



Contents lists available at SciVerse ScienceDirect



Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

A structural equation model to evaluate direct and indirect factors associated with a latent measure of mastitis in Belgian dairy herds

J. Detilleux ^{a,*}, L. Theron ^b, J.-M. Beduin ^b, C. Hanzen ^b

^a Quantitative Genetics Group, Department of Animal Production, Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium

^b Large Animal Clinics, Obstr. et path. de la reprod. des équidés, rumin. et porcs, Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium

ARTICLE INFO

Article history:

Received 29 March 2010

Received in revised form 12 June 2012

Accepted 14 June 2012

Keywords:

Mastitis

Structural equation model

Latent variable

Risk factors

ABSTRACT

In dairy cattle, many farming practices have been associated with occurrence of mastitis but it is often difficult to disentangle the causal threads. Structural equation models may reduce the complexity of such situations. Here, we applied the method to examine the links between mastitis (subclinical and clinical) and risk factors such as herd demographics, housing conditions, feeding procedures, milking practices, and strategies of mastitis prevention and treatment in 345 dairy herds from the Walloon region of Belgium. During the period January 2006 to October 2007, up to 110 different herd management variables were recorded by two surveyors using a questionnaire for the farm managers and during a farm visit. Monthly somatic cell counts of all lactating cows were collected by the local dairy herd improvement association. Structural equation models were created to obtain a latent measure of mastitis and to reduce the complexity of the relationships between farming practices, between indicators of herd mastitis and between both. Robust maximum likelihood estimates were obtained for the effects of the herd management variables on the latent measure of herd mastitis. Variables associated directly ($p < 0.05$) with the latent measure of herd mastitis were the addition of urea in the rations; the practices of machine stripping, of pre-and post-milking teat disinfection; the presence of cows with hyperkeratotic teats, of cubicles for housing and of dirty liners before milking; the treatment of subclinical cases of mastitis; and the age of the herd (latent variable for average age and parity of cows, and percentage of heifers in the herd). Treatment of subclinical mastitis was also an intermediate in the association between herd mastitis and post-milking teat disinfection. The study illustrates how structural equation model provides information regarding the linear relationships between risk factors and a latent measure of mastitis, distinguishes between direct relationships and relationships mediated through intermediate risk factors, allows the construction of latent variables and tests the directional hypotheses proposed in the model.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Public concern about farm animal welfare has steadily grown during recent years and many indicators have been proposed to evaluate its state in terms of the 5

freedoms, among which the freedom from disease (Dalmau et al., 2009). In dairy practice, disorders of the udder are among the most frequent clinical conditions encountered (Fourichon et al., 2001). For example, it was found in 2 independent surveys of Belgian dairy herds that 40% of quarter milk samples from subclinically infected cows contained mastitis pathogens (Detilleux et al., 1999; Piepers et al., 2007). Given this high prevalence, scientists have searched for and identified a great number

* Corresponding author. Tel.: +32 4 3664129; fax: +32 4 3664122.
E-mail address: jdetilleux@ulg.ac.be (J. Detilleux).

of risk factors for the development of clinical mastitis. These include factors at the herd, cow and quarter levels (Breen et al., 2009).

In this study, we restricted ourselves to the study of factors at the herd level. Factors such as inadequate nutrition, poor milking hygiene, deficient milking procedures or increased exposure of the herd to environmental pathogens may increase the risk of clinical and subclinical mastitis (de Vliegher et al., 2004; Haltia et al., 2006; Nyman et al., 2007; Schukken et al., 1991). But all these characteristics are interrelated so it is difficult to distinguish between spurious and factual associations, or between direct and mediated effects. To reduce the complexity of such situations, it is possible to use structural equation models (SEM) with latent variables (Bollen, 2002). Within this methodology, observed variables may be considered as the expression of a small number of latent constructs that cannot be directly measured but are inferred from the observed variables. In our setting, this would suggest that observed herd characteristics and mastitis indicators can be aggregated in a few numbers of underlying concepts. Then, the covariance or correlation matrices (or both) between latent constructs and observed variables, and between latent constructs, are modeled as a function of a set of parameters. It is then possible to test alternative hypotheses about the relationships between these variables. For example, relationships between somatic cell counts and milk yield have been explored with this method (de los Campos et al., 2006; Wu et al., 2008). The SEM handles collinearity among independent variables and controls spurious inflation of type I error which constitutes a risk in multiple univariate analyses when testing complex relationships (Pugesek, 2003). Another interesting feature of the SEM consists in the conjoint estimation of direct and mediated (or indirect) effects (Casal et al., 1990). Here, a direct effect will occur if a putative risk factor influences directly on mastitis. It is not passed on through intervening variables but remains when all other variables in the model are held constant. On the other hand, a mediated effect will occur when the influence of the risk factor on mastitis is through one (or more) intervening variable. If we denote Y as the state of mastitis, X as the risk factor and M as the intervening variable, then the path $X \rightarrow M \rightarrow Y$ represents the mediated effect by which X indirectly affects Y through M. Total effects are the sum of direct and mediated effects. The comparison of direct and mediated effects might allow us to substitute intervening variables to measures of mastitis when these are difficult to obtain or at a high cost. This could be the case if a treatment (X) was clearing a bacterial infection (Y) by modifying some biochemical milk profile (M). This example illustrates how it could be more cost-effective to evaluate the effect of a treatment by measuring the variation of milk components by modern rapid analytical methods (M) than by counting colonies on microbiological cultures (Y). The identification of intervening variables may serve also to clarify the nature of the relationships between mastitis and risk factor. This issue received renewed interest as a way to address the criticisms of 'black box' epidemiology (Hafeman and Schwartz, 2009).

The objective of this paper is thus to apply the SEM methodology to evaluate and to quantify the direct and

mediated effects of herd characteristics on mastitis in dairy herds from the Walloon region of Belgium.

2. Materials and methods

2.1. Animals and study area

A random stratified sample of 349 dairy farms was surveyed between January 2006 and October 2007 in the Walloon region of Belgium. This sample represented 25% of producers registered for national milk recording, for a total of 16,000 cows. As a rule of thumb, the sample size in SEM should be 10–20 times the number of exploratory variables, with a minimum of 200 (Hoe, 2008). The stratification was based on the results of a principal component analysis of the management and herd characteristics of all dairy herds present in the region ($n = 1303$). The first two principal components were divided in quartiles and the herds were randomly sampled within each of these quartiles (16 different strata) to ensure all farm characteristics were represented.

Up to 110 different farming practices and indicators of mastitis were recorded in each farm by two surveyors. A questionnaire was designed to collect information during the interview with the farmer, using closed questions (Delfosse et al., 2008). Farming practices included herd demographics ($n = 5$), productive and reproductive indicators ($n = 4$), feeding procedures ($n = 16$), types of housing ($n = 12$), strategies of mastitis prevention and treatment ($n = 48$), and milking methods ($n = 19$). Six different indicators of mastitis were considered based on SCC because information on bacteriological findings and on clinical cases of mastitis was not obtainable. These indicators were '3-mo SCC', 'cows actually with high SCC' (numbers and percentages), 'cows previously with high SCC', 'last SCC', and 'cows milked aside'. They are described in Table 1.

2.2. Statistical analysis

After a general introduction to SEM, the statistical analyses for our specific study are presented. They consisted in the identification of noteworthy associations between mastitis indicators and farming practices, the preparation of the data for the SEM (imputation and factor analysis), the establishment of the SEM itself (measurement and structural submodels), and finally the estimation of the direct, mediated and total effects.

2.2.1. Introduction to SEM

A SEM is composed of two different components. The first component is the structural submodel where endogenous latent variables are regressed on latent predictors according to:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},$$

where $\boldsymbol{\eta}$ is a vector of endogenous latent variables, $\boldsymbol{\xi}$ is a vector of exogenous latent variables and $\boldsymbol{\zeta}$ is a vector of error terms. The model contrasts exogenous (i.e., independent from the states of other variables) from endogenous variables. It is further assumed $E(\boldsymbol{\xi}) = \mathbf{0}$, $\text{var}(\boldsymbol{\xi}) = \boldsymbol{\Xi}$, $E(\boldsymbol{\zeta}) = \mathbf{0}$,

Table 1

Short description of indicators of high SCC and potential risk factors not eliminated after the preliminary univariate analyses.

Indicators	Recording	Description and classification
Somatic cell counts		
Last SCC	DHI-records	Log-transformed herd somatic cell count (SCC) at last recording before the visit ($\log_2 (\text{SCC}/100,000) + 3$, where SCC are counts/ml)
3-mo SCC	DHI-records	Arithmetic average over the last 3 months before the visit of log-transformed herd SCC ($\log_2 (\text{SCC}/100,000) + 3$, where SCC are counts/ml)
Cows actually with high SCC	DHI-records	Percentage and number of cows with ≥ 1 monthly sample $\geq 400,000$ cells/ml in the current lactation
Cows previously with high SCC	DHI-records	Number of cows with ≥ 1 monthly sample $\geq 400,000$ cells/ml in the previous lactation
Cows milked aside	Visual/interview	Number of cows treated with antibiotics and cows with high SCC for which milk was not included in the tank
Herd demographics		
Heifers	DHI-records	Percentage of non lactating primiparous cows
Age	DHI-records	Arithmetic average age in months of cows present at the farm visit
Parity	DHI-records	Arithmetic average parity of cows present at the farm visit
Lactating cows	DHI-records	Number of lactating cows
Dry cows	DHI-records	Number of dry cows
Productive indicators		
Milk	DHI-records	Arithmetic average test-day milk of lactating cows present at the farm visit
Milk fat	DHI-records	Arithmetic average test-day milk fat percentage of lactating cows present at the farm visit
Milk protein	DHI-records	Arithmetic average test-day protein percentage of lactating cows present at the farm visit
Milk quota	DHI-records	Total in tonnes of the herd milk quota
Milking methods		
Preparation time	Visual/interview	Time in seconds necessary to prepare the udder before milking
Delay	Visual/interview	Time between the end of the udder preparation and the installation of the milking machine (s)
Hand washing	Visual/interview	Whether or not milker(s) washed (or not) their hands after milking each cow
Milker	Visual/interview	Whether or not one person milked the cows
Teats pre-dipping	Visual/interview	Whether or not milker(s) disinfected teats before milking
Foremilk check	Visual/interview	Whether or not milker(s) examined foremilk of each quarter of each cow prior to milking
Machine stripping	Visual/interview	Whether or not milker(s) applied pressure on the claw at the end of milking to remove part of residual milk
Silicone liner type	Visual/interview	Whether or not liners of the milking machine were in silicone
Dirty liner	Visual	Whether or not liners were dirty inside at the beginning of milking
Pulsator type	Visual/interview	Whether or not pulsator was used per claw
Hyperkeratosis	Visual	Percentage of cows with lesion(s) of hyperkeratosis
Claws per milker	Visual	Number of claws per milker
Types of housing		
Calving pen	Visual/interview	Whether or not the farm had a pen in which cows may calve
Cubicles	Visual/interview	Whether or not lactating cows were housed in a free-stall on cubicles
Grated floor	Visual/interview	Whether or not lactating cows were housed on grated floor
Inside barn	Visual/interview	Whether or not lactating cows were housed inside barns
Mastitis prevention and treatment		
Partial insertion	Visual/interview	Whether or not cows received antibiotics with a short (<4 mm) or long (>54 mm) infusion cannula at drying
Treatment if subclinical mastitis	Visual/interview	Whether or not cows with SCC above threshold and/or positive CMT were infused with antibiotics
Treatment if clots	Visual/interview	Whether or not udders of cows with clots in milk were infused with antibiotics
Treatment if swelled udder	Visual/interview	Whether or not cows with swelled udder received antibiotics (systematic and intra-mammary administration)
Treatment if bad state	Visual/interview	Whether or not cows with clinical mastitis (bad general state) received antibiotics (systematic and intra-mammary administration)
Udder cleaning	Visual/interview	Whether or not milker(s) cleaned udder teats with a dry or humid tissue (paper or cloth)
Cup washing	Visual/interview	Whether or not milker(s) washed the cups of cows with subclinical or clinical mastitis
Teats post-dipping	Visual/interview	Whether or not milker(s) disinfected systematically teats before milking
Dry period	Interview	Time in days of the dry period
Feeding procedure		
Urea in feed	Visual/interview	Whether or not urea was added to the ration

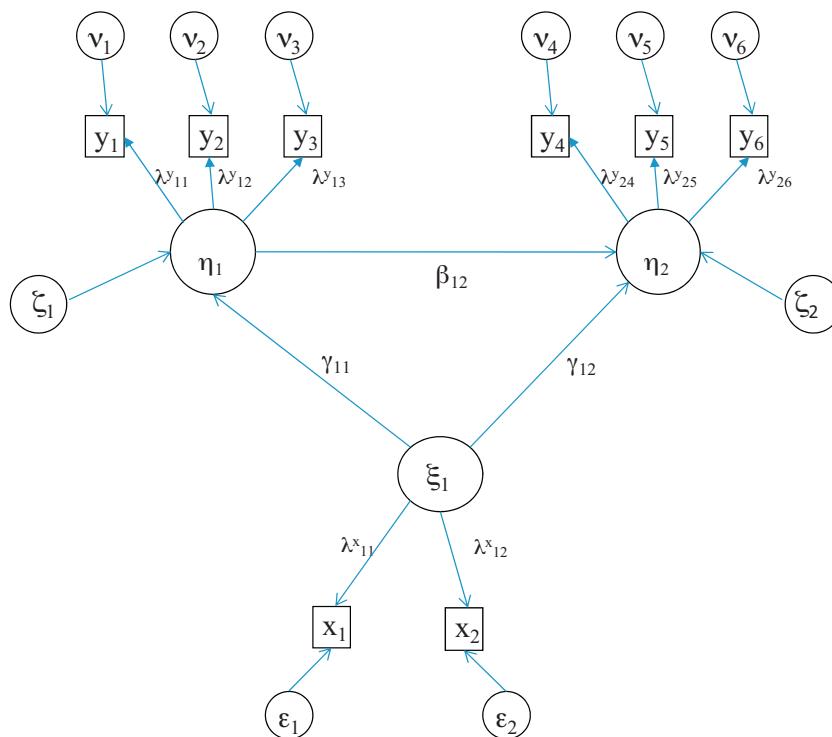


Fig. 1. An example of a structural equation model with 2 endogenous latent variables (η_1 and η_2), one exogenous latent variable (ξ_1) and 8 observed variables (y_1 to y_3 , y_4 to y_6 , x_1 and x_2). In the measurement submodel, regression coefficients relate observed and latent variables (λ_{y11}^x , λ_{y12}^x , and λ_{y11}^y to λ_{y26}^y). In the structural submodel, direct links between latent variables are represented by the regression coefficients β_{12} , γ_{11} and γ_{12} . Errors are represented by the symbols ν , ε , and ζ . See the text for more details.

$\text{var}(\zeta) = \Psi$, and $\text{cov}(\xi, \zeta) = \mathbf{0}$. Matrix \mathbf{B} contains recursive and simultaneous effects relating elements of η , and Γ contains regression coefficients between ξ and η . Equivalently, the structural model can be written as:

$$\eta = (\mathbf{I} - \mathbf{B})^{-1} \Gamma \xi + (\mathbf{I} - \mathbf{B})^{-1} \zeta.$$

The second component of the SEM contains the measurement submodels that link the vector for the observable exogenous (\mathbf{x}) and endogenous (\mathbf{y}) variables to their underlying (latent) variables:

$$\text{y-measurement submodel : } \mathbf{y} = \Lambda_y \eta + \mathbf{v},$$

$$\text{x-measurement submodel : } \mathbf{x} = \Lambda_x \xi + \mathbf{\epsilon},$$

with $E(\mathbf{v}) = \mathbf{0}$, $\text{var}(\mathbf{v}) = \Theta_y$, $E(\mathbf{\epsilon}) = \mathbf{0}$, $\text{var}(\mathbf{\epsilon}) = \Theta_x$ and $\text{cov}(\mathbf{v}, \mathbf{\epsilon}) = \mathbf{0}$. Thus, the elements of \mathbf{x} are indicators of the elements of ξ and the elements of \mathbf{y} are indicators of the elements of η . If $\Lambda_x = \mathbf{I}$ and $\mathbf{\epsilon} = \mathbf{0}$, then $\mathbf{x} = \xi$. Similarly, if $\Lambda_y = \mathbf{I}$ and $\mathbf{v} = \mathbf{0}$, then $\mathbf{y} = \eta$.

Combining the equations of the SEM, the \mathbf{y} -measurement model can be rewritten as:

$$\mathbf{y} = \Pi_x \delta,$$

where $\Pi = \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} \Gamma \Lambda_x^{-1}$ and $\delta = \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} [\zeta - \Gamma \Lambda_x^{-1} \mathbf{\epsilon}] + \mathbf{v}$. Through the matrix Π , a SEM provides a way of interpreting the relationship between observable variables (\mathbf{x} , \mathbf{y}) as a function of regression parameters of structural (\mathbf{B} and Γ) and measurement (Λ_x and Λ_y) submodels. Furthermore, the SEM reduces to a multivariate

multiple regression model when $\mathbf{B} = \mathbf{0}$, and \mathbf{x} and \mathbf{y} are measured without error (i.e., $\mathbf{y} = \Gamma \mathbf{x} + \zeta$). Fig. 1 shows the path diagram for a hypothetical SEM.

When no distinction is done between exogenous and endogenous variables, we obtained the so-called ‘all-y’ notation for both components of the SEM: $\mu = \Phi \mu + \omega$ for the structural submodel and $\mathbf{z} = \Lambda \mu + \theta$ for the measurement submodel, with:

$$\mu = \begin{bmatrix} \xi \\ \eta \end{bmatrix}, \quad \Phi = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \Gamma & \mathbf{V} \end{bmatrix}, \quad \text{var}(\omega) = \begin{bmatrix} \Xi & \mathbf{0} \\ \mathbf{0} & \Psi \end{bmatrix},$$

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_x \\ \Lambda_y \end{bmatrix}, \quad \text{and} \quad \text{var}(\theta) = \begin{bmatrix} \Theta_x & \mathbf{0} \\ \mathbf{0} & \Theta_y \end{bmatrix}.$$

The matrix Φ represents the direct effects between elements of μ . The mediated effects are Φ^2 and the total effects are $\Phi + \Phi^2$.

2.2.2. Identification of noteworthy associations

After a Box-Cox transformation of the continuous variables that were not normally distributed, we applied a battery of univariate analyses (correlation, chi-square and F-tests) and identified that, out of all 104 farming practices, 35 were potentially associated ($p \leq 0.10$) with each other and with at least one of the 6 indicators of mastitis. These variables were kept for the following analyses and are described in Table 1.

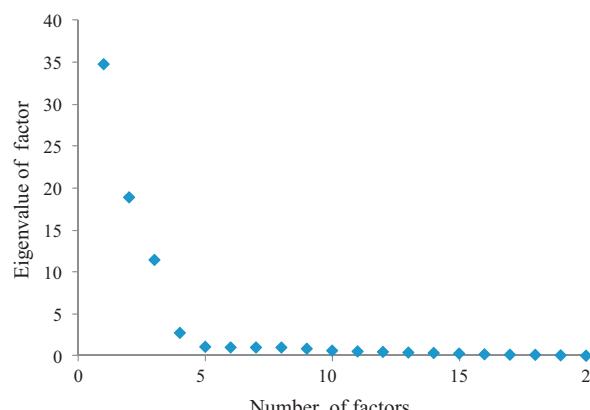


Fig. 2. Scree plot for the first 20 factors extracted from the data.

2.2.3. Multiple imputation and factor analysis

Missing data were imputed. The Markov chain Monte Carlo (MCMC) method algorithm was used to create 5 independent draws of missing data from multivariate normal distributions, and these draws were then averaged. The starting values for the MCMC algorithm were posterior modes of distributions generated by the expected-maximization algorithm from uninformative Jeffreys prior (procedures MI and MIANALYZE of SAS®). Averages of imputed categorical data were rounded to the nearest integer value. Simulation studies on data with binary and ordinal variables showed results using multiple imputation based on the multivariate normal distributions were not less robust than those based on the more flexible fully conditional specification (Lee and Carlin, 2010). However, data on 4 farms had to be discarded because information on all variables collected to describe the types of housing was lacking.

Next, we performed a factor analysis to identify whether factors were contributing to the common variance in the 41 variables measuring management practices and mastitis (procedure Factor on SAS®). Variables were not standardized for the analysis and the maximum likelihood method was chosen to estimate factor loadings, with squared multiple correlation of each observed variable regressed on all the others as the prior communalities. The results suggested that 18 out of the 41 variables could be grouped into 5 different factors. Indeed, the 5th factor was the first one to have an eigenvalue below 1 and it corresponded to the point at which the curve started bending on the scree plot (Fig. 2). The proportion of the common variance explained by the 5 factors was greater than 80%.

Variables of these 5 factors that had loadings >0.3 were considered in the latent variables of the SEM (see below: measurement submodel). The first latent variable was AGE regrouping the observed variables 'Parity', 'Heifers', and 'Age'; the second one was PROD regrouping the variables 'Milk', 'Milk protein', 'Milk fat', and 'Milk quota'; the third was NUM regrouping the variables 'Lactating cows' and 'Dry cows'; the fourth one was MLK regrouping the variables 'Preparation time', 'Delay', and 'Hand washing'; and the last one was MAM regrouping '3-mo SCC', 'Number of cows actually with high SCC', 'Number of cows already with

high SCC', 'Percentage of cows actually with high SCC', 'SCS at last recording', and 'Number of milked aside cows'. The remaining 23 variables were considered as independent risk factors in the SEM.

2.2.4. Identification of the measurement submodels

The measurement submodel indicates how observed variables are linked to underlying latent variables. Using the simple 'all-y' notation, all 41 observed variables (z_1 to z_{41}) were related to 28 latent variables (μ_1 to μ_{28}), with μ_1 = MAM regrouping z_1 to z_6 , μ_2 = AGE regrouping z_7 to z_9 , μ_3 = PROD regrouping z_{10} to z_{13} , μ_4 = NUMB regrouping z_{14} to z_{15} , μ_5 = MLK regrouping z_{16} to z_{18} , and μ_6 to μ_{28} are the error-free latent variables for the 23 remaining risk factors, i.e., considering z_{19} to z_{41} are random variables contaminated by errors (θ_{19} to θ_{41}). Models for μ_2 (variable AGE) and μ_{25} (=variable 'Dirty liners') are given as examples:

$$\begin{bmatrix} z_7 \\ z_8 \\ z_9 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_8 \\ \lambda_9 \end{bmatrix} [\mu_2] + \begin{bmatrix} \theta_7 \\ \theta_8 \\ \theta_9 \end{bmatrix} \quad \text{and} \quad z_{25} = \mu_{25} + \theta_{25},$$

with z_7 = 'Average age' in months, z_8 = 'Average parity', and z_9 = 'Percentage of heifers'. The value of λ_8 indicates the extent to which z_8 depends on AGE (μ_2), in units given by the reference, chosen here to be z_7 . The procedure was similar for MAM (unit = somatic cell score), PROD (units = kg milk), NUMB (units = number of lactating cows), and MLK (units = time in seconds to prepare the udder). Fixing one loading is required to identify the parameters of the model. Another constraint would consist in fixing the variance of the latent variable although Steiger (2002) argued this last constraint can interact with other model constraints and interfere with some hypothesis testing. Squared multiple correlations between latent and every observed variable, denoted R^2 , were used to determine the extent to which the observed variable adequately measures its respective underlying construct.

2.2.5. Identification of the structural submodels

Structural submodels indicate how latent variables are linked to each other. The premises for constructing this part of the SEM were that all 27 latent variables could affect MAM and be themselves affected by latent variables other than MAM and themselves. Then for the m th latent variable, we have:

$$[\mu_m] = [\phi_{m,1} \ \phi_{m,2} \ \phi_{m,3} \ \dots \ \phi_{m,28}] \cdot \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_{28} \end{bmatrix} + [\omega_m] \quad \text{with } m = 1-28$$

Values for $\phi_{m,1}$ and $\phi_{m,m}$ were fixed to 0 for $m=1-28$. We assigned (in the text below) the symbol RF^{MAM} to the latent variables for which partial regression coefficients with MAM ($\phi_{1,2}$, $\phi_{1,3}$, ..., or $\phi_{1,28}$) were significantly different from null ($p < 0.05$). Having identified these RF^{MAM}, we denoted as RF, the latent variables for which

partial regression coefficients ($\phi_{m,2}, \phi_{m,3}, \dots$, or $\phi_{m,28}$ with $m=2-28$) with RF^{MAM} were significantly different from null ($p < 0.05$). Variables other than RF^{MAM} and RF were not included in the final model. The direction of these relationships (from RF to RF^{MAM} to MAM) was based on our current knowledge of mastitis. We may think at the procedure as a classical backward stepwise regression analysis but, here, we went further by determining whether the RF^{MAM} kept at the end of the backward procedure were themselves influenced by RF variables.

The final theoretical model is:

$$\text{MAM} = \sum_{k=1}^n \phi_{1,k} \text{RF}_k^{\text{MAM}} + \omega_1 \quad \text{with } n \leq 28,$$

and

$$\text{RF}_m^{\text{MAM}} = \sum_{k=1}^{n_m} \phi_{m,k} \text{RF}_m + \omega_m \quad \text{for } m = 1, \dots, n.$$

2.2.6. Estimation of the parameters and fit of the models

Parameters of the final SEM were estimated with the “robust” maximum likelihood method (Boomsma and Hoogland, 2001). In this method, also called the maximum likelihood – mean adjusted, parameters are estimated by maximum likelihood and standard errors are estimated from the asymptotic covariance and correlation matrices. These “robust estimators” are supposed to be more robust against violations of the assumptions of large sample size and non-normality than regular estimators (Satorra and Bentler, 1988). The matrices were based on the Pearson correlations for continuous variables and on the tetrachoric and the polychoric correlations for ordinal indicators (LISREL® program 2.8, Jöreskog and Sörbom, 1996).

After the model parameters were estimated, the degree to which the model fits the data was assessed. Fit indices, all provided by LISREL, can be classified into several classes that include (1) discrepancy functions, such as the root mean square error of approximation (RMSEA = squared average differences between observed and model-implied covariances), (2) tests that compare the target model with the null model, such as the comparative fit index (CFI) or the normed fit index (NFI), (3) information theory goodness of fit measures, such as the Akaike Information criterion (AIC), and (4) non centrality fit indices. An RMSEA value of less than 0.08 suggests satisfactory fit, and a value of less than 0.05 suggests close fit (Browne and Cudeck, 1993). The indices NFI and CFI, all normally lie in the range 0.0–1.0, with higher values indicating better fit. As a benchmark for good fit, the value 0.90 is often used (Kline, 2005). In addition to omnibus goodness-of-fit measures (RMSEA, CFI, NFI, AIC), null hypotheses that individual parameters of the SEM are different from zero were tested through inspection of the Wald statistic (labeled *t*-value in LISREL), modification indices (estimates of the change in the likelihood-ratio chi-square statistic for the model if the parameter is specified as a free parameter), and standardized residuals (Bollen, 1989; Kline, 2005). Estimates of the parameters were obtained as if the variances of the variables were unity (complete standardization in LISREL).

2.2.7. Total, direct and indirect effects

The last step in the analyses consisted in determining, for the final SEM, the direct, mediated and total effect on MAM. Using the notation presented in the introduction, X is the risk factor or RF^{MAM}, Y is the MAM and M is an intermediate risk factor in the pathway from X to Y. The direct effects of X on Y were estimated by the corresponding elements of the matrix Φ , i.e., partial regression coefficients. The mediated effects were obtained from Φ^2 , as the product of the estimated regression coefficient of X on M times the estimated regression coefficient M on Y. The mediated effect of X on Y can also be estimated by standard regression methods, as the difference between estimated regression coefficient of X on Y ($X \rightarrow Y$) and estimated direct effect of X on Y ($X|M \rightarrow Y$). The two methods will lead equivalent estimates when no latent variable is used, no data is missing and covariates are the same in all equations. A simulation study illustrated the superiority of SEM (Iacobucci et al., 2007). The Sobel test (Sobel, 1982) was used to test the null hypothesis that the mediated effect is null.

3. Results

Descriptive statistics are given in Table 2 for the 345 farms used in the analyses. The average number of lactating cows was 50 per herd, ranging from 15 to 165. The average number of heifers was 6 per herd and ranged from 0 to 30. Milk quota varied from 105 to 1380 tonnes, with an average of 378 tonnes. Farm somatic cell score at last check ranged from 1.60 to 5.20 and the average over the last 3 months from 2.85 to 3.89. Management practices differed between farms but the majority of producers follow recommendations to prevent mastitis: they checked foremilk for early signs of infection (78%), cleaned the udder before milking (78%), practised post-milking teat disinfection (64%) and most of them treated cows with clinical mastitis (>90%). They dipped teats before milking (7%), stripped machine (37%) and washed the cups (34%). On average, 20% of sampled cows presented signs of hyperkeratosis and only 29% of farms had cows without signs of it.

Completely standardized estimates of the parameters of the final SEM are given in Table 3 for the measurement and in Fig. 3 for the final structural submodels. The goodness-of-fit statistics ($\text{RMSEA} < 0.01$; $\text{NFI} = 0.95$, and $\text{CFI} = 1.0$) suggested close fit of the final model. In Table 3, we see that 66% of the variance in the average parity number (z_8) is explained by the latent variable AGE (μ_2). The estimated value for λ_8 is 0.81 meaning that z_8 will increase by 0.81 in terms of standard deviations when AGE increases by one standard deviation. The ‘*t*-value’ of 20.4 suggests λ_8 exceeds the critical value of 2.58 at the significance level of 1% ($p < 0.01$). Parity was positively associated with the underlying AGE and the coefficient for the percentage of heifers was, logically, negative. Looking at MAM, the standardized coefficients showed that SCC at last check and animals with high SCC (number and percentage) were more highly correlated with MAM than cows milked aside. The variable MLK was positively related with the time necessary to clean the udder before milking and the practice to wash hand after milking. Finally, herd

Table 2

Percentage of herds characterized by the presence (vs. absence) of a binary risk factor (percent); mean and standard error (SE) for quantitative risk factor; average of 3-mo SCC for herds with (yes herds) or without (no herds) the risk factor; percentage of herds with no information on the risk factor (Pmiss).

Risk factors	Mean (SE)	Percent	Yes herds	No herds	Pmiss
Indicator of high SCC					
3-mo SCC	3.40 (0.01)				0.0
Cows actually with high SCC	13.76 (0.44)				2.3
Cows previously with high SCC	6.18 (0.25)				2.3
Cows milked aside	1.70 (0.11)				0.3
Last SCC	3.22 (0.03)				2.3
Herd demographics					
Heifers	33.03 (0.53)				2.3
Age	52.65 (0.36)				2.3
Parity	2.65 (0.02)				2.3
Lactating cows	50.48 (1.19)				0.0
Dry cows	6.16 (0.26)				1.4
Productive indicators					
Milk	24.12 (0.28)				2.3
Milk fat	4.06 (0.02)				2.3
Milk protein	3.33 (0.01)				2.3
Milk quota	377.99 (10.5)				2.6
Milking methods					
Preparation time	12.39 (0.70)				0.6
Delay	75.17 (4.18)				0.6
Hand washing	41.2	3.40 (0.01)	3.39 (0.01)		2.6
Milker	24.6	3.39 (0.02)	3.42 (0.03)		0.1
Teats pre-dipping	7.0	3.46 (0.04)	3.40 (0.01)*		0.3
Foremilk check	78.0	3.41 (0.01)	3.37 (0.02)		0.0
Machine stripping	37.4	3.43 (0.02)	3.38 (0.01)*		0.3
Silicone liner type	13.0	3.39 (0.03)	3.40 (0.01)		0.0
Dirty liner	7.8	3.51 (0.03)	3.39 (0.01)*		0.3
Pulsator type	19.1	3.43 (0.03)	3.39 (0.01)		0.6
Hyperkeratosis	19.62 (0.83)				8.3
Claws per milker	6.64 (0.17)				0.9
Housing					
Calving pen	34.8	3.36 (0.02)	3.42 (0.01)		3.5
Cubicles	45.2	3.37 (0.01)	3.43 (0.01)*		1.1
Grated floor	47.5	3.38 (0.01)	3.41 (0.01)		1.1
Inside barn	85.8	3.40 (0.02)	3.42 (0.03)		1.1
Mastitis treatment					
Subclinical mastitis	27.8	3.40 (0.02)	3.40 (0.01)		1.1
If presence of clots in milk	61.2	3.38 (0.04)	3.40 (0.01)		0.0
If swelled udder	93.3	3.40 (0.01)	3.41 (0.07)		0.3
If bad general state	94.2	3.39 (0.02)	3.40 (0.01)		1.4
Mastitis prevention					
Partial insertion	55.6	3.41 (0.01)	3.39 (0.01)		0.0
Udder cleaning	78.5	3.39 (0.01)	3.44 (0.02)		0.0
Cup washing	33.9	3.39 (0.02)	3.41 (0.01)		0.9
Post-dipping	64.1	3.38 (0.01)	3.44 (0.02)*		0.3
Dry period	50.05 (0.49)				0.1
Feeding procedure					
Urea in feed	10.4	3.46 (0.03)	3.39 (0.01)		1.7

* Significant difference between yes and no herds (*t*-test; $p < 0.05$).

average production was strongly associated with PROD contrary to fat and protein percentages.

For the structural submodel (Fig. 3), completely standardized coefficients showed that AGE, addition of urea in the rations, presence of dirty liners, machine stripping, presence of cows with hyperkeratosis, pre- and post-milking teat disinfection, treatment of subclinical cases of mastitis and housing of milking cows on cubicles are directly correlated with MAM. The estimates were negative when hyperkeratotic cows were observed in the farm, when farmers disinfect teats after milking and when

lactating cows were housed in a free-stall on cubicles, and positive for the others. Some of the risk factor with significant links with MAM were also associated with other variables, for example with MLK (post-milking teat disinfection, dirty liners), with NUMB (hyperkeratosis and machine strip), or with both PROD and NUMB (urea in feed, cubicles, dirty liners, AGE). Treatment for subclinical mastitis and post-dipping were positively associated. Variables such as partial insertion, foremilk check and udder cleaning were associated with MLK but not with MAM.

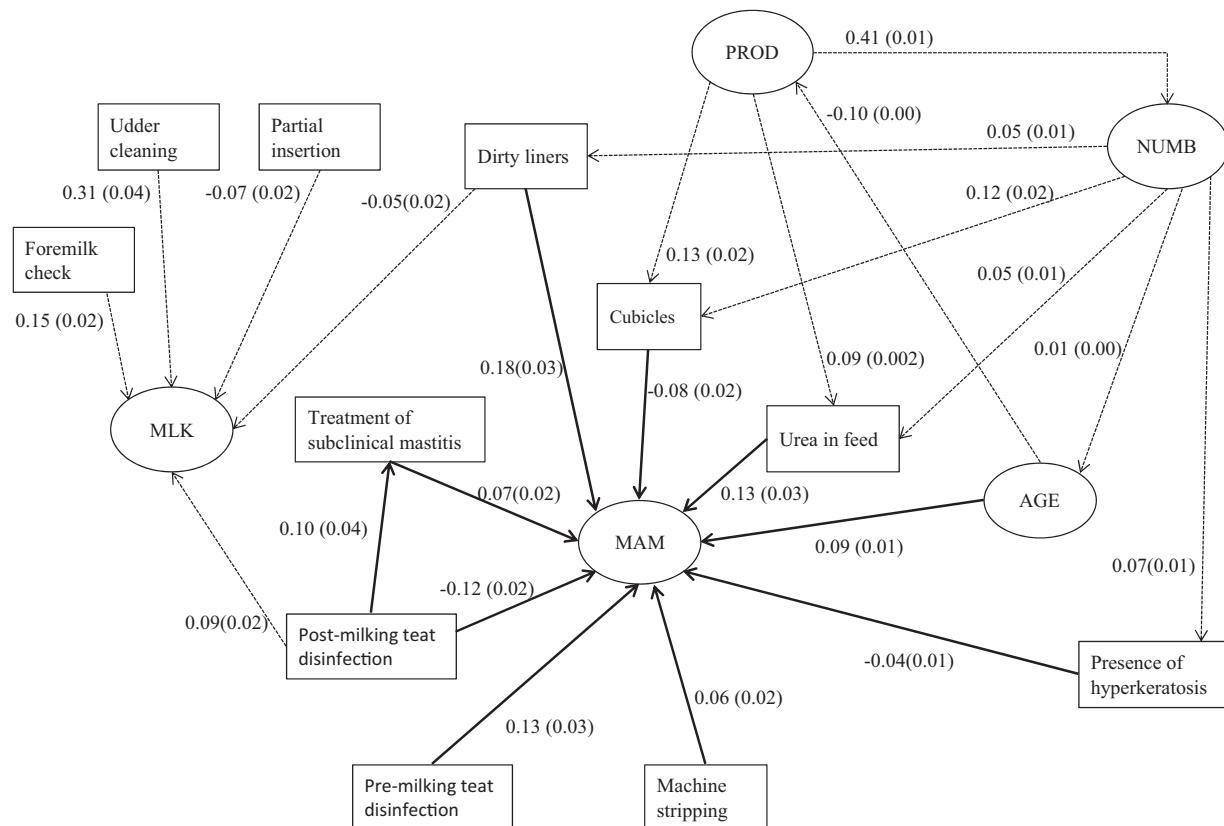


Fig. 3. Final structural model describing the direct (straight line) and indirect (dashed line) links between the latent state of udder health (MAM) and risk factors. See the text for more details.

Direct and mediated relationships were observed in the association 'post-milking teat disinfection – treatment of subclinical mastitis – MAM' (Fig. 3), in which treatment of subclinical mastitis was an intermediate in the pathway from teat disinfection to MAM. The part of the total association (-0.11 SCS; SE = 0.02) between MAM and teat disinfection that was mediated through subclinical treatment was estimated at $+0.01$ SCS (SE = 0.003). And, although this mediated effect was small, it was significantly different from null ($p < 0.05$).

4. Discussion

Among the numerous reports on the relationships between herd characteristics and mastitis, this one is particular in that it approached mastitis as a composite of several contributory indicators, represented by MAM. The MAM can be understood as a herd variable evaluating herd mastitis that is only observed through multiple indicators, each capturing different aspects of it. Statements about the associations between MAM and risk factors are therefore more comprehensive and less specific than those made for each individual mastitis indicator. The methodology is not adequate to find management strategies effective against particular mammary pathogens but it is suited to studies, like ours, with the objective of identifying factors influencing herd mastitis in a global manner, given that trends in

mastitis have changed over the last 50 years and that different mastitis pathogens are predominant in different herds, regions and countries (Zadoks and Fitzpatrick, 2009).

Looking at the completely standardized regression coefficients, all expressed in units of standard deviation, we observed the most influential variable on MAM was the presence vs. absence of dirty liners (μ_{25}). Indeed, MAM increased by 0.18 (SE = 0.03) when liners were dirty (Fig. 1). The presence vs. absence of dirty liners was itself increased by 0.05 (SE = 0.01) when NUMB increased by one standard deviation. Among other RF^{MAM} with standardized direct effects on MAM greater than 0.10, we also found dietary urea ($+0.13$, SE = 0.03) and the practice of pre-dipping ($+0.13$, SE = 0.03). Note, however, that controversial reports were retrieved in the literature as in Sterk et al. (1978) who observed a positive relationship between urea and mastitis and in Erb et al. (1976) in which no relationship was discovered.

The practice of post-milking teat disinfection was negatively associated with MAM, with a direct standardized link of -0.12 (SE = 0.02). This was expected given the numerous previous studies that demonstrated the beneficial impact of post-milking teat disinfection on clinical and subclinical mastitis (Barkema et al., 1998). But, the practice was also associated with MAM through its link with the variable 'treatment for subclinical case'. This mediated effect was estimated at $+0.01$ (SE = 0.00) and was in opposite

Table 3

Measurement models for the final structural equation model: robust maximum likelihood estimate (with the standard errors in parentheses) of the effects on each latent construct (AGE, PROD, MAM, MLK, and NUMB) along with the 't-value' (Wald statistic with critical value of 2.58 at $p < 0.01$ significance level) and the squared multiple correlation (R^2). See the text for more details.

	Estimate (SE)	't-Value'	R^2
AGE			
Parity	0.81 (0.04)	20.39	0.66
Heifers	-0.60 (0.10)	-6.54	0.36
Age	1 (ref)		
PROD			
Milk protein	-0.06 (0.01)	-13.9	0.00
Mil fat	0.02 (0.01)	3.37	0.00
Quota	1 (ref)		
Milk	0.27 (0.12)	2.31	0.09
MAM			
3-mo SCC	1 (ref)		
Cows actually with high SCC	0.46 (0.02)	22.19	0.25
Cows actually with high SCC (%)	0.55 (0.02)	21.90	0.35
Cows previously with high SCC	0.46 (0.04)	12.63	0.25
Cows milked aside	0.24 (0.02)	14.72	0.07
Last SCC	0.71 (0.03)	21.90	0.76
MLK			
Delay	0.46 (0.04)	5.98	0.20
Hand washing	0.24 (0.04)	5.44	0.04
Preparation time	1 (ref)		
NUMB			
Lactating cows	1 (ref)		
Dry cows	0.21 (0.08)	2.80	0.07

direction to the direct link, so the estimated total effect was estimated at -0.11 (SE = 0.02), lower than the direct effect. Even if they are not spectacular, these results illustrate the potential importance of decomposing the total relationship into its direct and indirect components. For example, if the mediated link had been greater than the direct link, post-dipping would have been incorrectly associated with no mastitis by standard regression techniques although its total effect would have been not protective.

Other RF^{MAM} had direct standardized effects lower than 0.10. For example, MAM was lower in farms that house lactating cows in winter in a tie-stall (-0.08, SE = 0.02) similarly to the association found with low clinical mastitis incidence (Kjæstad and Simensen, 2001) and herd average SCC (Köster et al., 2006). The latent variable AGE had a positive effect on MAM like in other studies that showed an effect of increasing lactation number on SCC (Busato et al., 2000). Another example was the slight negative association between the presence of hyperkeratosis and MAM. Others found severe teat lesions is associated with increased risk of clinical mastitis and higher SCC (Neijenhuis et al., 2001). Conversely, Gleeson (2004) found no correlation between teat hyperkeratosis score and SCC when teats were disinfected after milking.

Some latent variables, such as PROD and NUMB, influenced MAM only indirectly, through their links with other intervening variables. For example, they influenced MAM through their effect on the presence of cubicles. In this survey, farms that house lactating cows in winter in a tie-stall were indeed larger and had a greater milk quota: the average milk quota was 462,100 kg for 59.5 lactating cows

in farms with cubicles, to be compared to the 308,800 kg of milk quota and the average number of 43.1 lactating cows in farms without cubicles. The NUMB also influenced MAM through its effect on the presence of hyperkeratosis. The average percentage of cows with hyperkeratosis was higher in larger farms: in farms with more than 50 lactating cows, 21.7% of cows presented lesions of hyperkeratosis for 17.8% in farms with less than 50 lactating cows. Haskell et al. (2009) argued also that the relationship between herd size and SCC may be indirect.

Some latent variables had no effect on MAM. For example, the latent variable describing the time necessary to prepare the udder (MLK), had no effect on MAM but Kiiman et al. (2005) observed high bulk tank SCC when the total udder preparation was short (average of 30 s). The practice of examining foremilk to facilitate early detection of clinical mastitis observed in 78.2% of the farms was also not associated with MAM. Interestingly, Jayarao et al. (2004) observed that bulk tank SCC was higher in herds that practice fore-stripping before milking than in herds that do not.

5. Conclusions

This study described a structural equation model with latent variables to understand the relative influence of risk factors on a latent measure of the state of mastitis in herds from a particular region of Belgium. Variables associated directly with the latent measure of herd mastitis were the addition of urea in the rations; the practices of machine stripping, of teat disinfection before and after milking; the presence of cows with hyperkeratotic teats, of free-stall on cubicles to house lactating cows and of dirty liners before milking; the treatment of subclinical cases of mastitis; and an latent measure of the age of the herd. Teat disinfection after milking influenced also the latent measure of herd mastitis through its association with the treatment of sub-clinical cases of mastitis.

Acknowledgments

This study was supported by EADGENE (European Animal Disease Genomics Network of Excellence for Animal Health and Food Safety). The authors wished also to thank the OSAM team that collect and store the data.

References

- Barkema, H.W., Schukken, Y.H., Lam, T.J.G.M., Beiboer, M.L., Benedictus, G., Brand, A., 1998. Management practices associated with low, medium and high somatic cell counts in bulk milk. *J. Dairy Sci.* 81, 1917–1927.
- Bollen, K.A. (Ed.), 1989. Structural Equations with Latent Variables. Wiley, New York.
- Bollen, K.A., 2002. Latent variables in psychology and the social sciences. *Annu. Rev. Psychol.* 53, 605–634.
- Boomsma, A., Hoogland, J.J., 2001. The robustness of LISREL modeling revisited. In: Cudeck, R., du Toit, S., Sörbom, D. (Eds.), Structural Equation Modeling, Present and Future. Scientific Software International, Chicago, pp. 139–168.
- Breen, J.E., Green, M.J., Bradley, A.J., 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.
- Browne, M.W., Cudeck, R., 1993. Alternative ways of assessing model fit. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. SAGE, London, pp. 136–162.

- Busato, A., Trachsel, P., Schallibaum, M., Blum, J.W., 2000. Udder health and risk factors for subclinical mastitis in organic dairy farms in Switzerland. *Prev. Vet. Med.* 44, 205–220.
- Casal, J., Learte, P., Torre, E., 1990. A path model of factors influencing bovine leukemia virus transmission between cattle herds. *Prev. Vet. Med.* 10, 47–61.
- Dalmau, A., Temple, D., Rodriguez, P., Llonch, P., Velarde, A., 2009. Application of the Welfare Quality® protocol at pig slaughterhouses. *Anim. Welfare* 18, 497–505.
- Delfosse, C., Humblet, M.-F., Remience, V., Froidmont, E., Hanzen, C., Bartiaux-Thill, N., 2008. Identification et hiérarchisation des facteurs de risque liés aux mammites dans les élevages laitiers wallons. Rapport d'activité, Gembloux.
- de los Campos, G., Gianola, D., Heringstad, B., 2006. A structural equation model for describing relationships between somatic cell score and milk yield in first-lactation dairy cows. *J. Dairy Sci.* 89, 4445–4455.
- Detilleux, J.C., Arendt, J., Lomba, F., Leroy, P., 1999. Methods for estimating areas under receiver-operating characteristic (ROC) curves. Illustration with somatic-cell scores in subclinical intramammary infections. *Prev. Vet. Med.* 41, 75–88.
- de Vliegher, S., Maevens, H., Barkema, H.W., Dohoo, I.R., Stryhn, H., Opsomer, G., de Kruif, A., 2004. Management practices and heifer characteristics associated with early lactation somatic cell count of Belgian dairy heifers. *J. Dairy Sci.* 87, 937–947.
- Erb, R.E., Brown, C.M., Callahan, C.J., Moeller, N.J., Hill, D.L., Cunningham, M.D., 1976. Dietary urea for dairy cattle. II. Effect on functional traits. *J. Dairy Sci.* 59, 656–667.
- Fourichon, C., Seegers, H., Beaudeau, F., Verfaille, L., Bareille, N., 2001. Health-control costs in dairy farming systems in western France. *Livest. Prod. Sci.* 68, 141–156.
- Gleeson, D.E., 2004. Effect of teat hyperkeratosis on somatic cell counts of dairy cows. *Intern. J. Appl. Res. Vet. Med.* 2, 115–122.
- Hafeman, D.M., Schwartz, S., 2009. Opening the black box: a motivation for the assessment of mediation. *Int. J. Epidemiol.* 38, 838–845.
- Haltia, T., Honkanen-Buzalski, I., Spiridonova, A., Olkonen, A., Myllys, V., 2006. A study of bovine mastitis, milking procedures and management practices on 25 Estonian dairy herds. *Acta Vet. Scand.* 48, 22–28.
- Haskell, M.J., Langford, F.M., Jack, M.C., Sherwood, L., Lawrence, A.B., Rutherford, K.M.D., 2009. The effect of organic status and management practices on somatic cell counts on UK farms. *J. Dairy Sci.* 92, 3775–3780.
- Hoe, S.I., 2008. Issues and procedure in adopting structural equation modeling technique. *J. Appl. Quant. Meth.* 3, 76–83.
- Iacobucci, D., Saldanha, N., Deng, X., 2007. A meditation on mediation: evidence that structural equations models perform better than regressions. *J. Couns. Psychol.* 17, 139–153.
- Jayarao, B.M., Pillai, S.R., Sawant, A.A., Wolfgang, D.R., Hegde, N.V., 2004. Guidelines for monitoring bulk tank milk somatic cell and bacterial counts. *J. Dairy Sci.* 87, 3561–3573.
- Jöreskog, K.G., Sörbom, 1996. In: Stam, L. (Ed.), LISREL 8, User's Reference Guide. Scientific Software International, Chicago.
- Kiiman, H., Pärna, E., Leola, A., Kaart, T., 2005. Effect of pre milking procedures on milk somatic cell count. *Akad. Pöllum. Selts.*, 90–95.
- Kjæstad, H., Simensen, E., 2001. Cubicle refusal and rearing accommodation as possible mastitis risk factors in cubicle-housed dairy heifers. *Acta Vet. Scand.* 42, 123–130.
- Kline, R.B. (Ed.), 2005. Principles and Practice of Structural Equation Modeling. The Guilford Press, New York.
- Köster, G., Tenhagen, B.-A., Heuwieser, W., 2006. Factors associated with high milk test day somatic cell counts in large dairy herds in Brandenburg. I. Housing conditions. *J. Vet. Med. A* 53, 134–139.
- Lee, K.J., Carlin, J.B., 2010. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am. J. Epidemiol.* 171, 624–632.
- Neijenhuis, F., Barkema, H.W., Hogeweegen, H., Noordhuizen, J.P.T.M., 2001. Relationship between teat-end callosity and occurrence of clinical mastitis. *J. Dairy Sci.* 84, 2664–2672.
- Nyman, A.K., Ekman, T., Emanuelson, U., Gustafsson, A.H., Holtenius, K., Waller, K.P., Sandgren, C.H., 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.
- Piepers, S., De Meulemeester, L., de Kruif, A., Opsomer, G., Barkema, H.W., de Vliegher, S., 2007. Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. *J. Dairy Res.* 74, 478–483.
- Pugesek, B.H., 2003. Concepts of structural equation modeling in biological research. In: Pugesek, B.H., Tomer, A., von Eye, A. (Eds.), *Structural Equation Modeling, Applications in Ecological and Evolutionary Biology*. Cambridge University Press, Cambridge, pp. 42–59.
- Satorra, A., Bentler, P.M., 1988. Scaling corrections for chi-square statistics in covariance structure analysis. In: Proc. of Am. Stat. Assoc., Business and Economics Section, pp. 308–313.
- Schukken, Y.H., Grommers, F.J., van de Geer, D., Erb, H.N., Brand, A., 1991. Risk factors for clinical mastitis in herds with a low bulk milk somatic cell count. 2. Risk factors for *Escherichia coli* and *Staphylococcus aureus*. *J. Dairy Sci.* 74, 826–832.
- Sobel, M.E., 1982. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol. Methodol.* 13, 290–312.
- Steiger, J.H., 2002. When constraints interact: a caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychol. Methods* 7, 210–227.
- Sterk, V., Beslin, R., Anojcic, A., Pavlicevic, A., 1978. Effect of method of feeding on the defence capacity of the udder in dairy cows. *Vet. Glas.* 32, 899–903.
- Wu, X.-L., Heringstad, B., Gianola, D., 2008. Exploration of lagged relationships between mastitis and milk yield in dairy cows using a Bayesian structural equation Gaussian-threshold model. *Genet. Sel. Evol.* 40, 333–357.
- Zadoks, R.N., Fitzpatrick, J.L., 2009. Changing trends in mastitis. *Irish Vet. J.* 62, 59–70.