



Incidence validation and relationship analysis of producer-recorded health event data from on-farm computer systems in the United States

K. L. Parker Gaddis,^{*1} J. B. Cole,[‡] J. S. Clay,[†] and C. Maltecca^{*}

^{*}Department of Animal Science, North Carolina State University, Raleigh 27695

[†]Dairy Records Management Systems, Raleigh, NC 27603

[‡]Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

The principal objective of this study was to analyze the plausibility of health data recorded through on-farm recording systems throughout the United States. Substantial progress has been made in the genetic improvement of production traits while health and fitness traits of dairy cattle have declined. Health traits are generally expensive and difficult to measure, but health event data collected from on-farm computer management systems may provide an effective and low-cost source of health information. To validate editing methods, incidence rates of on-farm recorded health event data were compared with incidence rates reported in the literature. Putative relationships among common health events were examined using logistic regression for each of 3 timeframes: 0 to 60, 61 to 90, and 91 to 150 d in milk. Health events occurring on average before the health event of interest were included in each model as predictors when significant. Calculated incidence rates ranged from 1.37% for respiratory problems to 12.32% for mastitis. Most health events reported had incidence rates lower than the average incidence rate found in the literature. This may partially represent underreporting by dairy farmers who record disease events only when a treatment or other intervention is required. Path diagrams developed using odds ratios calculated from logistic regression models for each of 13 common health events allowed putative relationships to be examined. The greatest odds ratios were estimated to be the influence of ketosis on displaced abomasum (15.5) and the influence of retained placenta on metritis (8.37), and were consistent with earlier reports. The results of this analysis provide evidence for the plausibility of on-farm recorded health information.

Key words: dairy cattle, health, path analysis

INTRODUCTION

Production traits in dairy cattle are generally easy and inexpensive to measure. With the advent of artificial insemination and progeny testing, milk production per cow has more than tripled, along with an increase in protein and fat content, since the 1950s (National Agricultural Statistics Service, 2011). Conversely, health and fitness traits are difficult and expensive to measure. With a focus on production in the past, producers have had a larger incentive to increase profit by increasing production as opposed to decreasing management costs through improved health and fitness. With the great strides made in production, an antagonistic relationship between production and most disease traits has become apparent (Rauw et al., 1998).

A substantial limitation in the development of a system for genetic improvement of health traits in the past has been the lack of a central collection of health data. The United States has no mandatory or unified system for reporting health events on dairy farms. Relatively small studies have been completed using data collected from paper records. Kaneene and Hurd (1990) used data collected by specially trained veterinary officers during farm visits for the National Animal Health Monitoring System in Michigan. Specific worksheets were designed for the producers to use for data recording. Lyons et al. (1991) used data collected from forms given to producers to record incidences of health problems as they occurred. They analyzed the incidence of 22 individual health traits from 3,664 records supplied by producers in Wisconsin, Minnesota, and Iowa. Nonetheless, similar protocols are too labor intensive to be performed on a national level.

Health event data collected from on-farm computer management systems may provide an effective and low-cost source of health trait information. Studies have been completed using data recorded using computerized systems. Bartlett et al. (1986a,b,c) examined the incidence, descriptive epidemiology, and estimated economic impact of metritis (**METR**), mastitis (**MAST**), and cystic follicular disease. Data were collected from

Received March 26, 2012.

Accepted May 24, 2012.

¹Corresponding author: klparke4@ncsu.edu

22 herds in Michigan that participated in a computerized herd health program. Zwald et al. (2004) used data collected from on-farm computerized systems in the United States to determine the feasibility of genetic selection for health traits. Diseases analyzed included displaced abomasum (**DSAB**), ketosis (**KETO**), MAST, lameness (**LAME**), cystic ovaries (**CYST**), and METR occurring between 2001 and 2003. They concluded that using data from on-farm computerized recording systems would allow genetic selection to be used against common health disorders.

Collected health event data could be used as a source of information to gain insight into relationships between health events. Previous research has examined the relationships between diseases. Erb et al. (1981) used data from 20 commercial dairy herds participating in a herd health program of Ontario Veterinary College to construct path models including dystocia (**DYST**), retained placenta (**RETP**), METR, cystic follicles, and luteal cysts. An observational study using 34 Holstein herds in southwest Ontario was completed to examine associations between 11 health problems (Bigras-Poulin et al., 1990). van Dorp et al. (1999) used data from 32 registered Holstein herds in British Columbia to examine the effects of herd, age, year, season, and inter-relationships between diseases.

The above studies indicate that on-farm recorded data can indeed serve as a wealth of information to further understand the complexities of health traits; however, few studies have been conducted within the United States with more than 3 yr of data collection. The objective of this study included analyzing the reliability of health data recorded through on-farm recording systems throughout the United States. The overall goal when handling these data should be that the producer-recorded data accurately reflect the true incidence of health events in herds and accurately portray the relationships between health events that have previously been identified or explained biologically. Given that on-farm recorded health data sufficiently represented the true incidence of health events, phenotypic relationships between common health events were examined and compared with knowledge obtained from epidemiological studies.

MATERIALS AND METHODS

Editing Criteria

Two data sets were available for the study from Dairy Records Management Systems (Raleigh, NC): one consisting of health information and the other consisting of production information. General editing

was performed on the health data as summarized in Figure 1. The health information data set originally included 8,361,900 health records. After general editing, health records were classified into 80 categories as described in the supplementary material (Table S1, available at <http://www.journalofdairyscience.org/>). In total, 5,117,485 health records from 1996 to 2009 coincided with 1 of 80 categories, belonging to 544,573 cows across 1,524 herds. The original production data set consisted of 1,840,902 lactation records from 451,334 cows. Comparable general editing was also applied to the production data, resulting in a data set consisting of 1,427,435 records from 438,099 cows.

Health events of interest were those defined in "Format 6: Health Record" data exchange format (Animal Improvement Programs Laboratory, 2010). Standardized health codes were assigned to records in an effort to correct for improper spelling and inconsistent terminology based on the health identification code and health description reported by producers. Health events were assigned to a lactation, with lactations beginning with a calving. For each health event, heifer records and terminated lactations were excluded, the event had to be recorded within 365 d of calving, and the event had to occur within parities 1 through 5. Terminated lactations were defined as cows that were culled before the end of lactation. Termination codes were those found in "Format 4: Lactation" data exchange format (Animal Improvement Programs Laboratory, 2006). Minimum and maximum constraints were placed on the reporting frequency of each health event. Constraints were used to avoid herds that did not report a health event as well as herds that reported only sick cows. The minimum constraint was imposed by selecting records from herd-years with at least one reported incidence of the health event of interest and herd-years consisting of at least 5 cows. A maximum reporting frequency constraint was imposed by excluding herd-years with a reporting frequency greater than 2 standard deviations above the mean reporting frequency for that health event. Selected health events were further edited to ensure that **DYST** and **RETP** were recorded within 7 d of the calving date.

Summary statistics for each health event of interest are shown in Table 1, including number of herds reporting, number of cows, and total number of cases. Digestive problems (**DIGE**) included those reported as a general digestive problem by the producer and not those health events already explicitly included. Reproductive problems (**REPR**) included abortion, breech calf, cesarean section, ovary problem other than cystic ovaries, stillborn or mummified calf, uterine infusion, uterine infection or injury, and vaginal and uterine prolapse.

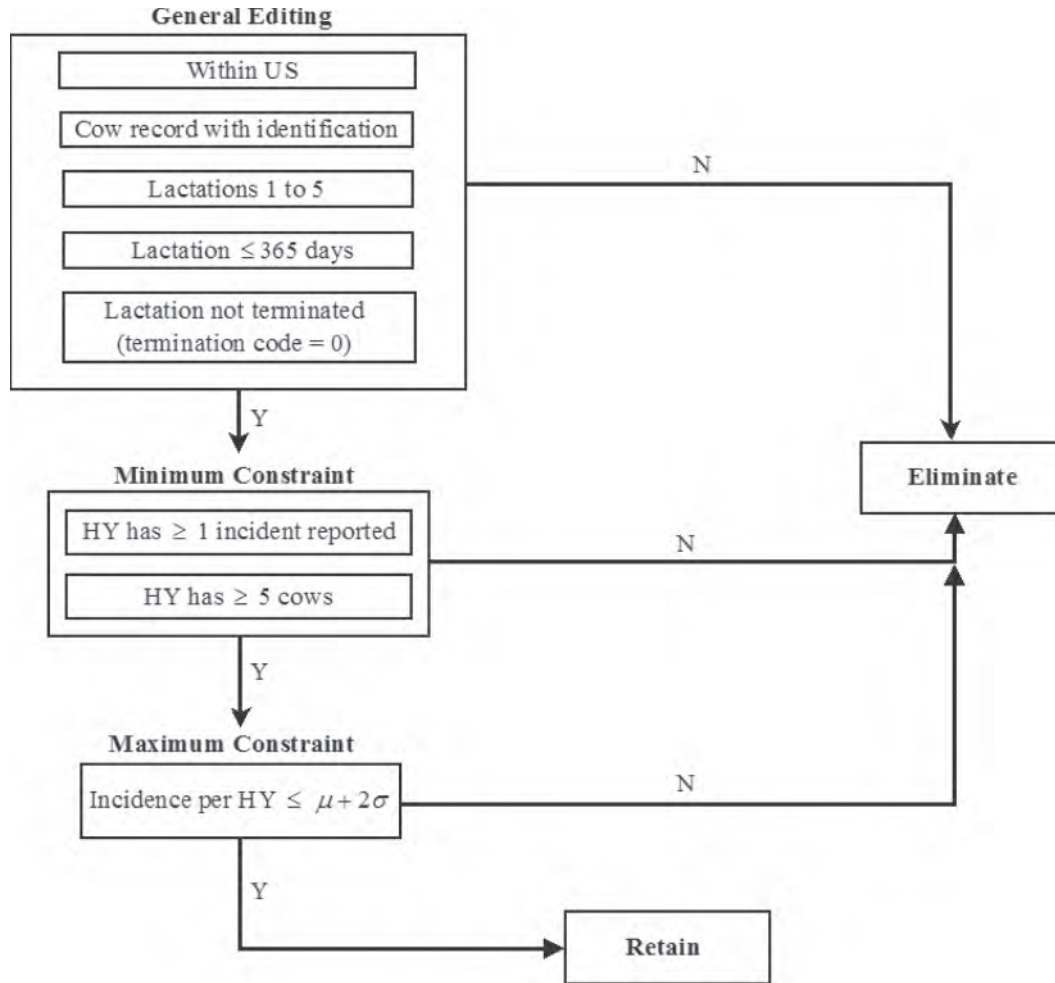


Figure 1. Data editing scheme for health events (HY = herd year; μ = mean incidence of health event; σ = standard deviation of health event; Y = yes; N = no).

Health Event Incidence

Disease frequency can be calculated as incidence (rate of occurrence of new cases of a disease per unit time) or prevalence (proportion of diseased animals at a given time). Disease incidences are more frequently reported in the literature (Kelton et al., 1998). The most common methods of reporting disease incidences are lactational incidence rate (**LIR**) for health events with short periods of risk or incidence density (**ID**) for health events with long periods of risk (Kelton et al., 1998). The LIR was obtained as the number of affected lactations per lactations at risk:

$$LIR = \frac{LAC_d}{LAC_t},$$

where LAC_d indicates the number of first occurrences of a specific health event in a lactation and LAC_t indicates the number of lactations at risk. The ID represents the number of new cases of a health event occurring during a lactation in a herd and was calculated as follows:

$$ID = \frac{LAC_d}{\left(\frac{LAC_t + LAC_e}{2} \right)},$$

where LAC_d indicates the number of first occurrences of a specific health event, LAC_t represents the number of cows at risk starting a lactation, and LAC_e indicates the number of cows at risk ending a lactation. The number of cows beginning a lactation was used for both

Table 1. Summary statistics of each health event of interest

Health event ¹	Health event 0 to 60 DIM			Health event 61 to 90 DIM			Health event 91 to 150 DIM		
	Herds (no.)	Cows (no.)	Total cases (no.)	Herds (no.)	Cows (no.)	Total cases (no.)	Herds (no.)	Cows (no.)	Total cases (no.)
DYST	317	123,552	5,024						
RETP	598	185,154	12,602						
KETO	211	75,458	5,280	211	74,150	36	211	74,146	27
DSAB	31	10,382	297	31	10,326	8	31	10,325	2
CALC	237	96,899	1,446	237	96,677	24	237	96,720	101
METR	600	200,875	25,525	599	192,862	385	599	193,057	1,071
DIAR	42	19,699	498	42	19,624	50	42	19,623	50
DIGE	442	152,730	5,352	442	151,591	415	442	151,652	667
RESP	457	169,561	3,013	457	169,060	277	457	169,106	384
REPR	723	225,558	7,445	720	223,709	508	721	224,142	1,850
MAST	759	233,368	26,403	759	227,624	5,174	759	228,424	8,410
CYST	699	190,227	2,519	699	190,449	2,557	707	211,338	3,690
LAME	632	205,531	7,848	631	204,090	2,521	632	204,952	4,863

¹DYST = dystocia; RETP = retained placenta; KETO = ketosis; DSAB = displaced abomasum; CALC = hypocalcemia; METR = metritis; DIAR = diarrhea; DIGE = digestive problem; RESP = respiratory problem; REPR = reproductive problem; MAST = mastitis; CYST = cystic ovaries; LAME = lameness.

the number of lactations at risk and the number of cows at risk starting a lactation. The number of cows at risk ending a lactation was obtained as the number of cows that did not have the health event of interest reported throughout that lactation ($LAC_t - LAC_d$). The standard error of the mean LIR or ID was calculated for each health event over lactations. Incidences calculated from the data were compared with those found in the literature. Between 5 and 30 studies were used to calculate the mean literature incidence and 95% incidence range for each health event, with a total of 46 studies used.

Phenotypic Analysis of Relationships Between Health Events

Logistic regression was used to analyze putative relationships among common health events. Logistic regression uses the logit link function. The logistic regression model was $\eta = \mathbf{X}\boldsymbol{\beta}$, where η represents the logit of observing the health event of interest, $\boldsymbol{\beta}$ represents a vector of fixed effects, and \mathbf{X} is the corresponding incidence matrix. If the probability of observing the health event of interest ($Y_t = 1$) is π_i , then the odds of observing the health event of interest is given by $\pi_i / (1 - \pi_i)$. The logit of observing the health event of interest is

$$\eta_i = \log \left(\frac{\pi_i}{(1 - \pi_i)} \right).$$

When using logistic regression, the coefficient estimates are equal to the log odds ratio given by

$$\boldsymbol{\beta} = \log \frac{\pi_i / (1 - \pi_i)}{\pi_j / (1 - \pi_j)}.$$

The odds ratio (**OR**) can be obtained by taking the exponential function of the coefficient estimates. An OR of 1.0 indicates no association between the independent and dependent variable, and OR further from 1.0 indicate stronger associations. Fixed effects included herd, parity, year, breed, and season of calving. Four seasons of calving were defined: January to March, April to June, July to September, and October to December. Breed effect included 5 levels: Brown Swiss, Holstein, Jersey, crossbred, and others, where this last level included the remaining minor breeds. The data were divided into 3 sets based on DIM at occurrence: 0 to 60 DIM, 61 to 90 DIM, and 91 to 150 DIM. Only the first occurrence of a health event for each cow within each timeframe was included in the analysis. Health events with an earlier mean DIM of occurrence were allowed to enter the model when significant as predictors. A record was given a “1” for a predictor health event if an incidence occurred before an incidence of the health event of interest and “0” otherwise. Analyses for the 0 to 60 DIM timeframe included incidences occurring within that timeframe. Similarly, analyses for the 61 to 90 and 91 to 150 DIM timeframes were completed with the health event of interest restricted to the specified timeframe, but predictor health events were allowed to range from 0 DIM up to the maximum DIM of the timeframe (i.e., either 90 or 150 DIM).

Because of the size of the data sets, sampling was used to arrive at a final model. Smaller data sets were created by sampling 100 herds without replacement

from the full data set for each health event of interest. For health events with fewer than 100 herds reporting [diarrhea (**DIAR**) and DSAB], smaller subsets were used. The DIAR data set sampled 30 herds from the full data set, and the DSAB data set sampled 20 herds from the full data set. This sampling procedure was repeated 20 times. For each sampled data set, forward and reverse stepwise regression was used to select prior health events that should enter the model using the step function of R (R Development Core Team, 2011). When a predictor entered at least 50% of models produced with the sampled data, that event was included as a predictor in the final model fitted with the full data set. A schematic of this process is shown in Figure 2. Analyses of full data sets were performed using the glm function of R (R Development Core Team, 2011). Path diagrams were constructed using OR estimates from logistic regression analyses.

Additional analyses were performed to further examine relationships among health events. An analysis was completed for the 61 to 90 DIM and 91 to 150 DIM timeframes for MAST and LAME. These health events are more likely to occur multiple times throughout a single lactation such that odds of a second incidence later in lactation may be increased. To analyze this, previous incidences were included as predictors for the models developed for MAST and LAME, as previously described. For example, incidences of MAST occurring within 0 to 60 DIM were included as predictors when modeling the occurrence of MAST within the 61 to 90 DIM timeframe. Incidences of MAST occurring within 0 to 90 DIM were included as predictors when modeling the occurrence of MAST within the 91 to 150 DIM timeframe. The same procedure was completed for LAME.

A separate analysis was completed to identify differences in cows having lactations longer than 365 d. Cows with extended lactations are generally cows that have not become pregnant again. It can be hypothesized that an incidence of a disease may result in a lack of conception. Overall incidence for each health event was compared between cows with lactations ending within 365 d and cows with lactations >365 d in length. Contingency tables were calculated and significance of the incidence difference for each health event was tested using a Chi-squared test. An additional path analysis was completed for cows with lactations >365 d for events occurring within the 0 to 60 DIM timeframe.

In the original models, parity was included as a fixed effect; however, it is reasonable to consider first and later parities separately due to physiological differences found in first-parity cows. A separate analysis was conducted for health events with a higher incidence rate in first-parity cows (METR and REPR). Separate

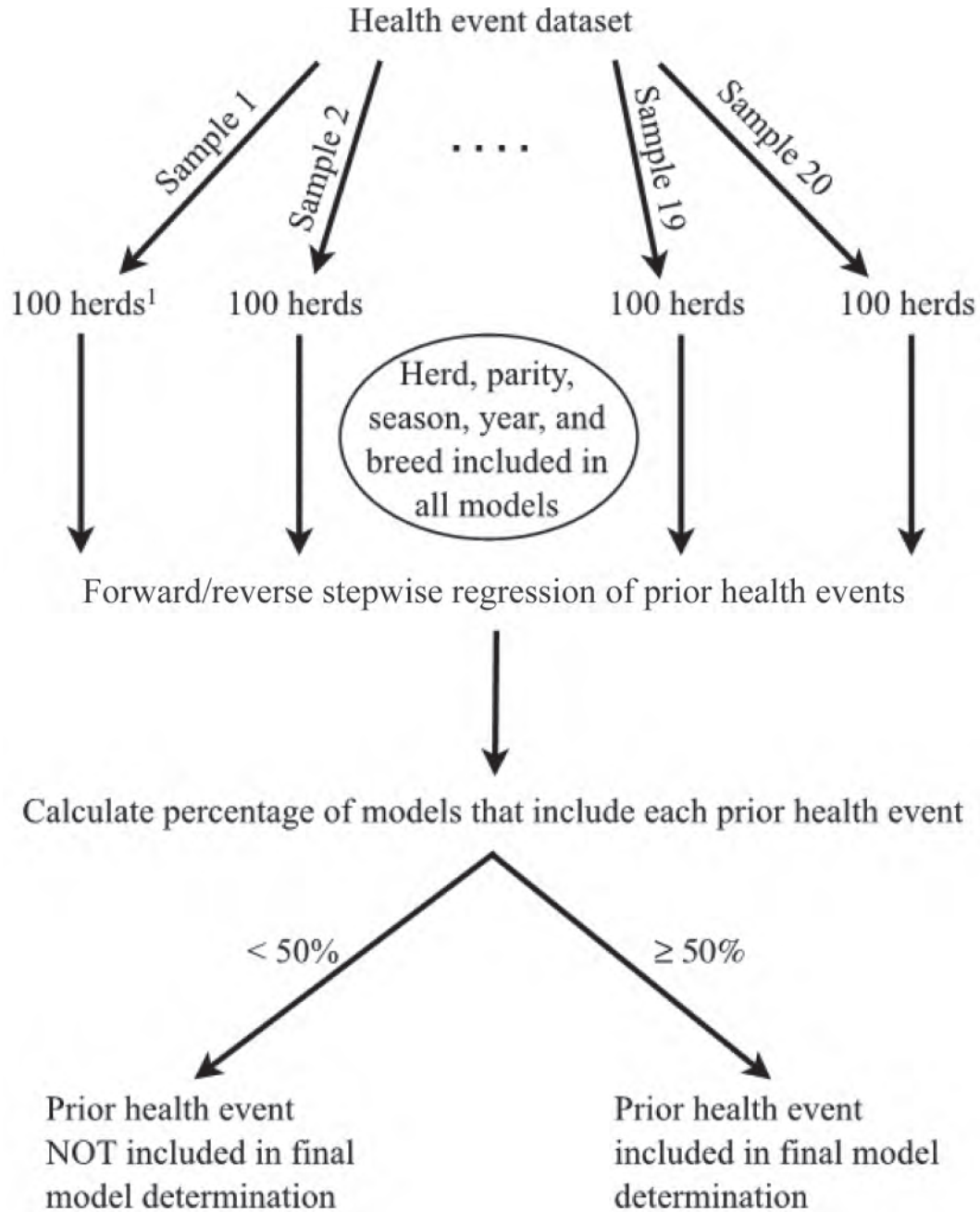
logistic models were analyzed for first-parity and later parities for the 0 to 60 DIM timeframe. Estimates from these models were compared with the estimates from the original models.

RESULTS

The LIR or ID of each health event by lactation is shown in Table 2, which also includes mean incidence and 95% incidence range of each health event as found in literature. Fewer reports were found for health events more commonly reported in calves such as DIAR, DIGE, and respiratory problems (**RESP**). For all other health events, at least 10 citations were used. The literature incidence for REPR is not reported in Table 2 because REPR represents a collection of reproductive disorders. All calculated incidences fell within the 95% range of incidences found in literature with the exception of DIAR. Each calculated mean incidence was less than the mean incidence found in literature except for KETO and DIGE. The incidence of most events increased from first lactation to fifth lactation with the exception of DYST, METR, and REPR.

Table 3 shows the results from logistic regression analyses for each timeframe of 1 to 60 DIM, 61 to 90 DIM, and 91 to 150 DIM, including the estimate, standard error, and probability for all significant results ($P < 0.05$). Each estimate represents the log OR when all other predictors are held fixed. Path diagrams are shown for each timeframe in Figures 3, 4, and 5. Health events are represented by acronyms, and are classified into 1 of 6 categories (reproductive, digestive, mammary, respiratory, locomotive, miscellaneous) depicted by different shapes in the diagrams. Edges represent putative relationships. Weights assigned to edges are OR that the health event at the arrowhead will occur given a prior incidence of the health event at the base of the arrow. Greater OR indicate stronger associations.

Within the 0 to 60 DIM timeframe, odds increased overall with increasing parity with the exceptions of DYST, REPR, DSAB, and RESP. Odds of DYST decreased with increasing parity, and METR and REPR had the highest odds in first parity. Jerseys were found to have decreased odds of DYST compared with other breeds, whereas Holsteins were found to have increased odds of REPR. Comparisons between breeds must be considered carefully because breeds such as Holstein and Jersey were more highly represented than the other breeds. Within the 61 to 90 DIM timeframe, odds increased overall with increasing parity except for CALC, DIAR, DSAB, REPR, and RESP. The lack of a pattern found for CALC, DIAR, and DSAB may be the result of few incidences reported within that timeframe. Jerseys were also found to have decreased odds of MAST



¹Smaller subsets were used for health events with fewer than 100 herds reporting (diarrhea dataset sampled 30 herds; displaced abomasum dataset sampled 20 herds).

Figure 2. Model construction schematic for each health event of interest.

and REPR within this timeframe. Within the 91 to 150 DIM timeframe, we observed an overall increase in odds with increasing parity as well. Exceptions were found

for CALC, DIGE, DSAB, KETO, and RESP. Similarly to the 61 to 90 DIM timeframe, CALC, DIAR, DSAB, and KETO had few incidences reported within the

Table 2. Health event incidence by lactation, mean over lactations, and mean literature incidence with 95% range

Health event ¹	Lactation	ID ² (%)	LIR ³ (%)	Mean (SE) over lactations (%)	Mean literature incidence ⁴ (95% range); [no. of citations]
CALC	1		0.23	2.40 (1.06)	7.44 (1.49, 21.75) [18]
	2		0.53		
	3		1.68		
	4		3.67		
	5		5.89		
CYST	1	2.65		3.50 (0.25)	9.05 (0.76, 21.70) [21]
	2	3.38			
	3	3.54			
	4	3.76			
	5	4.16			
DIAR	1		1.58	2.11 (0.35)	5.88 (2.77, 11.22) [5]
	2		1.66		
	3		2.13		
	4		1.72		
	5		3.48		
DIGE	1		2.21	3.04 (0.28)	2.60 (0.20, 6.89) [8]
	2		2.64		
	3		3.11		
	4		3.47		
	5		3.76		
DSAB	1		1.29	2.20 (0.42)	2.67 (0.56, 8.85) [11]
	2		1.26		
	3		3.17		
	4		2.18		
	5		3.11		
DYST	1		2.90	2.04 (0.23)	5.29 (0.80, 13.34) [14]
	2		1.77		
	3		1.70		
	4		1.74		
	5		2.07		
KETO	1		3.42	5.15 (0.78)	5.07 (0.32, 19.50) [21]
	2		3.10		
	3		6.03		
	4		6.61		
	5		6.61		
LAME	1	4.83		6.37 (0.46)	9.27 (2.54, 30.44) [17]
	2	5.80			
	3	6.75			
	4	7.22			
	5	7.24			
MAST	1	9.53		12.32 (1.06)	17.98 (0.96, 39.13) [29]
	2	10.24			
	3	12.98			
	4	13.75			
	5	15.09			
METR	1		8.71	6.86 (0.47)	12.34 (1.77, 35.50) [23]
	2		6.11		
	3		6.28		
	4		6.48		
	5		6.70		
REPR	1	3.06		2.83 (0.11)	
	2	2.50			
	3	2.62			
	4	2.98			
	5	2.97			
RESP	1		1.34	1.37 (0.04)	3.30 (0.21, 7.11) [12]
	2		1.26		
	3		1.41		
	4		1.50		
	5		1.36		

Continued

Table 2 (Continued). Health event incidence by lactation, mean over lactations, and mean literature incidence with 95% range

Health event ¹	Lactation	ID ² (%)	LIR ³ (%)	Mean (SE) over lactations (%)	Mean literature incidence ⁴ (95% range); [no. of citations]
RETP	1		2.49	4.60	8.02
	2		3.97	(0.63)	(2.33, 17.94)
	3		5.04		[30]
	4		5.44		
	5		6.07		

¹CALC = hypocalcemia; CYST = cystic ovaries; DIAR = diarrhea; DIGE = digestive problem; DSAB = displaced abomasum; DYST = dystocia; KETO = ketosis; LAME = lameness; MAST = mastitis; METR = metritis; REPR = reproductive problem; RESP = respiratory problem; RETP = retained placenta.

²ID = incidence density.

³LIR = lactational incidence rate.

⁴Calculated from Appuhamy et al. (2009); Barker et al. (2010); DeGaris and Lean (2008); Dubuc et al. (2010); Emanuelson et al. (1993); Faye (1992); Fleischer et al. (2001); Frei et al. (1997); Gay and Barnouin (2009); Groehn et al. (1992); Gröhn et al. (1989, 1995); Hamann et al. (2004); Heringstad et al. (1999); Miller and Dorn (1990); Mörk et al. (2009); Olde Riekerink et al. (2008); Stevenson (2000); Toni et al. (2011); Yániz et al. (2008).

91 to 150 DIM timeframe. Decreased odds for REPR events were found for parities 2 and 3.

The OR of an animal having an incidence of MAST within the 0 to 60 DIM timeframe and having a second incidence within the 61 to 90 DIM timeframe was 2.94. The OR of an animal having an incidence of MAST within the 0 to 90 DIM timeframe and having a second incidence within the 91 to 150 DIM timeframe was 2.83. Incidences of LAME within the 0 to 60 DIM timeframe did not have a significant effect on the occurrence of a second incidence of LAME within the 61 to 90 DIM timeframe. The OR of an animal having an incidence of LAME within the 0 to 90 DIM timeframe and having a second incidence of LAME within the 91 to 150 DIM timeframe was 1.93.

Cows with lactations >365 d had higher incidence rates of health events for all parities except CALC and DSAB. A higher LIR was found for CALC in third-parity cows with extended lactations. A higher LIR was found for DSAB in first-, second-, and fifth-parity cows with extended lactations. The Chi-squared tests for differences in overall incidence between lactations ending within 365 d and lactations >365 d were significant for all health events except CALC and DSAB ($P < 0.005$), indicating higher incidences for cows with extended lactations. The overall pattern of the path diagram did not change (results not shown). The overall pattern of putative relationships also did not change when comparing first-parity cows versus later-parity cows for METR and REPR.

DISCUSSION

Editing was performed to ensure data plausibility. Records reported outside the United States were removed, as the goal of this research was to evaluate the feasibility

of using producer-recorded data from US farms. Heifer records were removed because heifers have different incidence rates for health events (Uribe et al., 1995). The majority of data were from lactations 1 through 5. Lactations beyond this were excluded to avoid bias from fewer records of later lactations. This editing is also consistent with national genetic evaluations. Cow records terminated before the end of lactation were also removed to avoid any bias. Certain termination codes indicate a particular health event as the main reason a cow was culled. To ensure that a large amount of data was not being lost by excluding all terminated records, termination codes were examined for each health event. The percentage of cows terminated for each termination code was found to be similar between health events. Data from all breeds were included because the main goal of the research was to analyze the plausibility of producer-recorded data as opposed to thoroughly examining differences between breeds. Minimum and maximum frequency constraints were instead used to eliminate questionable data. The minimum constraint was used to identify herds that did not record a particular health event. The maximum constraint was used in an attempt to avoid records from herds that used recording systems to track treatments given to cows.

Incidence from literature were gathered from varied studies conducted from 1979 to 2011. Experimental design, population, and environment were different between most reports. Incidences from this data set are less than those previously calculated from a similar, although smaller data set with the exception of RETP (Cole et al., 2006). Incidences calculated from a meta-analysis were included when calculating mean literature incidence rates (Lean et al., 2006). For diseases more commonly found in calves (DIAR, DIGE, RESP), incidence rates calculated from studies involving calves

Table 3. Logistic regression results 0 to 60 DIM, 61 to 90 DIM, and 91 to 150 DIM

Health event of interest ¹	Prior health event	Estimate***	SE	Probability
Health event 0 to 60 DIM; predictors 0 to 60 DIM				
RETP	DYST	0.76	0.06	0.68
KETO	DYST	0.20*	0.10	0.55
	RETP	0.89	0.06	0.71
CALC	DYST	0.83	0.20	0.70
DSAB	RETP	0.96	0.02	0.72
	KETO	2.74	0.19	0.94
METR	DYST	0.86	0.05	0.70
	RETP	2.13	0.02	0.89
	KETO	0.52	0.06	0.63
	CALC	0.70	0.12	0.67
DIAR	KETO	1.30	0.19	0.79
	METR	0.88	0.14	0.71
DIGE	DYST	0.59	0.08	0.64
	RETP	0.60	0.06	0.65
	KETO	1.27	0.09	0.78
	METR	0.76	0.04	0.68
	DIAR	0.77**	0.24	0.68
RESP	DYST	0.39**	0.13	0.60
	RETP	0.53	0.08	0.63
	KETO	0.96	0.13	0.72
	CALC	1.10	0.25	0.75
	METR	0.50	0.06	0.62
	DIGE	0.96	0.08	0.72
MAST	DYST	0.25	0.05	0.56
	REPR	0.33	0.05	0.58
REPR	DYST	0.20*	0.10	0.55
	RETP	0.87	0.06	0.70
	KETO	0.42	0.13	0.60
Health event 61 to 90 DIM; predictors 0 to 90 DIM				
METR	RETP	1.28	0.21	0.78
DIGE	RETP	0.58**	0.22	0.64
RESP	DYST	0.77*	0.33	0.68
	DIGE	1.20	0.24	0.77
MAST	CALC	0.40*	0.20	0.60
	RESP	0.48	0.12	0.62
REPR	METR	1.22	0.15	0.77
CYST	MAST	0.30**	0.09	0.57
LAME	DIGE	0.63	0.13	0.65
	REPR	0.49	0.14	0.62
Health event 91 to 150 DIM; predictors 0 to 150 DIM				
METR	DYST	0.90	0.16	0.67
	RETP	1.53	0.12	0.99
DIGE	METR	0.61	0.12	0.65
RESP	METR	0.36*	0.18	0.59
	DIGE	0.97	0.21	0.73
REPR	RETP	0.66	0.10	0.66
	METR	1.04	0.07	0.74
	DIGE	0.69	0.12	0.67
MAST	METR	0.17**	0.05	0.54
	DIGE	0.23**	0.07	0.56
	RESP	0.38	0.09	0.59
CYST	RETP	0.21*	0.10	0.55
	MAST	0.25	0.06	0.56
LAME	RETP	0.25**	0.08	0.56
	METR	0.18**	0.06	0.54
	RESP	0.41	0.12	0.60

¹CALC = hypocalcemia; CYST = cystic ovaries; DIAR = diarrhea; DIGE = digestive problem; DSAB = displaced abomasum; DYST = dystocia; KETO = ketosis; LAME = lameness; MAST = mastitis; METR = metritis; REPR = reproductive problem; RESP = respiratory problem; RETP = retained placenta.

***All estimates were significant at $P < 0.001$ unless otherwise noted as **($P < 0.01$) or *($P < 0.05$).

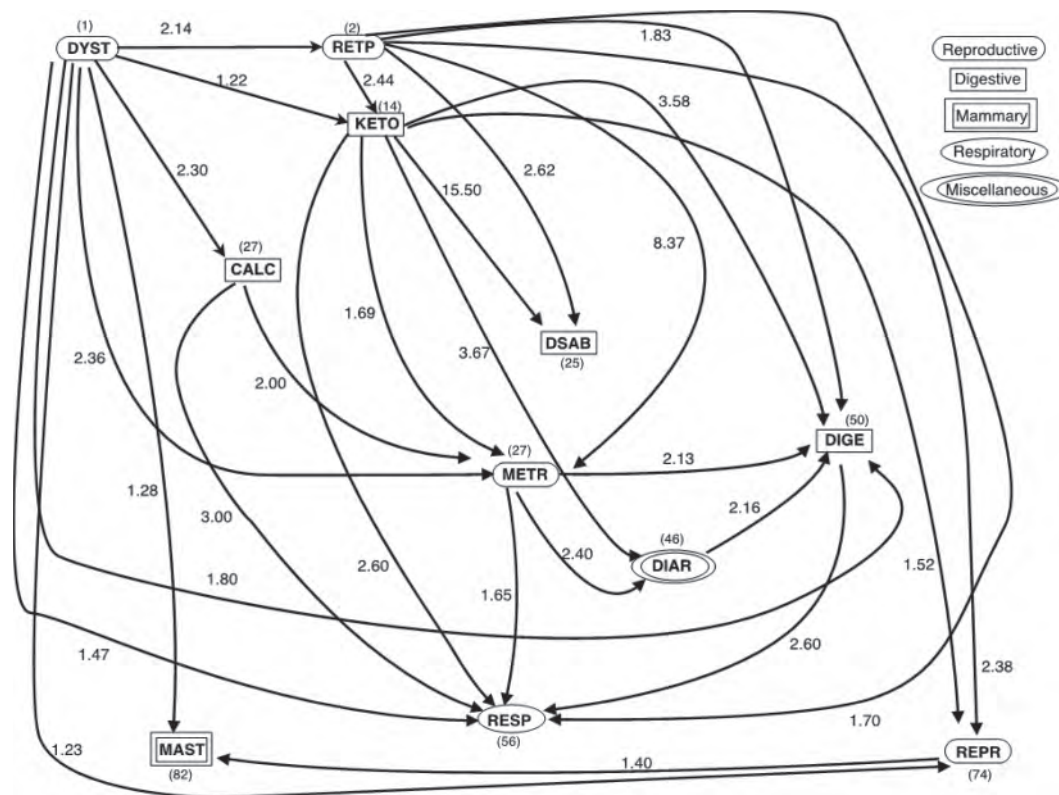


Figure 3. Path analysis of 0 to 60 DIM timeframe, with shapes representing event categories. Overall mean DIM at occurrence is shown in parentheses. CALC = hypocalcemia; CYST = cystic ovaries; DIAR = diarrhea; DIGE = digestive problem; DSAB = displaced abomasum; DYST = dystocia; KETO = ketosis; LAME = lameness; MAST = mastitis; METR = metritis; REPR = reproductive problem; RESP = respiratory problem; RETP = retained placenta.

were also included when calculating mean literature incidence rates (Gardner et al., 1990; Lundborg et al., 2005; Svensson et al., 2006; Gulliksen et al., 2009). Direct comparison across studies is difficult. Health event definitions varied between each study as did methods of calculating incidence. Specific values from each study cannot be used to directly compare results found in this study, although they do lend support that the values calculated from producer-reported data are within the range of previously calculated values. When compared with the mean incidence calculated from literature, producer-reported data are lower, with the exception of KETO and DIGE. This may indicate, at least in part, that producers are more likely to use data systems for recording treatments as opposed to diagnosis of health events. Health events deemed important by producers are likely to have more complete reporting when compared with events considered less important. van Dorp et al. (1999) reported that RETP, METR, MAST, CYST, and stable footrot were likely to be considered most important by producers. Consistent with this conclusion, the health events with the most herds reporting data in this analysis were MAST, REPR,

CYST, METR, LAME, and RETP. van Dorp et al. (1999) also reported that udder edema, milk fever, DSAB, and KETO were likely not considered priority diseases to producers. Again consistent with this conclusion, health events with the fewest number of herds reporting were DSAB, DIAR, KETO, and CALC.

The path diagrams allow putative relationships to be determined from an average timeline of occurrence. The relationships could then be compared with relationships previously described in literature as additional validation of the plausibility of producer-recorded data. The majority of health events occur within the first 60 d following parturition. Health events that occur early in lactation have the potential to influence the risk of experiencing a later health event. The 3 earliest health events (DYST, RETP, and KETO) all had numerous significant pathways leading to later health events within the 0 to 60 DIM timeframe. This indicates that a cow with an incidence of an early health event has an increased risk of experiencing a later health event. Results from DYST must be interpreted carefully because incidences are recorded through a different system. Later timelines were used to analyze the odds of an

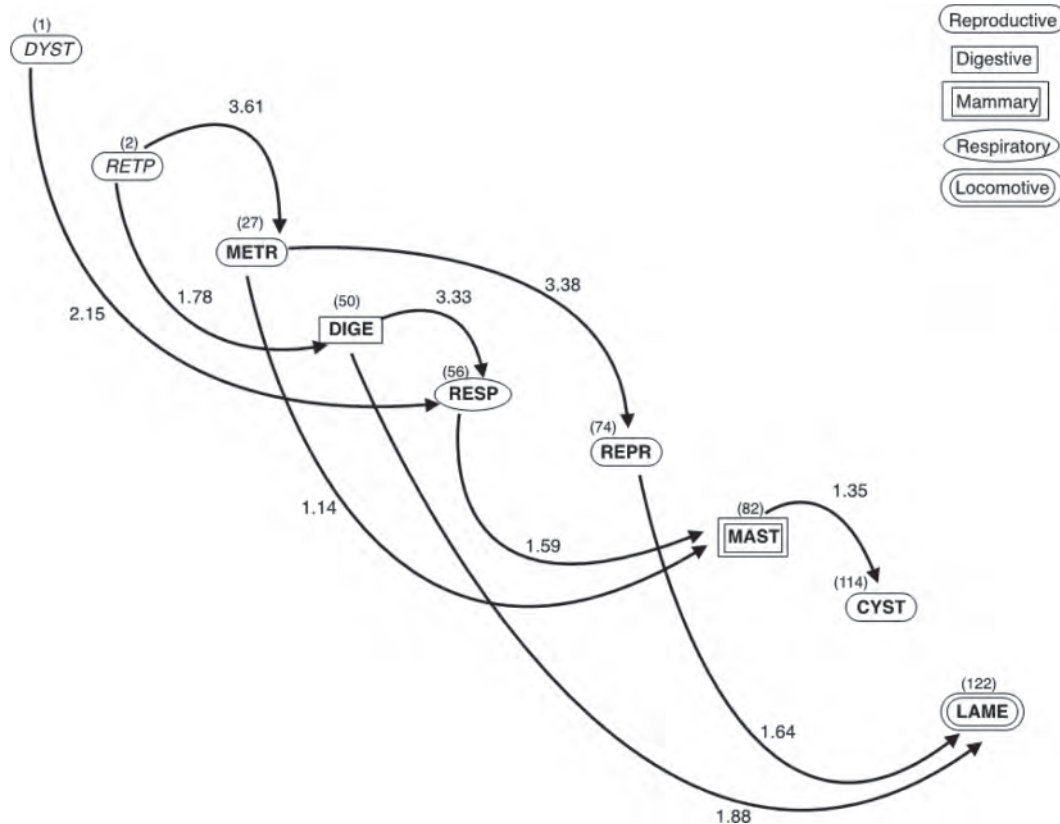


Figure 4. Path analysis of 61 to 90 DIM timeframe, with shapes representing event categories. Overall mean days in milk at occurrence is shown in parentheses. Italicized health events occurred before the 61 to 90 DIM period. CYST = cystic ovaries; DIGE = digestive problem; DYST = dystocia; LAME = lameness; MAST = mastitis; METR = metritis; REPR = reproductive problem; RESP = respiratory problem; RETP = retained placenta.

animal having a later incidence of a health event given a previous health event incidence. Odds ratios ranged from 1.10 to 15.50 in the 0 to 60 DIM timeframe. Odds ratios ranged from 1.14 to 3.61 in the 61 to 90 DIM timeframe and from 1.18 to 4.60 in the 91 to 150 DIM timeframe. The greatest OR was calculated for the influence of a prior incidence of KETO on DSAB. This was also the strongest relationship found by Correa et al. (1993) with an OR of 13.8. Correa et al. (1993) also included a 95% CI for each OR that ranged from 7.2 to 26.3 for the relationship between KETO and DSAB. van Dorp et al. (1999) also found a strong association of KETO on DSAB with an OR of 7.98.

The second greatest OR was 8.37 found between RETP and METR. A relationship between RETP and METR was found to have the strongest association in both later timeframes. The relationship between RETP and METR has been documented previously. Correa et al. (1993) calculated an OR of 6.0 with a 95% CI of 2.8 to 7.5, which is lower than the estimate calculated in this analysis. This CI does include the OR calculated for the 61 to 90 DIM and 91 to 150 DIM timeframes

equal to 3.61 and 4.60, respectively. van Dorp et al. (1999) calculated an OR of 3.53 between RETP and METR. This estimate is lower than the relationship of RETP on METR in any timeframe in this analysis. Erb et al. (1981) found an OR of 5.8 between RETP and METR. This estimate is lower than that calculated for the 0 to 60 timeframe but higher than that for later timeframes.

Several relationships involving the influence of DYST that were found in this analysis have also been found in previous studies. An association between DYST and RETP has been previously documented by Correa et al. (1993) with an OR of 2.2 and a 95% CI ranging from 1.7 to 2.8. The OR calculated in the 0 to 60 DIM timeframe of this analysis was 2.14. An influence of DYST on METR has been previously documented. Correa et al. (1993) calculated an OR of 2.1 with a 95% CI of 1.6 to 2.8. Erb et al. (1981) also documented a relationship between DYST and METR with an OR of 3.5. The OR between DYST and METR was equal to 2.36 in the 0 to 60 DIM timeframe and 2.45 in the 91 to 150 DIM timeframe.

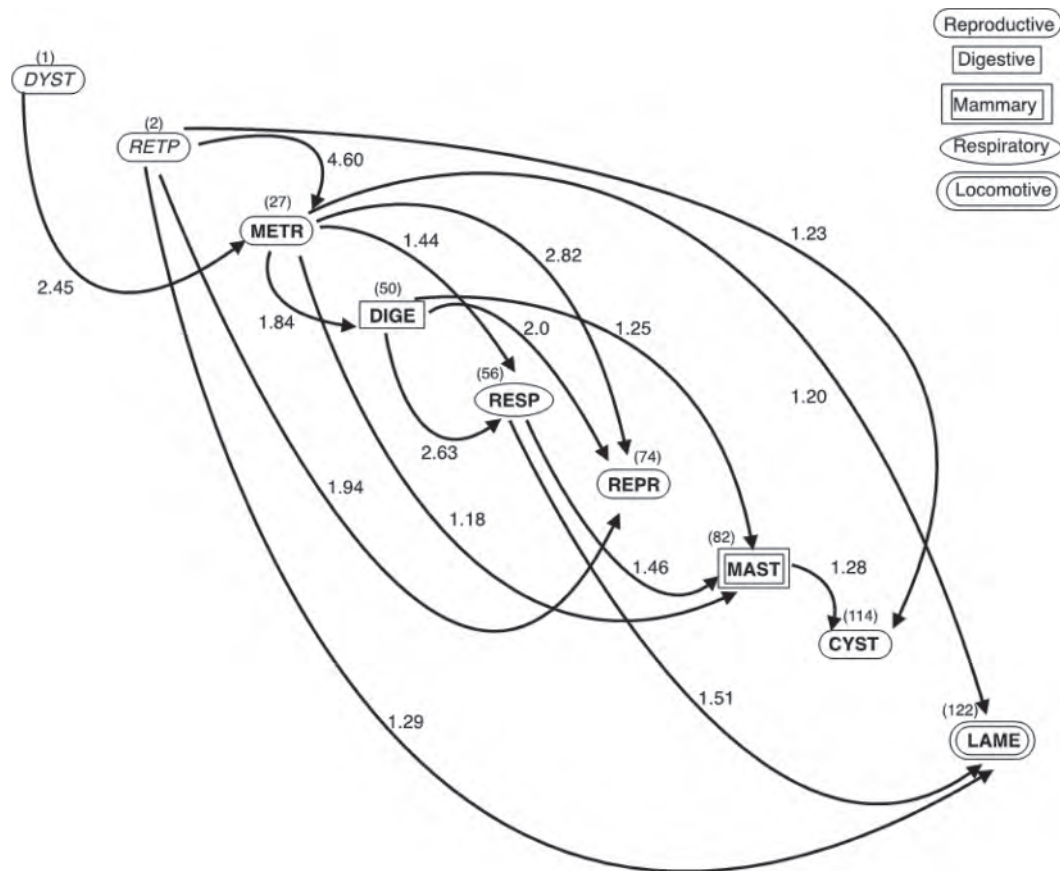


Figure 5. Path analysis of 91 to 150 DIM timeframe, with shapes representing event categories. Overall mean DIM at occurrence is shown in parentheses. Italicized health events occurred before the 91 to 150 DIM period. CYST = cystic ovaries; DIGE = digestive problem; DYST = dystocia; LAME = lameness; MAST = mastitis; METR = metritis; REPR = reproductive problem; RESP = respiratory problem; RETP = retained placenta.

Several relationships found due to a prior incidence of RETP have been previously documented by Dohoo and Martin (1984). An OR of 2.62 was found between RETP and DSAB in the 0 to 60 DIM timeframe. Dohoo and Martin (1984) calculated an OR of 3.8 for the influence of RETP on DSAB. The influence of RETP on KETO was calculated to have an OR of 2.44 in the 0 to 60 DIM timeframe. An OR of 1.9 was calculated for this relationship by Dohoo and Martin (1984).

Influences of KETO on later health events were consistent with previously documented results. Influence of a prior incidence of KETO on DIGE was calculated with an OR equal to 3.58 in the 0 to 60 DIM timeframe. Dohoo and Martin (1984) calculated an OR equal to 2.6 for the influence of KETO on DIGE; in their study, DIGE was defined as miscellaneous digestive tract disorders. An influence of KETO was also found to affect METR in the 0 to 60 DIM timeframe with an OR equal to 1.69. Correa et al. (1993) calculated an OR of 1.7 with a 95% CI spanning 1.0 to 3.0 for the influence of KETO on METR.

A large OR estimate of 3.0 was found from CALC to RESP in the 0 to 60 DIM timeframe. Within the 61 to 90 DIM timeframe, an influence of DIGE on RESP was found to have an OR of 3.33. Neither relationship has been previously documented, possibly because RESP is not usually a significant health event for cows compared with calves. In fact, RESP is more likely to be included in calf studies rather than studies examining cow health. Many other associations were found in this analysis that have not been previously documented. This may be the result of several factors including size of the data set and events recorded and used in analysis. Ordering of health events also varied slightly between studies affecting which results could be used for comparison.

A previous study included mastitis separated into distinct timeframes (van Dorp et al., 1999). The timeframes used were 0 to 30 d, 30 to 150 d, and 151 to 365 d. Significant OR between each of these timeframes were reported. The OR between the 0 to 30 d timeframe and the 30 to 150 d timeframe was 4.42; the

OR between the 30 to 150 d timeframe and the 151 to 365 d timeframe was 3.68. The values reported are higher than those reported here, perhaps because of the differing timeframes that were used. Previous studies were not found that examined the influence of prior incidences of LAME on later incidences; however, this analysis indicates increased odds of having a second incidence of LAME given a first incidence.

Empirical pathways based on order of occurrence were constructed at this time as opposed to using other, more complex methods to compare estimates with results reported in the literature. Phenotype networks could help explain complex biological systems, especially for traits with low heritabilities, such as health traits. Continued work should be conducted by incorporating genetic information with the use of, for example, recursive models to further analyze health event data. The path diagrams produced here could be used as a guide in constructing these models. For example, structural equation models could be used to analyze recursive relationships as well as incorporate genetic information (Rosa et al., 2011). Herd was included as a fixed effect in the analysis; however, we recognize that further research should examine various cow and herd level characteristics that may be used as indicators of health status. Nonparametric methods such as random forest could be used to select influential criteria from herd and cow characteristics. Further work should also be conducted to determine the most suitable editing criteria. The editing criteria used here produced a sufficiently reliable data set; however, other editing criteria could be used. Herd characteristics could be used to group similar herds and then flag those with a lower incidence rate than that of similar herds. Comparatively low intraherd heritability could also be used as an indicator of underreporting.

CONCLUSIONS

The results of this analysis provide evidence for the plausibility of on-farm recorded health information. Incidence rates of health events fell within the range of those found in literature; however, they were generally lower than the mean incidence of literature reports, especially for less common diseases. Path analyses allowed credible relationships to be constructed between health events. These relationships were then compared with previously identified relationships between health events, providing further validation for producer-recorded data. The information gathered may prove useful to producers by allowing the use of health event information from early lactation to help predict and prevent health events in later lactation. Knowledge of causal effects between health traits could also aid in the

development of breeding programs that more efficiently incorporate health information. More complete data recording along with standardized health event definitions would improve the credibility of the data.

ACKNOWLEDGMENTS

The authors thank Dairy Records Management Systems (Raleigh, NC) for providing the data. Partial funding for this research was provided by Genus plc (Hendersonville, TN) and Select Sires (Plain City, OH).

REFERENCES

- Animal Improvement Programs Laboratory. 2006. Format 4: Lactation. Accessed Feb. 23, 2011. <http://aipl.arsusda.gov/formats/fmt4.html>.
- Animal Improvement Programs Laboratory. 2010. Format 6: Health Record. Accessed May 19, 2011. <http://aipl.arsusda.gov/formats/fmt6.html>.
- Appuhamy, J. A. D. R. N., B. G. Cassell, and J. B. Cole. 2009. Phenotypic and genetic relationships of common health disorders with milk and fat yield persistencies from producer-recorded health data and test-day yields. *J. Dairy Sci.* 92:1785–1795.
- Barker, Z. E., K. A. Leach, H. R. Whay, N. J. Bell, and D. C. J. Main. 2010. Assessment of lameness prevalence and associated risk factors in dairy herds in England and Wales. *J. Dairy Sci.* 93:932–941.
- Bartlett, P. C., J. B. Kaneene, J. H. Kirk, M. A. Wilke, and J. V. Martenuik. 1986a. Development of a computerized dairy herd health data base for epidemiologic research. *Prev. Vet. Med.* 4:3–14.
- Bartlett, P. C., J. H. Kirk, M. A. Wilke, J. B. Kaneene, and E. C. Mather. 1986b. Metritis complex in Michigan Holstein-Friesian cattle: Incidence, descriptive epidemiology and estimated economic impact. *Prev. Vet. Med.* 4:235–248.
- Bartlett, P. C., P. K. Ntagize, J. B. Kaneene, J. H. Kirk, S. M. Anderson, and E. C. Mather. 1986c. Cystic follicular disease in Michigan Holstein-Friesian cattle: Incidence, descriptive epidemiology and economic impact. *Prev. Vet. Med.* 4:15–33.
- Bigras-Poulin, M., A. H. Meek, S. W. Martin, and I. Mcmillan. 1990. Health problems in selected Ontario Holstein cows: Frequency of occurrences, time to first diagnosis and associations. *Prev. Vet. Med.* 10:79–89.
- Cole, J. B., A. H. Sanders, and J. S. Clay. 2006. Use of producer-recorded health data in determining incidence risks and relationships between health events and culling. *J. Dairy Sci.* 89(Suppl. 1):10. (Abstr.)
- Correa, M. T., H. Erb, and J. Scarlett. 1993. Path analysis for seven postpartum disorders of Holstein cows. *J. Dairy Sci.* 76:1305–1312.
- DeGaris, P. J., and I. J. Lean. 2008. Milk fever in dairy cows: A review of pathophysiology and control principles. *Vet. J.* 176:58–69.
- Dohoo, I. R., and S. W. Martin. 1984. Disease, production and culling in Holstein-Friesian cows III. Disease and production as determinants of disease. *Prev. Vet. Med.* 2:671–690.
- Dubuc, J., T. F. Duffield, K. E. Leslie, J. S. Walton, and S. J. LeBlanc. 2010. Risk factors for postpartum uterine diseases in dairy cows. *J. Dairy Sci.* 93:5764–5771.
- Emanuelson, U., P. A. Oltenacu, and Y. T. Gröhn. 1993. Nonlinear mixed model analyses of five production disorders of dairy cattle. *J. Dairy Sci.* 76:2765–2772.
- Erb, H. N., S. W. Martin, N. Ison, and S. Swaminathan. 1981. Interrelationships between production and reproductive diseases in Holstein cows. Path analysis. *J. Dairy Sci.* 64:282–289.
- Faye, B. 1992. Interrelationships between health status and farm management system in French dairy herds. *Prev. Vet. Med.* 12:133–152.

- Fleischer, P., M. Metzner, M. Beyerbach, M. Hoedemaker, and W. Klee. 2001. The relationship between milk yield and the incidence of some diseases in dairy cows. *J. Dairy Sci.* 84:2025–2035.
- Frei, C., P. P. Frei, K. D. C. Stärk, D. U. Pfeiffer, and U. Kihm. 1997. The production system and disease incidence in a national random longitudinal study of Swiss dairy herds. *Prev. Vet. Med.* 32:1–21.
- Gardner, I. A., D. W. Hird, W. W. Utterback, C. Danaye-Elmi, B. R. Heron, K. H. Christiansen, and W. M. Sischo. 1990. Mortality, morbidity, case-fatality, and culling rates for California dairy cattle as evaluated by the national animal health monitoring system, 1986–87. *Prev. Vet. Med.* 8:157–170.
- Gay, E., and J. Barnouin. 2009. A nation-wide epidemiological study of acute bovine respiratory disease in France. *Prev. Vet. Med.* 89:265–271.
- Groehn, J. A., J. B. Kaneene, and D. Foster. 1992. Risk factors associated with lameness in lactating dairy cattle in Michigan. *Prev. Vet. Med.* 14:77–85.
- Gröhn, Y. T., S. W. Eicker, and J. A. Hertl. 1995. The association between previous 305-day milk yield and disease in New York state dairy cows. *J. Dairy Sci.* 78:1693–1702.
- Gröhn, Y. T., H. N. Erb, C. E. McCulloch, and H. S. Saloniemi. 1989. Epidemiology of metabolic disorders in dairy cattle: Association among host characteristics, disease, and production. *J. Dairy Sci.* 72:1876–1885.
- Gulliksen, S. M., E. Jor, K. I. Lie, T. Løken, J. Åkerstedt, and O. Østerås. 2009. Respiratory infections in Norwegian dairy calves. *J. Dairy Sci.* 92:5139–5146.
- Hamann, H., V. Wolf, H. Scholz, and O. Distl. 2004. Relationships between lactational incidence of displaced abomasum and milk production traits in German Holstein cows. *J. Vet. Med. A Physiol. Pathol. Clin. Med.* 51:203–208.
- Heringstad, B., G. Klemetsdal, and J. Ruane. 1999. Clinical mastitis in Norwegian cattle: Frequency, variance components, and genetic correlation with protein yield. *J. Dairy Sci.* 82:1325–1330.
- Kaneene, J. B., and H. S. Hurd. 1990. The national animal health monitoring system in Michigan. I. Design, data and frequencies of selected dairy cattle diseases. *Prev. Vet. Med.* 8:103–114.
- 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.
- Lean, I. J., P. J. DeGaris, D. M. McNeil, and E. Block. 2006. Hypocalcemia in dairy cows: Meta-analysis and dietary cation anion difference theory revisited. *J. Dairy Sci.* 89:669–684.
- Lundborg, G. K., E. C. Svensson, and P. A. Oltenacu. 2005. Herd-level risk factors for infectious diseases in Swedish dairy calves aged 0–90 days. *Prev. Vet. Med.* 68:123–143.
- Lyons, D. T., A. E. Freeman, and A. L. Kuck. 1991. Genetics of health traits in Holstein cattle. *J. Dairy Sci.* 74:1092–1100.
- Miller, G. Y., and C. R. Dorn. 1990. Costs of dairy cattle diseases to producers in Ohio. *Prev. Vet. Med.* 8:171–182.
- Mörk, M., A. Lindberg, S. Alenius, I. Vågsholm, and A. Egenvall. 2009. Comparison between dairy cow disease incidence in data registered by farmers and in data from a disease-recording system based on veterinary reporting. *Prev. Vet. Med.* 88:298–307.
- National Agricultural Statistics Service. 2011. Statistics by Subject: National Statistics for Milk. Accessed Mar. 19, 2011. http://www.nass.usda.gov/Statistics_by_Subject/.
- Olde Riekerink, R. G. M., H. W. Barkema, D. F. Kelton, and D. T. Scholl. 2008. Incidence rate of clinical mastitis on Canadian dairy farms. *J. Dairy Sci.* 91:1366–1377.
- R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- 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.
- Rosa, G. J. M., B. D. Valente, G. de los Campos, X. L. Wu, D. Gianola, and M. A. Silva. 2011. Inferring causal phenotype networks using structural equation models. *Genet. Sel. Evol.* 43:6.
- Stevenson, M. A. 2000. Disease incidence in dairy herds in the southern highlands district of New South Wales, Australia. *Prev. Vet. Med.* 43:1–11.
- Svensson, C., J. Hultgren, and P. A. Oltenacu. 2006. Morbidity in 3–7-month-old dairy calves in southwestern Sweden, and risk factors for diarrhoea and respiratory disease. *Prev. Vet. Med.* 74:162–179.
- Toni, F., L. Vincenti, L. Grigoletto, A. Ricci, and Y. H. Schukken. 2011. Early lactation ratio of fat and protein percentage in milk is associated with health, milk production, and survival. *J. Dairy Sci.* 94:1772–1783.
- 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.
- van Dorp, R. T. E., S. W. Martin, M. M. Shoukri, J. P. T. M. Noordhuizen, and J. C. M. Dekkers. 1999. An epidemiologic study of disease in 32 registered Holstein dairy herds in British Columbia. *Can. J. Vet. Res.* 63:185–192.
- Yániz, J., F. López-Gatiús, G. Bech-Sàbat, I. García-Ispuerto, B. Serano, and P. Santolaria. 2008. Relationships between milk production, ovarian function and fertility in high-producing dairy herds in northeastern Spain. *Reprod. Domest. Anim.* 43:38–43.
- Zwald, N. R., K. A. Weigel, Y. M. Chang, R. D. Welper, and J. S. Clay. 2004. Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values. *J. Dairy Sci.* 87:4287–4294.