matrix (**PA**). The second set used a modified relationship matrix that accounts for genetic relationships via pedigree and genomic data and predicts EBV with a single-step approach (**SSP**) (Misztal et al., 2009). In SSP, the **H** matrix replaces the numerator relationship matrix (**A**) with the following inverse (Aguilar et al., 2010; Christensen and Lund, 2010):

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

where **H** is a modified relationship matrix incorporating genomic information as described by (Legarra et al., 2009), **G** is a genomic relationship matrix (VanRaden, 2008) and $\mathbf{A}_{22}$ is the pedigree-based relationship matrix for genotyped animals. The genomic relationship matrix (**G**) was created as:

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{k},$$

For Peer Review

where **Z** is an incidence matrix, for each SNP genotype, with elements

$$z_{ij} = \begin{cases} 0-2p_j & \text{if homozygous 11} \\ 1-2p_j & \text{if heterozygous 12 or 21,} \\ 2-2p_j & \text{if homozygous 22} \end{cases}$$

for animal *i* and SNP genotype *j* with an assumed allele frequency of p_j=0.5 for all SNP alleles. The scaling parameter *k* was defined as

$$k = 2\sum p_j(1-p_j).$$

Predictions from the two methods were compared by the regressions:

$$DD = \mu + \delta EBV_{05} + e$$

where DD were deregressed evaluations from 154 genotyped bulls without daughter records in 2005 but with daughter records in 2009 that were computed with complete data but without genomic information; μ is a mean; δ is a regression coefficient; EBV_{05} are breeding values based on insemination records up to 2005; and e is residual error. The EBV_{05} were either PA (PA_{05}) or based on SSP ($SSP\text{-}EBV_{05}$).

RESULTS AND DISCUSSION

Table 2 contains estimates of genetic parameters in the first three parities. Heritability estimates of each parity were close to 2%, and genetic correlation between parities 1 and 2 was 0.877; between parities 2 and 3 was 0.808, and between parities 1 and 3 was 0.732.

R^2 and the coefficients of regression for bulls with no daughters in 2005 and at least 50 daughters in 2009 are presented in Table 3. The R^2 obtained with PA in 2005 were low because of limited data and the low heritability of CR. In the available data set, few breeding records were prior to 2003. Luan et al. (2009) assessed the accuracy of

For Peer Review

genomic selection in Norwegian Red Cattle and observed a strong relationship between accuracy and heritability of the trait; traits with low heritability had EBV with lower accuracy and greater bias. In the present study, the inclusion of genomic information (SSP-EBV) resulted in approximately doubling R^2 than when using relationships from pedigree information only (PA). The improvement of R^2 corresponded to approximately 10 additional effective daughter contributions. Regression of DD_{2009} on $SSP\text{-}EBV_{05}$ resulted in a higher coefficient δ than PA_{05}, indicating less bias SSP. Higher R^2 could be achieved with data over many generations and by adding heifer insemination data.

To investigate the increase in accuracy from using a three-parity multiple-trait model, the analyses were repeated for first parity service records only. Results comparing SSP and PA using single trait or multiple trait models are in Table 4. Use of multiple parity records results in a much higher R^2 and a less biased δ.

Combining a genomic-based relationship matrix with the pedigree-based relationship and the use of a multiple-trait model tripled the accuracy of estimated breeding values for CR in the first parity. This study shows the importance of using all available information (genomic markers and multiple parities) to achieve improved predictability of breeding values for a lowly heritable trait, such as CR. A simple repeatability model with service records from several parities could also be applied with some loss of accuracy. This topic remains to be addressed in further studies. Alternative methods and/or use of models incorporating genomic information, such as in VanRaden *et al.* 2009, could not be assessed in this study due to the limited time frame of our breeding services data. Further research on assessing the increase in accuracies of EBVs, as opposed to DD, with inclusion of genomic data by a SSP would be desirable. One

For Peer Review

reason why we observed an increase in accuracy is that the genomic data provides another valuable source of information. This can be seen by inspection of the diagonals of the inverse of the pedigree-based (**A**) and genomic-based (**G**) relationship matrices. While diagonal elements of $\mathbf{A}^{-1}$ for bulls are close to $2+n/2$, where n is the number of daughters, the corresponding elements of $\mathbf{G}^{-1}$ are $2+x/2$ where $x > n$. In that case, $x-n$ may be regarded as the additional number of daughter equivalents from the genomic information. The average (min, max) x for young bulls in this study was 10.7 (6.1, 82.2).

Computations were done in an Opteron 64-bit processor with a clock speed of 3.02 GHz, 64 Mbyte of memory and a cache size of 1 Mbyte. Initial computing with BLUP took 1.5 hrs. Computing with the added genomic information via SSP increased the time to 2 hrs. Computing with multiple-trait models and the genomic information is therefore realistic even if the data set is much larger.

CONCLUSION

Multiple trait genetic evaluation for conception rate using outcomes of all available inseminations is technically possible. Large improvements in accuracy are possible when genomic information is used. Computation with the single-step approach is straightforward. More accurate assessment of such an improvement would require either records over a longer period of time or a different methodology for comparisons.

ACKNOWLEDGMENTS

This study was partially funded by the Holstein Association USA Inc. and by AFRI grants 2009-65205-05665 and 2010-65205-20366 from the USDA NIFA Animal

For Peer Review

Genome Program. The authors thank the Cooperative Dairy DNA Repository (Beltsville, MD) for providing genotypic data. The Dairy Records Management Systems (Raleigh, NC, and Ames, IA), authorized use of the data included in this analysis which was extracted from the USDA national dairy database.

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.

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.

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

Gonzalez-Recio, O., Y. M. Chang, D. Gianola, and K. A. Weigel. 2005. Number of inseminations to conception in Holstein cows using censored records and time-dependent covariates. J. Dairy Sci. 88:3655–3662.

González-Recio, O., Y. M. Chang, D. Gianola, and K. A. Weigel. 2006. Comparison of models using different censoring scenarios for days open in Spanish Holstein cows. Animal Science 82:233–239.

Huang, C., I. Misztal, S. Tsuruta, and T. J. Lawlor. 2007. Methodology of evaluation for female fertility. Pages 156–160 Proceedings of the 2007 Interbull Meeting. Interbull Bulletin 37, Dublin, Ireland.

For Peer Review

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.

Kuhn, M. T., J. L. Hutchison, and H. D. Norman. 2008. Modeling nuisance variables for prediction of service sire fertility. J. Dairy Sci. 91:2823–2835.

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

Luan, T., J. A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T. H. E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. Genetics 183:1119–1126.

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.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Commun. No. 28–07 in 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

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, and I. Stranden. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166–1172.

For Peer Review

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

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for north american Holstein bulls. J. Dairy Sci. 92:16–24.

Veerkamp, R. F. and B. Beerda. 2007. Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68:S266–S273.

For Peer Review

Table 1. Descriptive summary of national and New York data by parity.

	National			New York State		
	1*	2	3	1	2	3
Insemination records	3,025,115	2,033,086	945,870	165,159	116,494	55,038
Herd-Year	14,581	14,322	12,203	862	855	745
Cows	1,186,451	790,354	380,776	67,083	46,248	22,634
Conception Rate (%)	33.0	31.0	30.7	35.4	32.8	33.2
Pedigree Animals	2,489,119			132,623		

For Peer Review

Table 2. Estimates of posterior means and standard deviations for genetic parameters for conception rate in the first three parities*

Traits	CR 1**	CR 2	CR 3
CR 1	0.018 ± 0.002	0.877 ± 0.045	0.732 ± 0.047
CR 2	0.288 ± 0.083	0.022 ± 0.002	0.808 ± 0.103
CR 3	0.162 ± 0.084	0.326 ± 0.07	0.016 ± 0.005

*Heritability estimates ± SD on the diagonal, genetic and permanent correlations above and below the diagonal, respectively.

** CR 1, CR 2 and CR 3, conception rate in first, second and third parity respectively.

For Peer Review

Table 3. Coefficients of determination (R^2) and coefficients of regression (δ) of daughter deviation on estimated breeding values using single-step approach (SSP-EBV_{05}) or parent average (PA_{05}).

	SSP-EBV_{05}		PA_{05}	
	R^2	δ	R^2	δ
CR 1*	0.15	0.84	0.07	0.72
CR 2	0.13	0.81	0.06	0.66
CR 3	0.10	0.96	0.05	0.82

*CR 1, CR 2 and CR 3, conception rate in first, second and third parity respectively.

For Peer Review

Table 4. Coefficients of determination (R^2) and coefficients of regression (δ) of daughter deviation on estimated breeding values using single-step approach (SSP-EBV_{05}) or parent average (PA_{05}) for first parity conception rate using single trait or multiple trait analysis.

	SSP-EBV_{05}		PA_{05}	
Model	R^2	δ	R^2	δ
Single Trait	0.07	0.86	0.02	0.57
Multiple Trait	0.15	0.84	0.07	0.72

For Peer Review