Genomic evaluation of health traits in dairy cattle

J. Cole[1], K.L. Parker Gaddis[2], J.S. Clay[3] and C. Maltecca[2]

[1]Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA
[2]Department of Animal Science, College of Agriculture and Life Sciences North Carolina State University, Raleigh, NC 27695-7621,USA
[3]Dairy Records Management Systems, Raleigh, NC, USA

Abstract

There is growing interest from dairy producers in traits related to health and fitness of cattle, which often have low heritabilities but high economic values. Traits with low heritability can be improved by genetic selection, but large numbers of daughter records are required to produce predicted transmitting abilities with high reliability. Producer-recorded health event data collected from on-farm computer systems were used to estimate variance components and compute traditional predicted transmitting abilities (PTA) for several health traits (digestive problems, displaced abomasum, ketosis, lameness, mastitis, metritis, reproductive problems, and retained placenta) using single-trait threshold sire models. Heritabilities ranged from 0.01 for lameness to 0.30 for displaced abomasum using only first lactation data. Results were similar when only first lactation or first through fifth parity data were used. Multiple trait models also were used to estimate genetic correlations among those traits, which ranged from -0.29 (ketosis, lameness) to +0.81 (displaced abomasum, ketosis). Only three traits (displaced abomasum, mastitis, metritis) had 300 or more bulls with traditional reliabilities of at least 0.50. A multiple-trait sire threshold model was used to compute genomic PTA for 2,649 genotyped bulls. The increase in reliability from including the genomic data ranged from 0.38 (displaced abomasum) to 0.48 (lameness). These results suggest that enough data may exist in on-farm computer systems to enable the routine calculation of genetic and genomic evaluations for the most common health disorders in US Holstein cattle.

Keywords: dairy cattle, genetic evaluation, genomic selection, health traits.

Introduction

A negative relationship of production with fitness traits, possibly in response to selection for increased dairy cattle production over the last 50 years, has become apparent (Rauw *et al*., 1998). Declining health of cows can impact the profitability of a herd in several way, including increased culling rates, decreased and withheld milk, veterinary expenses, and additional labor. Kelton *et al*. (1998) estimated the cost of several common health events, which ranged from $39 per lactation with an incidence of cystic ovaries to $340 per case of left displaced abomasum. Over the past fifteen years, however, these economic costs may have drastically changed.

Improvement of health traits by genetic selection is appealing because the approach is well understood and gains are cumulative. The potential for genetic improvement in health-related traits has been demonstrated in Scandinavian cattle breeds (Abdel-Azim *et al*., 2005), and mastitis incidence has been successfully improved in Norwegian cattle (Heringstad *et al*., 2003). However, there is no mandated or consistent data recording system for health traits in the United States.

Several previous studies have addressed the use of producer-recorded health information for genetic improvement. Zwald *et al*. (2004a) used producer-recorded health event records from 2001 through 2003 and concluded that those data are useable for genetic selection. Parker Gaddis *et al*. (2012) recently showed that similar data accurately reflected the true incidence of health events, and confirmed that phenotypic relationships among common health events were consistent with results from epidemiological studies. The amount of producer-recorded data stored in on-farm computer systems in the US is increasing, and may provide the records needed to implement routine genetic evaluations for health traits.

The objective of this study was to use genetic and genomic analyses and producer-recorded health event data to estimate variance components and heritability for common health traits in US dairy cattle. A multiple-trait genetic analysis was used to identify genetic relationships between health events. Single-step methodology was used to incorporate genomic information in a multiple-trait analysis of those traits.

Material and methods

Producer-recorded health event data from US farms between 1996 and 2012 were available from Dairy Records Management Systems (Raleigh, NC) (Table 1). The health events used for analysis were mastitis (**MAST**), metritis (**METR**), cystic ovaries (**CYST**), digestive disorders (**DIGE**), displaced abomasum (**DSAB**), ketosis (**KETO**), lameness (**LAME**), reproductive problems (**REPR**), and retained placenta (**RETP**) from cows of parities one through five. Previous editing was applied to the data for common health events as described in Parker Gaddis *et al*. (2012).

Table 1. Summary statistics for each health event of interest.

Health event	Number of records	Number of cows	Number of herd-years
Cystic ovaries	222 937	131 194	3 369
Digestive disorders	156 520	97 430	1 780
Displaced abomasum	213 897	125 594	2 370
Ketosis	132 066	82 406	1 358
Lameness	233 392	144 382	3 191
Mastitis	274 890	164 630	3 859
Metritis	236 786	139 818	3 029
Reproductive disorders	253 272	151 315	3 360
Retained placenta	231 317	138 457	2 930

Univariate analysis

A logistic sire model was used in ASReml (Gilmour *et al.*, 2009) due to the binary nature of the data. The model is given as follows:

$$\eta = X\beta + Z_h h + Z_s s$$

where η is the logit of observing the health event of interest, β is a vector of fixed effects including parity as first versus later parities and year-season, ***X*** is the corresponding incidence matrix of fixed effects, ***h*** represents the random herd-year effect, ***s*** represents the random sire effect where $s \sim N\left(0,\ A\sigma_s^2\right)$ with ***A*** representing the additive relationship matrix, and Z_h and Z_s represent the corresponding incidence matrices for the appropriate random effect. Variance components and heritabilities were estimated for each common health event individually. Accuracies and reliabilities of each sire's estimated breeding value (EBV) were calculated as:

$$rel = \frac{SE^2}{(1+f)\sigma_2^s}$$

where *rel* is the reliability, SE^2 is the squared standard error of the sire's EBV, *f* is the sire's inbreeding coefficient, and σ_s^2 is the estimated sire variance. Accuracy was calculated as the square root of reliability. The variance component estimates were then used as starting values of variance components in the multivariate analysis.

Multivariate analyses

A multiple trait threshold sire model was used to fit a seven-trait model for the following most common health events: MAST, METR, LAME, RETP, CYST, KETO, and DSAB. The model is given below:

$$\lambda = \mathbf{X}\beta + \mathbf{Z}_\mathbf{h} h + \mathbf{Z}_\mathbf{s} s$$

where λ represents a vector of unobserved liabilities to the given diseases, β is a vector of fixed effects including parity as first versus later parities and year-season, **X** is the corresponding incidence matrix of fixed effects, ***h*** represents the random herd-year effect, ***s*** represents the random sire effect where $s \sim N\left(0,\ \mathbf{A}\sigma_s^2\right)$ with **A** representing the additive relationship matrix, and $\mathbf{Z}_\mathbf{h}$ and $\mathbf{Z}_\mathbf{s}$ represent the corresponding incidence matrices for the appropriate random effect. Variance components and heritability were determined from parameter estimates calculated using THRGIBBS1F90 (Tsuruta and Misztal, 2006). A total of 100,000 iterations were completed with the first 10,000 discarded as burn-in, saving every 25 samples. Post-Gibbs analyses were completed using POSTGIBBSF90 (Misztal *et al.*, 2002). Posterior means of sire predicted transmitting abilities (**PTA**) were estimated on the liability scale as well as converted to probabilities of disease as described by Zwald (2006). Highest posterior densities for the 95% interval were calculated for each parameter. Reliabilities of estimated sire PTAs were calculated as shown above using the posterior mean of additive variance of each health event, standard deviation of each estimate distribution, and inbreeding coefficients of the sires.

Genomic data was incorporated through the use of a blended **H** matrix following single step methodology implemented with preGSf90 software (Aguilar *et al*., 2011). The software has a maximum number of genotyped animals that can be used, which was met by restricting the genotype data to only include sires with at least five daughters. Default editing conditions were applied as set by the software resulting in genomic data being included for 2,649 sires with 37,525 markers. The blended **H** matrix was incorporated into the same multiple trait threshold sire model as previously described using THRGIBBS1F90 (Tsuruta and Misztal, 2006). Difficulties were initially encountered with convergence using all seven traits. To obtain better starting values, 2 preliminary analyses were performed. One analysis contained four traits (MAST, METR, LAME, and KETO) and the second analysis contained the remaining three traits (RETP, CYST, and DSAB). The posterior means of these analyses were then used as starting values in the full, seven-trait analysis. Post-Gibbs analyses were completed with POSTGIBBSF90. Convergence was assessed using the Coda library (Plummer *et al*., 2006) of R (R Core Team, 2012). Reliability of genomic estimated breeding values (GEBV) was estimated following Misztal *et al*. (2013). The reliabilities from the pedigree-based multiple trait analysis were used as reliabilities calculated without genomic information. These reliabilities were then converted to the effective number of records for genotyped animals following the formula given below:

$$d_i = \alpha\left[1/\left(1-rel_i\right)-1\right]$$

where α is the ratio of residual variance to genetic variance calculated from the pedigree-based multiple trait analysis. The inverse matrix **Q** was calculated as:

$$\mathbf{Q}_i = \left[\mathbf{D} + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\alpha\right]^{-1}$$

where $\mathbf{G}^{-1}$ is the genomic relationship matrix and $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree-based relationship matrix for genotyped animals only. The genomic reliabilities were then approximated as shown below:

$$rel_i = 1 - \alpha q^{ii}$$

where q^{ii} is the diagonal element of $\mathbf{Q}^{-1}$ corresponding to the i^{th} animal.

Results and discussion

Heritabilities and standard errors estimated from the single trait analyses are shown in Table 2. All traits exhibited a genetic component, but most were lowly heritable. The highest heritability was found for DSAB at 0.20. This heritability is very close to that estimated with a similar but smaller dataset (Zwald *et al*., 2004a). The high heritability for DSAB may be at least partially explained by the severity of the event, often requiring veterinary intervention. Zwald *et al*. (2004b) found DSAB to be the most consistently recorded health event among producer recorded data. Lower heritabilities were found for traits such as CYST, LAME, REPR, and RESP. These are events that are generally much less likely to be recorded in a consistent manner. For example, producers may have differing opinions regarding what constitutes an incidence of lameness that needs to be recorded.

Table 2. Heritability estimates and standard errors from single-trait analyses using pedigree-based relationship matrix, ***A****.*

Health Event	Heritability	Standard Error
Cystic ovaries	0.03	0.006
Digestive disorders	0.06	0.02
Displaced abomasum	0.20	0.02
Ketosis	0.07	0.01
Lameness	0.03	0.005
Mastitis	0.05	0.006
Metritis	0.06	0.007
Respiratory disorders	0.04	0.01
Reproductive disorders	0.03	0.006
Retained placenta	0.07	0.01

Sire posterior mean of daughters' probability to each disease are shown in Figure 1. The mean probability of displaced abomasum was the highest equal to 0.53, though again, it is likely to be one of the diseases that are reported most consistently. The probability of daughters experiencing displaced abomasum ranged from 0.33 to 0.73. The mean probability of MAST was 0.515 and ranged from 0.29 to 0.66. These estimates are higher than those previously reported by Zwald (2004a). Probability of mastitis is more similar to those reported by Harder *et al.* (2006) when analyzing udder disorders as a group. Probabilities of experiencing a reproductive disorder are also similar to those reported by Harder *et al.* (2006).

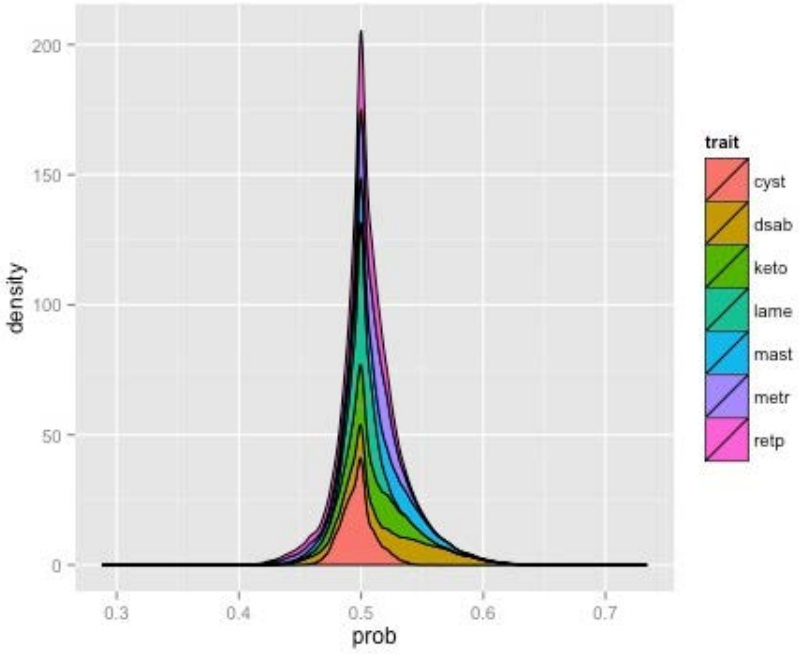

Figure 1. Sire posterior mean of daughters' probability to each disease (CYST = cystic ovaries; DSAB = displaced abomasum; KETO = ketosis; LAME = lameness; MAST = mastitis; METR = metritis; RETP = retained placenta).

Heritability estimates and 95% HPD from the multiple-trait threshold model are shown in Table 3. Genetic correlations between health events are included on the off-diagonals. Some traits with very low heritability estimates from the single-trait analyses were not included in the multiple-trait analysis. All heritability estimates were significantly different from zero. Heritability estimates of MAST and KETO increased in the multiple-trait model. Heritability estimates for METR, LAME, RETP, and DSAB decreased, whereas the estimate for CYST remained relatively constant. The heritability estimate for DSAB is similar to what has been reported previously (Zwald *et al.*, 2004b). Health events that were lowly heritable in the single trait analyses did not increase greatly through the use of a multiple-trait model. Several significant genetic correlations were found between health events. A genetic correlation of 0.81 [95% HPD = (0.70, 0.92)] was estimated between DSAB and KETO. Zwald *et al.* (2004b) estimated a genetic correlation between these two events equal to 0.14 (0.03) whereas a higher genetic correlation of 0.64 (0.10) was estimated by Koeck *et al*. (2012). This correlation also is consistent with previous analyses of these data using an informal path analysis that found an animal to have odds 15.5 times higher to have an incident of DSAB given that they previously had KETO (Parker Gaddis *et al.*, 2012). A high genetic correlation was also estimated between RETP and METR. This correlation is higher than a previous estimate found equal to 0.62 (0.11) (Koeck *et al.*, 2012). Significant, positive genetic correlations were also found between METR and KETO and METR and DSAB.

Table 3. Estimated heritabilities (95% HPD1) on the diagonal with estimated genetic correlations below the diagonal from multiple-trait analysis.

	Mastitis	**Metritis**	**Lameness**	**Retained placenta**	**Cystic ovaries**	**Ketosis**	**Displaced abomasum**
Mastitis	0.1 (0.09, 0.12)						
Metritis	-0.30 (-0.45,-0.15)	0.04 (0.03, 0.05)					
Lameness	-0.29 (-0.46,-0.11)	0.21 (0, 0.45)	0.019 (0.01, 0.03)				
Retained placenta	0.01 (-0.14, 0.16)	0.78 (0.68, 0.88)	-0.14 (-0.36, 0.07)	0.05 (0.03, 0.06)			
Cystic ovaries	-0.09 (-0.29, 0.13)	-0.17 (-0.37, 0.06)	-0.19 (-0.40, 0.06)	-0.12 (-0.34, 0.12)	0.026 (0.02, 0.03)		
Ketosis	-0.28 (-0.47,-0.07)	0.45 (0.26, 0.64)	0.08 (-0.17, 0.34)	0.10 (-0.17, 0.35)	-0.15 (-0.37, 0.13)	0.08 (0.05, 0.11)	
Displaced abomasum	0.001 (-0.15, 0.17)	0.44 (0.28, 0.60)	-0.10 (-0.29, 0.09)	0.06 (-0.12, 0.25)	-0.10 (-0.31, 0.10)	0.81 (0.70, 0.92)	0.13 (0.11, 0.16)

Heritability estimates from the multiple-trait analysis using single-step genomic BLUP (Table 4) were very similar to what was estimated using pedigree information, but the reliability of sire PTAs were improved. The addition of genomic information improved the reliabilities of sire PTAs for all health events as shown in Table 5. The reliabilities for these traits are low in comparison to production traits, however, the percent improvement that is obtained from the addition of genomic information is substantial. Percent improvement over reliabilities from single-trait analyses with pedigree information ranged from a 25% improvement in KETO to a 37% improvement in both MAST and METR.

Table 4. Estimated heritabilities (95% HPD1) on the diagonal with estimated genomic correlations below the diagonal from multiple-trait single-step analysis.

	Mastitis	**Metritis**	**Lameness**	**Retained placenta**	**Cystic ovaries**	**Ketosis**	**Displaced abomasum**
Mastitis	0.12 (0.10, 0.14)						
Metritis	-0.36 (-0.53, -0.19)	0.04 (0.027, 0.043)					
Lameness		0.13 (-0.1, 0.34)	0.026 (0.015, 0.034)				
Retained placenta				0.04 (0.03, 0.05)			
Cystic ovaries				-0.02 (-0.22, 0.16)	0.03 (0.01, 0.04)		
Ketosis	-0.16 (-0.31, 0.01)	0.44 (0.26, 0.64)				0.08 (0.05, 0.10)	
Displaced abomasum				0.01 (-0.21, 0.16)	-0.11 (-0.29, 0.13)		0.12 (0.09, 0.14)

Table 5. Mean reliabilities of sire PTA computed with pedigree information and genomic information.

Health event	**Pedigree information**	**Blended pedigree & genomic information**
Mastitis	0.30	0.41
Metritis	0.30	0.41
Lameness	0.28	0.37
Retained placenta	0.29	0.38
Ketosis	0.28	0.35
Displaced abomasum	0.30	0.40

Conclusions

These results suggest that enough data may exist in on-farm computer systems to enable the routine calculation of genetic and genomic evaluations for the most common health disorders in US Holstein cattle. Multiple-trait analysis is challenging because of demanding computational requirements, but the gain in information from correlated traits may be worth the additional time required for analysis.

List of References

Abdel-Azim, G.A., A.E. Freeman, M.E. Kehrli, S.C. Kelm, J.L. Burton, A.L. Kuck, & S. Schnell, 2005. Genetic basis and risk factors for infectious and noninfectious diseases in US Holsteins. I. Estimation of genetic parameters for single diseases and general health. J. Dairy Sci. 88: 1199–1207.

Aguilar, I., I. Misztal, A. Legarra, & S. Tsuruta, 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. J. Anim. Breed. Genet. 128: 422–428.

Bar, D., L.W. Tauer, G. Bennett, R.N. González, J.A. Hertl, Y.H. Schukken, H.F. Schulte, F.L. Welcome, & Y.T. Gröhn, 2008. The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. J. Dairy Sci. 91: 2205–2214.

Cha, E., J. a Hertl, D. Bar, & Y.T. Gröhn, 2010. The cost of different types of lameness in dairy cows calculated by dynamic programming. Prev. Vet. Med. 97: 1–8.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, & R. Thompson, 2009. ASReml User Guide Release 3.0.

Harder, B., J. Bennewitz, D. Hinrichs, & E. Kalm, 2006. Genetic parameters for health traits and their relationship to different persistency traits in German Holstein dairy cattle. J. Dairy Sci. 89: 3202–3212.

Heringstad, B., Y.M. Chang, D. Gianola, & G. Klemetsdal, 2003. Genetic analysis of longitudinal trajectory of clinical mastitis in first-lactation Norwegian cattle. J. Dairy Sci. 86: 2676–2683.

Kelton, D.F., K.D. Lissemore, & R.E. Martin, 1998. Recommendations for recording and calculating the incidence of selected clinical diseases of dairy cattle. J. Dairy Sci. 81: 2502–2509.

Koeck, A., F. Miglior, D.F. Kelton, & F.S. Schenkel, 2012. Health recording in Canadian Holsteins: data and genetic parameters. J. Dairy Sci. 95: 4099–4108.

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

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P.M. VanRaden, & T.J. Lawlor, 2013. Methods to approximate reliabilities in single-step genomic evaluation. J. Dairy Sci. 96:647–654.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, & D.H. Lee, 2002. BLUPF90 and related programs (BGF90). In Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France. 1–2.

Parker Gaddis, K.L., J.B. Cole, J.S. Clay, & C. Maltecca, 2012. Incidence validation and relationship analysis of producer-recorded health event data from on-farm computer systems in the United States. J. Dairy Sci. 95: 5422–5435.

Plummer, M., N. Best, K. Cowles, & K. Vines, 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 6: 7–11.

Rauw, W.M., E. Kanis, E.N. Noordhuizen-Stassen, & F.J. Grommers, 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livest. Prod. Sci. 56: 15–33.

Tsuruta, S. & I. Misztal, 2006. THRGIBBS1F90 for estimation of variance components with threshold linear models. In Proc. 8th World Congr. Genet. Appl. Livest. Prod. Belo Horizonte, Brazil. Commun. 27–31.

Zwald, N.R., K.A. Weigel, Y.M. Chang, R.D. Welper, & J.S. Clay, 2004a. Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values. J. Dairy Sci. 87: 4287–4294.

Zwald, N.R., K.A. Weigel, Y.M. Chang, R.D. Welper, & J.S. Clay, 2004b. Genetic selection for health traits using producer-recorded data. II. Genetic correlations, disease probabilities, and relationships with existing traits. J. Dairy Sci. 87: 4295–4302.

Zwald, N.R., K.A. Weigel, Y.M. Chang, R.D. Welper, & J.S. Clay, 2006. Genetic analysis of clinical mastitis data from on-farm management software using threshold models. J. Dairy Sci. 89: 330–336.