Genomic selection for producer-recorded health event data in US dairy cattle

K. L. Parker Gaddis,*[1] J. B. Cole,† J. S. Clay,‡ and C. Maltecca*
*Department of Animal Science, North Carolina State University, Raleigh 27695
†Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350
‡Dairy Records Management Systems, Raleigh, NC 27603

ABSTRACT

Emphasizing increased profit through increased dairy cow production has revealed a negative relationship of production with fitness and health traits. Decreased cow health can affect herd profitability through increased rates of involuntary culling and decreased or lost milk sales. The development of genomic selection methodologies, with accompanying substantial gains in reliability for low-heritability traits, may dramatically improve the feasibility of genetic improvement of dairy cow health. Producer-recorded health information may provide a wealth of information for improvement of dairy cow health, thus improving profitability. The principal objective of this study was to use health data collected from on-farm computer systems in the United States to estimate variance components and heritability for health traits commonly experienced by dairy cows. A single-step analysis was conducted to estimate genomic variance components and heritabilities for health events, including cystic ovaries, displaced abomasum, ketosis, lameness, mastitis, metritis, and retained placenta. A blended **H** matrix was constructed for a threshold model with fixed effects of parity and year-season and random effects of herd-year and sire. The single-step genomic analysis produced heritability estimates that ranged from 0.02 (standard deviation = 0.005) for lameness to 0.36 (standard deviation = 0.08) for retained placenta. Significant genetic correlations were found between lameness and cystic ovaries, displaced abomasum and ketosis, displaced abomasum and metritis, and retained placenta and metritis. Sire reliabilities increased, on average, approximately 30% with the incorporation of genomic data. From the results of these analyses, it was concluded that genetic selection for health traits using producer-recorded data are feasible in the United States, and that the inclusion of genomic data substantially improves reliabilities for these traits.
Key words: dairy cattle, health, genomic selection

INTRODUCTION

Previous emphasis on increased profit through increasing dairy cow production has made a negative relationship of production with fitness traits become apparent (Rauw et al., 1998). An alternative to increasing net profit is to decrease management costs by improving the overall health of the cows (Zwald et al., 2004a). Declining health of cows can affect the profitability of a herd by affecting several aspects, such as additional culling, decreased and lost milk sales, veterinary expenses, and additional labor (Hansen, 2000; Harder et al., 2006). Kelton et al. (1998) estimated the cost of several common health events ranging from $39 per lactation with an event of cystic ovaries up to $340 per case of left-displaced abomasum. Over the past 15 yr, however, these economic costs may have drastically changed. More recent studies have looked at the average cost per case of specific hoof and leg disorders such as sole ulcers, digital dermatitis, and foot rot. The average cost per case of these events was estimated to be $216.07, $132.96, and $120.70, respectively (Cha et al., 2010). These estimates accounted for factors such as milk loss, treatment cost, and decreased fertility. Other recent research estimated the factors that contribute to the cost of an incidence of mastitis. The average cost of clinical mastitis per case was approximately $179, with $115 of that the result of lost milk, $14 from increased mortality, and $50 from treatment costs (Bar et al., 2008).

Genetic selection is an appealing tool for improvement of health traits. Difficulty is encountered, however, as no mandated or consistent recording system of health traits exists in the United States. In some European countries, recording of health events is mandatory. Genetic selection for increased disease resistance has been performed for more than 30 yr and the potential for genetic improvement in health-related traits has been demonstrated in Scandinavian cattle breeds (Philipsson and Lindhé, 2003; Abdel-Azim et al., 2005). Genetic improvement of clinical mastitis incidence has also been demonstrated in Nordic cattle (Heringstad et al., 2003; Philipsson and Lindhé, 2003). The lack of health-related phenotypes in the United States creates

an obstacle to genetic improvement. Several previous studies have confirmed the possibility of using on-farm recorded health information for genetic improvement. Zwald et al. (2004a) used on-farm recorded health data from 2001 through 2003 and concluded that this data would allow genetic selection to be possible. Prior research was completed to analyze if producer-recorded data from a similar data set to the current study accurately reflected the true incidences of health events after several editing constraints were put in place. Phenotypic relationships were also examined between common health events and compared with results from epidemiological studies to further validate the data (Parker Gaddis et al., 2012).

Although genetic improvement in some health traits has been demonstrated, progress is slow, especially when compared with the improvements achieved in production traits. Health traits are typically categorized as being lowly heritable. Low sire reliabilities are also common for health traits due to a combination of low heritability and limited availability of phenotypes. Dense marker data have been shown in many studies to improve reliability of prediction (Harris and Johnson, 2008; Hayes et al., 2009; VanRaden et al., 2009). Increased availability of dense molecular marker data may allow progress to be achieved at a quicker rate, especially for lowly heritable traits. Marker information is attainable at birth, which could decrease the generation interval required to achieve an acceptable reliability. Genomic selection methodologies are currently being widely investigated and implemented in dairy cattle breeding (VanRaden et al., 2009; Veerkamp et al., 2011), as well as in other species (Ostersen et al., 2011; Simeone et al., 2012); however, most of this research has involved traditional traits, such as those related to production (VanRaden et al., 2009; Olson et al., 2011).

One method of including SNP marker data into genetic analyses is the single-step method. Misztal et al. (2009) and Legarra et al. (2009) proposed the single-step method, as an alternative to multi-stage approaches. The single-step procedure replaces the pedigree (**A**) and genomic (**G**) relationship matrices with a blended **H** matrix (Aguilar et al., 2010; Christensen and Lund, 2010) that combines the information from **A** and **G**. The **H** matrix can be implemented similarly to the **A** relationship matrix in BLUP analyses (Legarra et al., 2009). This allows a straight-forward application of genomic data to complicated models and complex data structures (Aguilar et al., 2010).

Several studies have incorporated functional traits along with production traits using genomic data (Brøndum et al., 2011; Koivula et al., 2012), although the vast majority of these were conducted outside the United States. The objective of the current study was to perform pedigree- and genomic-based analyses on producer-recorded health data to estimate variance components and heritabilities for health traits commonly encountered by dairy cows in the United States, thereby confirming a genetic component of major health events. A multiple-trait genetic analysis using pedigree data was completed to identify genetic relationships among common health events, including cystic ovaries (**CYST**), displaced abomasum (**DSAB**), ketosis (**KETO**), lameness (**LAME**), mastitis (**MAST**), metritis (**METR**), and retained placenta (**RETP**). Single-step methodology was used to incorporate genomic information into a multiple-trait analysis of common health events, using the estimates from pedigree-based analyses as starting values. Reliabilities were compared between the pedigree-based analyses and genomic-based analyses. Genetic correlations with more commonly reported fitness traits, including daughter pregnancy rate (VanRaden et al., 2004), productive life (VanRaden and Wiggans, 1995), SCS, net merit (Cole et al., 2010), and milk yield, were also approximated.

MATERIALS AND METHODS

Voluntary producer-recorded health event data were available from Dairy Records Management Systems (Raleigh, NC) from US farms from 1996 through 2012. The health events included in the analyses were MAST, METR, CYST, DSAB, KETO, LAME, and RETP from cows of parities 1 through 5. Cows with records in later parities were required to have records for all prior parities. Data quality edits were applied as described in Parker Gaddis et al. (2012). Minimum and maximum constraints were imposed on the data by herd-year to avoid using records from herd-years that over- or underreported an event. Extended lactations lasting up to 400 d postpartum were included in the analyses under the assumptions that cows with extended lactations were likely to be those that had not become pregnant. This decreased fertility could potentially be attributable to poor health, which could be reflected in the data. Production data included a variable indicating if a cow was removed from the herd during the lactation. Records being coded as anything other than a normal lactation were originally removed from the data set. These records included cows removed from the herd during lactation, potentially for health-related reasons. Analyses were later completed including these terminated records, as no significant difference was found when terminated records were included. After editing, there were 134,226 total first-parity records from 12,893 sires and 13,534 maternal grandsires. There were 174,069 total records from parities 2 through 5

Table 1. Summary statistics for each health event of interest

Health event	Number of records		Number of year-seasons		Number of herd-years	
	First parity	Later parities	First parity	Later parities	First parity	Later parities
Cystic ovaries	78,408	100,982	58	55	2,789	2,855
Displaced abomasum	78,025	100,108	56	54	2,049	2,073
Ketosis	47,101	60,791	55	51	1,154	1,172
Lameness	88,835	104,552	57	55	2,707	2,762
Mastitis	103,585	125,332	58	57	3,198	3,255
Metritis	83,691	108,825	58	57	2,580	2,626
Retained placenta	82,666	105,474	56	54	2,419	2,526

for 100,635 cows from 11,481 sires and 11,716 maternal grandsires. A summary of the data structure by health event is shown in Table 1.

Genomic data from the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) were available for 7,883 sires. Standard filters were previously applied to the marker data, including removing SNP with minor allele frequencies less than 0.05 and removing SNP that were in complete linkage disequilibrium with other SNP, resulting in a final set for analysis of 38,416 SNP (VanRaden et al., 2009). There were 4,814 genotyped sires that had daughters with at least 1 health record in the final data set.

Pedigree-Based Analyses

Two multivariate analyses were completed: one using only first-parity records, and a second using records from parities 2 through 5. This was performed considering the biological differences found in first-parity animals compared with multiparous animals (Curtis et al., 1985). The data set of later-parity records represents a selected subset including only cows that survived to their second calving. A multiple-trait threshold sire model was used to fit a 7-trait model for the following health events: MAST, METR, LAME, RETP, CYST, KETO, and DSAB. The model used for first parity records was

$$\boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z_h}h + \mathbf{Z_s}s + \text{e},$$

where $\boldsymbol{\lambda}$ represents a vector of unobserved liabilities to the given diseases; $\boldsymbol{\beta}$ is a vector of fixed effects including overall mean and year-season; $\mathbf{X}$ is the corresponding incidence matrix for the fixed effect; h represents the random herd-year effect, where $h \sim N\left(0, \mathbf{I}\sigma_h^2\right)$, with $\mathbf{I}$ representing an identity matrix and σ_h^2 representing the variance of herd-year; s represents the random sire effect, where $s \sim N\left(0, \mathbf{A}\sigma_s^2\right)$, with $\mathbf{A}$ representing the additive relationship matrix and σ_s^2 representing the sire variance; $\mathbf{Z_h}$ and $\mathbf{Z_s}$ represent the corresponding inci-

dence matrices for the appropriate random effects; and e represents the random residual, modeled following $N(\mathbf{0}, \mathbf{I})$, fixing the variance equal to 1 to attain identifiability. Herd-year and year-season were included as separate effects to avoid levels with very few or no records. A probit link was used to transform event incidence to liability. A Monte Carlo Markov chain approach through Gibbs sampling was used to obtain estimates of variance components. The model for later parities was similar, but included a fixed effect of parity (with levels 2 to 5) and a random permanent environmental effect:

$$\boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z_h}h + \mathbf{Z_s}s + \mathbf{Z_p}p + \text{e},$$

where $\boldsymbol{\beta}$ is a vector of fixed effects including mean, parity, and year-season; p represents the permanent environment effect; and $\mathbf{Z_p}$ represents the corresponding incidence matrix. All other variables remained the same as previously described. Variance components and heritabilities were determined from parameter estimates calculated using the THRGIBBS1F90 program (version 2.104; Tsuruta and Misztal, 2006). Repeatability was calculated as $\left(4\sigma_s^2 + \sigma_p^2\right)/\left(\sigma_s^2 + \sigma_p^2 + \sigma_e^2\right)$, where σ_p^2 represents the permanent environmental variance and σ_e^2 represents the residual variance, for later-parity records. 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 the POSTGIBBSF90 program (version 3.04; Misztal et al., 2002). Trace plots were inspected visually to ensure that convergence had been reached; in addition, Geweke's convergence statistic (Geweke, 1992) was calculated with the coda package (Plummer et al., 2006) in R version 2.15.1 (R Development Core Team, 2012). Posterior standard deviations were calculated for each estimate.

Posterior means of sire PTA were obtained on the liability scale and later converted to probabilities of disease as described by Zwald et al. (2006). Approximate reliabilities of estimated sire PTA were calculated us-

ing the ACCF90 program (version 1.67; Misztal et al., 2002). Genetic correlations between each health trait and other more commonly reported fitness traits were approximated using the reliabilities of sire PTA following Calo et al. (1973). Additional traits considered included daughter pregnancy rate (**DPR**), productive life (**PL**), milk yield (**MY**), SCS, and net merit (**NM**). Approximate genetic correlations were calculated using the method of Calo et al. (1973):

$$\hat{r}_{g_{1,2}} = \frac{\sqrt{\left(\sum_{i=1}^{n} RL_{1i}\right)\left(\sum_{i=1}^{n} RL_{2i}\right)}}{\sum_{i=1}^{n}\left(RL_{1i} \times RL_{2i}\right)} \times r_{1,2},$$

where $\hat{r}_{g_{1,2}}$ is the approximate genetic correlation between trait 1 and trait 2; RL_{1i} and RL_{2i} represent the reliabilities of trait 1 and trait 2, respectively, for sire i; and $r_{1,2}$ represents the correlation between PTA for traits 1 and 2. The standard error of the approximate genetic correlation was calculated as described by Sokal and Rohlf (1995):

$$SE = \sqrt{\frac{1 - \hat{r}_{g_{1,2}}}{n - 2}},$$

where n represents the number of sires with records.

Genomic-Based Analyses

Genomic data were incorporated using a blended **H** matrix in a single-step procedure as implemented in the preGSf90 software (version 1.142; Aguilar et al., 2011). Further editing was applied as set by the default software settings, resulting in genomic data being included for 7,883 sires with 37,713 markers. Default editing included exclusion of SNP with minor allele frequency less than 0.05, exclusion of SNP with call rate less than 0.90, and exclusion of individuals with call rate <0.90. The **G** matrix was calculated and scaled following VanRaden (2008), using allele frequencies calculated from the available genotypes. The blended **H** matrix was incorporated into the same multiple-trait threshold sire model as previously described above using the THRGIBBS1F90 program (version 2.104; Tsuruta and Misztal, 2006). A chain of 100,000 iterations was completed with 10,000 samples discarded as burn-in, saving every 25th sample. Post-Gibbs checks were carried out similarly to those described for the previous analysis.

Reliabilities of genomic PTA were estimated following Misztal et al. (2013). Reliabilities from pedigree-based multiple-trait analysis were used as reliabilities calculated without genomic information. These reliabilities were then converted to effective number of records for genotyped animals following the formula provided in Misztal et al. (2013):

$$d_i = \alpha\left[1\big/\left(1 - rel_{p_i}\right) - 1\right],$$

where α is the ratio of residual variance to genetic variance calculated from the pedigree-based multiple-trait analysis and rel_{p_i} represents approximated reliabilities based only on pedigree information. The inverse matrix $\mathbf{Q}^{-1}$ was calculated as

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

where $\mathbf{D}$ is a diagonal matrix composed of elements d_i, $\mathbf{G}^{-1}$ is the inverse genomic relationship matrix, and $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree-based relationship matrix for genotyped animals only (Misztal et al., 2013). The genomic reliabilities were then approximated as shown below:

$$rel_{g_i} = 1 - \alpha q^{ii},$$

where rel_{g_i} represented the approximate genomic reliability and q^{ii} was the diagonal element of $\mathbf{Q}^{-1}$, corresponding to the ith animal (Misztal et al., 2013).

RESULTS AND DISCUSSION

Heritabilities and genetic correlations (±posterior SD) from the pedigree-based analysis are shown in Table 2 for first-parity records. All traits exhibited a genetic component, although most had low heritabilities. The highest heritability in first-parity records was 0.22 found for both DSAB and RETP. The heritability of DSAB is similar to previously reported estimates (Neuenschwander et al., 2012; Uribe et al., 1995). Displaced abomasum was also found to be the most heritable health trait in a smaller data set that included data from fewer years and different organizations (Zwald et al., 2004a). The high heritabilities for DSAB and RETP may be partially explained by the severity of the event, with DSAB often requiring veterinary intervention. Zwald et al. (2004b) found DSAB to be the most consistently recorded health event among producer-recorded data. Discrepancies in diagnosis of DSAB and RETP are also minimal, which is likely to improve the consistency of reporting. Lower heritabilities were found for traits such as CYST and LAME, which are

PARKER GADDIS ET AL.

Table 2. Estimated heritabilities (SD) on the diagonal with estimated genetic correlations below the diagonal from multiple-trait pedigree-based analysis with first-parity records

Trait	Cystic ovaries	Displaced abomasum	Ketosis	Lameness	Mastitis	Metritis	Retained placenta
Cystic ovaries	0.03 (0.01)						
Displaced abomasum	−0.07 (0.15)	0.22 (0.03)					
Ketosis	−0.16 (0.16)	0.66 (0.07)*	0.09 (0.02)				
Lameness	−0.06 (0.24)	0.10 (0.18)	0.25 (0.19)	0.02 (0.005)			
Mastitis	0.16 (0.17)	0.04 (0.11)	0.10 (0.12)	0.26 (0.17)	0.06 (0.01)		
Metritis	−0.25 (0.18)	0.22 (0.12)	0.22 (0.14)	0.07 (0.18)	−0.22 (0.12)	0.04 (0.01)	
Retained placenta	0.24 (0.18)	0.42 (0.11)*	−0.01 (0.14)	−0.16 (0.20)	0.33 (0.11)*	0.56 (0.10)*	0.22 (0.04)

*Genetic correlations significant at $P < 0.05$.

much less likely to be recorded in a consistent manner. For example, producers may have different opinions regarding what constitutes an incidence of lameness that warrants being recorded.

Heritabilities and genetic correlations (±SD) from the pedigree-based analysis are shown in Table 3 for later-parity records. All heritabilities were similar to (CYST, LAME, and METR) or smaller than (DSAB, KETO, MAST, and RETP) the results from the first-parity analysis (Table 2). The highest heritability was found for DSAB and the lowest heritability was found for CYST. Posterior estimates of permanent environmental variance were small, ranging from 6.7×10^{-3} for RETP to 0.017 for CYST. Repeatabilities for the later-parity records ranged from 0.02 for CYST and LAME to 0.13 for DSAB, and were lower than those reported by Gernand et al. (2012), with the exception of RETP, which was very similar. Discrepancies in estimates from different studies may be related to event severity or variations in recording consistency.

The strongest genetic correlation in first-parity records was between DSAB and KETO (0.66 ± 0.07), which is similar to correlations reported previously (Zwald et al., 2004b; Koeck et al., 2012; Neuenschwander et al., 2012). The genetic correlation between DSAB and KETO in later parities was of similar magnitude (0.65 ± 0.15). Significant correlations between METR and RETP of 0.56 (±0.10) in first- and 0.69 (±0.10) later-parity records were found, which is smaller than a previously

reported value of 0.79 (±0.32; Neuenschwander et al., 2012). These correlation estimates were consistent with odds ratios previously reported from these data (Parker Gaddis et al., 2012).

Several diseases had negative correlations with CYST, but none were significant in first parity. The only trait in later parities with a significant (positive) correlation with CYST was METR. These spurious results may be the result of how the trait is recorded, the editing criteria used, or a combination of the 2. An incidence of CYST is likely to be reported following a veterinary visit, but such exams are not likely unless a cow has difficulty getting pregnant. It may be that CYST events actually affect the following lactation when the animal is being rebred, rather than the lactation when the event is recorded. Relationships between CYST and METR have been previously reported (Marion and Gier, 1968; Erb et al., 1981), supporting the significant genetic correlation estimated between CYST and METR from later-parity records. This may be an indication of the underlying reproductive health status of the cow, but questions about the consistency of reporting and data editing procedures introduce additional uncertainty to CYST-related results.

Heritabilities and genetic correlations (±SD) from the genomic-based analysis of first-parity records are shown in Table 4. The heritabilities of all traits were higher than in the pedigree-based analysis, with the exception of LAME, which remained the same. The

Table 3. Estimated heritabilities (SD) on the diagonal with estimated genetic correlations below the diagonal from multiple-trait pedigree-based analysis with later-parity records

Trait	Cystic ovaries	Displaced abomasum	Ketosis	Lameness	Mastitis	Metritis	Retained placenta
Cystic ovaries	0.01 (0.005)						
Displaced abomasum	0.25 (0.19)	0.12 (0.02)					
Ketosis	−0.02 (0.30)	0.65 (0.15)*	0.04 (0.01)				
Lameness	0.33 (0.32)	−0.04 (0.17)	−0.01 (0.20)	0.02 (0.006)			
Mastitis	0.15 (0.22)	0.02 (0.16)	−0.20 (0.24)	−0.15 (0.20)	0.03 (0.007)		
Metritis	0.69 (0.15)*	0.38 (0.13)*	0.22 (0.26)	0.29 (0.17)	−0.001 (0.18)	0.03 (0.006)	
Retained placenta	0.11 (0.20)	0.10 (0.16)	0.22 (0.20)	−0.16 (0.19)	0.02 (0.17)	0.69 (0.10)*	0.05 (0.01)

*Genetic correlations significant at $P < 0.05$.

Table 4. Estimated heritabilities (SD) on the diagonal with estimated genetic correlations below the diagonal from multiple-trait genomic-based analysis with first-parity records

Trait	Cystic ovaries	Displaced abomasum	Ketosis	Lameness	Mastitis	Metritis	Retained placenta
Cystic ovaries	0.05 (0.01)						
Displaced abomasum	0.01 (0.13)	0.32 (0.04)					
Ketosis	−0.06 (0.15)	0.65 (0.09)*	0.14 (0.03)				
Lameness	0.49 (0.16)*	−0.13 (0.14)	0.12 (0.17)	0.02 (0.005)			
Mastitis	0.24 (0.14)	0.06 (0.10)	0.12 (0.12)	0.05 (0.15)	0.10 (0.01)		
Metritis	−0.30 (0.15)	0.29 (0.10)*	0.21 (0.13)	0.20 (0.19)	−0.08 (0.12)	0.07 (0.01)	
Retained placenta	0.17 (0.22)	−0.09 (0.16)	−0.21 (0.19)	−0.28 (0.22)	0.05 (0.17)	0.36 (0.15)*	0.36 (0.08)

*Genetic correlations significant at $P < 0.05$.

largest heritabilities were again found for RETP and DSAB. The correlation of DSAB with KETO was similar in pedigree- and genomic-based analyses. The correlation of RETP with METR was smaller (0.36 ± 0.15) in genomic-based analysis, but remained significant. A significant correlation was found between DSAB and METR, which was significant only in later-parity records using pedigree-based data. A significant correlation was found between CYST and LAME (0.49 ± 0.16), but results should be interpreted with caution because LAME is a highly subjective event. Large discrepancies tend to exist in recording LAME; producers may only record certain cases, and these practices will depend largely on management routines (Zwald et al., 2004b).

Heritabilities and genetic correlations (±SD) for genomic-based analysis of later-parity records are listed in Table 5. Heritability was similar to prior analyses and ranged from 0.02 (±0.01) for CYST up to 0.17 (±0.03) for DSAB. All heritabilities estimated for later parities with genomic information were larger than heritability estimates using only pedigree information. However, all heritabilities were larger for first- than later-parity genomic analyses. Significant genetic correlations were again found between KETO and DSAB (0.61 ± 0.12) and between METR and RETP (0.81 ± 0.06). No other correlations were significant in this analysis.

Estimates of heritability including genomic information were similar to those estimated using pedigree

information. Differences in heritability estimates between pedigree- and genomic-based analyses may be the result of differences in scale of the relationship matrices. Direct comparisons of estimates between the analyses are not possible, however, because the **A** matrix and **H** matrix are produced using different base populations. The largest change between the 2 analyses was observed in reliability of sire PTA. The addition of genomic information improved the reliabilities of sire PTA for all health events, as shown in Table 6. The reliabilities for these traits were low in comparison with production traits; however, the improvement obtained from the addition of genomic information was substantial. The increases in average reliability ranged from 9 percentage points for RETP (55 to 64%) to 15 percentage points for LAME (24 to 39%), which is consistent with results of VanRaden et al. (2009). Reliabilities reported in other studies are comparable. Brøndum et al. (2011) reported a genomic reliability for diseases unrelated to the udder ranging from 0.25 to 0.43, depending on the population. Su et al. (2010) calculated an expected reliability of genomic EBV for diseases unrelated to the udder slightly higher at 0.59. A third study investigated different methods of calculating genomic predictions for production traits as well as MAST (Koivula et al., 2012). One model was similar to the single-step method; however, deregressed proofs were used as opposed to raw data. Validation reliabilities for genomic breeding values of MAST in that study

Table 5. Estimated heritabilities (SD) on the diagonal with estimated genetic correlations below the diagonal from multiple-trait genomic-based analysis with later-parity records

Trait	Cystic ovaries	Displaced abomasum	Ketosis	Lameness	Mastitis	Metritis	Retained placenta
Cystic ovaries	0.02 (0.01)						
Displaced abomasum	0.26 (0.23)	0.17 (0.03)					
Ketosis	0.09 (0.24)	0.61 (0.12)*	0.08 (0.02)				
Lameness	−0.36 (0.23)	0.001 (0.18)	0.03 (0.21)	0.03 (0.01)			
Mastitis	0.54 (0.35)	0.06 (0.20)	−0.01 (0.19)	−0.31 (0.18)	0.05 (0.01)		
Metritis	−0.04 (0.21)	0.24 (0.15)	0.28 (0.21)	0.11 (0.18)	−0.01 (0.15)	0.06 (0.01)	
Retained placenta	0.24 (0.23)	−0.07 (0.15)	0.26 (0.18)	−0.20 (0.20)	0.12 (0.16)	0.81 (0.06)*	0.07 (0.01)

*Genetic correlations significant at $P < 0.05$.

Table 6. Mean reliabilities of sire PTA computed with pedigree information and genomic information

Health event	Pedigree information			Blended pedigree and genomic information			
	Overall mean	Unproven sires[1]	Proven sires[2]	Overall mean	Unproven sires	Proven sires	Overall gain[3]
Displaced abomasum	0.44	0.22	0.65	0.55	0.38	0.71	0.11
Ketosis	0.35	0.18	0.52	0.48	0.35	0.61	0.13
Lameness	0.24	0.15	0.32	0.39	0.31	0.47	0.15
Mastitis	0.39	0.26	0.52	0.51	0.40	0.61	0.12
Metritis	0.35	0.24	0.46	0.48	0.38	0.57	0.13
Retained placenta	0.55	0.42	0.67	0.64	0.54	0.73	0.09

[1]Unproven sires considered sires with less than 10 daughters.

[2]Proven sires considered sires with at least 10 daughters.

[3]The increase in mean reliability calculated as the difference in overall mean reliability between the blended model and the traditional (pedigree data only) model.

ranged from 0.15 to 0.17 (Koivula et al., 2012), which is lower than the reliability estimated for MAST in the current study.

The amount of increase in reliability from pedigree-based analysis compared with genomic-based analysis varied among sires. The number of daughters with records varied for each sire and health event, with the maximum number of daughter records being 1,567 for MAST. The average number of daughters per sire across all health events was approximately 20. Although sires with at least 10 daughters had higher mean reliability overall, sires with less than 10 daughters had the greatest improvement in reliability from the addition of genomic data. These results are expected. Sires with numerous daughters generally have sufficient phenotypic data to achieve acceptable reliabilities. Young sires may not have had sufficient time to accrue the number of daughters needed to reach equivalent reliability levels. The trend for increase in reliability based on number of daughters for each sire is shown in Figure 1 for MAST. As the number of daughters increased, the amount of improvement in reliability decreased. Results from all other health events showed a similar pattern.

Sire posterior means of daughters' probability to each disease are shown in Figure 2 for first-parity records. The mean probability can be considered the average percentage of bulls' daughters expected to experience an incidence of a given health event under equivalent management conditions. The highest mean sire PTA of probability to a disease, 0.27, was found for RETP, which is similar to the value for METR reported by Zwald et al. (2004a), where METR included cases of either METR or RETP. The probability of MAST in first parity was lower than that reported by Zwald et al. (2004a), but was similar at 0.17 in later-parity records.

Approximate genetic correlations between health traits and more common fitness traits are listed in Table 7 for first-parity data. All correlations were significant

($P < 0.05$) except between CYST and SCS, METR and SCS, DSAB and MY, and KETO and MY. Significant negative correlations were found between DPR and PL, with all health events in both first- (Table 7) and later- (data not shown) parity groups. Negative correlations between health events and DPR and PL were also found by Zwald et al. (2004b), and support the proposition that increased genetic liability to disease is associated with decreased cow reproductive performance and longevity. Significant negative genetic correlations were also estimated between NM and all health events except CYST. Previous research has shown a positive cor-
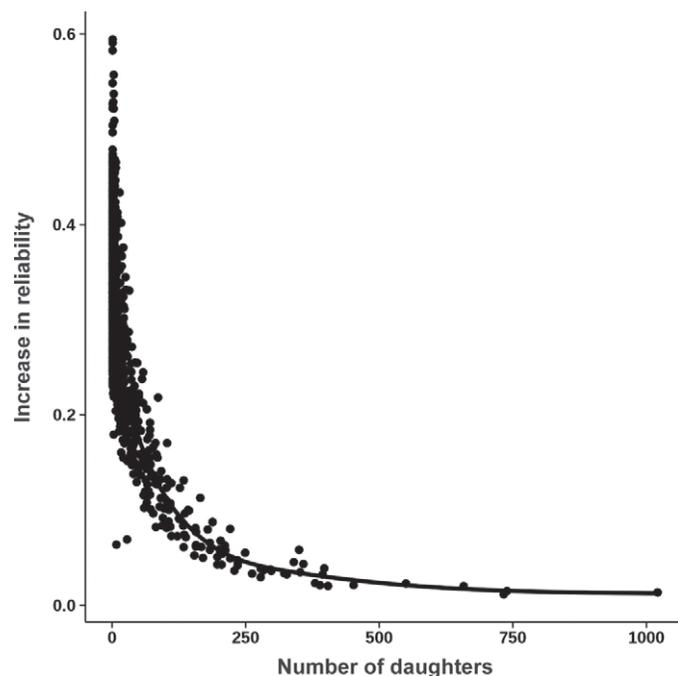

Figure 1. Trend for number of daughters plotted against increase in reliability for each sire in single-step analysis for mastitis.

Table 7. Approximated genetic correlations (SE) between fitness traits[1] and net merit (NM) with results from pedigree-based analysis of first-parity records

Trait	DPR	PL	MY	SCS	NM
Cystic ovaries	−0.18 (0.019)*	−0.20 (0.019)*	0.24 (0.015)*	0.002 (0.018)	0.28 (0.015)*
Displaced abomasum	−0.47 (0.021)*	−0.47 (0.021)*	0.02 (0.017)	0.19 (0.016)*	−0.35 (0.020)*
Ketosis	−0.48 (0.021)*	−0.48 (0.021)*	0.02 (0.017)	0.25 (0.015)*	−0.40 (0.020)*
Lameness	−0.19 (0.019)*	−0.32 (0.020)*	−0.09 (0.018)*	0.48 (0.013)*	−0.44 (0.021)*
Mastitis	−0.19 (0.019)*	−0.27 (0.020)*	0.09 (0.017)*	0.56 (0.012)*	−0.21 (0.019)*
Metritis	−0.35 (0.020)*	−0.18 (0.019)*	−0.21 (0.019)*	0.006 (0.017)	−0.37 (0.021)*
Retained placenta	−0.44 (0.021)*	−0.34 (0.020)*	−0.05 (0.018)*	0.24 (0.015)*	−0.30 (0.020)*

[1]DPR = daughter pregnancy rate; PL = productive life; MY = milk yield.
*Genetic correlation significant at $P < 0.05$.

relation between CYST and milk production (Johnson et al., 1966; Zwald et al., 2004b); however, milk volume receives zero emphasis in the index for NM (Cole et al., 2010). This could reflect a positive correlation between CYST and components of NM, such as protein yield. Although a consensus has not been reached, studies have estimated a positive genetic correlation between CYST and protein yield (Hooijer et al., 2001; Vanholder et al., 2006). The genetic correlations approximated in the current study also suggest a small, positive genetic correlation between MY and CYST. Somatic cell score was most highly correlated with MAST, having an ap-

proximate genetic correlation equal to 0.56 (±0.012). This was expected, given the well-known correlation of MAST with SCS (Carlén et al., 2004; Heringstad et al., 2006; Gernand et al., 2012).

When selecting a genomic evaluation method, many aspects need to be considered. Low heritability traits will need a larger number of records to reach reliabilities equivalent to those found for more heritable traits (Hayes et al., 2009). As more records are collected, it is also important that those records be consistent (Goddard and Hayes, 2009). Consistent recording of health data are more difficult than other traits due to subjectivity of diagnosis and reporting. Accumulation of more health records over time, as well as additional genotypes, is expected to improve genomic prediction, regardless of the method being used. This will allow more rapid genetic improvement to be made in lowly heritable yet economically important traits.

Advantages of single-step methodology, in addition to only requiring 1 step, include that traditional BLUP methodology can be used with only modification to the relationship matrix. This makes the single-step method easy to implement for complex data and models such as multivariate, threshold, and random regression models (Aguilar et al., 2010). The main disadvantage of the single-step method is that it can be more computationally expensive due to having to form the $\mathbf{H}^{-1}$ matrix, although further methods have been developed to more efficiently compute this matrix (Aguilar et al., 2010). Reliabilities of prediction also have to be approximated because direct matrix inversion is infeasible for large data sets. This will become especially important as the number of genotyped animals increases (Misztal et al., 2013).

Regardless of the method for recording and analyzing health events, the health status of cows in a herd can have a large effect on profitability. The cost of a health event will largely depend on the severity of the event and the treatments that are used, as well as other factors not related directly to the individual cow, such as the current cost of milk and herd pregnancy
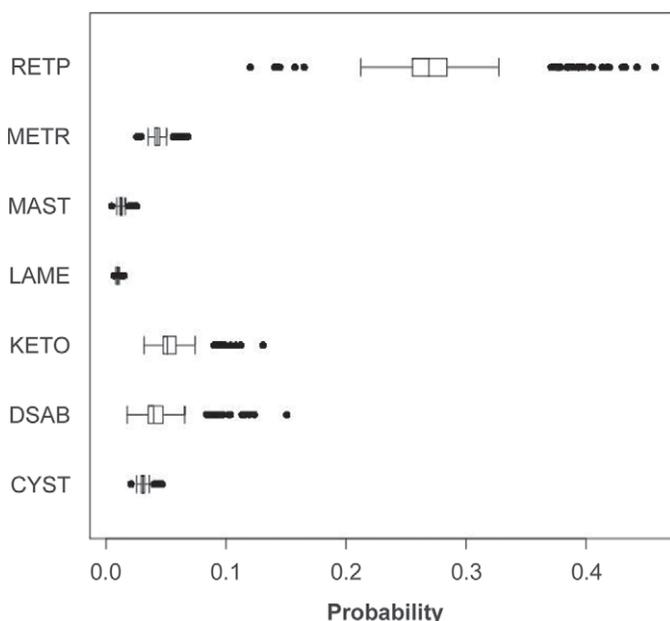

Figure 2. Sire posterior mean PTA of daughters' probability to each health event in first parity. CYST = cystic ovaries; DSAB = displaced abomasum; KETO = ketosis; LAME = lameness; MAST = mastitis; METR = metritis; RETP = retained placenta. The bottom and top bars of the boxes represent the first and third quartiles. The band within each box represents the median. The whiskers represent the lowest and highest data points within 1.5 times the interquartile range. Data points outside this range are represented by individual points.

rate. An effective strategy to keep these costs as low as possible, irrespective of all other factors, may be to incorporate genomic selection for improved cow health. Before further progress can be accomplished, however, many challenges still exist. A unified recording system would greatly improve the consistency of health event reporting. The incorporation of genomic data will allow progress to be made at a more rapid rate, but that does not diminish the necessity for many phenotypes. Additional research will also need to further investigate the incorporation of genomic data. Several other methods using a Bayesian framework exist that were not explored herein. Little research has been conducted at this time that investigates the performance of genomic methodologies when applied to data from animals with low reliabilities for lowly heritable traits. Based on the improvement in reliability estimated in this study, health traits have the potential to greatly benefit from genomic data, which will in turn lead to increased profitability for producers. Before this can occur, however, further research will need to explore the performance of the different methodologies.

CONCLUSIONS

This study demonstrated the potential for genetic improvement of health traits using producer-recorded data. Significant genetic components were estimated for all common health events investigated when evaluated using either pedigree data or pedigree data blended with genomic data. Health traits were lowly heritable, making consistent, long-term goals essential to achieve genetic improvement, regardless of the availability of genomic data. Significant correlations were found between RETP and METR, and between KETO and DSAB. The incorporation of genomic information using single-step methodology increased mean sire reliability by 9 to 15 percentage points. The largest improvement in sire reliability was found for sires with fewer than 10 daughters with health records. Based on this, it would be feasible to use genomic information on young bulls to achieve acceptable reliabilities in a shorter period of time.

ACKNOWLEDGMENTS

REFERENCES

Abdel-Azim, G. A., A. E. Freeman, M. E. Kehrli Jr., S. C. Kelm, J. L. Burton, A. L. Kuck, and 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, 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.

Aguilar, I., I. Misztal, A. Legarra, and 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, and 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.

Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtsen, W. F. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J. Dairy Sci. 94:4700–4707.

Calo, L. L., R. E. McDowell, L. D. Van Vleck, and P. D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. J. Anim. Sci. 37:676–682.

Carlén, E., E. Strandberg, and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows. J. Dairy Sci. 87:3062–3070.

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

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

Cole, J. B., and P. M. VanRaden, and Multi-State Project S-1040. 2010. Net merit as a measure of lifetime profit: 2010 revision. AIPL Research Report NM$4 (12-09). USDA Animal Improvement Programs Laboratory (AIPL), Beltsville, MD.

Curtis, C. R., H. N. Erb, C. J. Sniffen, R. D. Smith, and D. S. Kronfeld. 1985. Path analysis of dry period nutrition, postpartum metabolic and reproductive disorders, and mastitis in Holstein cows. J. Dairy Sci. 68:2347–2360.

Erb, H. N., S. W. Martin, N. Ison, and S. Swaminathan. 1981. Interrelationships between production and reproductive diseases in Holstein cows. Conditional relationships between production and disease. J. Dairy Sci. 64:272–281.

Gernand, E., P. Rehbein, U. U. von Borstel, and S. König. 2012. Incidences of and genetic parameters for mastitis, claw disorders, and common health traits recorded in dairy cattle contract herds. J. Dairy Sci. 95:2144–2156.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Pages 169–193 in Bayesian Statistics 4. J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, ed. Oxford University Press, Oxford, UK.

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381–391.

Hansen, L. B. 2000. Consequences of selection for milk yield from a geneticist's viewpoint. J. Dairy Sci. 83:1145–1150.

Harder, B., J. Bennewitz, D. Hinrichs, and 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.

Harris, B. L., D. L. Johnson, and R. J. Spelman. 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. Interbull Meeting, Niagara Falls, Canada.

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.

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

Heringstad, B., D. Gianola, Y. M. Chang, J. Ødegård, and G. Klemetsdal. 2006. Genetic associations between clinical mastitis and somatic cell score in early first-lactation cows. J. Dairy Sci. 89:2236–2244.

Hooijer, G. A., R. B. F. Lubbers, B. J. Ducro, J. A. M. van Arendonk, L. M. T. E. Kaal-Lansbergen, and T. van der Lende. 2001. Genetic parameters for cystic ovarian disease in Dutch Black and White dairy cattle. J. Dairy Sci. 84:286–291.

Johnson, A. D., J. E. Legates, and L. C. Ulberg. 1966. Relationship between follicular cysts and milk production in dairy cattle. J. Dairy Sci. 49:865–868.

Kelton, D. F., K. D. Lissemore, and 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, and F. S. Schenkel. 2012. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099–4108.

Koivula, M., I. Strandén, G. Su, and E. A. Mäntysaari. 2012. Different methods to calculate genomic predictions—Comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). J. Dairy Sci. 95:4065–4073.

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

Marion, G., and H. Gier. 1968. Factors affecting bovine ovarian activity after parturition. J. Anim. Sci. 27:1621–1626.

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, I. Aguilar, A. Legarra, P. M. VanRaden, and 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, and D. Lee. 2002. BLUPF90 and related programs (BGF90). Communication No. 28-07 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

Neuenschwander, T. F.-O., F. Miglior, J. Jamrozik, O. Berke, D. F. Kelton, and L. R. Schaeffer. 2012. Genetic parameters for producer-recorded health data in Canadian Holstein cattle. Animal 6:571–578.

Olson, K. M., P. M. VanRaden, M. E. Tooker, and T. A. Cooper. 2011. Differences among methods to validate genomic evaluations for dairy cattle. J. Dairy Sci. 94:2613–2620.

Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su, and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in purebred pigs. Genet. Sel. Evol. 43:38.

Parker Gaddis, K. L., J. B. Cole, J. S. Clay, and 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.

Philipsson, J., and B. Lindhé. 2003. Experiences of including reproduction and health traits in Scandinavian dairy cattle breeding programmes. Livest. Prod. Sci. 83:99–112.

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

R Development Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

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

Simeone, R., I. Misztal, I. Aguilar, and Z. G. Vitezica. 2012. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. J. Anim. Breed. Genet. 129:3–10.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: The Principles and Practice of Statistics in Biological Research. 3rd ed. W. H. Freeman, New York, NY.

Su, G., B. Guldbrandtsen, V. R. Gregersen, and M. S. Lund. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J. Dairy Sci. 93:1175–1183.

Tsuruta, S., and I. Misztal. 2006. THRGIBBS1F90 for estimation of variance components with threshold linear models. Communication No. 27-31 in Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil.

Uribe, H. A., B. W. Kennedy, S. W. Martin, and D. F. Kelton. 1995. Genetic parameters for common health disorders of Holstein cows. J. Dairy Sci. 78:421–430.

Vanholder, T., G. Opsomer, and A. de Kruif. 2006. Aetiology and pathogenesis of cystic ovarian follicles in dairy cattle: A review. Reprod. Nutr. Dev. 46:105–119.

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

VanRaden, P. M., A. H. Sanders, M. E. Tooker, R. H. Miller, H. D. Norman, M. T. Kuhn, and G. R. Wiggans. 2004. Development of a national genetic evaluation for cow fertility. J. Dairy Sci. 87:2285–2292.

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.

VanRaden, P. M., and G. R. Wiggans. 1995. Productive life evaluations: Calculation, accuracy, and economic value. J. Dairy Sci. 78:631–638.

Veerkamp, R. F., H. A. Mulder, R. Thompson, and M. P. L. Calus. 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. J. Dairy Sci. 94:4189–4197.

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