Farm-level risk factors for bovine mastitis in Dutch automatic milking dairy herds

Z. Deng,[1]* G. Koop,[1] T. J. G. M. Lam,[1,2] I. A. van der Lans,[3] J. C. M. Vernooij,[1] and H. Hogeveen[1,4]
[1]Department of Farm Animal Health, Utrecht University, Yalelaan 7, Utrecht 3584 CL, the Netherlands
[2]GD Animal Health, PO Box 9, 7400 AA Deventer, the Netherlands
[3]Marketing and Consumer Behaviour Group, Wageningen University, Hollandseweg 1, Wageningen 6706 KN, the Netherlands
[4]Chair Group Business Economics, Wageningen University, Hollandseweg 1, Wageningen 6706 KN, the Netherlands

ABSTRACT

Automatic milking systems (AMS) are installed on a growing number of dairy farms worldwide. Management to support good udder health might be different on farms with an AMS compared with farms milking with a conventional milking system, as risk factors for mastitis on farms using an AMS may differ. The aim of this study was to identify farm level factors associated with mastitis on Dutch dairy farms using an AMS. In 2008, risk factor data were collected using a questionnaire combined with on-farm recordings of cow, stall, and AMS hygiene on 135 farms. These risk factor data were linked to 4 udder-health-associated dependent variables: average herd somatic cell count (HeSCCav), variance of the average herd somatic cell count (SCC) on test days (HeSCCvar), the average proportion of new high SCC cases (NHiSCC), and the farmer-reported annual incidence rate of clinical mastitis (IRCM). We employed regression models using multiple imputation to deal with missing values. Due to the high dimensionality of the risk factor data, we also performed nonlinear principal component analysis (NLPCA) and regressed the dependent variables on the principal components (PC). Good hygiene of cows and of AMS were found to be related to a lower HeSCCav and less NHiSCC. Effective postmilking teat disinfection was associated with a lower NHiSCC. A higher bulk tank milk SCC threshold for farmers' action was related to more NHiSCC. Larger farm size was related to lower HeSCCvar but higher NHiSCC. Negative attitude of farmers to animal health, higher frequency of checking AMS, and more time spent on viewing computer data were all positively related to higher IRCM. An NLPCA with 3 PC explained 16.3% of the variance in the risk factor variables. Only the first 2 PC were associated with mastitis. The first PC reflected older and larger farms with poor cow hygiene and AMS hygiene, and was related to higher HeSCCav and NHiSCC, whereas the second PC reflected newly built smaller farms with poor cow hygiene and low milk production, and was associated with higher HeSCCvar and NHiSCC, but lower IRCM. Our study suggests that many of the risk factors on conventional milking system farms are applicable to AMS farms, specifically concerning hygiene of the cows and the milking machine, but on large AMS farms, udder health may need more attention than on smaller AMS farms. Multiple imputation is instrumental to deal with missing values and NLPCA is a useful technique to process high dimensional data in our study.

Key words: mastitis, risk factor, automatic milking system, nonlinear principal component analysis, principal component regression

INTRODUCTION

Bovine mastitis is a multifactorial disease, and numerous papers have been published on management factors related to the disease (e.g., Dufour et al., 2012; Santman-Berends et al., 2016; Taponen et al., 2017). Risk factors for mastitis, according to Barkema et al. (1999), can be grouped into 3 categories: factors related to (1) exposure to causal pathogens, (2) host resistance to infection, and (3) cure of infection. On farms with conventional milking systems (**CMS**), frequently reported risk factors within the first category are, for instance, associated with hygiene of cow, milking machine, and housing (Schukken et al., 1990; Schreiner and Ruegg, 2003; Breen et al., 2009), milking procedures (Barkema et al., 1999), postmilking teat disinfection (Lam et al., 1997; Santman-Berends et al., 2016), and milk leakage (Klaas et al., 2005). Risk factors related to host resistance to infection are, for example, breed (Heringstad et al., 2000), milk production level (Schukken et al., 1990), and nutrition (Hogan et al., 1993). Factors associated with cure of infection include cow, pathogen, and treatment-associated factors (Barkema

et al., 2006). For instance, parity, strain type (van den Borne et al., 2010a), and the time of applying treatment (van den Borne et al., 2010b) are related to the cure rate of *Staphylococcus aureus* infection.

Since the first introduction of automatic milking systems (**AMS**) on commercial dairy farms in 1992, the use of AMS has rapidly increased in the Netherlands and many other countries (Barkema et al., 2015). Many risk factors for mastitis on CMS farms are also probably applicable on AMS farms. However, because of differences in management between AMS and CMS farms, some of the determinants of mastitis may differ on AMS farms. For example, on AMS farms, detection of clinical mastitis is based on a screening with sensors, whereas on CMS farms detection of clinical mastitis is done cow side by the farmer during milking. Moreover, udder preparation is done automatically on farms with an AMS. That means that the process is always carried out with the same intensity, whereas at a CMS farm, the milker can adjust the process to the circumstances (e.g., dirtiness of the udder). Additionally, risk factors that are present in both AMS and CMS farms may not have the same importance in both systems. Compared with the large number of studies on CMS farms, only a limited number of studies have been published on risk factors for mastitis on AMS farms. Persson Waller et al. (2003) reported that the incidence of milk leakage is higher on AMS farms than on CMS farms, which increases the risk of mastitis. Dohmen et al. (2010) indicated that on AMS farms, the hygiene of the cow and the milking system were related to udder health. Mollenhorst et al. (2011), in a cow-level study, reported that the milking interval was related to mastitis occurrence in cows milked by an AMS.

Because of differences in management between CMS and AMS farms, and because of the important role mastitis plays on AMS farms (Hovinen and Pyörälä, 2011), it is useful to specifically identify factors associated with udder health on AMS farms. This is generally done using questionnaires and multiple regression analysis. This approach may result in inflated family-wise type I errors because of the high number of factors studied. Therefore, principal component analysis (**PCA**) is an appropriate method that can be used to replace the large number of factors by a much smaller number of components. In questionnaire data, the presence of categorical variables and nonlinear relationships among factors is not rare. Therefore, nonlinear PCA (**NLPCA**), which is a nonlinear equivalent to PCA, was developed to deal with categorical variables and nonlinear relationships between variables (Linting et al., 2007a).

In this paper, we performed a comprehensive exploratory study of risk factors for mastitis on AMS farms.

Due to the high dimensionality of the data resulting from the study design, we used NLPCA in addition to regression analysis. The aim of this study was to identify farm level risk factors for mastitis on Dutch dairy farms using AMS.

MATERIALS AND METHODS

Farm Selection and Data Collection

For this study, we used data from an extensive questionnaire about potential farm-level risk factors for mastitis on AMS farms and on farm measurements on hygiene of cows, milking machine, and barns, part of which was used by Dohmen et al. (2010) in a previous study.

About 2,400 (Reinemann, 2008) of the total 21,300 dairy farms (http://www.dutchfarmexperience.com/history-nl-dairy-farming/) were using AMS in the Netherlands in 2007. Based on other published risk factor studies on mastitis [e.g., Schukken et al. (1990) with 125 farms, Elbers et al. (1998) with 171 farms, and Santman-Berends et al. (2012) with 189 farms], a convenience sample of approximately 150 farms was aimed for. To reach that number and assuming a response rate of 25 to 35%, FrieslandCampina (Amersfoort, the Netherlands) approached 400 randomly selected farms with an AMS with the request to collaborate in this research in 2008. A total of 161 farms (40.3%) were willing to participate, and of these 161 farms, 10 farms failed to meet the inclusion criteria below, so 151 farms were visited. The following selection criteria were used: (1) the farm has had an AMS for more than 1 yr and performed no conventional milking anymore, and (2) the farm was participating in the Dutch milk production recording (**MPR**) system. Data were collected by trained veterinary master's students. In total, 5 students were involved in the data collection. All farms were visited by a team of 2 of these 5 students, in different combinations. The students collecting the data received a training session of 1 d. During that day, the questionnaire was explained to the students by the developers of the data collection protocol, including the background of the questions, the way the data should be retrieved from the various software systems, and on the data collection in the barns (e.g., hygiene parameters). During the farm visit, questionnaires were completed by farmers with the assistance of the students to clarify questions if needed. Additionally, hygiene scoring of the AMS and of the cows as well as an assessment of the functioning of the AMS were performed by pairs of students using standardized protocols. For data collection and hygiene scoring as well as for the evaluation of AMS units, protocols were developed in the following

steps: as a first step, 10 Dutch mastitis experts created a list of potential risk factors. The list of risk factors was summarized in a number of questions and in data collection protocols, where existing scoring schemes [e.g., described by Schreiner and Ruegg (2003)] were used as much as possible. Protocols described how to retrieve data from the software systems of the various AMS manufacturers. Since the students followed the training together and visited the farms in pairs, we assumed that this would guarantee a data collection procedure that was as consistent as possible. Clinical mastitis data were collected based on farmers' reports, and cow SCC data were based on MPR records. Milk production information, including cow identification, MPR test dates, and milk yield and SCC on test dates, were extracted from the Dutch national MPR system (CRV, Arnhem, the Netherlands). Milk production data in 2008, from 106 to 296 d prior to the farm visit (with an average number of MPR test dates per farm of 8.5, ranging from 3 to 14) were used in the analysis. Data collection protocols, questionnaires, and forms (in Dutch) are available from the last author.

Data Cleaning

Of the 151 farms that were visited, 135 farms were included in the final analysis. Of the 16 farms that were excluded from the analysis, 10 farms had no match of the farm identification with the questionnaire, MPR information, hygiene scores, or the data on AMS functioning, 4 farms turned out to have been using an AMS for less than 1 yr and 1 farm was also using CMS, and 1 farm had invalid records (no clinical mastitis cases during the study period) with regard to the number of clinical mastitis cases.

Before analysis, records with test-day milk yield or SCC equal to zero (201,557 records), records from cows with an impossible parity (parity as 56 or 57; 27 records), and records from MPR test days with ≤10 records (93 records) were excluded. Variables with more than 5% of missing values (29 variables; Schafer, 1999) and binary variables with a category having less than 8 observations were discarded (24 variables; Linting et al., 2007b), resulting in a total of 113 risk factor variables.

Explanatory Factors

The 113 risk factor variables that were captured using the questionnaire, hygiene scoring of the AMS and of the individual cows and of the barns, as well as from the assessment of the functioning of AMS, were evaluated, of which 97 risk factors were derived from the questionnaire, 8 risk factors were about AMS hygiene,

3 about cow hygiene, and 5 about AMS functioning. Hygiene scoring was performed on 8 AMS parts for each robot, scores ranging from 1 to 4 (1 = clean, 2 = slightly dirty, 3 = dirty, 4 = very dirty). The hygiene of udder, of thighs and of legs for at least 10 randomly selected milking cows was scored (from the cows milked by the AMS, 25% were randomly selected before entering the barn, based on the barn list, to be scored for hygiene, with a minimum of 10 cows per farm). The hygiene of teats before and after cleaning was scored during milking (1 = completely free of dirt or has very little dirt, 2 = slightly dirty, 3 = mostly covered in dirt, and 4 = completely covered, caked-on dirt) as described by Schreiner and Ruegg (2003). The functioning of the AMS was evaluated during 10 milkings on each farm during the farm visit by scoring 6 different procedures of the AMS, all with different scoring protocols. The scores for procedure and each farm were averaged for the 10 milkings (Dohmen et al., 2010).

Recoding of the hygiene-related variables was performed as follows:

(1) For hygiene scores of the AMS units, the hygiene scores were transformed as dirty (average hygiene score of the units greater than 1) or clean (average score of 1).

(2) The hygiene scores of individual cows were aggregated into 3 farm level variables denoting the proportion of dirty thighs, udders, or legs per farm, with dirty defined as hygiene score >2. The threshold for dirty was set as hygiene score >2 because only less than 6% of the hygiene scores were equal to 1 in each of the variables.

Dependent Variables

The following dependent variables associated with udder health were defined: the average herd SCC (**HeS-CCav**), the variance of the average herd SCC on each MPR test date (**HeSCCvar**), the average proportion of new high SCC cases (**NHiSCC**), and the farmer-reported incidence rate of clinical mastitis (**IRCM**). To determine HeSCCav, first, for each farm and each test-day the arithmetic mean of SCC was calculated and then HeSCCav was calculated by taking the arithmetic mean of these test-day arithmetic means. Finally, it was $\log_{10}$ transformed. In the Netherlands, farms with a geometric mean BMSCC above 400,000 cells/mL in the last 3 mo receive a penalty. The HeSCCvar was calculated for each farm by first calculating the farm average of SCC for each test day, then the variation of these test day farm-average SCC was calculated for each farm and then log10 transformed. New high SCC cows were defined as cows with a normal SCC on the previous

test day and a high SCC (>150,000 cells/mL for heifers and >250,000 cells/mL for multiparous cows) on the current test day. The proportion of NHiSCC cows on each test day was calculated by dividing the number of NHiSCC cows by the number of cows with a SCC ≤150,000 cells/mL for heifers and ≤250,000 cells/mL for multiparous cows on the previous test day. Cows newly entered the farm or cows dried off during the study period were not considered in the calculation. The NHiSCC for each farm was calculated as the arithmetic mean of the proportion of NHiSCC cows of each of the test days. The annual average IRCM was calculated as the farmers' reported total number of cases of clinical mastitis in the past year, divided by the number of lactating cows in that year reported by the farmers. We calculated the Pearson correlation coefficient between each of the 4 dependent variables to check for collinearity between these dependent variables and also knowing the correlations among the dependent variables may facilitate the interpretation of the β-estimates of the same independent variable for different dependent variables (i.e., the β-estimates to be similar or not).

Statistical Analysis

Risk factors were identified using 2 approaches: (1) linear and negative binomial regression models (after pre-screening for variable selection in univariable regression analysis) and (2) principal component regression analysis (**PCRA**; Massy, 1965). The PCRA was performed by regressing the 4 udder-health-associated dependent variables to the principal components (**PC**) constructed by NLPCA on the evaluated risk factors.

Regression Analysis with Risk Factor Variables. In the regression analysis, the 4 udder-health-associated dependent variables were used as dependent variables in 4 different models, and the 113 risk factor variables were used as independent variables. Statistical significance was considered as $P < 0.1$ in univariable regression analysis, using one of the udder health variables as the dependent variable and one of the risk factor variables as the independent variable, and as $P < 0.05$ in multiple regression analysis. Because of missing values, and therefore differences in the number of observations between nested models, leading to incomparability of the model fit, we applied multiple imputation. Before imputing the missing values in the data set, we used the aggr function from the VIM package version 4.7.0 (Kowarik and Templ, 2016) in R to check if the missing-values pattern fits the assumptions of multiple imputation. Multiple imputation with 10 imputations for 10 iterations, with other parameters set to default, was performed on all variables that were associated

with one or more of the 4 udder-health-associated dependent variables in the univariable analysis together with the 4 udder-health-associated dependent variables by using the mice function from the MICE package version 2.46.0 (van Buuren and Groothuis-Oudshoorn, 2011) in R. The default methods for imputation for different types of variables were predictive mean matching for numeric data, logistic regression imputation for binary data or factor with 2 levels, polytomous regression imputation for unordered categorical data (factor ≥2 levels), and proportional odds model for ordered factors with ≥2 levels. The imputed values for each variable with missing values were inspected by plotting the imputed and nonimputed values against iteration number. Convergence was evaluated by plotting the mean of the imputed values for each variable against the iteration number.

Collinearity was evaluated for continuous independent variables after the univariable analyses. If 2 continuous variables had a Pearson correlation coefficient >0.7, one of the variables was selected to be included in multivariable regression analysis based on the biological interpretability of the variables. A stepwise forward method based on *P*-value was used for model selection for each dependent variable in the multivariable regression analysis. Confounding was examined by evaluating the change in the β-estimates of variables in the model when adding a new variable. A change in the β-estimate of more than 20% was deemed indicative of confounding. Confounders were retained in the final model. To check robustness, the final models that were constructed using multiple imputation were rerun using the original data set with missing values and the corresponding estimates were compared with the models constructed from the imputed data set. We used linear regression models for HeSCCav, HeSCCvar, and NHiSCC, and a negative binomial regression model for IRCM because of the negative binomial distribution of IRCM with overdispersion (Schukken et al., 1991). The linear and negative binomial regression models are described as follows:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i + \varepsilon_i, \text{ and}$$

$$\log(\text{CM}) = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i$$
$$+ \text{offset}[\log(\text{herd\_size})],$$

where Y = the continuous dependent variables (HeSCCav, HeSCCvar, or NHiSCC), α = the intercept, β_i = the regression coefficient for each of the independent variables X_i, ε_i = the residual random error, CM = the farmer-reported number of clinical mastitis cases last year, and $\log(\text{herd\_size})$ = the offset term.

The model assumptions were evaluated as follows: for linear regression models, homoscedasticity was evaluated by plotting the standardized residuals against the predicted values, normality of residuals was examined by Q-Q plots, and linearity was evaluated by plotting residuals against the fitted values for continuous variables; for the negative binomial regression model, residuals were evaluated by plotting standardized deviance residuals against predicted values.

Principal Component Regression Analysis. Because of the low variance explained in the dependent variables in ordinary regression analysis, and also to reduce the high dimensionality and to capture the underlying structure of the explanatory variables, we also did principal components regression (Massy, 1965) for which we applied NLPCA on the 113 explanatory variables, with missing values imputed as the averages, from the Gifi package version 0.3.7 (Mair et al., 2017; https://rdrr.io/rforge/Gifi/) in R. In NLPCA, the categorical variables were transformed into numeric values through optimal scaling. These numeric values, referred to as category quantifications, for one variable together form that variable's transformation. Nominal or ordinal analysis levels were used for categorical variables and spline analysis levels for continuous variables. The number of PC was determined by inspecting the scree plots of NLPCA solutions with different numbers of PC, the interpretability of the PC (Jolliffe, 2002; Linting et al., 2007b; Linting and van der Kooij, 2012). We tested NLPCA with a 3, 4, 5, 6, and 7 dimensional solution. The scree plots for the solutions with 4 and 5 components showed a consistent elbow at the 3rd PC in NLPCA with a different number of PC, and the eigenvalues of the first 4 PC were all larger than 1. We tried to interpret the PC resulting from the NLPCA with a different number of PC, and the first 3 PC were interpretable. We did not base our selection of the PC on the cumulative variance accounted for (**VAF**) criterion, because with 113 explanatory variables even smaller cumulative VAF capture the variance that is present in a considerable number of variables. Outliers were defined as observations with object scores exceeding the range between −3.5 to 3.5 (Linting and van der Kooij, 2012). The NLPCA was repeated after deleting outliers until no outliers were present anymore. The 4 dependent udder health variables were then regressed on the first 3 PC using negative binomial regression for IRCM and linear regression for the other 3 dependent variables, respectively. Model selection was performed as mentioned above. We did k-fold cross-validation with 10-fold for the 3 PCRA linear regression models using the caret package version 6.0–80 in R. All statistical analyses were conducted in R version 3.4.4 (R Core Team, 2018).

RESULTS

Descriptive Statistics

In total, 135 farms were included in the analysis, of which 69 farms had complete records and 66 farms had missing values in at least one variable. The percentage of missing values on the 66 farms with missing values was on average 1.4%, ranging from 0.8 to 5.1%. Of the 113 explanatory variables, 70 variables had no missing values, whereas 43 variables did have missing values, on average 1.9%, ranging from 0.7 to 4.4%. Two farms did not have data on IRCM, but no missing values were present in other dependent variables.

The general farm characteristics of the 135 participating farms are provided in Table 1. The correlations among HeSCCav, HeSCCvar, and NHiSCC ranged from 0.44 to 0.74, whereas the correlation between IRCM and other 3 dependent variables ranged from 0.097 to 0.17 (Figure 1).

Regression Analysis with Risk Factor Variables

In the univariable regression analysis, a total of 48 risk factor variables were found to be associated with

Table 1. Descriptive statistics of 135 Dutch dairy farms using an automatic milking system (AMS)

Variable	Number	Mean	Minimum	Maximum
Number of cows	135	80	22	365
Number of AMS	131	1.6	1	6
Milk quota (× 1,000 kg)	132	796	154	5,000
305-d milk yield (kg)	135	8,952	5,500	11,000
Average herd SCC[1] (× 1,000 cells/mL)	135	271.8	83.4	505.3
Variance of average of herd SCC[2] (× 10^7)	135	928	69	5,873
Average proportion of new high SCC cases (%)	135	13.4	4	29.3
Annual incidence rate of clinical mastitis (cases/100 cows per year)	133	27	1.4	135

[1]Arithmetic mean.
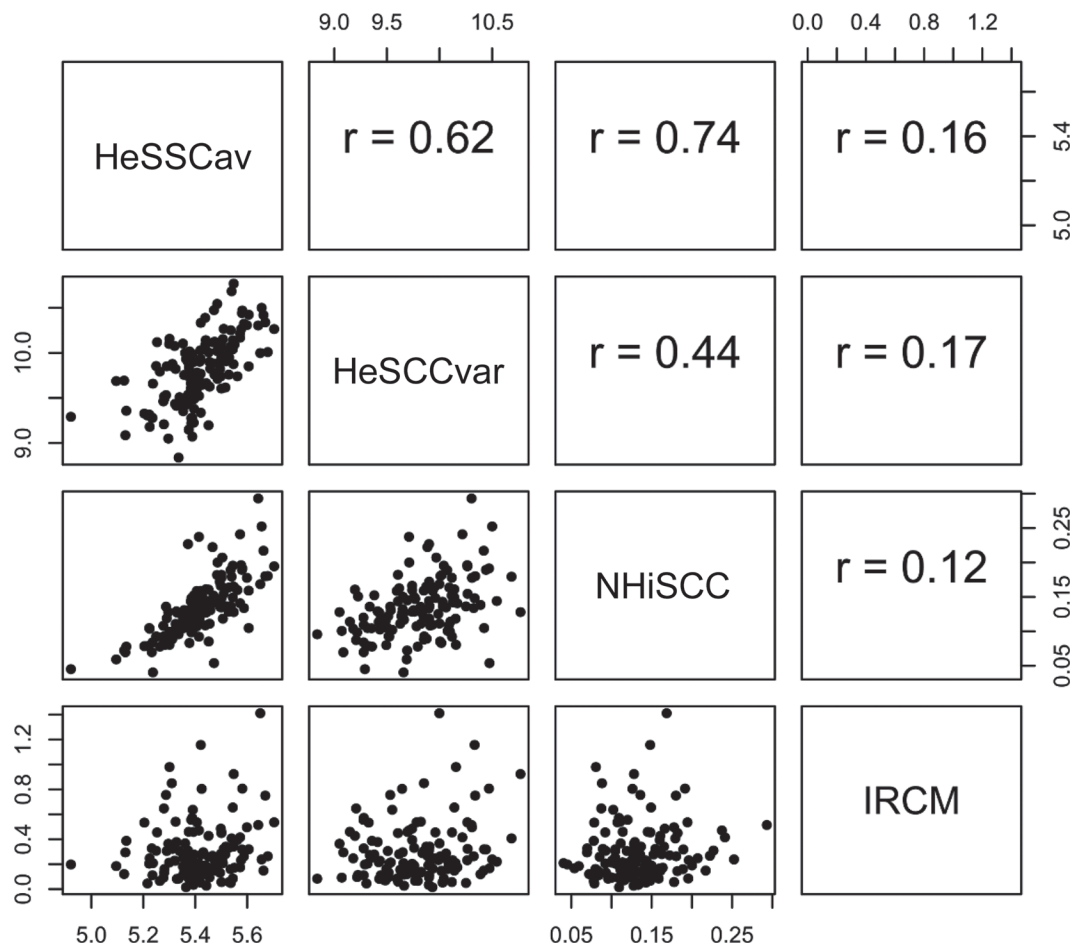
[2]Variance of the average herd SCC between test days.

Figure 1. Pair-wise scatter plot with Pearson correlation coefficient (r) between $\log_{10}$-transformed average herd SCC (HeSCCav), $\log_{10}$-transformed variance of average herd SCC on each test day (HeSCCvar), average proportion of new high SCC cases (NHiSCC), and farmer-reported annual incidence rate of clinical mastitis (IRCM) per cow per farm, based on data collected on 135 Dutch dairy farms using an automatic milking system.

one or more of the dependent udder health variables (with $P < 0.1$), which are presented in Supplemental Table S1; https://doi.org/10.3168/jds.2018-15327. In Table 2, the results of the multiple regression models are given, showing that several risk factors were significantly associated with multiple dependent variables, namely participation in a *Salmonella* control program, the percentage of cows lying outside cubicles, including hay in the diet of dry cows, and the moment that the farmer stops antibiotic treatment of mastitis cases. None of the variables that remained significant in the IRCM model were present in any of the other models. The adjusted coefficient of determination values of the regression models for HeSCCav, HeSCCvar, NHiSCC, and IRCM based on the nonimputed data set were 0.21, 0.18, 0.46, and 0.44, respectively. In general, the estimates of models using multiple imputation were very similar to the ones using nonimputed data.

Principal Component Regression Analysis

The 3 PC in the NLPCA explained 16.3% of the variance in the risk factor variables. The component loadings for variables with loadings >0.35 or <-0.35 (Hair et al., 2009) in 1 or more of the 3 PC are presented in Table 3. Most of these variables were related to farm size, hygiene of the cows, and hygiene of the AMS. The loadings of variables with loadings >0.35 or <-0.35 are shown in Figure 2. The transformation plots, which display the shape of the relationship between the original variable and the transformed variable, are in Supplemental Figure S1 (https://doi.org/10.3168/jds .2018-15327) for the first PC and Supplemental Figure S2 (https://doi.org/10.3168/jds.2018-15327) for the second PC. Many of the transformation plots showed linear relationships, but some of the transformation plots demonstrated nonlinear relationships.

Table 2. Multivariable regression models for 4 udder health variables, based on data collected on 4 dependent variables and 113 independent variables on 135 Dutch dairy farms using an automatic milking system (AMS)[1]

Variable	HeSCCav[2] β[6]	2.5%	97.5%	HeSCCvar[3] β	2.5%	97.5%	NHiSCC[4] β	2.5%	97.5%	IRCM[5] β	2.5%	97.5%
Intercept	5.22	5.09	5.34	9.96	9.71	10.21	−0.22	−0.52	0.09	−1.75	−1.99	−1.51
Farm size												
Milk quota (× 1,000 kg)				−0.01	−0.03	0.01	0.01	0.00	0.02			
Number of AMS	0.03	0.00	0.06									
Hygiene												
Percentage of cows lying outside cubicles							0.01	0.00	0.02			
Percentage of teats not completely covered by spray							0.03	0.01	0.05			
Farm average teat hygiene score before cleaning	0.10	0.05	0.16				0.02	0.00	0.03			
Animal health program												
Participate in *Salmonella* program												
No	Referent			Referent			Referent					
Yes	−0.08	−0.12	−0.03	−0.21	−0.36	−0.07	−0.02	−0.03	−0.00			
Participate in bovine diarrhea virus control program												
No							Referent					
Yes							−0.03	−0.05	−0.01			
Participate in bacterial culture program												
No										Referent		
Yes										0.60	0.21	0.99
Nutrition												
Including hay in dry cow diet												
No	Referent						Referent					
Yes	0.05	0.01	0.09				0.01	0.00	0.03			
Leftovers from milking cow used as feed for dry cows												
No							Referent					
Yes							0.03	0.01	0.05			
Other												
305-d milk yield (× 1,000 kg)				−0.00	−0.00	0.000	−0.01	−0.01	0.00	0.47	0.01	0.94
Time spent viewing data in computer												
Time spent nondairy farming activities												
Significant changes on farm recently												
No				Referent								
Yes				0.13	0.01	0.25						
When do you stop antibiotic treatment for clinical mastitis cases?												
According to the prescription				Referent			Referent					
Until no flakes present in milk				0.31	0.12	0.50	0.02	0.00	0.04			
Other				0.13	−0.04	0.31	0.01	−0.01	0.02			
Log10 (BMSCC)[7] that results in farmers' action							0.06	0.01	0.12			
Check AMS in addition to regular check												
No										Referent		
Yes										0.37	0.09	0.65
Ever checked AMS while milking?												
No										Referent		
Yes										0.29	0.01	0.56

[1] Models were constructed using the data set with multiple imputation.
[2] Log10-transformed average herd SCC.
[3] Log10-transformed variance of herd SCC on each test day.
[4] Average proportion of new high SCC cows.
[5] Farmer-reported annual incidence rate of clinical mastitis.
[6] Estimate of regression coefficient.
[7] Log10-transformed bulk tank milk SCC.

Table 3. Variables with loadings >0.35 or <−0.35 from nonlinear principal component analysis with 3 dimensions on 113 risk factor variables, based on data collected on 135 Dutch dairy farms using an automatic milking system (AMS)

Category	Abbreviation	Variable	Dimension			Total VAF[1]	Sum
			1	2	3		
Farm size	N_cows	Number of cows on farm	0.68	−0.54	—[2]	0.76	
	Milk_quota	Milk quota	0.59	−0.56	—	0.66	
	N_eat_place	Number of eating places	0.49	—	—	0.29	
	N_cubicle	Number of cubicles	0.57	−0.51	—	0.59	
	N_AMS	Number of AMS units	0.60	−0.41	—	0.53	
Hygiene of AMS	Camera_clean	Camera clean or dirty	0.40	—	0.40	0.37	
	Tube_clean	Tube clean or dirty	0.55	—	—	0.41	
	Cup_clean	Milk cups clean or dirty	0.51	—	—	0.32	
	Airintake_clean	Air intake clean or dirty	0.51	—	—	0.35	
	Arm_clean	AMS arm clean or dirty	0.59	—	—	0.46	
	Manger_clean	Manger clean or dirty	0.47	—	—	0.28	
	Floor_clean	Floor clean or dirty	0.56	—	0.42	0.49	
	Brush_clean	Brush clean or dirty	0.50	—	—	0.35	
Hygiene of cow	Dirty_udder	Percentage of cows with udder hygiene score >2	0.46	0.43	−0.48	0.62	
	Dirty_thigh	Percentage of cows with thigh hygiene score >2	0.37	—	−0.43	0.43	
	Dirty_leg	Percentage of cows with leg hygiene score >2	0.44	—	−0.48	0.45	
	Teat_clean_before	Average teat hygiene scores before cleaning	0.44	—	—	0.41	
	Dirty_teat_before	Percentage of teats with hygiene score >2 before cleaning	0.40	—	−0.36	0.38	
	Dirty_teat_after	Percentage of teats with hygiene score >2 after cleaning	0.36	0.42	−0.41	0.47	
	Per_teat_not_covered_spray	Percentage of teats not spraying after milking	—	0.39	—	0.22	
Nutrition	Fresh_grass_milking	Fresh grass for milking cows	—	0.38	—	0.17	
	Maize_dry_cow	Maize for dry cows	—	−0.38	—	0.16	
	305_MY	305-d milk yield	—	−0.41	—	0.26	
Other	When_AMS_installed	When did AMS get installed	—	−0.39	—	0.24	
	Feed_system_milking	Feeding system for milking cows	—	−0.36	—	0.18	
	Type_liner	Type of liners	—	−0.36	0.37	0.27	
	Control_daily_work	Follow to daily work protocol	—	—	−0.35	0.27	
	Time_nondairy	Time spent on nondairy farming	0.43	—	—	0.34	
	Use_checklist	Use checklist	—	−0.38	—	0.17	
	Fre_replace_filter	Frequency of replacing milk filter	—	—	—	0.10	
	When_replace_filter	When do you replace milk filter	—	—	0.38	0.19	
	Fre_manure_remove	Frequency of manure removal	—	—	−0.35	0.20	
	Additional_hand_manure	Remove manure by hand in addition to by robot	—	−0.43	—	0.28	
	N_cow_collect_wait	Number of cows collected in wait area	0.46	—	—	0.40	
	N_cow_connect_hand	Number of cows connected to robot by hand	—	—	−0.38	0.19	
	Time_monitoring_cow	Time spent on checking cows in barn	—	−0.35	—	0.16	
Eigenvalue			7.38	6.36	4.66		18.4
Variance accounted for (%)			6.53	5.63	4.12		16.28

[1]Cumulative variance accounted for in the first 3 principal components for each variable.

[2]Dash means the absolute value of the loading is >−0.35 and <0.35.

Figure 2. Loading plot for the first principal component (PC1) and the second principal component (PC2) of a nonlinear principal component analysis on 113 risk factor variables based on data collected on 135 Dutch dairy farms using an automatic milking system, showing all variables with loadings >0.35 or <−0.35 in PC1 or PC2 or both. The explanation of the variable names is given in Table 3.

The variables that loaded high in the first PC were mostly related to farm size (5 variables), hygiene of the cows (8 variables), and the AMS (8 variables). The variables that loaded high in the second PC were related to farm size (4 variables), hygiene of cows (3 variables), and some other variables. The variables that loaded high in the third PC were mostly related to hygiene of cows (7 variables). Based on the loadings and the cor-

responding transformation plots of these variables, the first PC represented larger farms with poor hygiene of cows and AMS, the second PC represented newly built smaller farms with poor cow hygiene and low milk production, and the third PC represented farms with good hygiene of cows. Regression model results of the PC are presented in Table 4, showing that PC1 was positively associated with HeSCCav and NHiSCC, and PC2 was

Table 4. Multivariable regression model results for 4 udder health variables with principal components (PC1 and PC2) from nonlinear principal component analysis on the 113 risk factor variables, based on data collected on 134 Dutch dairy farms using an automatic milking system (after excluding 1 outlier farm)

Dependent variable	Independent variable	Number of farms	β^1	95% CI	
				2.50%	97.50%
HeSCCav[2]	Intercept		5.416	5.394	5.438
	PC1	134	0.028	0.007	0.05
HeSCCvar[3]	Intercept		9.809	9.745	9.873
	PC2	134	0.081	0.017	0.144
NHiSCC[4]	Intercept		0.134	0.127	0.141
	PC1	134	0.011	0.004	0.018
	PC2	134	0.007	0.000	0.014
IRCM[5]	Intercept		−1.241	−1.362	−1.117
	PC2	134	−0.122	−0.24	−0.005

[1]Estimates of regression coefficient.

[2]Log_{10}-transformed average herd SCC.

[3]Log_{10}-transformed variance of average herd SCC on each test day.

[4]Average proportion of new high SCC cases.

[5]Farmer-reported annual incidence rate of clinical mastitis.

positively associated with HeSCCvar, NHiSCC, but negatively associated with IRCM. We plotted the PC with 4 dependent variables and did not find proof of a nonlinear relationship. The adjusted coefficient of determination for the regression models with PC for HeSCCav, HeSCCvar, NHiSCC, and IRCM were 0.041, 0.041, 0.086, and 0.089, respectively. The standard deviation of RMSE for PCRA models with log_{10} (HeSCCav) was 0.0358, for log_{10} (HeSCCvar) was 0.0308, and for NHiSCC was 0.0078.

DISCUSSION

In this study, we explored farm level risk factors for mastitis on AMS farms by using regression analysis and NLPCA. Cow hygiene, as well as hygiene of the AMS, were important factors related to various udder health parameters. In addition, we found several factors related to the farmers' behavior as well as herd size to be determinants of udder health. A previous study by Dohmen et al. (2010) was a first analysis on a part of the data that specifically aimed at hygiene variables. Other factors might also play a role in mastitis on AMS farms, so we carried out this study on many more potential risk factors. The results, however, confirmed that hygiene is important and, in addition, pinpoint that farm size is relevant to mastitis on AMS farms. The average farm size was 86 in our study, which was somewhat lower than 96 in 2002 and 105 in 2003 for dairy farms using AMS in Netherlands (Bijl et al., 2007). The udder health situation on a farm can be described by various parameters, including the 4 dependent variables used in our study, HeSCCav, HeSCCvar, NHiSCC, and IRCM. Even though bulk tank SCC is a widely used parameter

for herd level udder health, the bulk tank SCC was only available as farmers' reported values; due to the high proportion of missing values in farmers' reported bulk tank SCC (14.8% missing) and also to have a more precise estimation of the herd SCC, we calculated the herd SCC using DHI data. These herd SCC were not weighted by milk production, and even though it did not mimic the bulk tank milk SCC completely, the calculated average herd SCC still gives a good approximation of the BMSCC. The average HeSCCav was 271,828 cells/mL in our sample, which is higher than average HeSCCav of 214,000 cells/mL on Dutch AMS farms and 196,000 cells/mL on Dutch CMS farms (Steeneveld et al., 2015). We found NHiSCC in our study to be on average 13%, ranging from 4 to 29%, which is somewhat higher than the 9% (ranging from 0 to 18%) reported in a study on 1,903 Dutch CMS dairy farms (Nor et al., 2014). The IRCM was based on farmers' retrospectively reporting the number of clinical mastitis cases in the past year, which may well have been influenced by recall bias. Still, the average IRCM of 0.27 in our study is in line with the IRCM of 0.28 reported for Dutch CMS farms (Lam et al., 2013) and 0.29 for Dutch AMS farms (Huijps et al., 2008), although considerably higher than the IRCM of 0.20 reported by Mollenhorst et al. (2012) on Dutch AMS farms. These parameters, together with the general farm characteristics in Table 1, gave us the indications that the farms included in our study were representative for Dutch dairy farms using AMS.

We calculated HeSCCvar as the variance of the average herd SCC across test days, a parameter to our knowledge not previously studied. The HeSCCvar reflects temporal variation of herd SCC, thereby giving an estimate of the stability of the udder health situ-

ation. The 3 SCC-based udder health variables were significantly associated with each other, but not with the IRCM (Figure 1). In our study, we did not find an association between IRCM and HeSCCav, which is in line with previous research (Barkema et al., 1998; Olde Riekerink et al., 2008).

The missing values in the data hampered regression modeling, as the number of observations differs between nested models, leading to incomparability of the model fit. To be able to run all models on the total sample, we therefore applied multiple imputation (Sterne et al., 2009). Although we did not find patterns in the missing values in our analysis, it is possible that such patterns were present. However, the effect of incorrect assumptions on patterns in missing values on the estimates is expected to be limited (Collins et al., 2001). A comparison of the regression models based on the imputed data with the models based on the nonimputed data showed that the differences between the estimates were negligible. The variables and the direction of the association between dependent variables and independent variables were the same, only the β-estimate changed slightly (with a mean change of β-estimate being 25%, with a 95% CI of 14 to 36%). Moreover, with multiple imputation, we can use all farms in the model building process, which suggests that multiple imputation is a useful technique to facilitate the model building process, without substantially altering the model results. In the data collection process, we did not evaluate the agreement between the 2 students in performing interview or scoring. Both of them trained and worked together, so even though potential discrepancy in data collection between the 2 students might exist, we assume that this would guarantee an data collection process as consistent as possible. Because the questionnaire data were collected by the farmers' recall, lack of accuracy might exist. However, we had 28 numerical explanatory variables and 1 dependent variable (number of clinical mastitis cases last year) based on the farmers' recall. Most of these variables were related to farm size and frequency of cleaning AMS units, which were quite familiar to farmers. Farmers' recall of the number of clinical mastitis cases might have been problematic; however, not all farmers register clinical mastitis cases in the Netherlands, so having all farmers recall the number of clinical mastitis cases provided us with an equal level of uncertainty for all farms. The final regression models contained several variables, several of which were hard to interpret in terms of causal associations. Some of these associations may be confounded, but given the high dimensionality of our data, several these associations may also have resulted from an inflated familywise type I error and expecting to explain more variance in the dependent variables, we also ap-

plied principal components regression (Massy, 1965) with NLPCA, expecting the PC could explain more variance in the dependent variables and avoid problems in the interpretation of the regression coefficients that are due to multicollinearity. However, the PC in the regression models for 4 dependent variables only explained less than 1% of the variance in the dependent variables, and the variables that loaded high in each of the PC were quite similar to the variables remained in the ordinary regression models, which suggested that NLPCA was useful in reducing the dimension of the data set, whereas it did not add much to the ordinary regression analysis in terms of explaining variance in the dependent variables. In the data set, we had variables measured in different scales (units), and without a rescaling to make sure that the all variables have zero mean and unit variance, PCA will be biased toward giving higher loadings to variables with larger variance. Reasons for this low cumulative VAF could be (1) randomness: to some extent, which is always existent in epidemiological studies; (2) resolution and accuracy of the measurements: we had 28 numerical variables based on the farmers' memory, which could contribute to a substantial amount of noise in the data, but on the other hand because these numerical variables were mostly related to farm size and frequency of cleaning AMS units (i.e., frequency of cleaning milk tube, frequency of cleaning teat cup, and frequency of cleaning camera), which were quite familiar to the farmers, we assumed that the accuracy of the measurements was acceptable; and (3) unobserved factors: we had 10 Dutch mastitis experts working together during the study design, and all obvious potential risk factors for mastitis on AMS farms were included in the questionnaire. But indeed, it is still possible that other factors outside our scope were still present and unmeasured in our study design. There are several potential reasons for the low amount of variance in the dependent variables explained by the PC. First, the PC extracted from the explanatory variables of course have not been strongly associated with the dependent variables (Hadi and Ling, 1998). In the ordinary multivariable analyses, however, several variables that are present in the PC did show a strong association with the dependent variables (e.g., hygiene of udder and teats). It seems more likely that the strength of the causal association of the combined set of variables in a PC is lower than that of single variables. For instance, in the ordinary regression analysis, hygiene of the udder and the teats is strongly associated with udder health, but when these variables were combined in PC1 (21 variables with loadings >0.35 or <−0.35, of which 6 variables were related to hygiene of cows) and PC2 (16 variables with loadings >0.35 or <−0.35, of which 2 variables were related to hygiene of

cows) with other hygiene-related variables, it may be that this causal relation is diluted by other, nonassociated, variables. In addition, it seems that our analyses point out that there are likely to be important drivers of udder health in AMS farms other than the factors that we measured. Such an unobserved factor can lead to a low explained variance. Even though we collected data on a large number potential risk factors known to be related to udder health in CMS farms as well as risk factors that could potentially be related to udder health in AMS farms, it is possible that in AMS farms there are still factors that have not yet been identified in conventionally milking dairies by our expertise in the design of this study. This is intriguing and requires further research on factors that are specific for AMS farms.

Cleaner teats and fewer cows lying outside the cubicles were associated with lower HeSCCav in the regression model. The NLPCA results indicated that not only is cow hygiene associated with lower HeSCCav, but also good hygiene of the AMS is associated with lower HeSCCav as well as with less NHiSCC. It is known that good hygiene of cows is important. For instance, Schreiner and Ruegg (2003) and Sant'anna and Paranhos da Costa (2011) found good hygiene of cows to be associated with better udder health on CMS farms. A dirty udder and a dirty milking machine likely increase the chances of new IMI (Cardozo et al., 2015), resulting in an overall increase in SCC and more NHiSCC. Hygiene of the AMS was positively associated with hygiene of the cows, suggesting that they have determinants in common, such as the farmers' attitude toward cleaning. Better and more frequent manure removal, a sufficient number of cubicles for the number of cows, and effective and timely cleaning of the milking machine are likely to be effective interventions to improve udder health (Ruegg, 2006).

The percentage of teats not covered by spray at all was positively associated with more NHiSCC (Table 2). This is in line with previous work that showed a strong effect of postmilking teat disinfection on transmission of mastitis pathogens (Lam et al., 1997). In our study, the average percentage of teats not covered by spray at all was 15.6%. On 13 out of 135 farms (9.6%), all teats that were inspected during the farm visit were completely missed by the spraying device. This clearly suggests that postmilking teat disinfection deserves specific attention on AMS farms.

Participating in a *Salmonella* control program was found to be negatively associated with HeSCCav, HeSCCvar, and NHiSCC, and participating in a bovine diarrhea virus control program was negatively related to NHiSCC. Participation in an animal health program may have a direct beneficial effect on udder health through a reduction of the total disease burden. Alternatively, participation in animal health programs may also suggest that the farmer is ambitious to improve animal health, including udder health, which merely reflects an attitude that is likely associated with a lower HeSCCav, lower HeSCCvar, and less NHiSCC.

Other variables that reflect to some extent the farmers' attitude toward herd health, such as the use of bacteriological culturing of milk samples, checking the functioning of the AMS during milking on a regular basis, and checking the AMS in addition to the regular checks, were surprisingly positively associated with IRCM. It is unlikely that more frequent checking of the milking machine leads to more clinical mastitis. Rather, a more frequent checking of the milking machine may be the result of a high IRCM. On the other hand, the higher IRCM to some extent reflects the accuracy of the farmer in diagnosing and recording CM, where farmers with a stronger motivation to improve udder health may actually report a higher IRCM. As farmers' attitudes have been reported to be related to udder health on CMS farms (Jansen et al., 2009), interventions that contribute to a more positive attitude toward udder health are likely to be effective in also improving udder health on AMS farms.

We found that a larger milk quota, which essentially represents a larger herd size, was associated with a lower HeSCCvar, whereas the number of AMS, also a proxy for herd size, was positively related to more NHiSCC in the regression analysis. The same associations were also found with NLPCA. Herd size has been studied as a potential risk factor for clinical mastitis on CMS farms, but was not found to be significantly associated with IRCM (Schukken et al., 1990; Nyman et al., 2007). To the best of our knowledge, this is the first time herd-size-related variables were found to be associated with lower HeSCCvar and more NHiSCC. On larger farms, the influence of temporary changes in SCC in some animals may be less visible because of the large farm size, resulting in a lower HeSCCvar. A larger number of animals seems to result in a HeSCCav that is less prone to fluctuations because of incidental effects on individual animals. Still, farmers with a large herd may have less time available per cow for udder health preventive measures, which may result in more frequent new infections, reflected in a higher NHiSCC. We did not find any herd-size-related variables to be associated with HeSCCav in the regression models. In the NLPCA, however, the first PC summarized information on various herd-size-related factors, such as the number of cows, milk quota, number of feeding and lying places, and number of AMS, and this PC was associated with

a higher HeSCCav. Although these results are hard to interpret, our results suggest that udder health in larger AMS farms is more at risk than in smaller AMS farms. Because in many countries currently both the herd size and the proportion of farms with an AMS are increasing, further studies on dynamics of udder health in large AMS herds are needed.

The data that we used for this study are from 2008 and are therefore relatively old. Naturally, since 2008 there have been changes in farm structure and mastitis management. The farm size of the farms in the current study was on average 86 cows. Since 2008 the farm size has increased but not considerably. A study using data from 2013 found an average farm size of 103 (Steeneveld et al., 2015) and very recent data of 42 farms associated with the University Veterinary Practice (Harmelen, the Netherlands) showed an average farm size of 93 cows (unpublished data). Similarly the udder health status, as represented by the average farm SCC did not change much since 2008. The farm average SCC of the farms in our study was 271,800 cells/mL, whereas this was 214,000 cells/mL in 2013 (Steeneveld et al., 2015) and 175,840 cells/mL (geometric mean) in 2018 (unpublished data). This means that the results of the current paper, based upon data from 2008, still seem to be valid for current AMS farms.

CONCLUSIONS

Multiple imputation is instrumental to deal with the missing values and NLPCA was shown to be a useful technique to process high dimensional data in our study, which are common problems in questionnaire-based risk factor studies. Many of the risk factors identified in this study are comparable to those factors described in CMS farms, but farm size seems to be a factor that plays a specific role on AMS farms. Thus, our findings suggest that most of the management for mastitis control on CMS farms can be applied to AMS farms, whereas on large AMS farms, udder health may need more attention. Further studies on drivers of mastitis transmission in large AMS herds should aim to identify why larger AMS herds seem to have poorer udder health.

ACKNOWLEDGMENTS

REFERENCES

Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, M. L. Beiboer, G. Benedictus, and A. Brand. 1999. Management practices associated with the incidence rate of clinical mastitis. J. Dairy Sci. 82:1643–1654. https://doi.org/10.3168/jds.S0022-0302(99)75393-2.

Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, M. L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81:411–419. https://doi.org/10.3168/jds.S0022-0302(98)75591-2.

Barkema, H. W., Y. H. Schukken, and R. N. Zadoks. 2006. Invited review: The role of cow, pathogen, and treatment regimen in the therapeutic success of bovine *Staphylococcus aureus* mastitis. J. Dairy Sci. 89:1877–1895. https://doi.org/10.3168/jds.S0022-0302(06)72256-1.

Barkema, H. W., M. A. G. von Keyserlingk, J. P. Kastelic, T. J. G. M. Lam, C. Luby, J.-P. Roy, S. J. LeBlanc, G. P. Keefe, and D. F. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–7445. https://doi.org/10.3168/jds.2015-9377.

Bijl, R., S. R. Kooistra, and H. Hogeveen. 2007. The profitability of automatic milking on Dutch dairy farms. J. Dairy Sci. 90:239–248. https://doi.org/10.3168/jds.S0022-0302(07)72625-5.

Breen, J. E., M. J. Green, and A. J. Bradley. 2009. Quarter and cow risk factors associated with the occurrence of clinical mastitis in dairy cows in the United Kingdom. J. Dairy Sci. 92:2551–2561. https://doi.org/10.3168/jds.2008-1369.

van Buuren, S., and K. Groothuis-Oudshoorn. 2011. mice: Multivariate imputation by chained equations in *R*. J. Stat. Softw. 45. https://doi.org/10.18637/jss.v045.i03.

Cardozo, L. L., A. Thaler Neto, G. N. Souza, L. C. A. Picinin, N. C. Felipus, N. L. M. Reche, F. A. Schmidt, D. Werncke, and E. E. Simon. 2015. Risk factors for the occurrence of new and chronic cases of subclinical mastitis in dairy herds in southern Brazil. J. Dairy Sci. 98:7675–7685. https://doi.org/10.3168/jds.2014-8913.

Collins, L. M., J. L. Schafer, and C. M. Kam. 2001. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol. Methods 6:330–351. https://doi.org/10.1037/1082-989X.6.4.330.

Dohmen, W., F. Neijenhuis, and H. Hogeveen. 2010. Relationship between udder health and hygiene on farms with an automatic milking system. J. Dairy Sci. 93:4019–4033. https://doi.org/10.3168/jds.2009-3028.

Dufour, S., I. R. Dohoo, H. W. Barkema, L. Descoteaux, T. J. DeVries, K. K. Reyher, J.-P. Roy, and D. T. Scholl. 2012. Manageable risk factors associated with the lactational incidence, elimination, and prevalence of *Staphylococcus aureus* intramammary infections in dairy cows. J. Dairy Sci. 95:1283–1300. https://doi.org/10.3168/jds.2011-4711.

Elbers, A. R., J. D. Miltenburg, D. De Lange, A. P. Crauwels, H. W. Barkema, and Y. H. Schukken. 1998. Risk factors for clinical mastitis in a random sample of dairy herds from the southern part of The Netherlands. J. Dairy Sci. 81:420–426. https://doi.org/10.3168/jds.S0022-0302(98)75592-4.

Hadi, A. S., and R. F. Ling. 1998. Some cautionary notes on the use of principal components regression. Am. Stat. 52:15–19. https://doi.org/10.1080/00031305.1998.10480530.

Hair, J. F., W. C. Black, J. B. Babin, and R. E. Anderson. 2009. Multivariate Data Analysis. 7th ed. Pearson, Hoboken, NJ.

Heringstad, B., G. Klemetsdal, and J. Ruane. 2000. Selection for mastitis resistance in dairy cattle: A review with focus on the situation in the Nordic countries. Livest. Prod. Sci. 64:95–106. https://doi.org/10.1016/S0301-6226(99)00128-1.

Hogan, J. S., W. P. Weiss, and K. L. Smith. 1993. Role of vitamin E and selenium in host defense against mastitis. J. Dairy Sci. 76:2795–2803. https://doi.org/10.3168/jds.S0022-0302(93)77618-3.

Hovinen, M., and S. Pyörälä. 2011. Invited review: Udder health of dairy cows in automatic milking. J. Dairy Sci. 94:547–562. https://doi.org/10.3168/jds.2010-3556.

Huijps, K., T. J. G. M. Lam, and H. Hogeveen. 2008. Costs of mastitis: Facts and perception. J. Dairy Res. 75:113–120. https://doi.org/10.1017/S0022029907002932.

Jansen, J., B. H. van den Borne, R. J. Renes, G. van Schaik, T. J. G. M. Lam, and C. Leeuwis. 2009. Explaining mastitis incidence in Dutch dairy farming: The influence of farmers' attitudes and behaviour. Prev. Vet. Med. 92:210–223. https://doi.org/10.1016/j.prevetmed.2009.08.015.

Jolliffe, I. T. 2002. Principal Component Analysis. 2nd ed. Springer, New York, NY.

Klaas, I. C., C. Enevoldsen, A. K. Ersboll, and U. Tolle. 2005. Cow-related risk factors for milk leakage. J. Dairy Sci. 88:128–136. https://doi.org/10.3168/jds.S0022-0302(05)72670-9.

Kowarik, A., and M. Templ. 2016. Imputation with the R Package VIM. J. Stat. Softw. 74. https://doi.org/10.18637/jss.v074.i07.

Lam, T. J. G. M., B. H. P. van den Borne, J. Jansen, K. Huijps, J. C. L. van Veersen, G. van Schaik, and H. Hogeveen. 2013. Improving bovine udder health: A national mastitis control program in the Netherlands. J. Dairy Sci. 96:1301–1311. https://doi.org/10.3168/jds.2012-5958.

Lam, T. J. G. M., J. H. van Vliet, Y. H. Schukken, F. J. Grommers, A. van Velden-Russcher, H. W. Barkema, and A. Brand. 1997. The effect of discontinuation of postmilking teat disinfection in low somatic cell count herds. II. Dynamics of intramammary infections. Vet. Q. 19:47–53. https://doi.org/10.1080/01652176.1997.9694806.

Linting, M., J. J. Meulman, P. J. Groenen, and A. J. Van der Kooij. 2007a. Stability of nonlinear principal components analysis: An empirical study using the balanced bootstrap. Psychol. Methods 12:359–379. https://doi.org/10.1037/1082-989X.12.3.359.

Linting, M., J. J. Meulman, P. J. Groenen, and A. J. Van der Kooij. 2007b. Nonlinear principal components analysis: Introduction and application. Psychol. Methods 12:336–358. https://doi.org/10.1037/1082-989X.12.3.336.

Linting, M., and A. van der Kooij. 2012. Nonlinear principal components analysis With CATPCA: A tutorial. J. Pers. Assess. 94:12–25. https://doi.org/10.1080/00223891.2011.627965.

Mair, P., J. De Leeuw, and P. Groenen. 2017. Gifi: Multivariate analysis with optimal scaling. Accessed May 16, 2018. http://gifi.stat.ucla.edu/gifi/_book/.

Massy, W. F. 1965. Principal components regression in exploratory research. J. Am. Stat. Assoc. 60:234–256.

Mollenhorst, H., M. M. Hidayat, J. van Den Broek, F. Neijenhuis, and H. Hogeveen. 2011. The relationship between milking interval and somatic cell count in automatic milking systems. J. Dairy Sci. 94:4531–4537. https://doi.org/10.3168/jds.2011-4244.

Mollenhorst, H., L. J. Rijkaart, and H. Hogeveen. 2012. Mastitis alert preferences of farmers milking with automatic milking systems. J. Dairy Sci. 95:2523–2530. https://doi.org/10.3168/jds.2011-4993.

Nor, N. M., W. Steeneveld, and H. Hogeveen. 2014. The average culling rate of Dutch dairy herds over the years 2007 to 2010 and its association with herd reproduction, performance and health. J. Dairy Res. 81:1–8. https://doi.org/10.1017/S0022029913000460.

Nyman, A. K., T. Ekman, U. Emanuelson, A. H. Gustafsson, K. Holtenius, K. P. Waller, and C. H. Sandgren. 2007. Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. Prev. Vet. Med. 78:142–160. https://doi.org/10.1016/j.prevetmed.2006.10.002.

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. https://doi.org/10.3168/jds.2007-0757.

Persson Waller, K., T. Westermark, T. Ekman, and K. Svennersten-Sjaunja. 2003. Milk leakage - An increased risk in automatic milking systems. J. Dairy Sci. 86:3488–3497. https://doi.org/10.3168/jds.S0022-0302(03)73953-8.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Reinemann, D. J. 2008. Robotic Milking: Current Situation. Pages 75–80 in National Mastitis Council 47th Annual Meeting Proc., New Orleans, LA. National Mastitis Council, Madison, WI.

Ruegg, P. 2006. The role of hygiene in efficient milking. WCDS Adv. Dairy Technol. 18:285–293.

Sant'anna, A. C., and M. J. Paranhos da Costa. 2011. The relationship between dairy cow hygiene and somatic cell count in milk. J. Dairy Sci. 94:3835–3844. https://doi.org/10.3168/jds.2010-3951.

Santman-Berends, I. M. G. A., R. G. M. Olde Riekerink, O. C. Sampimon, G. van Schaik, and T. J. G. M. Lam. 2012. Incidence of subclinical mastitis in Dutch dairy heifers in the first 100 days in lactation and associated risk factors. J. Dairy Sci. 95:2476–2484. https://doi.org/10.3168/jds.2011-4766.

Santman-Berends, I. M. G. A., J. M. Swinkels, T. J. G. M. Lam, J. Keurentjes, and G. van Schaik. 2016. Evaluation of udder health parameters and risk factors for clinical mastitis in Dutch dairy herds in the context of a restricted antimicrobial usage policy. J. Dairy Sci. 99:2930–2939. https://doi.org/10.3168/jds.2015-10398.

Schafer, J. L. 1999. Multiple imputation: A primer. Stat. Methods Med. Res. 8:3–15. https://doi.org/10.1177/096228029900800102.

Schreiner, D. A., and P. L. Ruegg. 2003. Relationship between udder and leg hygiene scores and subclinical mastitis. J. Dairy Sci. 86:3460–3465. https://doi.org/10.3168/jds.S0022-0302(03)73950-2.

Schukken, Y. H., G. Casella, and J. van den Broek. 1991. Overdispersion in clinical mastitis data from dairy herds: A negative binomial approach. Prev. Vet. Med. 10:239–245. https://doi.org/10.1016/0167-5877(91)90007-o.

Schukken, Y. H., F. J. Grommers, D. Van de Geer, H. N. Erb, and A. Brand. 1990. Risk factors for clinical mastitis in herds with a low bulk milk somatic cell count. 1. Data and risk factors for all cases. J. Dairy Sci. 73:3463–3471. https://doi.org/10.3168/jds.S0022–0302(90)79045–5.

Steeneveld, W., J. C. Vernooij, and H. Hogeveen. 2015. Effect of sensor systems for cow management on milk production, somatic cell count, and reproduction. J. Dairy Sci. 98:3896–3905. https://doi.org/10.3168/jds.2014-9101.

Sterne, J. A. C., I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. 2009. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. Br. Med. J. 338:b2393. https://doi.org/10.1136/bmj.b2393.

Taponen, S., E. Liski, A. M. Heikkilä, and S. Pyörälä. 2017. Factors associated with intramammary infection in dairy cows caused by coagulase-negative staphylococci, *Staphylococcus aureus*, *Streptococcus uberis*, *Streptococcus dysgalactiae*, *Corynebacterium bovis*, or *Escherichia coli*. J. Dairy Sci. 100:493–503. https://doi.org/10.3168/jds.2016-11465.

van den Borne, B. H. P., M. Nielen, G. van Schaik, M. B. Melchior, T. J. G. M. Lam, and R. N. Zadoks. 2010a. Host adaptation of bovine *Staphylococcus aureus* seems associated with bacteriological cure after lactational antimicrobial treatment. J. Dairy Sci. 93:2550–2558. https://doi.org/10.3168/jds.2009-2971.

van den Borne, B. H. P., G. van Schaik, T. J. G. M. Lam, and M. Nielen. 2010b. Therapeutic effects of antimicrobial treatment during lactation of recently acquired bovine subclinical mastitis: Two linked randomized field trials. J. Dairy Sci. 93:218–233.