Creating a model to detect dairy cattle farms with poor welfare using a national database

Krug, C; Haskell, MJ; Nunes, TP; Stilwell, G

Published in:
Preventive Veterinary Medicine

DOI:
10.1016/j.prevetmed.2015.10.014

First published: 01/01/2015

Title Page

Creating a model to detect dairy cattle farms with poor welfare using a national database

C. Krug[a,]* M. J. Haskell[b], T. Nunes[a], G. Stilwell[a]

[a] *Centro de Investigação Interdisciplinar em Sanidade Animal,Faculdade de Medicina Veterinária, Universidade de Lisboa, 1300-477 Lisboa, Portugal*

[b] *Scotland's Rural Agricultural College, EH9 3JG Edinburgh, UK*

*Catarina de Noronha Krug Marques da Silva

3200, rue Sicotte, C.P. 5000 Saint-Hyacinthe, Québec, Canada J2S 7C6

E-mail: catarina.de.noronha.krug.marques.da.sil@umontreal.ca / denoronhakrug@hotmail.com

Tel: +1 450 773 8521 x. 8494 / +1 207 558 0083

*Manuscript

Creating a model to detect dairy cattle farms with poor welfare using a national database

C. Krug[a],* M. J. Haskell[b], T. Nunes[a], G. Stilwell[a]

[a] *Centro de Investigação Interdisciplinar em Sanidade Animal,Faculdade de Medicina Veterinária, Universidade de Lisboa, 1300-477 Lisboa, Portugal*

[b] *Scotland's Rural College, EH9 3JG Edinburgh, UK*

*Catarina de Noronha Krug Marques da Silva

3200, rue Sicotte, C.P. 5000 Saint-Hyacinthe, Québec, Canada J2S 7C6

E-mail: catarina.de.noronha.krug.marques.da.sil@umontreal.ca / denoronhakrug@hotmail.com

Tel: +1 450 773 8521 x. 8494 / +1 207 558 0083

ABSTRACT

The objective of this study was to determine whether dairy farms with poor cow welfare could be identified using a national database for bovine identification and registration that monitors cattle deaths and movements. The welfare of dairy cattle was assessed using the Welfare Quality® protocol (**WQ**) on 24 Portuguese dairy farms and on 1,930 animals. Five farms were classified as having poor welfare and the other 19 were classified as having good welfare. Fourteen million records from the national cattle database were analysed to identify potential welfare indicators for dairy farms. Fifteen potential national welfare indicators were calculated based on that database, and the link between the results on the WQ evaluation and the national cattle database was made using the identification code of each farm. Within the potential national welfare indicators, only two were significantly different between farms with good welfare and poor welfare, 'proportion of on-farm deaths' ($p < 0.01$) and 'female/male birth ratio' ($p < 0.05$). To determine whether the database welfare indicators could be used to distinguish farms with good welfare from farms with poor welfare, we created a model using the classifier J48 of Waikato Environment for Knowledge Analysis. The model was a decision

tree based on two variables, ‘proportion of on-farm deaths’ and ‘calving-to-calving interval’, and it was able to correctly identify 70% and 79% of the farms classified as having poor and good welfare, respectively. The national cattle database analysis could be useful in helping official veterinary services in detecting farms that have poor welfare and also in determining which welfare indicators are poor on each particular farm.

Key words: dairy cattle; animal welfare; Welfare Quality; national cattle database

INTRODUCTION

In the last 50 years, the main goal of dairy farming has been to increase milk production through genetic selection and management, thereby increasing farm profit and reducing cost for consumers. However, this one-sided selection for increased yield has brought, along with other issues, lower ability to reproduce, higher incidence of several production diseases, decreased longevity and modification of normal behaviour, which may contribute to a decline in the welfare of dairy cows (Oltenacu and Broom, 2010).

Consumer demands are the most important drivers of change in breeding and management practices, and although there has been a growing body of legislation on animal welfare within the European Union, there are only a few member states that have specific legislation on adult dairy cattle (e.g. Denmark, Austria, Sweden; European Commission, 2015). Therefore, scientific attention has been drawn to finding practical, accurate and measurable indicators of animal welfare for use on dairy cattle farms. To this end, various on-farm welfare assessment protocols have been developed. Most recently, the European Welfare Quality® (**WQ**) project developed protocols for dairy cattle and for other domestic species that resulted in reliable on-farm monitoring systems. The WQ assessment protocol for dairy cows includes 30 measures, 12 criteria and four principles (good feeding, good health, good housing and appropriate behaviour) that contribute to the final classification of a dairy farm. In contrast to previous protocols that focused mainly on resource-based measures (Sørensen et al., 2001; Main et al., 2007; Calamari and Bertoni, 2009), the WQ protocols focus mainly on animal-based measures, or outcome measures, which reflect the interaction between the animal and its environment (Veissier, 2007). However, application of the WQ protocol is time-consuming and expensive and there are concerns about whether it can feasibly be implemented in all farms (Knierim and Winckler, 2009).

The welfare of cattle on dairy farms is generally assessed for two main reasons: for quality assurance or for detection of poor welfare conditions. For the former, all farms should be evaluated but for the latter, reducing the number of farms that must be inspected by using a system that identifies a smaller sample of 'at risk' farms from pre-existing data from national cattle databases, would be advantageous.

National herd identification and registration databases for cattle contain a list of records that have become more comprehensive since the Bovine Spongiform Encephalopathy crisis, and that have the potential to be part of the future welfare monitoring systems (Fraser, 2004; European Food Safety Authority, 2012). To our knowledge, only three studies have explored the use of databases to identify dairy herds with poor or good welfare (Sandgren et al., 2009; Nyman et al., 2011; de Vries et al., 2014). Sandgren et al. (2009) and Nyman et al. (2011) used the same data set to detect dairy herds with poor and good welfare, respectively, but they used only nine animal-based measures to assess welfare at farm level. De Vries et al. (2014) employed a larger data set and the WQ protocol to assess welfare at farm level, then used the potential welfare indicators to predict specific WQ measures (e.g. severely lame cows, avoidance distance, very lean cows).

In the current study the objective was to identify routinely collected records from the national cattle database that would allow to predict the overall welfare at the farm level. These indicators could then be used to facilitate the identification of farms for which a complete WQ audit is necessary (i.e. those with a relatively high probability of insufficient welfare).

MATERIALS AND METHODS

Farms, animals and the Welfare Quality® protocol

Data from 24 dairy herds were included in this cross-sectional study. The convenience-based selection of farms was done by using contacts that had already been established for another study on culling strategies (for which farms were selected because they had reliable and available records; Barros, 2013) or through veterinary practitioners. Thirteen of these farms were located in the centre of Portugal and 11 were located in the north of the country. Holstein–Friesian was the predominant cow breed. All farms used free-stalls with the exception of one, which was based on an open bedded system. Two of the herds had 400 - 680 milking cows, seven had 200 - 399, nine had 100 - 199 and six had 20 - 99 milking cows. A total of 1,930 cows were assessed. Each farm was visited once between January 2013 and March 2013 by the first author (CK), spending an average of one day per farm.

The WQ assessment protocol for dairy cattle was conducted (Welfare Quality®, 2009). The protocol consists of 30 measures that cover four principles – health, feeding, housing and behaviour. The sample size of cows on each farm was selected according to the WQ protocol, being determined by herd size. As suggested by the WQ protocol, cows in each farm were selected randomly, in the milking parlour. In the case of one farm that had a robotic milking system, animals were selected in the feeding rack, choosing every n^{th} cow in the rows. No dry cows, or animals housed away from the milking herd were included.

Data collected on farm (30 welfare measures) were used to calculate scores for the 12 animal welfare criteria, which in turn were used to score the four welfare principles – Good feeding, Good health, Good housing and Appropriate behaviour (Table 1) – and these contribute to the final welfare classification of a dairy farm. Each farm has four possible classifications: excellent, enhanced, acceptable and not classified (poor). To be assigned to one of these levels of welfare, a farm must reach the assigned value for that particular classification (≥75 for excellent, ≥50 for enhanced, ≥15 for acceptable) on 2 or 3 of the 4 principles, and not score below that value for the lowest category on the other principle(s). For example, if a farm has an excellent classification in two principles, and the other two are acceptable, the farm is considered enhanced (Welfare Quality®, 2009).

In our project, only one of the 24 farms was scored as having enhanced welfare, while the majority (18 farms) was scored as acceptable and five farms were considered not classified because they did not reach the minimum requirements. Following this classification the farms were divided into two groups: farms scored ‘enhanced’ or ‘acceptable’ (n = 19) were classified as having ‘good welfare’ (**GW**) and farms with score ‘not classified’ (n=5) were categorized as having ‘poor welfare’ (**PW**).

Potential welfare indicators from national cattle database

A subset of data concerning the time between January 2008 and December 2011 was extracted from the Portuguese national cattle database (Sistema Nacional de Identificação e Registo de Bovinos, SNIRB), with the exception of animal movements for which data until October 2012 were available. The data subset included the following tables: live cattle; births; herd movement records; and records at slaughter. From these tables, that contained a total of 14,558,563 records, variables for analysis were generated (see Table 2 for calculations).

Variables were selected based on a literature review on animal welfare (Fraser and Broom, 1990; European Food Safety Authority, 2006), on potential welfare indicators already identified by Sandgren et al. (2009) and on the data that were available in the national database. The variables calculated were: median age at first calving (**AFC**); proportion of calving intervals lower than the biologically acceptable (**CCI < 345**); proportion of calving intervals higher than 430 days (**CCI > 430**); calf mortality rate (until six months; **MtC**); mortality rate (**Mt**); proportion of on-farm deaths (**OFD**); proportion of emergency slaughter (**EmgSl**); median total life span (**TLS**); proportion of cows slaughtered before 30 days post-partum (**30ppSl**); proportion of cows slaughtered before 60 days post-partum (**60ppSl**); proportion of partial carcass rejection (**PartCRej**); proportion of total carcass rejection (**TotCRej**); proportion of carcasses weighting less than 272 kg (**C < 272**); proportion of carcasses with fat class 'very thin' (class 1; **VTC**); and female/male births ratio (**SexRatio**). Both CCI < 345 and SexRatio were considered as potential indicators of record completeness. Calving to calving intervals below 345 days are very unlikely and so may indicate poor record keeping. Similarly, an unbalanced ratio between males and females (with 'unbalanced' defined as being when a ratio fell in the upper or lower quartile - see Table 3) may indicate that calves of one of the sexes were not recorded in dairy producers' records.

Variables were computed for all Portuguese dairy farms. This allowed 17,649 individual dairy farms to be identified, but farms with less than five cows were then excluded, resulting in a total of 6,605 dairy farms.

Data analysis

Data extracted from SNIRB were used to calculate potential welfare indicators using Microsoft Office Access 2007, with the exception of CCI, which were calculated using R i386 3.0.0.

Histograms for the different variables were generated and normality was evaluated through visual inspection. In most instances data were not normally distributed (Table 3 and Table 4 show mean, SD, median and IQR for each potential welfare indicator), therefore, for simplicity, a non-parametric test, the two-sample Wilcoxon test was used to determine if there was a significant difference in potential welfare indicators between GW and PW farms. The same test was also used to compare the 24 dairy farms assessed using the WQ protocol with the 6,605 Portuguese dairy farms.

Data mining techniques were used to develop the model relating WQ welfare classification to the potential welfare indicators. Data mining techniques enable to automatically evaluate large datasets containing various variables and decide which variables are most relevant (Abernethy, 2010a). Specifically, we created the model using the classification method. This involves using a database with dairy cattle farms whose welfare state is known, and then building a model that is able to automatically classify the level of welfare of new farms (whose level of welfare is unknown) through their attributes (i.e. potential welfare indicators; Abernethy, 2010b). We used the classifier J48 from WEKA 3.6.9 (WEKA, 2015), an open source data mining software in Java. J48 is a preferred method for small datasets (Ali et al., 2012) and acted by: 1) building a decision tree based on the entire given dataset; 2) splitting the data into smaller subsets by testing for a given potential welfare indicator; 3) identifies the potential welfare indicators that discriminate the various cases of good or poor level of welfare most clearly (those that 'have the highest information gain').

In the current study the number of farms with PW and GW was disproportionate (5 to 19), therefore we used the Synthetic Minority Over-sampling Technique (**SMOTE**) to reduce disproportion, which resulted in a total number of 29 observations, ten with PW and 19 with

GW. This method over-samples the minority group and under-samples the majority group. By doing that, when classifying the minority class, it is possible to get an increment on its sensitivity. The over-sampling of the minority class is done by producing 'synthetic' examples (Chawla et al., 2002).

To evaluate our model, sensitivity, specificity and predictive values were calculated. Sensitivity was defined as the proportion of PW farms (according to the WQ audit) that were correctly identified by the model (using potential welfare indicators from the database). Specificity was defined as the proportion of GW farms (according to the WQ audit) that were correctly identified (using potential welfare indicators). Positive predictive value (**PPV**) and negative predictive value (**NPV**) were defined as true positives divided by total number of predicted positives; and true negatives divided by total number of predicted negatives, respectively. Hence, the PPV represents the probability of truly being a PW farm (based on the WQ audit) among the farms identified by the model as having PW (based on the potential welfare indicators from the national database), while the NPV represents the probability of truly being a GW farm among the farms identified by the model as having GW.

RESULTS

Potential welfare indicators from national cattle database

Distributions of the different variables created using data from the national database for all 6,605 Portuguese farms with more than five cows and, more specifically, for the 24 farms with WQ assessment are presented in Table 3. Most variables differed significantly ($P < 0.05$; Wilcoxon test) between national dairy farms and our study sample, with the exception of AFC, CCI<430, VTC and SexRatio.

Identifying farms with poor welfare using indicators from the national cattle database

The Wilcoxon test revealed that only two potential welfare indicators differed significantly between both groups: OFD and SexRatio (Table 4). The classification tree is presented in Figure 1. This decision tree had two potential welfare indicators: OFD and CCI > 430. The model had a good overall accuracy, correctly classifying 22 observations while classifying seven observations erroneously. Moreover, it had a sensitivity of 70.0% (7/10), a specificity of 78.9% (15/19), a PPV of 63.6% (7/11) and a NPV of 83.3% (15/18) (Table 5).

DISCUSSION

The current study sample can hardly be considered a representative sample of the Portuguese dairy cattle population, due to the sampling process (i.e. convenient selection) and its small size (i.e. accounts for only 0.14% of all Portuguese dairy farms (n = 17,649) and 0.36% of farms with greater than five dairy cows (n = 6,605)). Only four potential welfare indicators were not significantly different between sample and source population.

In the current study, due to travel limitations, only farms from central and northern Portugal were evaluated with the WQ audit, while the national cattle database also includes farms from the south and the islands. It is, therefore, possible that the differences between the national population of farms and the study sample are due to differences between Portuguese regions, or due to the effect of individual farms. As explained by Dohoo et al. (2009), a sample that is based on convenience should not be used in a descriptive study aiming to describe population parameters (i.e. results from the 24 herds should not be used to describe the distribution of the welfare indicators in Portuguese dairy farms). This was not a problem in the current study, since the distribution of the welfare indicators in the population could be estimated from the 6,605 herds (Table 3). Although the sample size of the current study is small, we could make the assumption that the model is appropriate to farms from north and centre of Portugal, however, caution should be taken in generalization of these results to farms from southern Portugal and from its islands. On the other hand, since studies from other countries have found that mortality (Sandgren et al., 2009; Nyman et al., 2011; de Vries et al., 2014) and fertility variables (Sandgren et al., 2009; Nyman et al., 2011) are good welfare indicators within national databases, it suggests that maybe our model could be applied to the other Portuguese regions and even to other countries.

The model includes two variables, OFD and CCI > 430 (Figure 1), while OFD and SexRatio were found to be statistically different between GW and PW farms (Table 4). This is probably due to the fact that the Wilcoxon test compares only one indicator at a time (i.e. univariable analysis) between GW and PW farms. The J48 methodology, on the other hand, uses the best group of indicators to detect farms with GW and PW. It is likely that the variable SexRatio did not add much value to the model when OFD was already considered for predicting the welfare of dairy farms, whereas CCI>430 did.

The small and unbalanced sample size (5 PW vs. 19 GW farms) makes it difficult to detect small differences in potential welfare indicators between GW and PW farms and may explain why only two statistically different potential welfare indicators (OFD and SexRatio; Table 4) could be highlighted by the univariable analyses.

Proportion of on farm deaths was both statistically different between GW and PW farms and present in the decision tree (Table 4; Figure 1). A high mortality rate is an important indicator of poor welfare but this rate should be appraised with caution and always correlated with the culling indices. For instance, a high mortality rate with a low culling rate may indicate that some very sick animals were not culled early enough and were kept in miserable conditions for an inappropriate length of time. In the current study, OFD was 55% and 85% for farms with GW and PW, respectively. There were, therefore, more cows dying on-farm vs being culled and sent to the slaughterhouse in PW farms. However, cows that were sold to other farms before being sent for slaughter were not considered when computing this index. Including them could possibly provide a more accurate index to represent the total number of culled cows. In a study performed in Portugal using farm-based records, Barros (2013) concluded that within all culled cows (n = 2,476), 6% (n = 156) were sold to other farms, 26% (n = 641) died on-farm and 68% (n = 1,679) were sent to the abattoir. Therefore numbers of cows that died on-farm or went to slaughter, were quite different from the current study. Once again, the small sample size used in the current study or the fact that it is a convenience sample might be the reason for the differences observed.

Female/male birth ratio was statistically different between GW and PW farms (Table 4). It was calculated to understand if farms with GW would have better record completeness than farms with PW. In this case, it is hypothesized that male calves' births are not fully reported on some farms, resulting in an unbalanced SexRatio. Results from the current study seem to support this hypothesis, since the median value for SexRatio obtained for farms with PW (1.8) was substantially and statistically significantly higher than the GW farm ratio (1.0). This could mean that PW farms are less likely to keep good and comprehensive records, the care of males may also be poorer or males may be more likely to be killed at birth and not registered. Another possibility is that stress experienced during pregnancy may skew the sex ratio as has been shown in other farm animal species (Baxter et al., 2012), and that might be a risk factor in dairy cattle too. Finally, the use of sexed semen to obtain more heifer calves, could also be a possible explanation. There are no scientific data on the use of sexed semen in Portugal, however, according to Portuguese bovine practitioners its use is scarce (5%), which makes that hypothesis quite unlikely (George Stilwell, personal communication, October 1, 2015).

The decision tree revealed OFD and CCI > 430 as welfare indicators (Figure 1). Hartigan (1995) stated that the percentage of cows in a herd outside the range of 365 to 415 days of calving interval should be less than 5%. According to that author, values outside this range will possibly mean lower production benefits for the farmer and might suggest poor reproductive management but also the presence of nutrition, health and welfare problems that result in reproductive failure. High percentages of CCI > 430 might be related to factors not directly related to welfare, such as low oestrus detection or ineffective artificial insemination. In contrast, some studies have shown that farmers may voluntarily seek higher calving intervals, which might be sometimes advantageous, particularly in high yielding cows (Ratnayake et al., 1997; Arbel et al., 2001; Österman, 2003). However, the targeted calving interval should be shorter to that found in our dataset (both WQ and national database herds).

To the authors' knowledge, only three studies explored the use of routinely collected data to identify dairy herds with poor or good welfare (Sandgren et al., 2009; Nyman et al., 2011; de Vries et al., 2014). The current study's results are in agreement with those of Sandgren et al. (2009) and Nyman et al. (2011) in which indicators referring to mortality and to reproduction were statistically significant potential welfare indicators. In the study of de Vries et al. (2014), on-farm mortality of cows less than 60 DIM was the variable most frequently included in the final models. These results and those from the current study lead to the conclusion that, despite the different husbandry practices in dairy farms of each country (Netherlands, Sweden, Portugal), and the different contents of the national databases, mortality seems to be an important welfare indicator.

Although the WQ is considered as the most comprehensive welfare assessment protocol in farm animals, there are still some points to improve. One point is the need to reduce the workload to ensure feasibility (de Vries et al., 2013b). Heath et al. (2014a) tried to solve this problem by using the animal-based ‘iceberg indicators’ method, in which the criterion ‘Absence of prolonged thirst’ was showed to have a deterministic role in the overall farm classification. Their results revealed another problem with the WQ protocol that had already been identified by de Vries et al. (2013a): the aggregated scores lead to a higher relative importance of the welfare principles good feeding and good housing compared to the other principles, which goes against the goals of the WQ protocol. Heath et al. (2014a) also questioned the usefulness and validity of the overall aggregation of the single welfare measures. Other problems are inter-evaluator repeatability and the lack of information on how to deal with missing data (Heath et al., 2014b). The application of the WQ protocol has uncovered some issues, and the results of the current study should be considered in this light. We should also consider the lag between the time the national data was collected (2008-2013) and the time the WQ audit was conducted (2013). When developing the model we made the assumption that the factors affecting welfare (such as feeding, bedding and stocking density) on the farm remained constant throughout the period (2008-2013). In general, farms are slow to change their overall management systems, but farmers do make small minor changes to their practices (Dufour et al., 2010). This assumption is, therefore, likely to be valid. A more prudent interpretation of the current results, however, would be that the identified welfare indicators (OFD and CCI > 430) are good predictors of the subsequent herd WQ score. Given that a delay in the availability of the national data is always expected, the developed model would still be very applicable and practical for screen out GW farms.

In different countries, the implementation and enforcement of national cattle registries might differ and therefore the application of some of the proxy indicators that were used in this study should be evaluated. The process should also be adaptive and sensitive to changes in farmers' attitudes to welfare and record keeping, therefore the application of these approaches for the identification of farms at risk should consider systematic model updates based on on-farm controls results. Although this study was conducted with dairy cows, similar methods could possibly be applied to other kind of animal production systems which have a rich and comprehensive national database.

Finally, in the future, further work using a larger sample size, a more powerful study design (with a balanced number of farms within each group) and a sample of farms that includes farms graded as 'excellent' or 'enhanced', will be needed. Moreover, more attention should be drawn to the validity of the welfare assessment classification. The novelty of our decision tree, classifying farms whose welfare level is unknown, is that it is easy to apply and to interpret. However, the model could be improved by the inclusion of other data such as veterinarian treatments, milk quality and organs rejected at the slaughter house. These could be useful in identifying farms with high levels of lameness or mastitis, for example, which are known to be major welfare problems in dairy farms.

CONCLUSIONS

This study allowed for the development of a model to identify herds with poor welfare through two variables: 'proportion of on-farm deaths' and 'calving-to-calving interval'.

The national cattle database analysis proved to be an important tool in a stepwise dairy cow welfare evaluation and it could be useful in helping official veterinary services in detecting farms that are more likely to have good welfare in order to focus the more time-consuming WQ audit on farms that are more likely to be problematic.

ACKNOWLEDGMENTS

The authors acknowledge the financial support from Fundação para a Ciência e Tecnologia (Project Pest-OE/AGR/U10276/2014). The authors also gratefully acknowledge Dr. Simon Dufour for his comments and insights on an earlier version of the manuscript, and Dr. Tiago Barros, Dr. António Ventura, Portuguese official veterinary services and the Portuguese farmers that participated in the project.

REFERENCES

Abernethy, M., 2010a. Data mining with WEKA, Part 1: Introduction and regression. Series: Data Mining with WEKA, Part1. DeveloperWorks, IBM Technical Library. Retrieved 01-10-2015, from http://www.ibm.com/developerworks/library/os-weka1/os-weka1-pdf.pdf.

Abernethy, M., 2010b. Data mining with WEKA, Part 2: Classification and clustering. Series: Data Mining with WEKA, Part2. DeveloperWorks, IBM Technical Library. Retrieved 01-10-2015, from http://www.ibm.com/developerworks/library/os-weka2/os-weka2-pdf.pdf.

Ali, J., Khan, R., Ahmad, N., Maqsood, I., 2012. Random forests and decision trees. I. J. C. S. I., 9(5), 3.

Arbel, R., Bigun, Y., Ezra, E., Sturman, H., Hojman, D., 2001. The effect of extended calving intervals in high lactating cows on milk production and profitability. J. Dairy Sci., 84(3), 600-608.

Barros, T. M. L. C., 2013. Análise dos padrões de refugo em explorações leiteiras do sul de Portugal. MS Thesis. Universidade Técnica de Lisboa, Lisboa.

Baxter, E. M., Jarvis, S., Palarea-Albaladejo, J., Edwards, S. A., 2012. The weaker sex? The propensity for male-biased piglet mortality. PLoS One, 7 (1), e30318.

Calamari, L., Bertoni, G., 2009. Model to evaluate welfare in dairy cow farms. Ital. J. Anim. Sci., 8 (1), 301-323.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intelli. Res., 16, 321-357.

de Vries, M., Bokkers, E., Dijkstra, T., Van Schaik, G., De Boer, I., 2011. Invited review: Associations between variables of routine herd data and dairy cattle welfare indicators. J. Dairy Sci. 94, 3213-3228.

de Vries, M., Bokkers, E. A. M., van Schaik, G., Botreau, R., Engel, B., Dijkstra, T., de Boer, I. J. M., 2013a. Evaluating results of the Welfare Quality multi-criteria evaluation model for classification of dairy cattle welfare at the herd level. J. Dairy Sci., 96(10), 6264-6273.

de Vries, M., Engel, B., den Uijl, I., van Schaik, G., Dijkstra, T., de Boer, I., Bokkers, E., 2013b. Assessment time of the Welfare Quality® protocol for dairy cattle. Anim. Welf., 22(1), 85-93.

de Vries, M., Bokkers, E. A. M., van Schaik, G., Engel, B., Dijkstra, T., de Boer, I. J. M., 2014. Exploring the value of routinely collected herd data for estimating dairy cattle welfare. J. Dairy Sci., 97(2), 715-730.

Dohoo, I.R., Martin, S.W., Stryhn, H., 2009. Veterinary Epidemiologic Research. AVC-Inc Charlottetown, Canada.

Dufour, S., Barkema, H. W., DesCôteaux, L., DeVries, T. J., Dohoo, I. R., Reyher, K., Roy, J. P., Scholl, D. T., 2010. Development and validation of a bilingual questionnaire for measuring udder health related management practices on dairy farms. Prev. Vet. Med., 95 (1), 74-85.

European Commission, 2015. Retrieved 01-10-2015, from http://ec.europa.eu/citizens-initiative/public/documents/23.

European Food Safety Authority, 2006. Basic information for the development of the animal welfare risk assessment guidelines. Italy: EFSA.

European Food Safety Authority, 2012. Scientific opinion on the use of animal-based measures to assess welfare of dairy cows. The EFSA Journal, 10 (1), 2554.

Fraser, D., 2004. Applying science to animal welfare standards. In: Proceedings of Global Conference on Animal Welfare: an OIE initiative, Paris, France, pp. 121-134.

Fraser, A.F., Broom D.M., 1990. Farm animal behaviour and welfare. (3rd ed.). Scotland, UK: Bailliere tindall – WB Saunders, pp. 256-279, 294-308, 350-357.

Hartigan, P. J., 1995. Cattle breeding and infertility. In: Meredith, M. J. (Ed.), Animal breeding and infertility. London: Blackwell Science Ltd., pp. 88-168.

Heath, C., Browne, W., Mullan, S., Main, D., 2014a. Navigating the iceberg: reducing the number of parameters within the Welfare Quality® assessment protocol for dairy cows. animal, 8(12), 1978-1986.

Heath, C., Lin, Y-C., Mullan, S., Browne, W.J., Main, D.C.J., 2014b. Implementing Welfare Quality® in UK assurance schemes: evaluating the challenges. Anim. Welf., 23., 95-107.

Knierim, U., Winckler, C., 2009. On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. Anim. Welf., 18 (4), 451-458.

Main, D.C.J., Whay, H.R., Leeb, C. Webster, A.J.F., 2007. Formal animal-based welfare assessment in UK certification schemes. Anim. Welf., 16, 233-236.

Nyman, A. K., Lindberg, A., Sandgren, C. H., 2011. Can pre-collected register data be used to identify dairy herds with good cattle welfare. Acta Vet. Scand, 53 (1), S8.

Oltenacu, P.A., Broom, D.M., 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Anim. Welf., 19(S), 39-49.

Österman, S., 2003. Extended calving interval and increased milking frequency in dairy cows. PhD Thesis. Swedish University of Agricultural Sciences, Uppsala.

Ratnayake, D., Berglund, B., Bertilsson, J., Forsberg, M., Gustafsson, H., 1997. Fertility in dairy cows managed for calving intervals of 12, 15 or 18 months. Acta Vet. Scand., 39 (2), 215-228.

Sandgren, C. H., Lindberg, A., Keeling, L., 2009. Using a national dairy database to identify herds with poor welfare. Anim. Welf., 18 (4), 523-532.

Sørensen, J.T., Sandøe, P., Halberg, H., 2001. Animal welfare as one among several values to be considered at farm level: the idea of an ethical account for livestock farming. Acta Agr. Scand., 30, 11-17.

Veissier, I., 2007. Rationale behind the Welfare Quality® assessment of animal welfare. In: Veissier, I., Forkman, B., Jones, B. (Eds.), Assuring animal welfare: from societal concerns to implementation, Second Welfare Quality stakeholder conference, 3–4 May 2007, Berlin, Germany, p.9-12.

WEKA (Waikato Environment for Knowledge Analysis), 2015. Retrieved 05-06-2015, from http://www.cs.waikato.ac.nz/~ml.

Welfare Quality®., 2009. Welfare Quality® assessment protocol for cattle. Lelystad, Netherlands: Welfare Quality® Consortium.

Table 1. Welfare principles, criteria and indicators of Welfare Quality® protocol for dairy cattle (WQ, 2009)

Principles	Welfare criteria	Welfare measures
Good feeding	Absence of prolonged hunger	Very lean cows
	Absence of prolonged thirst	Water points conditions
Good housing	Comfort around resting	Lying behavior; dirtiness
	Thermal comfort	As yet, no indicator is developed
	Ease of movement	Presence of tethering
Good health	Absence of injuries	Lameness; integument alterations
	Absence of disease	Cough; nasal discharge; ocular discharge; vulvar discharge; diarrhea; hampered respiration; subclinical mastitis; on-farm mortality; dystocia; and downer cows
	Absence of pain induced by management procedures	Disbudding/dehorning and tail docking
Appropriate behaviour	Expression of social behaviours	Agonistic encounters
	Expression of other behaviours	Access to pasture
	Good human-animal relationship	Avoidance distance
	Positive emotional state	Scores of 20 terms of the Qualitative Behaviour Assessment

Table 2. Source tables, selection parameters and formula of the potential welfare indicators derived from the Portuguesedatabase SNIRB

Potential welfare indicators[a]	Source tables	Selection parameters	Formula
AFC	Live cattle, births	Calved cows	Date at first calving minus date of birth, herd median
CCI < 345	Births	Calved cows	No. of calving intervals lower than 345 d divided by total no. of calving intervals
CCI > 430	Births	Calved cows	No. of calving intervals higher than 430 d divided by total no. of calving intervals
MtC	Births, movements, slaughter	Female calves	Incidence of on-farm deaths and emergency slaughter reported in death / 100 animal-year at risk[b]
Mt	Live cattle, movements, slaughter	Female cattle	Incidence of on-farm deaths and emergency slaughter reported in death / 100 animal-year at risk
OFD	Movements		No. of on-farm deaths divided by the total no. of slaughtered animals, on-farm deaths and animal disappearances[c]
EmgSl	Slaughter	Calved cows	No. of emergency slaughters divided by no. of regular and emergency slaughters
TLS	Live cattle, slaughter	Calved cows	Date of slaughter minus date of birth, herd median
30ppSl	Births, slaughter	Calved cows	No. of cows slaughtered between 0 - 30 d post-partum divided by no. of slaughtered animals
60ppSl	Births, slaughter	Calved cows	No. of cows slaughtered between 0 - 60 d post-partum divided by no. of slaughtered animals
PartCRej	Live cattle, slaughter	Calved cows	No. of partially rejected carcasses divided by the total of no. of approved carcasses, totally rejected carcasses and partially rejected organs or carcasses
TotCRej	Live cattle, slaughter	Calves cows	No. of totally rejected carcasses divided by the sum of no. of approved carcasses, totally rejected carcasses and partially rejected organs or carcasses
C < 272	Slaughter	Calved cows	No. of carcass weighting less than 272kg divided by no. of slaughtered animals
VTC	Slaughter	Calved cows	No. of carcasses with fat class 1 divided by no. of carcasses with fat class 1 (very thin) or 3 (normal)
SexRatio	Births		No. of female calves divided by no. of male calves

[a] AFC = age at first calving; CCI < 345 = proportion of calving intervals lower than the biological acceptable; CCI > 430 = proportion of calving intervals higher than 430 d; MtC = calf mortality (until six months); Mt = mortality rate; OFD = proportion of on-farm deaths; EmgSl = proportion of emergency slaughter; TLS = total life span; 30ppSl = proportion of cows slaughtered before 30 post-partum; 60ppSl = proportion of cows slaughtered before 60 d post-partum; PartCRej = proportion of partial carcass rejection; TotCRej = proportion of total carcass rejection; C < 272 = proportion of carcasses with less than 272 kg; VTC = proportion of carcasses with fat class ‘very thin’ (class 1); SexRatio = female/male births ratio.

[b] animal-days at risk was computed as the number of animals that remained on farm during the complete study period multiplied by length of the study period (i.e. 180 d or 365 d) plus sum of the exact number of days spent on farm for the animals that left the farm (transfer in life, slaughter, depart to EU, exportation, death on farm, disappearance, deactivation); result was multiplied by 36500, so that instead of animal-days, result could be presented as 100 animal-year.

[c] animal disappearances is referred to animals that were stolen or not found in the farm.

Table 3. Results of potential welfare indicators from national cattle database on national dairy farms and on farms using the Welfare assessed Quality protocol

	National dairy farms (n = 6,605)				Farms assessed using WQ[d] (n = 24)				
Potential welfare indicators[a]	Mean	SD[b]	Median	IQR[c]	Mean	SD	Median	IQR	*p*[e]
AFC (mo.)	27	3.82	27	25 - 29	26	1.09	26	25 - 27	0.06
CCI < 345 (%)	17.0	12.92	14.3	8.7 - 21.3	10.4	5.36	9.9	7.9 - 15.8	**0.01**
CCI > 430 (%)	48.1	20.79	44.8	33.3 - 58.1	45.6	9.66	45.2	38.3 - 49.3	0.99
MtC (deaths / 100 animal-year)	20.8	113.80	0	0 - 20.3	28.2	23.41	23.1	8.1 - 45.7	**<0.01**
Mt (deaths / 100 animal-year)	5.3	71.91	0	0 - 3.8	4.5	3.19	3.6	2.6 - 6.0	**<0.01**
OFD (%)	40.8	29.89	37.5	14.3 - 62.5	61.5	24.47	62.1	40.0 - 81.1	**<0.01**
EmgSl (%)	1.9	6.61	0	0 - 0	1.4	1.84	0.1	0 - 2.6	**<0.01**
TLS (mo.)	81	29.24	75	63 - 95	69	11.29	68	63 - 74	**0.03**
30ppSl (%)	6.1	13.64	0	0 - 7.7	10.5	20.68	3.6	0.8 - 11.7	**<0.01**
60ppSl (%)	10.9	18.03	0	0 - 16.7	16.3	20.21	10.7	6.5 - 16.4	**<0.01**
PartCRej (%)	1.2	4.83	0	0 - 0	1.4	1.70	0.9	0 - 2.8	**<0.01**
TotCRej (%)	4.0	9.40	0	0 - 5.2	6.8	9.67	4.6	2.7 - 6.3	**<0.01**
C < 272 (%)	61.1	28.24	60.9	40.0 - 84.3	38.1	13.36	42.8	28.1 - 47.3	**<0.01**
VTC (%)	47.3	33.90	48.2	20.0 - 75.0	36.1	15.14	39.0	27.8 - 47.9	0.13
SexRatio	1.4	1.50	1.0	0.8 - 1.5	1.4	0.69	1.1	0.9 - 1.7	0.25

[a] AFC = age at first calving; CCI < 345 = proportion of calving intervals lower than the biological acceptable; CCI > 430 = proportion of calving intervals higher than 430 d; MtC = calf mortality rate (until six months); Mt = mortality rate; OFD = proportion of on-farm deaths; EmgSl = proportion of emergency slaughter; TLS = total life span; 30ppSl = proportion of cows slaughtered before 30 post-partum; 60ppSl = proportion of cows slaughtered before 60 d post-partum; PartCRej = proportion of partial carcass rejection; TotCRej = proportion of total carcass rejection; C < 272 = proportion of carcasses with less than 272 kg; VTC = proportion of carcasses with fat class 'very thin' (class 1); SexRatio = female/male births ratio.

[b] SD = standard deviation.

[c] IQR = interquartile range.

[d] WQ = Welfare Quality protocol.

[e] Wilcoxon tests between national dairy farms and farms assessed using the Welfare Quality protocol.

Table 4. Results of potential welfare indicators from national cattle database on national dairy farms and on farms with good and poor welfare

	Farms with good welfare (n = 19)				Farms with poor welfare (n = 5)				
Potential welfare indicators[a]	Mean	SD[b]	Median	IQR[c]	Mean	SD	Median	IQR	p[d]
AFC (mo.)	26	0.92	26	25 - 27	25	1.67	25	24 - 26	0.32
CCI < 345 (%)	10.0	5.46	9.9	7.4 - 15.0	12.3	5.07	9.3	8.5 - 16.5	0.59
CCI > 430 (%)	45.8	10.67	45.2	37.6 - 50.6	44.9	4.83	47.8	40.4 - 48.6	0.83
MtC (deaths / 100 animal-year)	26.3	22.98	20.7	9.9 - 39.3	35.5	26.30	45.5	9.0 - 54.0	0.43
Mt (deaths / 100 animal-year)	4.7	3.38	3.6	3.2 - 6.0	3.8	2.52	2.9	2.2 - 4.1	0.53
OFD (%)	55.2	23.16	52.2	35.9 - 73.0	85.4	11.16	87.2	78.0 - 95.1	**0.01**
EmgSl (%)	1.4	1.75	0.7	0 - 2.6	1.1	2.35	0	0 - 0	0.32
TLS (mo.)	70	12.17	68	63 - 72	67	7.80	67	64 - 74	0.97
30ppSl (%)	7.9	8.46	5.0	2.1 - 11.7	20.3	44.54	0	0 - 1.7	0.16
60ppSl (%)	13.7	10.13	9.4	7.0 - 16.4	26.3	41.60	12.0	5.1 - 14.3	0.86
PartCRej (%)	1.7	1.78	1.2	0 - 2.9	0.6	1.17	0	0 - 0.5	0.19
TotCRej (%)	5.1	3.31	4.4	2.7 - 6.3	13.3	20.55	4.9	3.6 - 5.3	0.68
C < 272 (%)	38.9	10.88	42.9	28.1 - 46.6	35.1	21.90	36.8	32.1 - 51.4	1.00
VTC (%)	37.4	13.34	39.0	29.8 - 45.9	31.2	21.88	27.0	25.0 - 48.3	0.94
SexRatio	1.2	0.56	1.0	0.9 - 1.6	2.1	0.81	1.8	1.4 - 2.9	**0.02**

[a] AFC = age at first calving; CCI < 345 = proportion of calving intervals lower than the biological acceptable; CCI > 430 = proportion of calving intervals higher than 430 d; MtC = calf mortality rate (until six months); Mt = mortality rate; OFD = proportion of on-farm deaths; EmgSl = proportion of emergency slaughter; TLS = total life span; 30ppSl = proportion of cows slaughtered before 30 post-partum; 60ppSl = proportion of cows slaughtered before 60 d post-partum; PartCRej = proportion of partial carcass rejection; TotCRej = proportion of total carcass rejection; C < 272 = proportion of carcasses with less than 272 kg; VTC = proportion of carcasses with fat class 'very thin' (class 1); SexRatio = female/male births ratio.

[b] SD = standard deviation.

[c] IQR = interquartile range.

[d] Wilcoxon tests between farms with poor and good welfare.

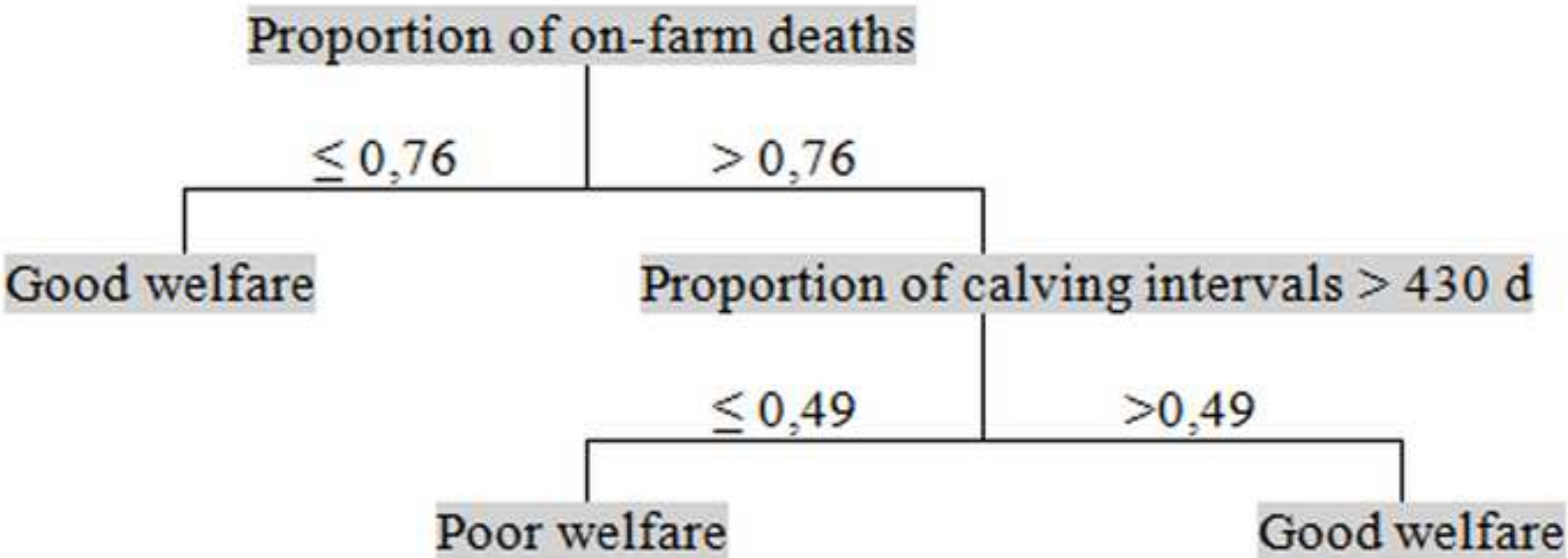
Table 5. Contingency table in the classifier J48 from Waikato Environment for Knowledge Analysis (WEKA) to detect farms with poor welfare

Accuracy[a]: 75.86%			
	Predicted positive	Predicted negative	Total
Actual positive	7	3	10
Actual negative	4	15	19
Total	11	18	29

[a] Accuracy is a ratio of no. of correctly classified instances divided by the total no. of instances; therefore, this model has an accuracy of (22 / 29) = 0.76

Figure 1. Classification tree created with J48 classifier from Waikato Environment for Knowledge Analysis (WEKA) to detect farms with poor welfare

Figure 1



Proportion of on-farm deaths
≤ 0,76
> 0,76
Good welfare
Proportion of calving intervals > 430 d
≤ 0,49
>0,49
Poor welfare
Good welfare